

Coimbatore Institute of Technology
Department of Computing (Decision and Computing Sciences)
2018 Batch
VII Semester Project Work and Viva Voce I

S. No.	Roll No.	Name	Project Title
1	18 33 001	AADHITHIAN.L.	Enzyme Kinetics Analysis
2	18 33 002	AKKSHAYA SRI.J.	Web Portal Development for Vehicle Insurance
3	18 33 004	ASHOK ARVIND T.N.	Car rental booking system
4	18 33 005	BARATH VIGNESH.S.	Business Analytics and Reporting on CPG product
5	18 33 006	BOBBY PRATHIKSHANA.M.	Trade Promotion and Sales Analytics for Consumer Packaged Goods Industry
6	18 33 008	CHITHIRAI JOTHI.U	Network Intrusion Detection and System Using Machine Learning'
7	18 33 009	DEEPAK KUMAR.C.R.	COMPARING MACHINE LEARNING ALGORITHMS FOR PREDICTING TRANSPORT OF SPECIES IN POROUS MEDIA USING PHYSICS INFORMED MACHINE LEARNING APPROACH
8	18 33 010	DEETCHIKA. R	IOT Cyber Security using machine learning
9	18 33 011	DHANUSH KANNAN.A.	Software Development in Neurosurgical Simulator
10	18 33 012	DINESH KUMAR .T	Sign language detection using computer vision
11	18 33 013	GAUTAM.T.	Creation of website and engaging with customers for divine layout promoters
12	18 33 014	GOKULAVASAN.K.R.	Dissecting Insider threat using data science - predictive modelling
13	18 33 015	GUHANESVAR.M.	Anomaly Detection and Data Analysis in Biocide production
14	18 33 016	HARI PRIYA.H.	Digital marketing strategies for the development of Treat Pa in Search Engine
15	18 33 017	HARINI.C.	DIGITAL MARKETING PLANS TO BOOST THE ORGANIC TRAFFIC OF THETILEBROS
16	18 33 018	HARIRAMAKRISHNAN.S	Web Project Proposal, Educational Websites (Startup Coach)
17	18 33 019	JAVAGAR.M.	Real- Time hand gesture detection and sign recognition
18	18 33 020	KANISHK.P.	Snipzz - Personalized News Application
19	18 33 021	KAVYA.B	AUDIT ANALYTICS AUTOMATION WITH ALTERYX AND POWER BI
20	18 33 022	KAVYA.P.	Voice gender recognition and Add-on extensions of gmeet for gesture control and emotion analysis
21	18 33 023	KEERTHIKA.J.	Digital marketing strategies for the development of Tiny Twig in Search Engine
22	18 33 024	KOSAL RAM.G.	Automatic Number Plate Recognition
23	18 33 025	MEGAVARNIYA.S.	Digital marketing Plan to increase users for Befable
24	18 33 026	MOHAMMED ISMAIL.A.	Early Detection of Autism Spectrum Disorder in Children - Facial Emotion Recognition and survey data analysis
25	18 33 027	MOHINDRAN.S.R.	Deployment Notifier Tool and Election Data Management Tool
26	18 33 028	MRITHIKA.S.	Dynamic Test Suites Selector
27	18 33 029	NARESHKUMAR.G.	Analysis and Identification of Fraudulent Users in E-commerce
28	18 33 030	NAVEENKUMAR.R.D	Zila Shop Management System
29	18 33 031	NITHIN.S.V.	Car Insurance Claim Fraud Detection
30	18 33 032	NITHISH G.	Natural Infusions
31	18 33 033	NIVETHA.V.	Digital Marketing for FACE Prep
32	18 33 034	NIVETHENI.C.G.	Security Bot System
33	18 33 035	PRANAAV.A.R.	Suspicious network activity detection and classification in IOT devices using Machine Learning
34	18 33 036	PRANEETA.T.R.	Website enhancement and social media marketing
35	18 33 037	PREETHI.S.	Social media marketing and SEO
36	18 33 038	PRIYADHARSHINI.B.	Digital Marketing Plan for Promoting Sales for Premier No1

37	18 33 039	PRIYADHARSHINI.R.	Patient Healthcare Risk Identification and Prediction Using Machine learning algorithms
38	18 33 040	RINUBHA.P.	Digital and Social Media Marketing
39	18 33 041	ROHINTH.S.	Early Detection of Autism Spectrum Disorder in Children
40	18 33 042	ROSHINI.G.	Business Analytics and Reporting on Sales of CPG Goods
41	18 33 043	SAAGARIKA.S.	User Experience Enhancement and Web Development
42	18 33 044	SACHIN DHANA PAUL.J.	Detection of DDOS attacks using ML approach
43	18 33 045	SADHVI ANUNAYA.R.	Digital Marketing techniques to increase the visibility of the website Nidra Nutrition
44	18 33 046	SANCHEZ INNOCENCIA .D.	Analytical Development in Neurosurgical Simulator
45	18 33 047	SANGAVI.S.	Credit Card Fraud Detection
46	18 33 048	SANJEETH.S.	Intrusion Detection and Log File Analysis in Cybersecurity
47	18 33 049	SARATHRAM.M	Credit Model Risk Analysis
48	18 33 051	SRIDHAR.S.	Real Time Facemask Detection
49	18 33 052	SUBASH.S.	Early Detection of Autism Spectrum Disorder in Children - Gaze Tracking and Object Detection
50	18 33 053	SUBAVARSSHINI.R.T	User Experience Development System
51	18 33 054	SWETHA SRINIVASAN	A Tool for the Study of Impact of Big Data Analytics on Firm Financial Performance and Analysing Online Brand Interactions and Conversations using Natural Language Processing
52	18 33 055	TILAK VIJAYARAGHAVAN.S.	Qualitative analysis on Employee Performance
53	18 33 056	VENKATESHAN.D.	Phishing Website Detection By Machine learning
54	18 33 057	VENKETESZ.R.R.	Geospatial Image Classification and Change detection
55	18 33 058	VIGNESH.N.	Dissecting Insider Threat Using Data Science For Behavioral Analysis and Predictive Modelling
56	18 33 059	VIKNEESH.C.S	Performance analysis of medical representatives



CERTIFICATE OF INTERNSHIP

We present this to

Viknesh C S

of COIMBATORE INSTITUTE OF
TECHNOLOGY for successfully completing the
Internship of Data Science with the Society of AI
from 05th July 2021 to 05th January 2022

Sakshi Ganda
HR-Manager



www.societyofai.in



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Vignesh N

to certify that he/she has completed Intensive internship
on

**Dissecting Insider Threat with Data Science for Behavioral
Analytics and Predictive Modelling**

During the period

June - December 2021

Mamta Verma

Dr. Mamta Verma

Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislaborg>



LAUNCHTRAX

GST - 33AAECL0372P1ZC
Email - info@launchtrax.com

INTERNSHIP COMPLETION CERTIFICATE

From,
Ramji Srimurugan
Director
LaunchTrax Private Limited
51/A Link Road
Virudhunagar - 626001

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Venkatesz RR, underwent his internship at LaunchTrax Private Limited. During the period of his internship between June 2021 to December 2021 as an Application Engineer, his performance was found to be Excellent. He worked on a project titled "**Geospatial Image Classification and Change Detection**" and his key project contributions were in the area of Geospatial analytics.

For any further clarifications, please reach out to us over email at info@launchtrax.com

Date : February 03, 2022

 
Authorized Signatory :
Name of the Person : Ramji Srimurugan
Designation : Director
Firm Name & Seal : Launchtrax Private Limited



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Venkateshan D

to certify that he/she has completed Intensive internship
on

Phishing Website Detection By Machine Learning

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislabs.org>



> Expertise > Resource > Value >

Appreciation Letter

Trichy
21st December 2021

From

The Management,
DSM Soft (P) Ltd,
Cantonment,
Trichy - 01.

To

Mr. Tilak Vijayaraghavan,
IV - Decision & Computing Science,
CIT,
Coimbatore.

Dear Mr. Tilak Vijayaraghavan,

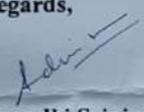
SUB: Appreciation to the candidate

This letter of appreciation is to applaud your dedicated efforts and commitment towards the completion of the internship program in **Quality Assurance - Data Quality Analysis**. Your understanding of the process is well appraised.

Congratulations once again on the completion of your internship with us.

Wishing you all the very best!

Regards,


Anandhi Srinivasan
Senior Manager - HR

D-07999-TRC-PSD-01333-V2.0

DSM Soft (P) Ltd., First Floor, No. 4, VOC Road, Trichy - 620 001, India.

+91 80567 86691 info@dsmsoft.com www.dsmsoft.com

CIN No. : U72300TN1991PTC020923

MITACS Globalink CERTIFICATE OF COMPLETION

Swetha Srinivasan

This certifies that Swetha Srinivasan has successfully completed a Mitacs Globalink Research Internship effective October 2021.

Project Title: The Study of Impact of Big Data Analytics and firm financial performance

Host University: Université du Québec à Chicoutimi – Chicoutimi

Host Professor: Myriam Ertz

Globalink Research Internships are 12 weeks in duration and interns are required to work 40 hours per week.

October 10th, 2021



Gail Bowkett
Vice President of Programs





Dated: October 25, 2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Ms. Swetha Srinivasan did her internship with us from July 26, 2021 to October 25, 2021 at Gurgaon office.

During the period of her internship, we found her dutiful, hardworking and a good team player.

We take this opportunity to wish her all the best in her future endeavours.

For Edelman India Private Limited

Pankaj Suri
Executive Vice President- Human Resources, India

Edelman India Private Limited

Vatika Triangle, 6th Floor, Sushant Lok - 1, Block A
Gurugram, Haryana 122 002, India

Tel +91 124 4131 400 | Fax +91 124 4130 499 | www.edelman.com

A Daniel J. Edelman Company

CIN: U74140MH1993PTC071100

CERTIFICATE

CoverNest Insurance Web Aggregator Private Limited

1st Floor, No – 6/20, Bharathiar Street, Vanuvampet, Chennai – 600091.
Ph – 044 22603661, Mail – info@covernest.com, www.covernest.com

30 December 2021

Internship Completion Certificate

This is to certify that **Subavarsshini R T (Roll No - 1833053)**, pursuing her pre-final year M.Sc Decision and Computing Sciences (5 years Integrated) at Coimbatore Institute of Technology, Coimbatore has successfully completed her internship for a duration of 6 months from 14 June 2021 to 30 November 2021 at our organization. During the period of internship, she was found hard working, professional and has shown good commitment towards the task assigned to her. Below are her project specifications:

Project Title: User Interface Development System

We wish her all success in her future endeavors.

For CoverNest Insurance Web Aggregator Pvt Ltd,



Rajasekaran R G,
Managing Director





December 20, 2021

INTERNSHIP CERTIFICATE

To whomever it may concern

This is to certify that Subash S worked under my supervision during his internship period at Cornerstohn Solutions during the period June 7, 2021 to December 7, 2021. His internship was focused on Early Detection of Autism Spectrum Disorder in Children.

Subash possesses a good moral character, pleasing personality, attention to detail and the intent to complete his work on time. I wish him success for his future endeavors.

Best Wishes,

Kirthi Murthiraj

Cornerstohn Solutions

1.1 ORGANIZATION PROFILE

Dr. E. Chandra Blessie, Assistant Professor
Department of Computing (Artificial Intelligence and Machine Learning)
Coimbatore Institute of Technology
Civil Aerodrome Post, Coimbatore 641014.
Ph - 9842266044, Mail - blessie@cit.edu.in

08/02/2022

Internship Completion Certificate

This is to certify that **T.N.Ashok Arvind (Roll No.:1833004)** pursuing M.Sc. (Decision and Computing Sciences) at Coimbatore Institute of Technology, Coimbatore has successfully completed his internship under me for a duration of 4 months from 01 September 2021 to 30 December 2021. During the period of internship, he was sincere and dedicated towards the task assigned to him. The title of the project is "**CAR RENTAL BOOKING SYSTEM**".

Wish him all success in his future endeavors.


Dr. E. Chandra Blessie
Supervisor

Dr. E. Chandra Blessie, MCA, M.Phil, Ph.D
Assistant Professor
Department of Computing
Coimbatore Institute of Technology
Coimbatore - 641 014.



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Sridhar S

to certify that he/she has completed Intensive internship
on

REAL TIME FACE MASK DETECTION

During the period

June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislaborg>



05 FLOOR, DEVTHASTHALI CORPORATE
TOWER, A-42/5 VISHWAKARMA RD, BLOCK A, INDUSTRIAL
AREA, SECOR 62, NOIDA, UTTAR PRADESH - 201309

HUEY TECH PVT LTD

Experience Letter

18th January 2022

Name – M Sarathram

Location – Noida

Dear Sarathram,

This is to confirm that you were employed with us from **14th June 2021 to 14th December 2021** and was employed as **Intern- Data Science**.

You stand relieved from the services of the company from the closing hours **14th December 2021**.

We would like to remind you of your obligation under the confidentiality agreement signed by you at the time of your joining and the confidentiality and Non-Disclosure clause in your appointment letter. Specific attention is drawn to your declaration not to divulge or use such confidential information failure of which the company would be liable to take legal action against you.

We thank you for your contributions to the organization and wish you the very best for your future endeavors.

Yours Sincerely,

Huey Tech Pvt Ltd

This letter does not give you any legal authority to represent the company or its subsidiaries and you are hereby notified that any use, disclosure, copying, distribution, or reliance upon the contents of letter is strictly prohibited. The company may take necessary legal action in case of any breach.

This is a computer generated letter. No signature is required.

CERTIFICATE



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Sanjeeth

to certify that he/she has completed Intensive internship
on

**Dissecting Insider Threat with Data Science for behavioural
Analytics and Predictive modelling**

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION
A Not-For-Profit Entity
<https://Sycurislab.org>



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Sangavi S

to certify that he/she has completed Intensive internship
on

CREDIT CARD FRAUD DETECTION

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION
A Not-For-Profit Entity
<https://Sycurislaborg>



Curneu MedTech Innovations Private Limited

Dated: 10th December 2021

TO WHOM IT MAY CONCERN

This is to certify that **Ms. Sanchez Innocencia D** has done her internship with **Curneu MedTech Innovations Private Limited** as a **Junior Data Science Developer** dated from **June 1, 2021 to November 30, 2021**.

She has worked on a project titled **Analytical Development in Neurosurgical Simulator** during the internship and she has shown full Sincerity, Dedication and Hard work towards her concerned job, which has helped in improving the management of the company.

We wish her all the best for her future endeavors.

For, Curneu MedTech Innovation Pvt Ltd.

David Roshan Cyril

CEO

Curneu MedTech Innovations Private Limited, C-2 Kurichi Housing Unit, Phase 1, SIDCO Post,
Coimbatore - 641021

PAN: AAHCC4229K TAN: CMBC06382F

Contact Number: +91 97157 07807 / +91 82177 87644

CERTIFICATE



Date: 30.11.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Ms. Sadhvi Anunaya R** student of **Coimbatore Institute of Technology** has successfully finished her internship in "**Digital Marketing**". During the period from 17th May 2021 to 30th November 2021 at **Bright Bridge InfoTech Pvt Ltd**, Coimbatore.

During the period of her internship with us, she was found punctual, hardworking, and inquisitive.

We wish her all the best in future endeavors.

For Bright Bridge InfoTech Pvt. Ltd.,

Jeffery Jones P

Jeffery Jones P

Senior HR Executive

jeffery@brightbridge.co

Registered Office

No.17, Shriram Vijaya Hyvde park, Duraisamy Layout, Peelamedu, Coimbatore-641004.



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Sachin Dhana Paul J

to certify that he/she has completed Intensive internship
on

Detection of DDOS Attacks Using ML Approach

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislabs.org>

CoverNest Insurance Web Aggregator Private Limited

1st Floor, No – 6/20, Bharathiar Street, Vanuvampet, Chennai – 600091.
Ph – 044 22603661, Mail – info@covernest.com, www.covernest.com

30 December 2021

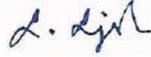
Internship Completion Certificate

This is to certify that **Saagarika S (Roll No - 1833043)**, pursuing her pre-final year M.Sc Decision and Computing Sciences (5 years Integrated) at Coimbatore Institute of Technology, Coimbatore has successfully completed her internship for a duration of 6 months from **14 June 2021** to **30 November 2021** at our organization. During the period of internship, she was found hard working, professional and has shown good commitment towards the task assigned to her. Below are her project specifications:

Project Title: User Experience Enhancement and Web Development

We wish her all success in her future endeavors.

For CoverNest Insurance Web Aggregator Pvt Ltd,



Rajasekaran R G,
Managing Director



www.thorogood.com
info@thorogood.com

Bangalore
ACR Towers
2nd Floor
32 Residency Road
Bangalore 560 025
Tel: +91 (0)80 4123 3700
Fax: +91 (0)80 4123 3800

London
Tel: +44 (0) 20 8231 0800
Philadelphia
Tel: +1 877 905 5010
Singapore
Tel: +65 3138 8068
São Paulo
Tel: +55 134040 4150

4 February 2022

Bangalore, India

THOROGOOD®

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Roshini Gopalan (1833042)** of 2018 - 2023, Integrated M.Sc. Decision and Computing Sciences, Coimbatore Institute of Technology, Coimbatore has successfully completed her 7th Semester Internship as a Business Intelligence and Analytics Consultant from June 1st, 2021, to January 31th, 2022 at Thorogood Associates India Private Limited, Bangalore.

She worked on the project **Business Analytics and Reporting on Sales of CPG Goods** under the guidance of Ms. Jessina Kolady, Business Intelligence and Analytics Consultant. She was very sharp in picking up technologies like Azure Databricks, Azure Data Factory, Power BI etc. She played a significant role in the project by analysing the customer requirement, developing the technical components, and completing deliverables on time. Her contribution was well appreciated by the customer.

Regards,



Jessina Kolady
Business Intelligence & Analytics Consultant
Thorogood Associates India Private Ltd.
Bangalore, India



December 20, 2021

INTERNSHIP CERTIFICATE

To whomever it may concern

This is to certify that Rohinth S worked under my supervision during his internship period at Cornerstohn Solutions during the period June 7, 2021 to December 7, 2021. His internship was focused on Early Detection of Autism Spectrum Disorder in Children.

Rohinth possesses a good moral character, pleasing personality, attention to detail and the intent to complete his work on time. I wish him success for his future endeavors.

Best Wishes,

Kirthi Murthiraj

Cornerstohn Solutions

FINE TEX ENGINEERING

29, Periyar Nagar, Nehru Nagar East,
Civil Aerodrome Post, Coimbatore-641 014.

Phone: 80563 22457

E-Mail: finetexengg@gmail.com

Coimbatore

15.12.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Ms. Rinubha P(1833040)** student of Coimbatore Institute of Technology has successfully finished her internship program in "**Digital and Social Media Marketing**", during the period from 2nd June 2021 to 15th December 2021 at Fine Tex Engineering, Coimbatore.

For Fine Tex Engineering,



Proprietor.



Virtusa Consulting Services Private Limited
Registered Office: No.54 IT Highway, Nandambur, Chennai, T.N.; 600 130 IN, T | +91 443987 3000 | F: +91 44 2743 5138

EXPERIENCE & SERVICE CERTIFICATE

7th February 2022

Dear **Priyadarshini R,**

Please find your Internship Record and Relieving date with Virtusa Consulting Services Private Limited as per the details below.

Name	Priyadarshini R
Employee Code	8098718
Designation on Relieving	Intern-Technology
Date of Joining	14-06-2021
Date of Relieving	14-12-2021

We wish you all the best for your future endeavors.

For Virtusa Consulting Services Private Limited

Rahul Sahay,
Senior Vice President (Level II)- Human Resources





we care what you drink

05.02.2022

TO WHOM IT MAY CONCERN

This is to certify that Ms. Preethi Sethumadhavan, a student of Msc.,(DCS), of Coimbatore Institute of Technology ,has successfully Seven(07) Months of Internship (From 16th of June 2021to 31st January 2022) long internship programme at Aguacian Water Purifiers Private Limited . During the period of his internship programme with us she was found punctual, hardworking and inquisitive in nature. Her contribution to Aguacian is appreciable .

We Wish her all success in her future endeavors.

For Aguacian Water Purifiers Private Limited

Mrs.S.Karunya John

Department of Human Resources

Aguacian Water Purifiers Pvt. Ltd.

4/104 - G, Cheran Co-operative Ind. Estate, Somayampalayam (Po), Coimbatore - 641 108

☎ +91 9364260030 ☎ +91 92444 12344 (Service) ✉ info@aguacian.com 🌐 www.aguacian.com

CERTIFICATE



Date: 30.11.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Ms. Priyadharshini B** student of **Coimbatore Institute of Technology** has successfully finished her internship in "**Digital Marketing**". During the period from 17th May 2021 to 30th November 2021 at **Bright Bridge InfoTech Pvt Ltd**, Coimbatore.

During the period of her internship with us, she was found punctual, hardworking, and inquisitive.

We wish her all the best in future endeavors.

For Bright Bridge InfoTech Pvt. Ltd.,

Jeffery Jones P

Jeffery Jones P

Senior HR Executive

jeffery@brightbridge.co



5K Car Care,
54A Bharathi Colony
Main Road,
Peelamedu,
Coimbatore- 641004

DATE: 31st Jan,2022

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Miss. Praneeta has completed her internship with 5k Car Care for 6 (six) months from July 1st, 2021 to December 25th, 2022 in Website Enhancement and Social Media Marketing.

During her internship period we found her to be sincere and hard-working and we wish her all the best for her future endeavors.

Yours Sincerely


Authorized Signatory

Supervisor



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Pranaav A.R

to certify that he/she has completed Intensive internship
on

**SUSPICIOUS NETWORK ACTIVITY DETECTION AND CLASSIFICATION
IN IOT DEVICES USING MACHINE LEARNING**

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma

Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislaborg>



December 31, 2021

Ms Nivetha V,
26, Smg Residence, Niranjan Garden, Mahalingapuram Post
T.Kottampatti, Pollachi.

WORK EXPERIENCE LETTER

This is to acknowledge and certify that **Ms Nivetha V** (Associate Id - 144452) has worked with our organisation Focus 4-D Career Education Pvt Ltd from **May 3, 2021** to **December 31, 2021**. Her designation during her exit was **Intern - Digital Marketing** and her base location was **Coimbatore**.

Ms Nivetha V has rendered her services with the highest degree of responsibility with a professional attitude and we wish all the best in her future endeavours.

Warm Regards,
For Focus 4-D Career Education Pvt. Ltd

A handwritten signature in black ink, appearing to read "Arumugam N Vadivelu".



Arumugam N Vadivelu
HR Business Partner



Date: 30.11.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Mr. Nithish G** student of **Coimbatore Institute of Technology** has successfully finished his internship in “**Digital Marketing**”. During the period from 17th May 2021 to 17th November 2021 at **Bright Bridge InfoTech Pvt Ltd**, Coimbatore.

During the period of his internship with us, he was found punctual, hardworking, and inquisitive.

We wish him all the best in his future endeavors.

For Bright Bridge InfoTech Pvt. Ltd.,

Jeffery Jones P

Jeffery Jones P

Senior HR Executive

jeffery@brightbridge.co

Registered Office

No.17, Shriram Vijaya Hyyde park, Duraisamy Layout, Peelamedu, Coimbatore-641004.

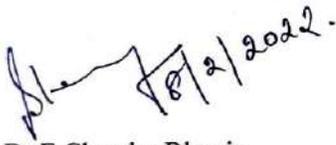
Dr. E. Chandra Blessie, Assistant Professor
Department of Computing (Artificial Intelligence and Machine Learning)
Coimbatore Institute of Technology
Civil Aerodrome Post, Coimbatore 641014.
Ph – 9842266044, Mail – blessie@cit.edu.in

08/02/2022

Internship Completion Certificate

This is to certify that **Nithin S.V. (Roll No.:1833031)** pursuing M.Sc. (Decision and Computing Sciences) at Coimbatore Institute of Technology, Coimbatore has successfully completed his internship under me for a duration of 4 months from 01 September 2021 to 30 December 2021. During the period of internship, he was sincere and dedicated towards the task assigned to him. The title of the project is **“CAR INSURANCE CLAIM FRAUD DETECTION”**.

Wish him all success in his future endeavors.



Dr.E.Chandra Blessie
Supervisor

Dr. E.Chandra Blessie, MCA., M.Phil, Ph.D
Assistant Professor
Department of Computing
Coimbatore Institute of Technology
Coimbatore - 641 014.



ZILA (G iscle System Pvt.Ltd.)
Video shopping app.

TO WHOM IT MAY CONCERN

Date: 10-12-2021

This is to certify that Naveenkumar has successfully completed 6 months of Internship with the title ZILA Shop Management under the guidance of Suman Kumar Jha with ZILA app.

The duration of his/her employment was from 8th June 2021 to 8th December 2021.

During this internship he/she has shown exceptional skills and contributed towards the growth of the company.

We wish you all the best for the future.



Suman Kumar Jha
Fonder

H No.: 46, VPO: Dasaut, Bhramantala,
Singhia, Samastipur, Bihar, 848209
info@giscle.com



Kitabeli Pte Ltd
160 Robinson Road
#24-09
SINGAPORE (068914)

Date: 27th Jan, 2022

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. Nareshkumar has completed his internship with KitaBeli Pte Ltd for 8 (Eight) month from June 1st, 2021 to January 28th, 2022 in Fraud Analysis and Detection Project.

During his internship period we found him to be sincere and hard-working and we wish him all the best for his future endeavors.

Your Sincerely

Prateek Chaturvedi

Co Founder

Striim Engineering Services India Private Limited

CIN: U72200TN2017FTC119970

R/o: Spaces Olympia Tech Park, 10TH Floor, Citius Block No.-1, SIDCO Industrial Estate,
Guindy, Chennai- 600032

Mobile : +91 9789989054, Landline : 044-6145 9013 , Email : maha@striim.com

17th Dec 2021

MRITHIKA SARAVANABABU

7/19B opp C S I colony,

Kottaipalayam,

Coimbatore - 641110

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Ms Mrithika Saravanababu, student from Coimbatore Institute of Technology has completed her Internship as a Software intern in our organization during the period 1st June 2021 to 30th November 2021.

We found her sincere and result-oriented; and wish her success in her career.

Best Regards,

DocuSigned by:
Mahadevan Lakshminarayanan

738A7EC85A7A4E9

Mahadevan Lakshminarayanan Vice President, India Head.



Date: 20-Dec-21

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Mr. Mohindran S R** has completed his Internship with VerSe Innovation Private Limited from **June 21, 2021 to Nov 30, 2021**.

He was part of the DH – Engineering under the mentorship of Mr. Chandranath Patra.

His project details as below:

Full stack development and implementation of internal tool for deployment notifications; and improvements and development of new features for H2I module for election results display.

During his internship period we found him to be sincere & hard working & we wish him all the best for his future endeavors.

For VerSe Innovation Private Limited

Bhumika Khona
Assistant Manager – Human Resources



December 20, 2021

INTERNSHIP CERTIFICATE

To whomever it may concern

This is to certify that Mohammed Ismail A worked under my supervision during his internship period at Cornerstohn Solutions during the period June 7, 2021 to December 7, 2021. His internship was focused on Early Detection of Autism Spectrum Disorder in Children.

Mohammed Ismail possesses a good moral character, pleasing personality, attention to detail and the intent to complete his work on time. I wish him success for his future endeavors.

Best Wishes,

Kirthi Murthiraj

Cornerstohn Solutions



Befable LLP
befable.com
hello@befable.com

December 18, 2021

To,
Megavarniya

Subject: Completion of Internship

This is to certify that **Megavarniya** has successfully completed the internship with Befable LLP as a **Social Media Marketing Intern** to Create and implement a digital marketing plan to increase users, from June 21st, 2021 to December 18th, 2021.

During the course of her internship, she was sincere, hard working, target-oriented and a good team player and proved to be an overall asset to the company.

We sincerely wish her the best in all her future endeavors and a successful career ahead.

Thank you

A handwritten signature in black ink, appearing to read "Ashutosh".

Sincerely,
Ashutosh Mujumdar
Founder

Certificate of Internship

To whomever it may concern

This is to certify that KOSAL RAM G (Reg. No: 1833024) from Coimbatore Institute of Technology pursuing M.Sc. Decision and Computing Sciences fourth year has successfully completed the internship program from July-2021 to December-2021 at Assent Technologies

During the internship, KOSAL RAM G has worked on the project Automatic Number Plate Recognition.

During the internship, KOSAL RAM G was a quick learner and a great team player with an eagerness to contribute. We wish him good luck for all the future endeavours.

Warm Regards,
From Assent Technologies,



N. Rajesh Kannan

Manager - Human Resource

CERTIFICATE



Date: 30.11.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Ms. Keerthika J** student of **Coimbatore Institute of Technology** has successfully finished her internship in "**Digital Marketing**". During the period from 02nd June 2021 to 30th November 2021 at **Bright Bridge InfoTech Pvt Ltd**, Coimbatore.

During the period of her internship with us, she was found punctual, hardworking, and inquisitive.

We wish her all the best in future endeavors.

For Bright Bridge InfoTech Pvt. Ltd.,

Jeffery Jones P

Jeffery Jones P

Senior HR Executive

jeffery@brightbridge.co

Registered Office

No.17, Shriram Vijaya Hyyde park, Duraisamy Layout, Peelamedu, Coimbatore-641004.



MulticoreWare India (P) Ltd Sixth Floor, Block 9A

DLF IT Park, Manapakkam Chennai – 600 089

www.multicorewareinc.com

15 Dec 2021

To whomever it may concern

This is to certify Kavya Ponnusami doing M.Sc - Decision and Computing Sciences at Coimbatore Institute of Technology, Coimbatore has completed her internship for a period of six months effective from Jun 15, 2021 to Dec 15, 2021. She was reporting to Benuraj Sharma on a project titled "Voice gender recognition & Google meet add-on extension for Gesture control and Emotion Analysis" at Multicoreware India (P) Ltd.,

We wish her all success in her future endeavors.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Shashikanth Jayaraman'.

Shashikanth Jayaraman

Vice President - Human Resources

S.R. BATLIBOI & ASSOCIATES LLP

S. R. Batliboi & Associates LLP
6th Floor – "A" Block
Tidel Park
No. 4, Rajiv Gandhi Salai Taramani
Chennai – 600 113, India
Phone: +91 44 6632 8000
Fax: +91 44 2254 0120

28 January 2022

Kavya B
Coimbatore Institute of Technology, Coimbatore.

To Whomsoever It May Concern

This is to certify that **Ms. Kavya B (Register No: 1833021)**, student of MSc Decision and Computing Science from **Coimbatore Institute of Technology**, has done her Project Internship at *S. R. Batliboi & Associates LLP*.

Duration: 26 July 2021 – 28 January 2022

Her project title is "**Audit Analytics Automation with Alteryx and Power BI**"

During her internship period, it is found that her work is committed and good.

We wish her all the best for her future endeavors.

For *S. R. Batliboi & Associates LLP*,



VIKRAM KUMAR - D
DIRECTOR, ASSURANCE ANALYTICS

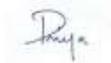
CERTIFICATE

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. Kanishk P (1833020) a student of M.Sc Decision and Computing Sciences at Coimbatore Institute Of Technology had successfully completed his internship titled "**Snipzz - Personalized News Application**" at Red Ant Media LLP from **May 26th, 2021 to January 31st, 2022**

We wish him the best in his future endeavors.

For Red Ant Media,



Priya
Supervisor



TO WHOM IT MAY CONCERN

10th December 2021

This is to certify that **Mr. Javagar M** has successfully completed his Internship program with **OptiSol Business Solutions Private Limited** located at Baid Hi-Tech Park, Thiruvanniyur, Chennai. His internship tenure was from **01st June 2021** to **10th December 2021**.

His performance during this internship has been exemplary. He showed good aptitude and motivation to learn new technology stacks and had been a great addition to our team. We wish him all the very best in his future endeavours.

For OptiSol Business Solutions Private Limited

A handwritten signature in black ink, appearing to read "Antony Justine D Julian".

Antony Justine D Julian

Senior Manager - Human Resources

OptiSol Business Solutions Pvt.Ltd.,

Regd.Office : #1295, Baid Hi Tech Park, 4th & 5th Floor, ECR, Thiruvanniyur, Chennai-600041, TN, India.Ph:+91-44-24512296
Email : www.optisolbusiness.com

Web : www.optisolbusiness.com

CERTIFICATE



SARA INNOVATIONS

Office Address: Siva Nagar, Ponnusamy Nagar, Coimbatore, Tamil Nadu 641007.
Phone No: 9842486037 E-mail: sivaraghavi91@gmail.com Website: www.sarainnovations.in

Ref: SIS-000311669-SCP

Date: 29/11/2021

CERTIFICATE OF COMPLETION

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. **HARIRAMAKRISHNAN S**, student of Coimbatore Institute of Technology has successfully completed the internship with **Sara Innovations** as a **Web Developer** and worked on **Educational websites**, from April 8th, 2021 to November 30th, 2021.

During the course of his internship he demonstrated a self-motivated attitude to learn new things and his performance exceeded expectations.

We wish him the best in all his future endeavors.

For, **SARA INNOVATIONS**.

Sincerely,

Siva Raghavi, Founder.

CERTIFICATE



123coimbatore.com
79,3rd Floor, Atawarya Complex,
Nehruji Road, P.N.Palayam,
Coimbatore-641037.
Tele : +91 0422 435 0451

To Whomsoever It May Concern

This is to certify that **Miss. Harini C (Reg No: 1833017)**, student at Coimbatore Institute of Technology Pursuing her "M.Sc (Decision and Computing Science)" has successfully completed her Internship Training program at **123Coimbatore and Project entitled as "SEO/ DIGITAL MARKETING"** The tenure of her Internship is from **July 2021 to Dec 2021**.

We at **123Coimbatore** wish her all success in her future endeavors.

For **123COIMBATORE**

Name: **Monika**
(Manager) Human Resource

Website : www.123coimbatore.com

E-mail : info@123coimbatore.com

CONTENTS (For Management Projects)

CHAPTER	PAGE NO
ACKNOWLEDGEMENT	i

CERTIFICATE



Date: 30.11.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Ms. Hari Priya H** student of **Coimbatore Institute of Technology** has successfully finished her internship in "**Digital Marketing**". During the period from 02nd June 2021 to 30th November 2021 at **Bright Bridge InfoTech Pvt Ltd**, Coimbatore.

During the period of her internship with us, she was found punctual, hardworking, and inquisitive.

We wish her all the best in future endeavors.

For Bright Bridge InfoTech Pvt. Ltd.,

Jeffery Jones P

Jeffery Jones P

Senior HR Executive

jeffery@brightbridge.co

Registered Office

No.17, Shriram Vijaya Hyyde park, Duraisamy Layout, Peelamedu, Coimbatore-641004.



Certificate of Completion

This is to certify that

Guhanesvar Manikandan

has successfully completed the

Buckman Internship Program

at Buckman Insights Lab, Chennai from 24-May 2021 to 30-November 2021.

His research project was focused on Data Analytics.

His performance during the period was Outstanding.

L. Divagar

Dr. Divagar Lakshmanan
Digital Innovation Manager

Narasimha Rao

Dr. Narasimha Rao
Chief Digital Officer

CERTIFICATE:



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Gokulavasan

to certify that he/she has completed Intensive internship
on

**Dissecting Insider Threat with Data Science for behavioural
Analytics and Predictive modelling**

During the period
June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION
A Not-For-Profit Entity
<https://Sycurislabs.org>

1st Feb 2022

PROJECT COMPLETION CERTIFICATE

This is to certify that the following student of COMIBATORE INSTITUTE OF TECHNOLOGY, M.SC. DECISION AND COMPUTING SCIENCES, 4TH YEAR, 7TH SEMESTER has completed his internship and project at DeepVisionTech.AI.

1. DINESH KUMAR T (1833012)

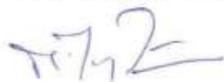
Project Title: SIGN LANGUAGE DETECTION USING COMPUTER VISION

Project / internship Duration: 01.10.2021 to 31.01.2022

During his internship / project work at our organization, we found him extremely committed and diligent. He took ownership of the entire project, learnt several new technologies & techniques in AI like Machine Learning and Deep Learning, tried various options and delivered optimum solution.

His association with us was very fruitful and we wish him **ALL THE BEST** in his future endeavors.

For DeepVisionTech.AI,



(Mr. Jayasudan Munsamy)

Founder & CEO at DeepVisionTech.AI

DeepVisionTech is Associated with:



INCEPTION
PROGRAM

ORACLE
for Startups





Curneu MedTech Innovations Private Limited

Dated: 10th December 2021

TO WHOM IT MAY CONCERN

This is to certify that **Mr. Dhanush Kannan A** has done his internship with **Curneu MedTech Innovations Private Limited** as **Junior Data Science Developer** dated from **June 1, 2021 to November 30, 2021**.

He has worked on a project titled **Software Development in Neurosurgical Simulator** during the internship he has shown full Sincerity, Dedication and Hard work towards his concerned job, which has helped in improving the management of the company.

We wish him all the best for his future endeavours.

For Curneu MedTech Innovation Pvt Ltd.

A handwritten signature in black ink, appearing to read "David Roshan Cyril", with a stylized flourish at the end.

David Roshan Cyril

CEO

Curneu MedTech Innovations Private Limited, C-2 Kurichi Housing Unit, Phase 1, SIDCO Post, Coimbatore - 641021

PAN: AAHCC4229K TAN: CMBC06382F

Contact Number: +91 97157 07807 / +91 82177 87644



CERTIFICATE
OF ACHIEVEMENT

THIS ACHIEVEMENT IS PRESENTED TO

R. DEETCHIKA

for Successful completion of Internship on
IOT CYBER SECURITY USING MACHINE LEARNING

Mamta Verma
(Dr. Mamta Verma)
Director
Sycuris Lab Foundation

From

Dr. Sivasankar P.
Assistant Professor,
Department of Petroleum Engineering and Earth Sciences,
Indian Institute of Petroleum and Energy,
Visakhapatnam – 53003, A.P.

To

The Head
Department of Computing Sciences
Coimbatore Institute of Technology,
Coimbatore – 640014, Tamil Nadu.

Sub: Internship completion certificate for Mr. C R Deepak Kumar (1833009), 7th Semester of Integrated M.Sc. Decision and Computing Sciences, Coimbatore Institute of Technology, Coimbatore.

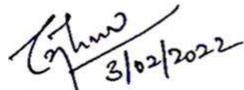
Dear Madam,

Warm greetings!

This letter is to certify that C R DEEPAK KUMAR has successfully completed his Internship under my guidance as a Research Intern at Indian Institute of Petroleum and Energy, Visakhapatnam. He worked from June 1st to December 23rd, 2021.

During his internship, he worked on the project titled "Comparing Machine Learning Algorithms for Predicting Transport of Species in Porous Media using Physics Informed Machine Learning Approach". I found that Mr. C R Deepak Kumar was sincere and dedicated in his duties. For this project, he contributed in obtaining the data, analysis of the data, integrating machine learning algorithms with physics based equations used for species transport in porous media and in producing the expected results. His knowledge in Machine Learning is exceptional and he has the potential to become a well-known expert in this domain, if he could continue with this effort. I wish him all the success for his future endeavours.

With best regards,


3/02/2022

Dr. Sivasankar P
Dr. Sivasankar P
Assistant Professor
Petroleum Engineering
Indian Institute of Petroleum & Energy



Sycuris
Lab Foundation

Certificate of Completion

is hereby granted to

Chithirai jothi U

to certify that he/she has completed Intensive internship
on

**NETWORK INTRUSION DETECTION AND CLASSIFICATION
USING MACHINE LEARNING**

During the period

June - December 2021

Mamta Verma

Dr. Mamta Verma
Director

SYCURIS LAB FOUNDATION

A Not-For-Profit Entity

<https://Sycurislabs.org>

www.thorogood.com
info@thorogood.com

Bangalore
ACR Towers
2nd Floor
32 Residency Road
Bangalore 560 025
Tel: +91 (0)80 4123 3700
Fax: +91 (0)80 4123 3800

London
Tel: +44 (0) 20 8231 0800
Philadelphia
Tel: +1 877 905 5010
Singapore
Tel: +65 3138 8068
São Paulo
Tel: +55 134040 4150

-- February 2022
Bangalore

THOROGOOD®

This is to certify that Bobby Prathikshana M of 2018-2023, Integrated M.Sc. Decision and Computing Science, Coimbatore Institute of Technology, Coimbatore has successfully completed her 7th semester internship as a Business Intelligence and Analytics Consultant from June 1st, 2021 to January 31th, 2022 at Thorogood Associates India Private Limited, Bangalore. She worked on the project 'Trade Promotion and Sales Analytics for Consumer Packaged Goods Industry. She was very sharp in picking up technologies like Azure Data Factory, Power BI, Python, MLflow, Databricks etc. She played a significant role in the project by analyzing the customer requirement, developing the technical components and delivering deliverables on time. Her contribution was well appreciated by the customer.



Afrid Sheik,
Business Intelligence and Analytics Consultant,
Thorogood Associates India Private Ltd,
Bangalore

Thorogood Associates India Pvt Limited

Registered Number: U72900KA2004PTC070740

Registered Address: Unit 201 ACR Towers, No 32 Residency Road, Bangalore, Karnataka 560025

Country of Registration: India

www.thorogood.com

info@thorogood.com

Bangalore

ACR Towers
2nd Floor
32 Residency Road
Bangalore 560 025
Tel: +91 (0)80 4123
Fax: +91 (0)80 4123

London

Tel: +44 (0) 20
8231 0800

Philadelphia

Tel: +1 877 905
5010

Singapore

Tel: +65 3138 8068

04 February 2022

Bangalore

THOROGOOD®

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Barath Vignesh S (1833005)** of 2018-2023, Integrated M.Sc. Decision and Computing Science, Coimbatore Institute of Technology, Coimbatore has successfully completed his 7th semester internship as a Business Intelligence and Analytics Consultant from June 1st, 2021 to January 31st, 2022 at Thorogood Associates India Private Limited, Bangalore. He worked on the project **Business Analytics and Reporting on CPG Product** under the guidance of **Ms. Akshaya Sivanantham**, Business Intelligence and Analytics Consultant. He was very sharp in picking up technologies like Angular, .NET Core, DevOps, Power BI etc. He played a significant role in the project by analyzing the customer requirement, developing the technical components and delivering deliverables on time. His contribution was well appreciated by the customer.



Ms. Akshaya Sivanantham,
Business Intelligence and Analytics Consultant,
Thorogood Associates India Private Ltd,
Bangalore

Thorogood Associates India Pvt Limited

Registered Number: U72900KA2004PTC070740

Registered Address: Unit 201 ACR Towers, No 32 Residency Road, Bangalore, Karnataka 560025

Country of Registration: India

CoverNest Insurance Web Aggregator Private Limited

1st Floor, No – 6/20, Bharathiar Street, Vanuvampet, Chennai – 600091.
Ph – 044 22603661, Mail – info@covernest.com, www.covernest.com

30 December 2021

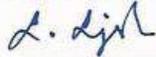
Internship Completion Certificate

This is to certify that **Akkshaya Sri J (Roll No - 1833002)**, pursuing her pre-final year M.Sc Decision and Computing Sciences (5 years Integrated) at Coimbatore Institute of Technology, Coimbatore has successfully completed her internship for a duration of 6 months from 14 June 2021 to 30 November 2021 at our organization. During the period of internship, she was found hard working, professional and has shown good commitment towards the task assigned to her. Below are her project specifications:

Project Title: Web Portal Development for Vehicle Insurance

We wish her all success in her future endeavors.

For CoverNest Insurance Web Aggregator Pvt Ltd,



Rajasekaran R G,
Managing Director





Date: 7th February 2022

Subject: Internship Certificate

To whomever it may concern

This is to certify that Mr. Aadhithian L from Coimbatore Institute of Technology has successfully completed his internship at Inpixon as a Software Intern for the period of six months from July 2021 – December 2021.

He worked under the direct supervision of his mentor Mr. Amuthukkiniyavel and got the opportunity to work on various projects. Along with his other duties, he was responsible for managing the timelines of the projects he was working on. His performance exceeded our expectations, and he was able to complete the project on time. We wish him all the best for his upcoming career.

For Inpixon India Limited

A handwritten signature in black ink, appearing to read "Shirish Tangirala".

**Shirish Tangirala
Member of the Board and Authorized Signatory
Inpixon India Limited**

Coimbatore Institute of Technology, Coimbatore - 14
Department of Computing - DCS
2017 Batch DCS - 17MDC101 - Project Work and Viva Voce II - 2021-22 Even

S. No	Reg. No	Name	Title of Project
1	1733001	Abinaya K	Digital Analytics
2	1733002	S.Agaramudhalvan	Crop Fertigation recommendation based on Soil Nutrient Value
3	1733003	ASHWIN KUMAR.P.	QUALDO-MQX
4	1733004	Christopher Immanuel	Credit Scoring Model of claims in Healthcare Insurance
5	1733006	Gowtham E	Eye-Gaze Product Tracking System
6	1733007	Harini Shre S	Bank Transaction Validation using Checks and Question Answering System
7	1733008	Harish Gowtham S	Unified Data Collection and Analytics Framework for Trimble applications and services
8	1733010	Jeevika G	Customer Sentiment Analysis
9	1733011	Karthikeyan	Freight Rate Sheet Processing and Forecasting
10	1733012	Khishore V S	Precision farming
11	1733013	LAKSHMAN C V	Automation and Capacity Analysis
12	1733014	MEHA.R	Recommendation of Most Valued Stocks
13	1733015	Padmavarshini C V	Cheque Digitization
14	1733016	Pavithra S	Migration and Enhancement of Population Health Monitoring and Tracking Report
15	1733017	Pradeep Kumar S	Validation Modules in Distributed Database
16	1733018	Priyadharshini N	Audit Analytics of Financial Data
17	1733019	Ragaavi D	Predicting Windmill Power Generation Using Regression Algorithm
18	1733020	Ragul M	OneSelf-iOS Application
19	1733021	Rithika Raj	Augmentation using AI
20	1733022	Sadhana N	Automated bot for website testing and UI validation
21	1733023	Sandhru D	Identity Resolution using ML Match
22	1733024	Sanjula K R	Claim analysis automation and reporting in healthcare insurance
23	1733025	Sowmya D	Report Generation for Tobacco Selling Company (Module 1)
24	1733026	Sri Dhanuja R	Authentication, Authorization and Auditing (AAA) System
25	1733027	Sri dhurkesh S	Utility Pole Maintenance and Automation
26	1733028	Sri Hari K V	Contactless KYC Process Automation
27	1733029	Srinandhini	Automated medical record audit using Natural Language Processing
28	1733030	Sridharan	Digest Email - A Feature for Collabrr
29	1733031	Srinidhi D	Model Development Process Streamlining
30	1733032	Sriram R M	CI/CD pipeline for IaC with Best Practices
31	1733033	Srivignesh S	Pariksha - Cloud Enabled Hiring Platform
32	1733034	Sussma S	The Time Tracker for Workforce
33	1733035	Suyamprakash RM	Artificial Code Summarizer
34	1733036	Swethaa Shri S	Sat2Farm Data Management System
35	1733037	Thenmalar S P	Automatic categorization of articles
36	1733038	Varshini S	Automated bots for web scrapping
37	1733039	Vijay M	OptiScan - Barcode Detection and Enhancement
38	1733040	Yaswanthan A	Resume builder



Infarmsys Technologies Private Limited.
#1, NABARD's Madurai Agribusiness Incubation Forum
Agricultural College and Research Institute,
Madurai - 625104
Ph. No: 0452 - 2424550

INTERNSHIP CERTIFICATE
TO WHOMSOEVER IT MAY CONCERN

THIS IS TO CERTIFY THAT **MR. S. AGARAMUDHALVAN (1733002)**, STUDENT OF **COIMBATORE INSTITUTE OF TECHNOLOGY** STUDYING **M.SC DECISION AND COMPUTING SCIENCES**, IS PURSUING HIS INTERNSHIP PROGRAMME ON THE PROJECT **CROP FERTIGATION RECOMMENDATION BASED ON SOIL NUTRIENTS** .

HE IS DOING HIS PROJECT TO OUR SATISFACTION UNDER OUR GUIDANCE. DURING THIS PERIOD OF INTERNSHIP HIS CONDUCT IS FOUND TO BE SATISFACTORY AND WE APPRECIATE HIM FOR THIS CONTRIBUTION. WE WISH HIM SUCCESS IN ALL THE FUTURE ENDEAVORS.

Warm Regards

A handwritten signature in blue ink, appearing to read "Ramasamy Esvar".

Ramasamy Esvar

Director,

Infarmsys Technologies Private Limited

Date: 11-May-2022

TO WHOMSOEVER IT MAY CONCERN

This letter is to confirm that **Ms. Ragaavi D** has done her internship in Ganit Business Solutions Private Limited, Chennai from 21st Feb 22.

During her internship she worked on a project named "**Predicting Windmill Power Generation using regression algorithm**". She demonstrated good skills with a self-motivated attitude to learn new things. Her performance was good

Wish you the best of your career with us.

Yours Sincerely,

For **Ganit Business Solutions Private Limited**



Vikneshwar J
Project Lead



16 May 2022

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Khishore V S** is currently pursuing his internship at **Trimble Information Technologies India Private Limited- Chennai** with the project titled **Precision Farming** from **Feb 21, 2022**.

For Trimble Information Technologies India Private Limited



Sheeja Mathews

Regional People Operations Manager - APAC

CIN: U72300TN1999PTC043647

Registered Office: Lakshmi Tech Park - No.5/639, Old Mahabalipuram Road, Kandanchavadi, Perungudi, Chennai – 600096- India
Tel: +91 44 2254 1941 / 43 / 61 / 63 - Fax: +91 44 2254 1942 - www.trimble.com

Branch Offices: Lakshmi Tech Park - No.5/639, Old Mahabalipuram Road, Kandanchavadi, Perungudi, Chennai – 600096.
Vikram Monarch, CTS No. 1115-A/1, 10th floor, GaneshKhind Road, Shivajinagar, Pune – 411 016, Maharashtra, India. Tel: +91 20 4917 4900
1st Floor, No.23, Sankey Road, Lower Palace Orchards, Sadashivanagar, Bangalore 560 003, Karnataka, India. Tel: +91 80 23348088

OSP Consulting India Pvt Ltd

Manyata Embassy Tech Park,
Ground Floor, Beech, E-1,
Bangalore, Karnataka



To

Sadhana N

Intern Data Analyst

To whomsoever it may concern

Dear Sadhana,

This is to certify that Miss Sadhana has undergone an Internship at OSP Consulting India Pvt Ltd, Bangalore as an Intern- Data Analyst from 8 November 2021 to 25 April 2022.

During the internship period, her attendance & conduct was good.

She has a keen ability to learn and achieve objectives set as a part of her project requirements. As a result, Sadhana has successfully completed the internship Project as Data Analyst.

We thank you for your contribution to the company and wish the very best in all your future endeavours.

For OSP Consulting India Pvt Ltd.

A handwritten signature in blue ink, appearing to read "K.P. Radhakrishnan Nambiar".

K.P. Radhakrishnan Nambiar

Director

18 May 2022

Bangalore

OSP Consulting India Pvt Ltd

Manyata Embassy Tech Park,
Ground Floor, Beech, E-1,
Bangalore, Karnataka



To

Varshini S

Intern Data Analyst

To whomsoever it may concern

Dear Varshini,

This is to certify that Ms. Varshini has undergone an Internship at OSP Consulting India Pvt Ltd, Bangalore as an Intern- Data Analyst from 8 November 2021 to 7 May 2022.

During the internship period, her attendance & conduct was good.

She has a keen ability to learn and achieve objectives set as a part of her project requirements. As a result, Varshini has successfully completed the internship Project as Data Analyst.

We thank you for your contribution to the company and wish the very best in all your future endeavours.

For OSP Consulting India Pvt Ltd.

A handwritten signature in blue ink, appearing to read "K.P. Radhakrishnan Nambiar".

K.P. Radhakrishnan Nambiar

Director

17 May 2022

Bangalore



Zoho Corporation Private Limited

Plot 140, 151, Estancia IT Park, Vallancheri,
Chengalpattu District, Tamilnadu, 603 202.

Ph: +91 - 44 - 6744 7070

www.zohocorp.com

SEZ Unit

PL-1243/HR/2022

INTERNSHIP CERTIFICATE

This is to certify that Mr/Ms. **Hariharan T - PT-4816/22** has undergone his/her internship training in **Zoho Corporation Private Limited - Unit 1**, from 20-Dec-2021 to 16-May-2022. During this period, his/her performance and conduct were found to be good.

Yours Sincerely,

For Zoho Corporation Private Limited

Raji R

Rajarajeswari R

Senior Associate - HR

Date of issue:





Zoho Corporation Private Limited

Plot 140, 151, Estancia IT Park, Vallancheri,
Chengalpattu District, Tamilnadu, 603 202.

Ph: +91 - 44 - 6744 7070

www.zohocorp.com

SEZ Unit

PL-1243/HR/2022

INTERNSHIP CERTIFICATE

This is to certify that Mr/Ms. **Hariharan T - PT-4816/22** has undergone his/her internship training in **Zoho Corporation Private Limited - Unit 1**, from 20-Dec-2021 to 16-May-2022. During this period, his/her performance and conduct were found to be good.

Yours Sincerely,

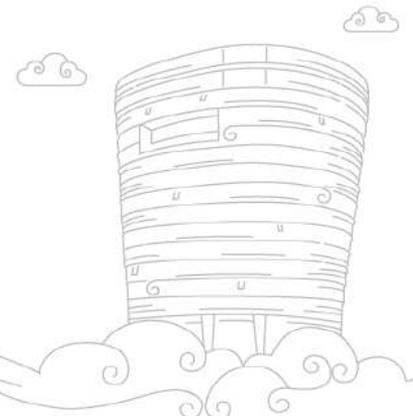
For Zoho Corporation Private Limited

Raji R

Rajarajeswari R

Senior Associate - HR

Date of issue:



12-May-2022

INTERNSHIP LETTER

This letter is to verify that MEHA.R M.Sc. Decision and computing sciences at Coimbatore Institute of Technology undergoing 6-months Internship at HSBC, Bangalore and working under Risk & Compliance, Global Risk Analytics of Wholesale Analytics.

Thanks & Regards,

G Alekhya

HR Executive– Campus Recruitment

Mobile: 9908764474

Email: g.alekhya@hsbc.co.in

URLs: Intranet: <http://www.hsbc.com> [Help on Careerlink](#)

HSBC East Campus; 1A, RMZ EcoSpace,Marathahalli - Sarjapur Outer Ring Rd, Devarabisanahalli, Bellandur, Bengaluru, Karnataka 560103, India.

INTERNAL

OSP Consulting India Pvt Ltd

Manyata Embassy Tech Park,
Ground Floor, Beech, E-1,
Bangalore, Karnataka



To

Varshini S

Intern Data Analyst

To whomsoever it may concern

Dear Varshini,

This is to certify that Ms. Varshini has undergone an Internship at OSP Consulting India Pvt Ltd, Bangalore as an Intern- Data Analyst from 8 November 2021 to 7 May 2022.

During the internship period, her attendance & conduct was good.

She has a keen ability to learn and achieve objectives set as a part of her project requirements. As a result, Varshini has successfully completed the internship Project as Data Analyst.

We thank you for your contribution to the company and wish the very best in all your future endeavours.

For OSP Consulting India Pvt Ltd.

A handwritten signature in blue ink, appearing to read "K.P. Radhakrishnan Nambiar".

K.P. Radhakrishnan Nambiar

Director

17 May 2022

Bangalore

OSP Consulting India Pvt Ltd

Manyata Embassy Tech Park,
Ground Floor, Beech, E-1,
Bangalore, Karnataka



To

Sadhana N

Intern Data Analyst

To whomsoever it may concern

Dear Sadhana,

This is to certify that Miss Sadhana has undergone an Internship at OSP Consulting India Pvt Ltd, Bangalore as an Intern- Data Analyst from 8 November 2021 to 25 April 2022.

During the internship period, her attendance & conduct was good.

She has a keen ability to learn and achieve objectives set as a part of her project requirements. As a result, Sadhana has successfully completed the internship Project as Data Analyst.

We thank you for your contribution to the company and wish the very best in all your future endeavours.

For OSP Consulting India Pvt Ltd.

A handwritten signature in blue ink, appearing to read "K.P. Radhakrishnan Nambiar".

K.P. Radhakrishnan Nambiar

Director

18 May 2022

Bangalore

May 16, 2022

To Whomever it may concern

This is to certify that **Miss. Padma Varshini C V** is pursuing her Internship in our organization from **20th December 2021**. She is hard working and her performance has been satisfactory.

For SCIOInspire Consulting Services (India) Private Limited



Payal N Chhabria

Assistant Vice President – Human Resource

MIGRATION AND ENHANCEMENT OF POPULATION HEALTH MONITORING AND TRACKING REPORT

PAVITHRA S
Register No: 1733016

DISSERTATION SUBMITTED IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
M.Sc. (Decision and Computing Sciences)
OF ANNA UNIVERSITY



MAY 2022

**DEPARTMENT OF COMPUTING
COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution affiliated to Anna University)
COIMBATORE – 641014**

COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution Affiliated to Anna University)
COIMBATORE 641014

(Bonafide Certificate)

Project Work – II

Tenth Semester

MIGRATION AND ENHANCEMENT –
POPULATION HEALTH MONITORING AND TRACKING REPORT

Bonafide record of work done by

PAVITHRA S

(Register No: 1733016)

Submitted in partial fulfilment of the
requirements for the degree of
M.Sc. (Decision and Computing Sciences)
of Anna University

MAY 2022

Faculty Guide

Head of the Department

Submitted for the viva-voce held on _____

Internal Examiner

External Examiner



Date: 12-05-2022

This is to certify that **Ms. Pavithra.S**, student of M.Sc. Decision and Computing Sciences (Integrated) from “**Coimbatore Institute of Technology**” has been working with **EXL Service.com (India) Private Limited** as an **Intern** for the period of 13th December 2021 to 13th June 2022. She has undergone academic training program at our organization as part of completion of requirement for the award of M.Sc. She is working on project work entitled “**Migration and Enhancement of Population Health Monitoring / Tracking Report**” during the said duration.

For, EXL Service.com (India) Private Limited

A handwritten signature in black ink, appearing to read 'Vikram Pragadeesh'.

Vikram Pragadeesh
Assistant Vice President – Advanced Analytics / Data Science, EXL Health

CONTENTS

CHAPTER	PAGE NO
ACKNOWLEDGEMENT	i
SYNOPSIS	ii
PREFACE	iii
I INTRODUCTION	1
1.1 ORGANIZATION PROFILE	1
1.2 PROBLEM DEFINITION	2
1.3 SYSTEM ENVIRONMENT	2
II SYSTEM ANALYSIS	7
2.1 SYSTEM DESCRIPTION	7
2.2 USE CASE MODEL	11
2.3 SOFTWARE REQUIREMENTS SPECIFICATION	13
2.4 TEST PLAN	20
III SYSTEM DESIGN	22
3.1 ARCHITECTURAL DESIGN	22
3.2 STRUCTURAL DESIGN	26
3.3 BEHAVIOURAL DESIGN	29
3.4 CODE DESIGN	31
IV SYSTEM TESTING	35
4.1 TEST CASES AND TEST REPORTS	37
V SYSTEM IMPLEMENTATION	38
VI CONCLUSION	39
BIBLIOGRAPHY	

ACKNOWLEDGEMENT

Apart from the efforts of myself, the success of this project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

I am extremely grateful to **Dr. A. RAJESWARI**, Principal, Coimbatore Institute of Technology, for giving me this opportunity to do my project at EXL Service Private Limited, Chennai.

I am deeply indebted to **Dr. K. SHAKTHI MALA**, Dean of Computing, Coimbatore Institute of Technology, Coimbatore, for her continual support and ardent motivation.

I extend my deep gratitude to **Dr. A. KANNAMMAL**, Head, Department of Computing, Decision and Computing Sciences, Coimbatore Institute of Technology, for her continual support

I extend my heartfelt gratitude to my internal guide, **Ms. V. SHANTHI**, Assistant Professor, Software Systems, Department of Computing, Coimbatore Institute of Technology, for her continual support and enduring guidance throughout my project tenure

I express my deep sense of gratitude to **Mr. VIKRAM PRAGADEESH**, Senior Manager, Assistant Vice President – Advanced Analytics/Data Science, EXL Service Private Limited, Chennai, for his invaluable guidance, support and suggestions throughout the course of this project work

I wish to acknowledge all the other members in my team, who made the work challenging and were very helpful in growing my learning curve. Finally, I express my gratefulness to all the staff of Department of Computing, Coimbatore Institute of Technology, Coimbatore and my family and friends for their encouragement and support

SYNOPSIS

The project entitled **“Migration and Enhancement of Population Health Monitoring and Tracking Report”** aims to create a shift in care management’s focus from highly reactive care management to a more proactive approach that encompasses managing members across the risk spectrum, including those who are yet to become high risk. The objective is to build a monitoring report which consists of clinical metrics regarding the population, including the cost trend, utilizations, compliance to gaps in care or visits to the healthcare service provider/client. The existing system is built using SAS, which computes the components of the final dataset to be given as an input to the monitoring report. This system is found to have several drawbacks and is inefficient to be able to easily run for a prolonged period of time. Thus, it is migrated to PySpark, which has direct connection or ease of access to the big data ecosystem. Altogether, the new system focuses on prioritizing of care, to achieve maximum efficiency and outcomes from care management teams and suggest an intervention approach to mitigate against undesired health outcomes, increased costs and also overcoming the drawbacks of earlier version of the system.

The process of migration includes thorough analysis of the existing system and the logic of computations of clinical metrics, in order to come up with an equivalent efficient way to attain the results. The input data is unstructured and are from different sources, namely, HRA, Claims, Pharmacy data etc. They are put together in a member level format to perform a member level analysis to derive the clinical metrics and evaluating their reasonability using pyspark.

The system works on the basis of ‘technology closed loop framework’, which works closely with the Hadoop ecosystem. The access and manipulation of data is performed in here. The computations performed using PySpark; are uploaded to the ecosystem, there by saving the vital checkpoints, which would be useful in the future.

PREFACE

Chapter I - **INTRODUCTION AND ORGANIZATION PROFILE** – Specifies the purpose, the objective and the environment in which the project works

Chapter II – **SYSTEM ANALYSIS** – Describes the existing tool along with its limitations, proposed updates and its features, and the software requirements specifications

Chapter III – **SYSTEM DESIGN** - Elucidates the architectural, structural and behavioural design of the dsystem

Chapter IV – **SYSTEM TESTING** – Specifies the testing techniques used, with test scenarios

Chapter V – **SYSTEM IMPLEMENTATION** - Lays out the implementation of how the system is being implemented

Chapter VI - **CONCLUSION** – Puts forth the special features and suggestions for future enhancements of this project

BIBLIOGRAPHY

CHAPTER I

INTRODUCTION

This chapter contains the introduction about the organization, about the population health monitoring report and system specifications, under which the project was developed and worked on.

1.1 ORGANIZATION PROFILE

SCIO for more than 14 years, provides its healthcare partners actionable insights and scalable solutions through deep expertise, flexible analytics, innovative services and technology that drive measurable outcomes improvement. It focuses on the areas of population health, payment integrity, risk adjustment, care management, opportunity analysis, incentive design, consumer segmentation and engagement, network performance and commercial effectiveness. Each year actionable insights reach more than 150 million covered lives, 406 million medical claims, 1.3 billion prescription claims, and over 60,000 pharma dashboards with 300 unique data sources, refreshed monthly.

EXL's acquisition of SCIO in 2018 expands the expertise as a strategic digital transformation partner and deepens analytics, service and technology capabilities across the healthcare continuum.

TEAM PROFILE

Data and Analytics Services provides rich data assets to support client data needs, delivers end-to-end data-enabled solutions and offers enterprise data and analytics. The team helps clients identify, optimize, create, capture and grow business value by leveraging data assets, analytic capabilities, and data management expertise. Care Optimization (CO) in data and analytics services is responsible for the following services.

- **Population Health**

Identify and stratify current and emerging risk, target highest area of impact and propensity to engage, intervene early, fuel engagement and outreach programs.

- **Risk adjustment and Quality**

Support quality outcomes measures, documentation accuracy, and optimized revenue

- **Real world insights**

Deliver persona and patient-centred insights and outcomes-based contracting

1.2 PROBLEM DEFINITION

To migrate and enhance the population health monitoring report which provides key insights about the population to assist the care management analytics team of the health care service provider organization located in Michigan; also aid them in comparing performance against past and upcoming scenarios so that the current or potential problems can be identified, analysed and prevented.

1.3 SYSTEM ENVIRONMENT

The following are the hardware and software specifications that are used in the development of the project.

1.3.1 HARDWARE SPECIFICATION

Processor: Intel® Core™ i5-8500M CPU

Clock Speed: 2.40 GHz

Hard Disk: 500 GB

RAM: 8.00 GB 1.2.2

1.3.2 SOFTWARE SPECIFICATION

Operating System: Windows 10 PRO 64-bit OS

Tools: Zeppelin, Ambari, Jupyter Notebook

Database: Hive

Language: PySpark, Python, SAS

Tools: Ambari, Zeppelin, SAS, Hive

1.3.3 TOOLS AND TECHNOLOGIES USED

The technologies and tools that are used in the project and a brief description about each of them are discussed in the following subsections.

- **Python Language**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's elegant syntax and dynamic typing, together with its interpreted nature, makes it an ideal language for scripting and rapid application development in many areas on most platforms.

- **Jupyter Notebook**

The Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebooks documents.

Jupyter is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media, usually ending with the “.ipynb” extension. Jupyter Notebook can connect to many kernels, (by default Jupyter Notebook ships with the IPython kernel) to allow programming in many languages.

- **PySpark**

Apache Spark is a lightning-fast cluster computing technology, designed for fast computation. It is based on Hadoop MapReduce and it extends the MapReduce model to efficiently use it for more types of computations, which includes interactive queries and stream processing. The main feature of Spark is its in-memory cluster computing that increases the processing speed of an application. Spark is designed to cover a wide range of workloads such as batch applications, iterative algorithms, interactive queries and streaming. Apart from supporting all these workloads in a respective system, it reduces the management burden of maintaining separate tools. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs.

- **Zepplin**

Apache Zeppelin is a web-based notebook which brings data exploration, visualization, sharing and collaboration features to Spark. It supports Python, but also a growing list of programming languages such as Scala, Hive, SparkSQL, shell and markdown. Zeppelin provides a “Modern Data Science Studio” that supports Spark and Hive out of the box. It supports multiple language backends which has support for a growing ecosystem of data sources.

- **Hive**

Apache Hive is an open-source data warehouse system built on top of Hadoop. It is used for querying and analyzing large datasets stored in Hadoop files. Hive gives a SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data.

Hive provides the necessary SQL abstraction to integrate SQL-like queries (HiveQL) into the underlying Java without the need to implement queries in the low-level Java API.

- **Ambari**

Apache Ambari is an open-source administration tool deployed on top of Hadoop clusters, and it is responsible for keeping track of the running applications and their status. Apache Ambari can be referred to as a web-based management tool that manages, monitors, and provisions the health of Hadoop clusters.

It provides a highly interactive dashboard that allows administrators to visualize the progress and status of every application running over the Hadoop cluster. Its flexible and scalable user interface allows a range of tools such as Pig, MapReduce, Hive, etc. to be installed on the cluster and administers their performances in a user-friendly fashion.

- **SAS**

SAS (Statistical Analysis System) is a software system for data inspection and report writing.

Statistical Analysis System is a batch of programs that work together to reclaim them and to reserve data values, reform data, evaluate complex and single exponential analyses and generate reports. SAS is a powerful and flexible statistical package that runs on many platforms, including Windows and Unix. The SAS Programming involves first creating/reading the data sets into the memory and then doing the analysis on this data.

CHAPTER II

SYSTEM ANALYSIS

This section deals with the overall description, which involves the dissection of the system into its component pieces, to investigate further upon how these components interact and work.

2.1 SYSTEM DESCRIPTION

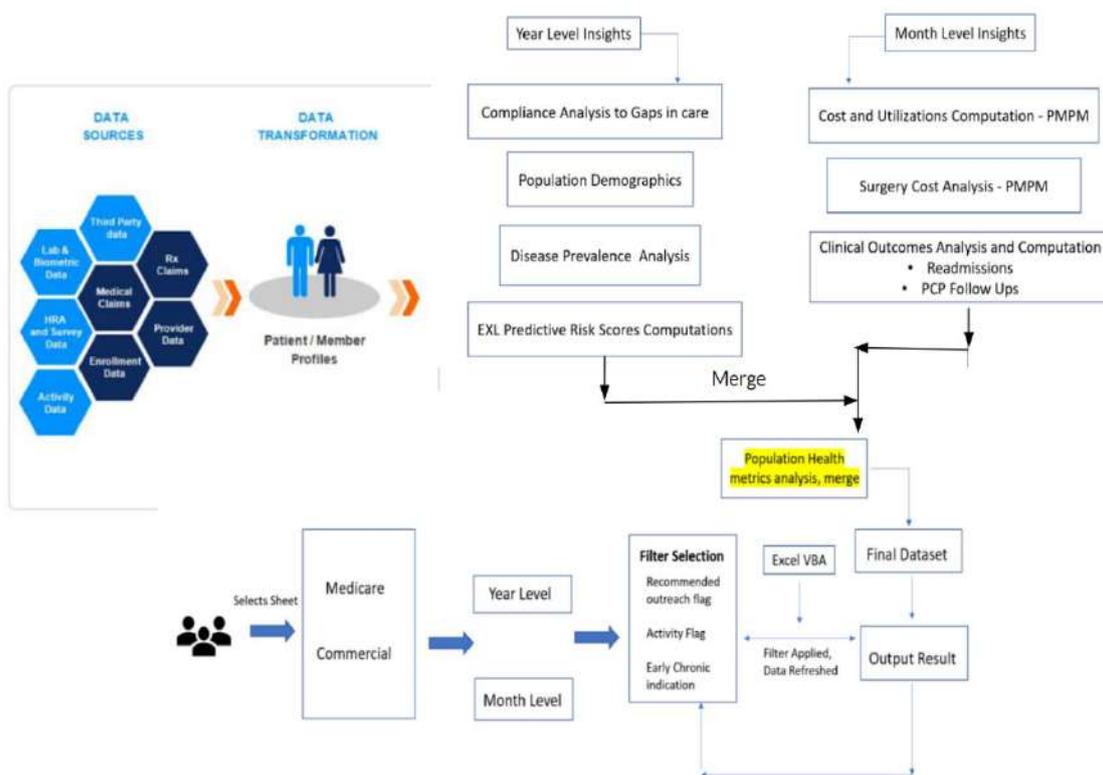


Figure 2.1 General Overview of Population Health Monitoring Report

As given in figure 2.1, the system mainly focuses on deriving the final input dataset for the Population health monitoring report. This dataset is the union of individual components which is eventually derived by member level analysis and computations.

Existing System Overview:

- The input data is fetched from the big data ecosystem / Hadoop
- The Individual analysis of metrics in the system is performed in SAS
- The user level view of the report as shown in figure 2.3 is followed earlier and currently as well
- The user makes a selection in the excel sheet, chooses the line of business
- Once the LOB (Line of Business) is chosen, the year or month level view sheet is selected
- The filters are applied in the filters section in the sheet
- Once the filters are selected, the data gets refreshed and applies to the filter of selection accordingly. This is achieved using the Excel VBA scripts, which when applied filters, gets triggered and fetches the output view

Drawbacks of the existing system:

- Drive Storage - Every computation made on the component of the system is stored locally, which takes up a lot of space
- Time Consuming - SAS by itself takes more time to perform computations
- Data Loss or Data tampering is more likely to take place

2.1.1 DATA SOURCES

In figure 2.1, the data to be used in the system is gathered together in the initial stage, which involves the pharmacy, lab, HRA (Health Risk Assessment) survey data, medical claims data, enrolment, provider and third-party data. All these data are used in a manner that the vital metrics from each are utilized for analysis and computations.

HRA data involves the Health Risk Assessment data, which is gathered together through a survey. This data allows clients to be proactive, rather than reactive. It identifies and predicts risk factors for preventable chronic conditions. Combined with readiness to change data, clients can pinpoint and prioritize the best programs and interventions for their populations.

Medical Claims Data gives an overview of the data of claims requested by patients. This data helps identify the patient's conditions and hospital visits. Similarly, the lab data and third party data is gathered for better analysis of the population health.

2.1.2 DATA TRANSFORMATION

The data is converted into member level profiles for a better patient centric approach of deliverable. It is known that almost 93 percent of the US population takes up claims to support their healthcare expenses. A member could take any number of claims and every claim is identified by a unique claim id. This member level information is used up for further analysis, having the primary key consideration as the member id.

2.1.3 ANALYTICS ENGINE

This section constitutes the Analysis performed on the data, involving evaluation of risk scores, such as the impactability, intervenability and clinical outcomes such as PCP follow ups and Readmissions.

The computations are done for both member and year level of insights. The existing system which is conventionally built using SAS, is converted into a system completely built using pyspark.

2.1.4 OVERVIEW ON THE DELIVERY

The report delivers the vital insights for the care management team, enabling to make decisions based on the same to prevent upcoming threats or scenarios with respect to patient health or engagements.

Care Management:

A Care Management team performs the set of activities intended to improve patient care and reduce the need for medical services by enhancing coordination of care, eliminate duplication and helping patients to effectively manage health conditions

Role of Care management in healthcare:

- Provide Quality Healthcare
- Patient Outreach and follow ups
- Providing Constant monitoring and support
- Patient-Centred Approach

Client Benefits from the Report:

Membership Insights

Informing what members truly need and how they need that care based on known social and clinical factors.

Care Management Efficiencies and Effectiveness

Focusing on the right members most likely to engage and improve health outcomes.

Risk and Cost Management

Moving down the risk funnel to mitigate potential future risk/costs.

Member Engagement & Health Outcomes

Creating far deeper member views inform effective intervention and engagement strategies.

Proactive Care Management

That addresses varying levels of member needs

2.2 USE CASE MODEL

The system use case model defines how the members engage and interact with the system. The usage of the system varies from the user to the developer. The way in which the user or client interacts with the system is different from the manner in which the developer or a member who builds the report interacts.

The use case model shown in figure 2.2, of the system gives an overview of vital tasks that is being performed by the members using the system. The data is prepared by the data management team and the vantage team, where in the data is prepared member level. The user selects or applies filter to the report, where the data gets refreshed according to the same. The testing is performed to the components of the system.

Ad Hoc testing is used to test the system, which doesn't require the system to be completely in order or to be testing ready. Since ad hoc testing is unstructured, there is no predefined order involved. Testing the component any point of development stage, gives more advantage over structured systematic testing.

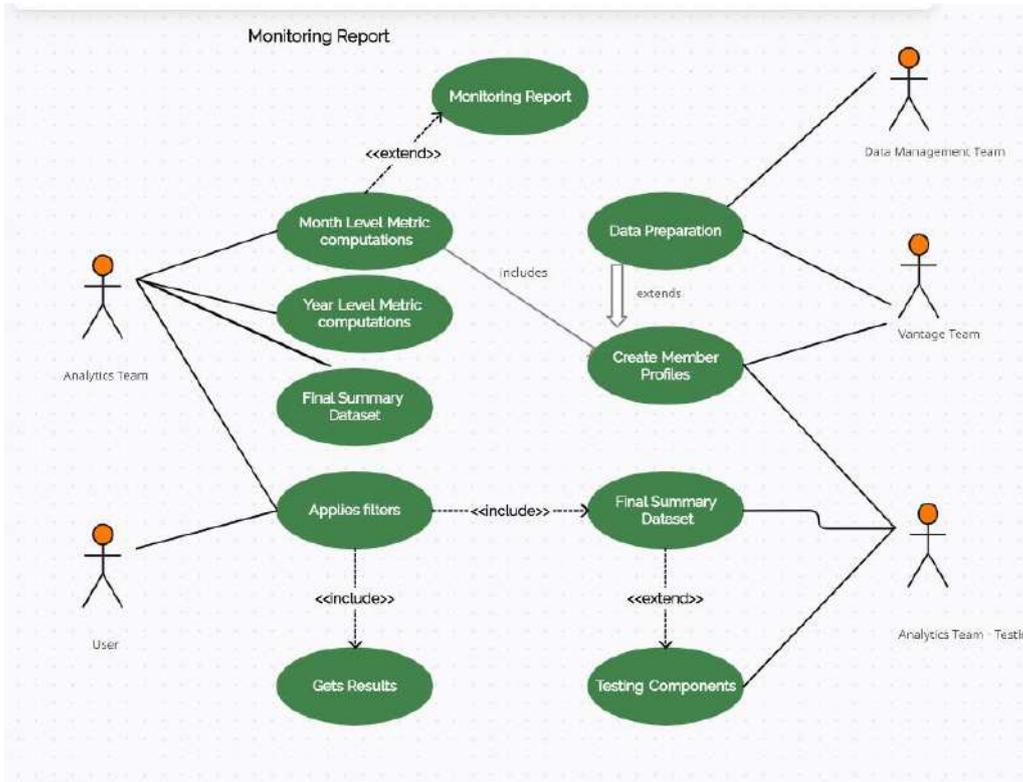


Figure 2.2 Use Case Diagram

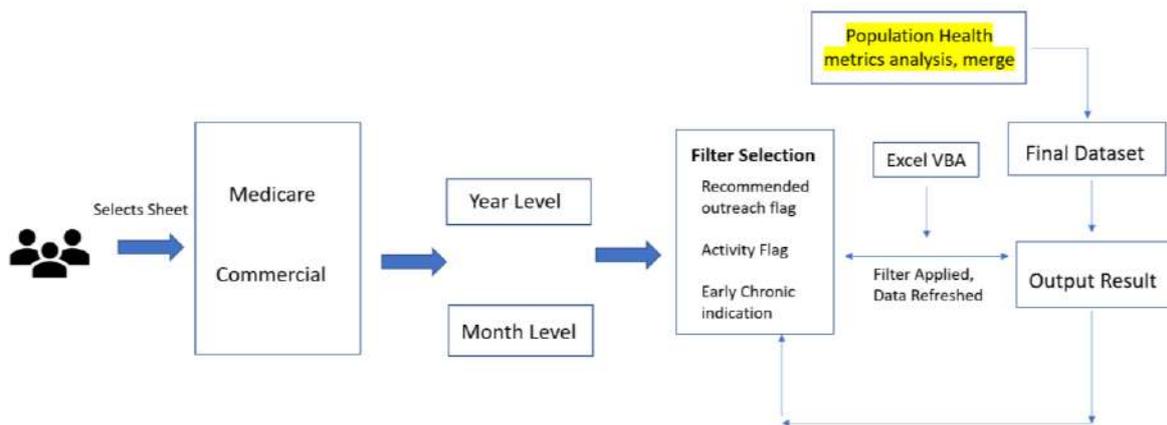


Figure 2.3 Client/User Interaction Flow

The above figure 2.3 shows the user interaction with the system. This user is the client who runs uses the monitoring report. There are two lines of business namely, Medicare and Commercial, which the user can select at the beginning stage. There are both year and month level of insights being provided for both the categories. The filters are applied to the report, which there by triggers the excel macro and calling the VBA function, to refresh and load the data according the filter applied.

2.3 SOFTWARE REQUIREMENT SPECIFICATION

This section focuses on the software requirement specifications, which covers the functional and non-functional requirements of the system

2.3.1 FUNCTIONAL REQUIREMENT SPECIFICATION

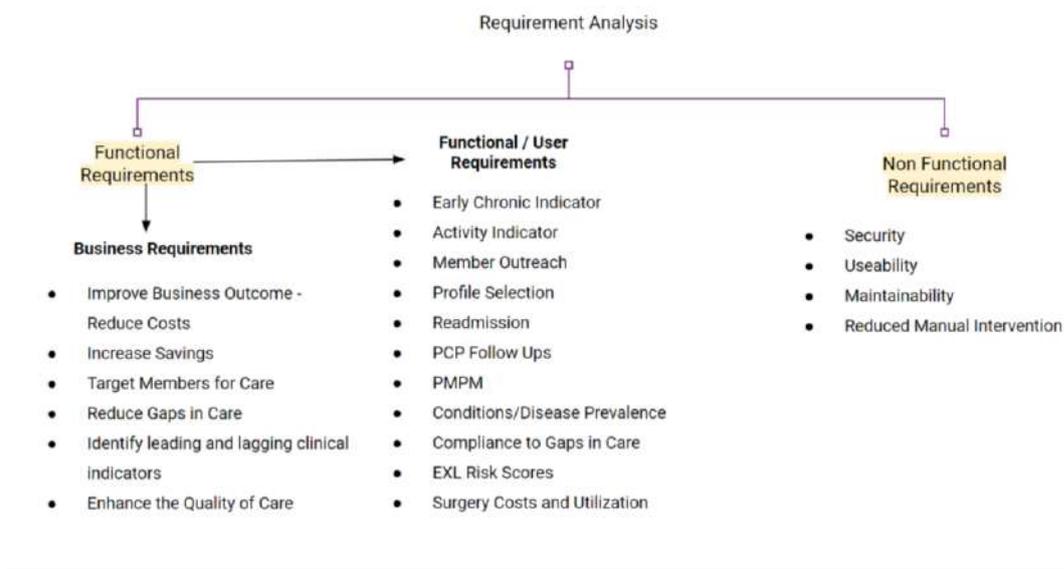


Figure 2.4 Requirement Analysis

The functional requirements denote the major working components of the system which constitutes the major part.

1. Early Chronic Indicator

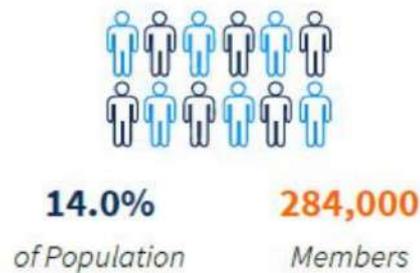


Figure 2.5 Early Chronic Indicator Percentage

Early Chronic Indicator is a flag which is being applied to the entire dataset which filters out the members having early chronic conditions. The excel VBA performs together with the Pivot to fetch the members under this category. Figure 2.5 shows how much percentage of the population shows an early chronic indication, since the analysis is performed in the member level.

Early Chronic indication is a condition or early symptom which depicts that the member is more likely to get sick in the near future or in the next 2-3 years. Members at both lines of business together comes up to around 8 million. Members in commercial or Medicare alone, comes up approximately around 4.5 million. The following metrics helps analyse and compute the chronic indication of members.

- Member Condition/Disease
- Member Claims
- Compliance to Gaps
- Barriers
- Activity information

2. Activity Indicator

Activity Indicator indicates the members who are been reached out by the primary care physicians. This helps or gives an idea of members who have been already reached out. The activity flag is used for further analysis like member outreach prediction and intervenability scores.

3. Member Outreach

Members who are more likely to interact or reply back to the primary care physicians are those who fall under this category. This metric or indication is a solution provided by EXL, using the following information.

- a. Member Activity Data/Indicator
- b. Demographics
- c. Health Literacy
- d. Barriers to Healthcare
- e. Lifestyle
- f. Income margin

4. Profile Selection

The profile selection is a feature in the monitoring report that acts like a filter. A profile generally in this case indicates the month in a year. Once the profile is selected, past 11 months from the selected profile is viewed in the report. This gives an idea of how the metrics have been changing over the period of time. The trend diagram is also shown for the same.

5. Readmission

Readmission is termed as an episode when a patient discharged from the hospital gets admitted again in a specific period of time interval. Readmission is considered as a vital metric in the report which gives the care management team an overview of percentage or rate of people who have been readmitted.

Readmission is computed based on the patient first date of service, last date of service, and several other vital metrics. This computation is performed for the member level of report for both lines of businesses namely, Medicare and commercial. Overall Population consists of members in both Medicare and commercial. The computations are performed on the whole data (8 million approximately) and then separated based on the lines of business, for the client.

Readmission could occur due to the following reasons:

- Early Discharge
- Poor Coordination
- Inadequate communication during discharge

The essential metrics considered while computing Readmission are as follows:

- Claim Date
- Last Date of Service
- Admission Date
- Claim ID

6. PCP Follow Ups

Follow ups are termed as the activity of PCP to reach out the patients, post to their discharge. Follow Ups and Readmission are both an important metric with respect to the care management strategy.

In order to provide preventive health care, the readmission and follow ups are monitored closely, to track how the care management team is performing. As the follow ups increase, the readmission is likely to drop. With this strategy, the care management team comes up with an innovative approach to handle any discrepancies in the care.

The clinical metrics considered for the computation of Readmission rate, are as follows:

- Claim Date
- PCP Date
- Discharge date
- Claim ID

7. PMPM – Costs and Utilizations

Costs incurred during the healthcare services offered are computed based on each category of expense. Surgery costs, contributes as a major source of costs among other costs. Healthcare in US, is expensive. The tracking of costs incurred during a visit to the hospital or during the episode of care, is monitored, which helps in reimbursement of the same. Per month per member deals with the cost for one person, during the month for the health insurance claims.



Figure 2.6 PMPM – Per Member Per Month

$$\text{PMPM} = \frac{\text{Sum (Total Amount)}}{\text{Sum (Total members)}}$$

The PMPM is calculated for one member based on the total amount incurred for claims, pooled together for a group of members in a certain category

8. Conditions/Disease Prevalence

Disease prevalence shows the percentage score of how much the condition is present in the population of that specific time period and selection. This gives an insight of the overall percentage of disease prevalence in the population. Guides the client to identify potential areas to focus on.

9. Compliance to Gaps in Care

Gaps in Care:

It is defined as the discrepancy between the care provided to patients and the recommended best practices in healthcare. Gaps in care are a vital issue for both patients and payers like self-insured employers. Closing these gaps can help support the goal of offering better health and wellness to patients.

Few Examples of Gaps in Care:

- A female patient age 50 with a family history of breast cancer who doesn't get annual mammograms
- A hypertension patient who doesn't refill their medication for blood thinners
- A patient's new diagnosis isn't shared with their PCP (Primary Care Physician) or specialist for another related condition
- A patient doesn't follow recommended prescription guidelines for a condition
- A patient isn't able to find the care they need in their network or area

Patient Compliance to Gaps:

Compliance rates to the gaps in care, denotes how much a condition is being complied against the gaps.

This information is vital for the care management team to better understand the gaps to make the prior arrangements to prevent or overcome the same, if the compliance rate is low for a condition. The report shows the compliance rates for the gaps in all conditions

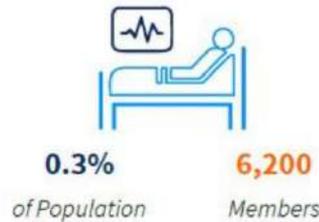


Figure 2.7 Members Compliance Rate

Members who take up claims and visit the hospital for treatment, fail to revisit or to follow the medications prescribed. For example, as given in figure 2.7, if 6,200 members take up claims for a medical condition, around 0.3% of the people tend to comply or fill the gaps in care.

10. EXL Risk Scores – Impactability and Intervenability



Figure 2.8 Risk Management Strategy

The risk scores serve an essential insight on the population health. The risk scores are a major resource for decision making in the care management team of a health care service provider organization.

As shown in figure 2.8, when there is an increased intervention and engagement with the patient, there is an improved health outcome. This is taken care by the primary care physicians, who are present for every patient as a personal care taker. Identifying the early chronic or risky patients in the cohort enables care management to take preventive care measures to mitigate any severity in the health and the overall costs.

Population at impactable stage, or having the impactability risk score. Impactability deals with identification of patients who are most likely to have a specific future event.

This score helps health care service providers to focus on an analytically driven approach, which is more patient-centered. Analyzing impactability can help organizations understand where to best focus on. Intervention is defined as the process of preventing and treating disease to improve health. Medical interventions could thereby be either preventive or therapeutic interventions.

2.3.2 BUSINESS REQUIREMENTS

- **Improve Business Outcome – Reduce Costs**

When illness is detected beforehand, the expense incurred could be less compared to the treatment taken place when the condition gets critical.

- **Increase Savings**

Great savings in cost could happen when the patient who are likely to be sick in near future are detected early. This concept brings in a lot of advantage in terms of savings, healthier patients and improved business outcomes.

- **Target Members for Care**

Using the report, the members who require care could be focused more. The focus could be streamlined and this saves up more cost and time.

- **Reduce Gaps in Care**

The gaps in care denotes the non-compliance to the medications or hospital visits suggested. The gaps have to be reduced so that the conditions could be prevented at an earlier stage. The reduction of gaps in care could be established by constant intervention by analysing the barriers for the patients with respect to access to care, demographics, health literacy and educational status

- **Identify leading and lagging clinical indicators**

Identifying the leading and lagging metrics in the population will help the business to better understand the population health and the performance of the business, in terms of profits, loss and savings during the period of time.

- **Enhancing the Quality of Care**

The overall objective for a care management team is to have a constant integration with the patients to be aware of their lifestyle, health conditions and health literacy. Enhancing the quality of care is focused to bring in maximum efficiency in the overall care management team and the healthcare insurance service provider, by improving cost savings and mitigating high expenses.

2.3.3 NON-FUNCTIONAL REQUIREMENT SPECIFICATION

- **Performance**

Performance of the system should meet the expectations

- **Security**

Data Security remains the most vital non-functional requirement since the data could be lost or replaced at any circumstance. In order to prevent this, it is vital for the system to have a safe data storage.

- **Reliability**

Data reliability should be ensured prior to any analysis. Proceeding with the analysis or computation with false data leads to incorrect results which could there by deteriorate the entire system result. A misinterpretation of a piece of information could lead to the whole system being incorrect.

- **Useability**

The usage of the system is a vital key. The system must be in a manner that it is easy to use and generate results as per required.

- **Maintainability**

The system must be easy to maintain for a long run. The components involved in the system must perform effectively in order to make sure that the system in properly maintained and working as expected.

- **Reduced Manual Intervention**

There should be reduced or minimal manual intervention as possible, since this could save up a lot of time and errors

2.4 TEST PLAN

System Testing:

- System Testing is testing conducted on a complete integrated system to evaluate the system's compliance with its specified requirements
- System testing takes, all of its components or individual components to validate the results
- The purpose of a system test is to evaluate the end-to-end system specifications

2.4.1 AD HOC TESTING

Ad hoc analysis is generally performed by business users on an as-needed basis to address data analysis needs not met by the business's established, recurring reporting that is already being produced on a daily, weekly, monthly or yearly basis.

Testing the current system on an ad hoc basis is more adequate and apt since the system consists of multiple individual components that ultimately contributes to the final result.

Thus, it is important for the components of the system to be more accurate and individual testing of the same could be done as and when needed.

Ad hoc testing is un-structured and convenient, since it doesn't require any pre-defined complete set of codes to test.

2.4.2 TESTING SCOPE

Ad hoc testing is widely practiced by many software houses both product and project-based organizations as it allows testers to test the product as per client's requirements. Can get more detailed insight on the software product resulting in high product quality and customer satisfaction. Main aim of this testing is to find defects by random checking

2.4.3 TOOLS USED FOR TESTING

Ad Hoc testing for this system requires the following tools:

- Microsoft SQL Server Management Studio
- Minerva Application – PySpark Script
- Ad Hoc testing requires connection to specific servers and databases to trigger the spark session to test
- The snippet or part of the system to be tested is scripted in the Minerva application, which serves as a platform to connect to specific databases used in the code snippet
- The PySpark engine is triggered using the SQL server connection which has the link specification to connect to the script

CHAPTER III

SYSTEM DESIGN

This chapter deals with the overview of designs which includes architectural design, Structural design, behavioural design, navigation design and code design of the system.

3.1 ARCHITECTURAL DESIGN

Architectural design focuses on the components or elements of the structure. This design gives an overview of the process outlook in brief.

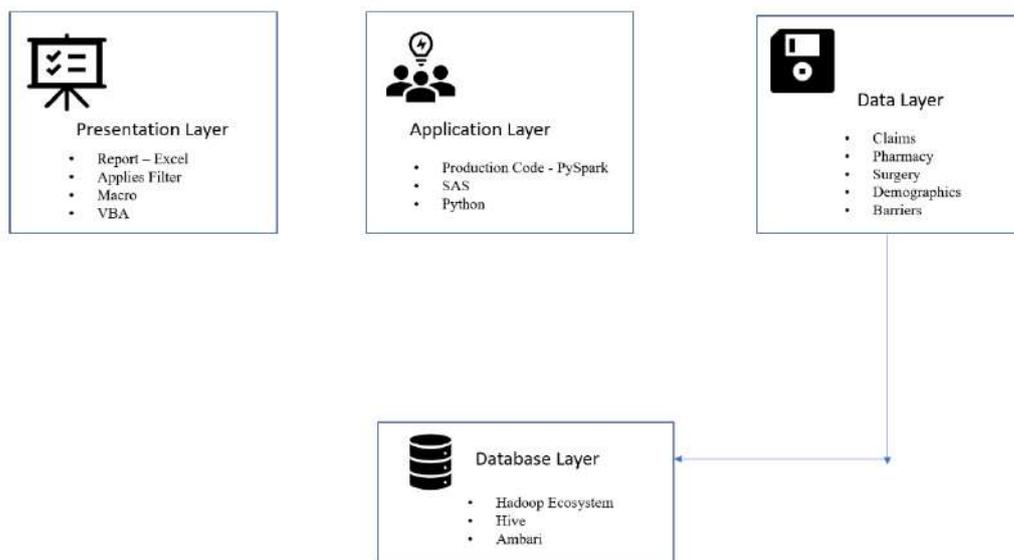


Figure 3.1 Three Tier Architecture Overview

3.1.1 ARCHITECTURAL OVERVIEW

As shown in figure 3.1, the database layer provides access to the client data and other databases required for analysis. The Hadoop ecosystem enables applications to work parallelly as containers.

The data layer gathers up together all the segments to be used for the monitoring report. Further, post to analysis and computations which is done using pyspark, in the application layer, is closely linked to the database layer where in the results are stored back into it. Earlier, this system was developed in SAS, which has results post to computations, being stored in local drives.

The Vantage and data management teams are responsible for creating profiles and organizing the data. The newer version of the system works closely with the EXL Hadoop ecosystem, which has more accessibility and saves a lot of drive space. The newer system is designed to be able to produce results which would thereby be loaded into the Hadoop ecosystem for future usage.

New System Overview:

- The new system involves computations performed and tested in PySpark
- Has a structured, organized code which generates the final resultant data set for the report
- This new system follows the principle of ‘Technology Closed Loop Framework’ which works closely with the existing technologies and complies to them
- Built in a manner that it is easier for data access and usage, also mitigating the drawbacks encountered in the earlier version of the system
- The intermediate outputs generated from the newer system is saved back into the big data ecosystem
- The workflow comparison of the earlier, or SAS version or system and the newer in the later part of this chapter

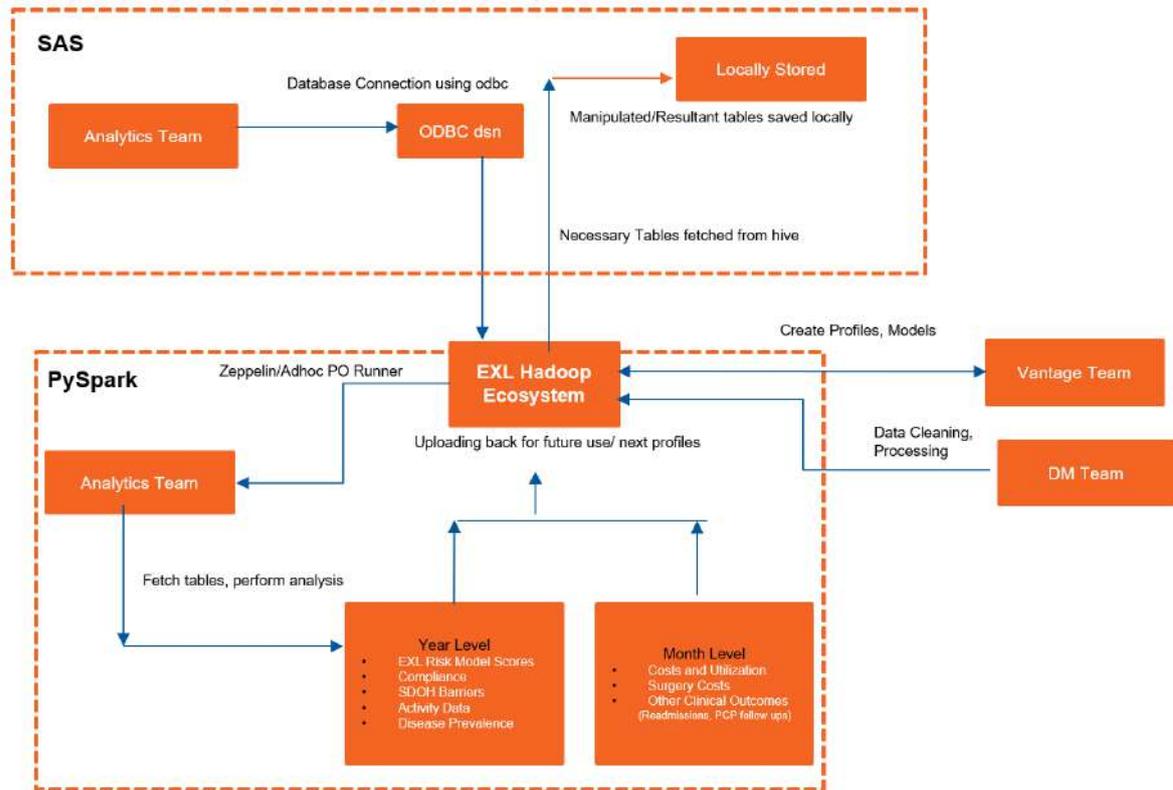


Figure 3.2 Workflow Comparison

3.1.2 BENEFITS OF MIGRATION FROM SAS TO PYSPARK

The main features that drive in migrating the SAS codes to apache spark are mentioned in the following.

- **In-Memory Computation in Spark**

With in-memory processing, we can increase the processing speed. Here the data is being cached so we need not fetch data from the disk every time thus the time is saved.

Spark has DAG execution engine which facilitates in-memory computation and acyclic data flow resulting in high speed

- **Swift Processing**

Using Apache Spark, we achieve a high data processing speed of about 100x faster in memory and 10x faster on the disk. This is made possible by reducing the number of read-write to disk.

- **Dynamic in Nature**

Parallel applications can be easily developed, as Spark provides 80 high-level operators Lazy Evaluation Outstanding feature of Spark which is call by need or memorization. It waits for instruction before providing final result which saves significant time.

- **Fault Tolerance in Spark**

Apache Spark provides fault tolerance through Spark abstraction-RDD. Spark RDDs are designed to handle the failure of any worker node in the cluster. Thus, it ensures that the loss of data is reduced to zero.

- **Real-Time Stream Processing**

Spark has a provision for real-time stream processing. Earlier the problem with Hadoop MapReduce was that it can handle and process data which is already present, but not the real-time data. but with Spark Streaming this problem can be solved.

3.2 STRUCTURAL DESIGN

Structural design deals with the solution design. This design focuses on picturizing the well-organized elements of the solution.

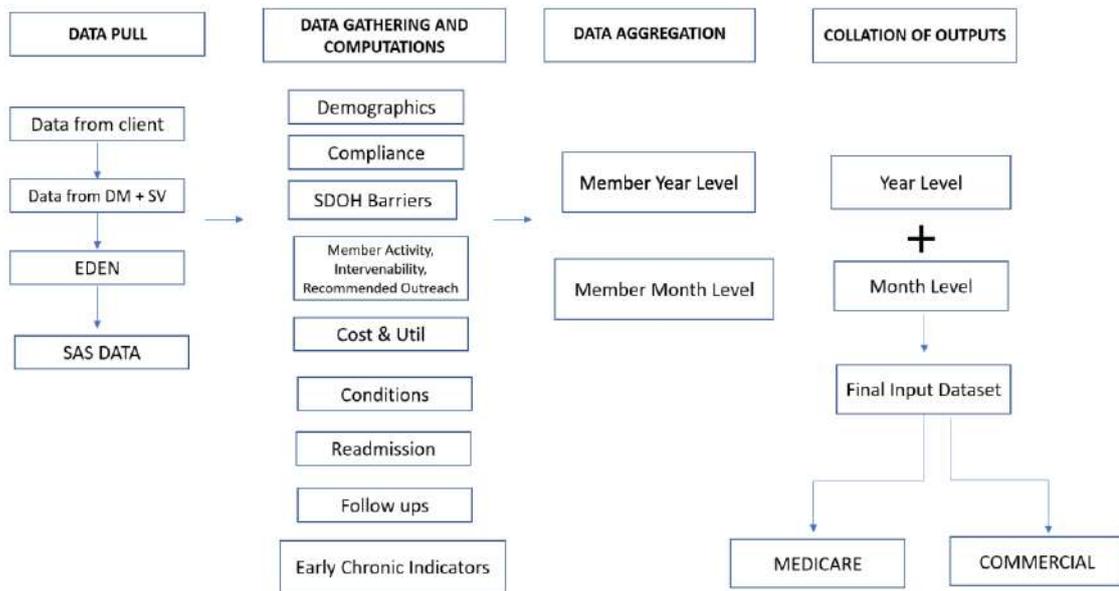


Figure 3.3 Structural Design

Structural Design Overview:

The computations and aggregations of data is categorized into both year and month level. There is a difference in both the levels, since the client/healthcare service provider gets separate insights in the individual levels as shown in figure 3.3. PMPM (Per Month per member) is a concept widely used in the healthcare insurance claims, which requires focusing population health over a period of month.

The clinical outcomes include information related to the primary care physician follow ups and readmissions which infers valuable insights to the care management team. Risk scores are better analysed with a period of 12 months and over, along with disease/condition prevalence and compliance rates.

3.3 BEHAVIOURAL DESIGN

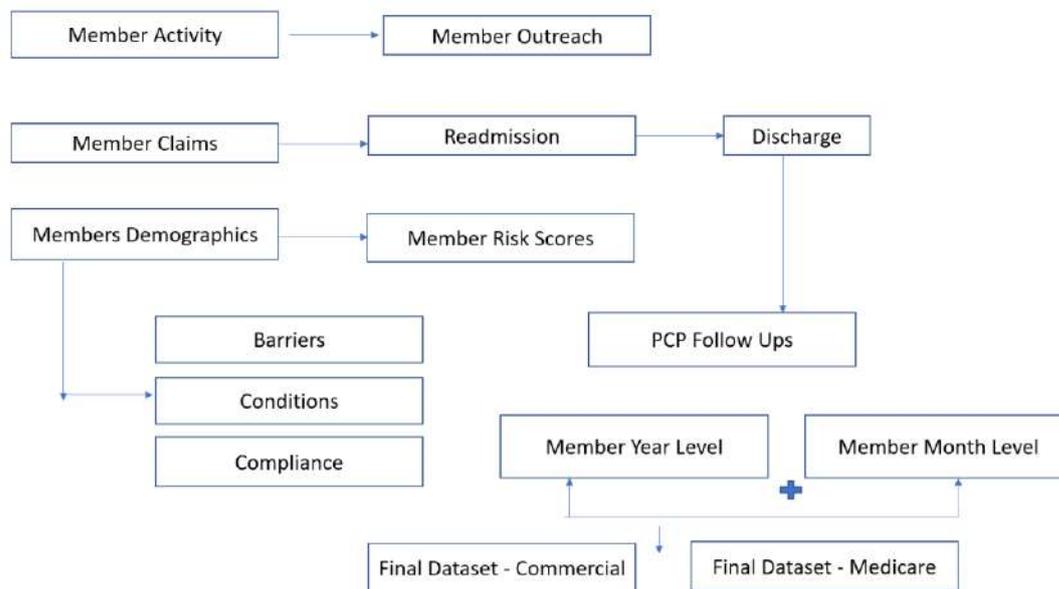


Figure 3.4 Behavioural Design

Discharge (Uses Readmission) + PCP → **Follow Ups**

Cost and Utilizations → Surgery Costs Analysis + Utilizations

Member Year Level → Member Activity + Conditions+ Risk Scores + Demographics + Barriers + Compliance

Member Month Level → Member year Level + Readmissions + Follow Ups + Costs and Utilizations

- The Model has components that has relationship within each other, which consists of deriving the later from the previous parts
- This relationship among the components and the way they interact with each other makes them reliable on each of them
- Incorrect or false computation of readmission could lead to a false value of discharge
- Which again, could affect the follow up rate which depends on the discharge data output

- Thus, the relation and interaction of the components in the system is important to recognize and to give all component the same priority
- The member month level data uses the year level data result to get the activity data insights from it. Thus, the year level data should be made sure that it denotes the right information, and should be testing for the accuracy of results generated by it
- The accuracy and adequacy of the whole system depends on its individual component's accuracy

Figure 3.3 shows how the source data is fetched for the analysis of individual components. The data is first pulled from the big data warehouse and gathered for computations for each level. Once the individual components are computed, they are tested for their accuracy. Later which, they are merged together and the output is collated in the form in which it is given as an input to the excel.

3.4 CODE DESIGN

The code generated for the system is made sure that it is efficient compared to the earlier version. Good coding practices and guidelines are followed for achieving the same

Object Oriented Designs followed:

- The system follows the object-oriented methodology in the code, which there by enables to structure and organize the individual components
- Enables reusing of code using inheritance
- Flexibility through polymorphism

- **Abstract Factory Method:**

It is a Creational design pattern which solves the problem of creating product objects without specifying their concrete classes. This technique is used in a part of the system. The abstract factory method enables interaction of methods from difference classes, which makes it easier to access a functionality or logic present in a component.

```
import pyspark
from pyspark import SparkContext
from pyspark.sql import SparkSession
sc = SparkContext.getOrCreate()

profileid = '202201'

class readmission:

    def __init__(self, id):
        self.id = profileid
        return self.id

    def compute(self):
        return 'readmission'

class followups:

    def __init__(self, id):
        self.id = profileid
        return self.id

    def compute(self):
        return "followups"

if __name__ == "__main__":

    rdms = readmission()
    flw = followups()

    print('{rdms.compute()}')
    print('{flw.compute()}')
```

Figure 3.5 Abstract factory method- Design Pattern Example

- **Adapter Design Pattern:**

This is a structural design pattern that allows objects with incompatible interfaces to collaborate.

This technique is being used in the system where the objects of classes of components which are not related, interacts. The adapter pattern has the advantage of code reusability and flexibility. Allows interface of existing class to be used as another interface

Coding practices followed:

- **Well Documented**

The code has to be well documented, so that it is well understood regarding the classes and methods used in it

- **Inline Comments**

Comments are very vital for the purpose of understanding the code, or while going through the code. Adding a line or two at each checkpoints helps users identify and understand the code flow better

- **Structured Well Organized using Object Oriented Techniques**

The code has to be organized and structured well, in order to avoid confusion. Since the current system comprises of various parts, it is vital for the code to be structured

- **Does not Produce Errors**

The code has to be error free and should generate the results as desired

- **Does not have duplication of steps/codes**

There should be no duplication of steps or functions in the code.

If so any, all of it that is similar has to be merged or made into one common function to access information from it. The existing system is made sure that it does not have any duplications

Chapter IV

SYSTEM TESTING

4.1 TESTING TECHNIQUES

Unit Testing

Unit Testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. It is performed by both developers and testers.

Integration Testing

Integration Testing is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units. Test drivers and test stubs are used to assist in Integration Testing. It is performed by testers.

Acceptance Testing

Acceptance testing, a testing technique performed to determine whether or not the software system has met the requirement specifications. The main purpose of this test is to evaluate the system's compliance with the business requirements and verify if it is having met the required criteria for delivery to end users

4.2 CURRENT SYSTEM TESTING:

Ad Hoc testing is used for the testing and validating the results generated from the system. This type of testing is a component level testing which enable better integration of components by ensuring its individual accuracy.

The flow of Ad Hoc testing for the system is as follows, in addition to the testing discussed in the earlier section (2.4)

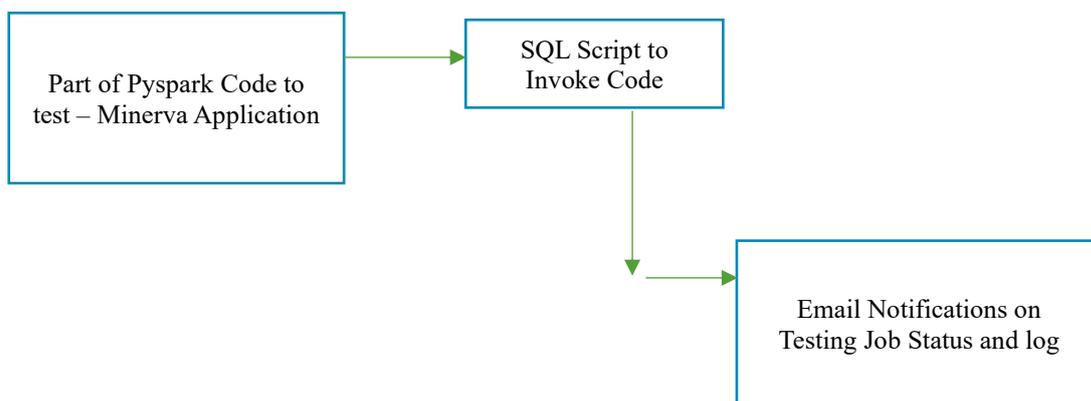


Figure 4.1 Testing Flow for the System

Purpose and Advantages of Ad Hoc Testing:

- This type of testing is performed when there is a limited amount of time
- This testing does not require pre planning, documentation or process to be followed
- Ad Hoc Testing enables detection of more defects than planned testing
- Since the system constitutes of individual components, thus the system is testing for each of its components individually for its accuracy

4.3 TEST CASES AND SCENARIOS

Test Scenario	Data Used	Expected Result	Result
Compliance	Member Claims	To Match SAS Output	Pass
Follow Ups	Member Claims	To Match SAS Output	Pass
Readmissions	Member IP claims	To Match SAS Output	Pass
Member year level	Merged - metrics	To Match SAS Output	Pass
Member month level	Merged - metrics	To Match SAS Output	Pass
Summary Dataset	Year+ Month	To Match SAS Output	Pass

TESTING CHECKPOINTS/ VALIDATIONS

- To validate the total members, count in Medicare. It should comply with the previous total Distinct Members trend
- To validate the total members, count in Commercial. It should comply with the previous total Distinct Members trend or results
- Rows Count must be validated in the summary dataset
- Presence of Clinical Metrics or Columns should be validated
- The Accuracy or reasonability of Clinical Measures are checked, Member Wise
- Total Average Costs Comparison should have minimum difference or similar to the past trends

CHAPTER V

SYSTEM IMPLEMENTATION

The System currently comprises of the Excel Report which is delivered to the client.

Client Perspective:

- The implementation of the system by the client is easier, and the pre requisite is to have Microsoft Excel
- The data or report gets refreshed manually by clicking the user created update button in the sheet prepared, which will trigger the Macro to call the VBA functions to populate the newer data, for every month of delivery
- This system could be implemented by anyone inside the organization, but implementing it outside the organization is not possible

Developer Perspective:

- The Pyspark code is executed, which has connection to the big data ecosystem
- The output datasets are uploaded to the ecosystem
- The final summary dataset for a profile is uploaded in Excel, using a macro auto update user defined button
- Using pivot, macro and VBA in Excel, the dataset is presented as a monitoring report

CHAPTER VI

CONCLUSION

Overview of System Developed:

The System thus created using pyspark generates the overall input summary dataset for the monitoring report, for both levels and lines of businesses. The System is made to test using Ad hoc for accurate generation of outputs matching with the expected ones. The newer system mitigates problems faced by the conventional system, like time taken and drive space constraints. The newer system works closely with the technology closed loop framework, to make ease of computations for further analysis. The data loss scenarios is taken care in the new system, where the data is safely stored in the big data ecosystem. The system is developed in a more organized way, to avoid confusions while generating the resultant input dataset for the report during every month of delivery. Clinical Outcome evaluation and Quality check is performed at the later stage to ensure that the system is complying to the regular value trend

Future Scope and Enhancements:

- The Excel Report could be converted into a Tableau/ Power BI Dashboard for a better interpretation and user experience for the care management team in the healthcare service provider organization/ Client – If the client wishes to switch
- The System can include another flag as a filter which would fetch the members who have engaged with the PCPs

- Introducing a separate analysis for instate and outstate members of Michigan
- Introducing and addition of new gaps in compliance rates, to make the care management team get a better overview of conditions

BIBLIOGRAPHY

1. American Academy of Family Physicians Government Relations. Government Affairs Weekly.
http://www.aafp.org/dam/AAFP/documents/advocacy/washington_update/2015/grweekly-jan-30.pdf. Accessed June 16, 2015
2. Alexandar JA, Markovitz AR, Paustian ML, et al. Implementation of patient-centered medical homes in adult primary care practices. *Med Care Res Rev.* 2015;72(4):438-467.
3. U.S. Department of Health and Human Services. Better, smarter, healthier: in historic announcement, HHS sets clear goals and timeline for shifting Medicare reimbursements from volume to value.
<http://www.hhs.gov/about/news/2015/01/26/better-smarter-healthier-in-historic-announcement-hhs-sets-clear-goals-and-timeline-for-shifting-medicarereimbursements-from-volume-to-value.html>. Accessed January 20, 2016.

Website References:

- <https://www.exlservice.com/healthcare>
- <https://www.nlm.nih.gov/research/umls/>
- <https://machinelearningmastery.com/>
- <https://www.datacamp.com/tracks/big-data-with-pyspark/>
- <https://medium.com/>
- <https://support.sas.com/documentation/onlinedoc/>

MIGRATION AND ENHANCEMENT OF POPULATION HEALTH MONITORING AND TRACKING REPORT

PAVITHRA S
Register No: 1733016

DISSERTATION SUBMITTED IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
M.Sc. (Decision and Computing Sciences)
OF ANNA UNIVERSITY



MAY 2022

**DEPARTMENT OF COMPUTING
COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution affiliated to Anna University)
COIMBATORE – 641014**

COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution Affiliated to Anna University)
COIMBATORE 641014

(Bonafide Certificate)

Project Work – II

Tenth Semester

MIGRATION AND ENHANCEMENT –
POPULATION HEALTH MONITORING AND TRACKING REPORT

Bonafide record of work done by

PAVITHRA S

(Register No: 1733016)

Submitted in partial fulfilment of the
requirements for the degree of
M.Sc. (Decision and Computing Sciences)
of Anna University

MAY 2022

Faculty Guide

Head of the Department

Submitted for the viva-voce held on _____

Internal Examiner

External Examiner



Date: 12-05-2022

This is to certify that **Ms. Pavithra.S**, student of M.Sc. Decision and Computing Sciences (Integrated) from “**Coimbatore Institute of Technology**” has been working with **EXL Service.com (India) Private Limited** as an **Intern** for the period of 13th December 2021 to 13th June 2022. She has undergone academic training program at our organization as part of completion of requirement for the award of M.Sc. She is working on project work entitled “**Migration and Enhancement of Population Health Monitoring / Tracking Report**” during the said duration.

For, EXL Service.com (India) Private Limited

A handwritten signature in black ink, appearing to read 'Vikram Pragadeesh'.

Vikram Pragadeesh
Assistant Vice President – Advanced Analytics / Data Science, EXL Health

CONTENTS

CHAPTER	PAGE NO
ACKNOWLEDGEMENT	i
SYNOPSIS	ii
PREFACE	iii
I INTRODUCTION	1
1.1 ORGANIZATION PROFILE	1
1.2 PROBLEM DEFINITION	2
1.3 SYSTEM ENVIRONMENT	2
II SYSTEM ANALYSIS	7
2.1 SYSTEM DESCRIPTION	7
2.2 USE CASE MODEL	11
2.3 SOFTWARE REQUIREMENTS SPECIFICATION	13
2.4 TEST PLAN	20
III SYSTEM DESIGN	22
3.1 ARCHITECTURAL DESIGN	22
3.2 STRUCTURAL DESIGN	26
3.3 BEHAVIOURAL DESIGN	29
3.4 CODE DESIGN	31
IV SYSTEM TESTING	35
4.1 TEST CASES AND TEST REPORTS	37
V SYSTEM IMPLEMENTATION	38
VI CONCLUSION	39
BIBLIOGRAPHY	

ACKNOWLEDGEMENT

Apart from the efforts of myself, the success of this project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

I am extremely grateful to **Dr. A. RAJESWARI**, Principal, Coimbatore Institute of Technology, for giving me this opportunity to do my project at EXL Service Private Limited, Chennai.

I am deeply indebted to **Dr. K. SHAKTHI MALA**, Dean of Computing, Coimbatore Institute of Technology, Coimbatore, for her continual support and ardent motivation.

I extend my deep gratitude to **Dr. A. KANNAMMAL**, Head, Department of Computing, Decision and Computing Sciences, Coimbatore Institute of Technology, for her continual support

I extend my heartfelt gratitude to my internal guide, **Ms. V. SHANTHI**, Assistant Professor, Software Systems, Department of Computing, Coimbatore Institute of Technology, for her continual support and enduring guidance throughout my project tenure

I express my deep sense of gratitude to **Mr. VIKRAM PRAGADEESH**, Senior Manager, Assistant Vice President – Advanced Analytics/Data Science, EXL Service Private Limited, Chennai, for his invaluable guidance, support and suggestions throughout the course of this project work

I wish to acknowledge all the other members in my team, who made the work challenging and were very helpful in growing my learning curve. Finally, I express my gratefulness to all the staff of Department of Computing, Coimbatore Institute of Technology, Coimbatore and my family and friends for their encouragement and support

SYNOPSIS

The project entitled **“Migration and Enhancement of Population Health Monitoring and Tracking Report”** aims to create a shift in care management’s focus from highly reactive care management to a more proactive approach that encompasses managing members across the risk spectrum, including those who are yet to become high risk. The objective is to build a monitoring report which consists of clinical metrics regarding the population, including the cost trend, utilizations, compliance to gaps in care or visits to the healthcare service provider/client. The existing system is built using SAS, which computes the components of the final dataset to be given as an input to the monitoring report. This system is found to have several drawbacks and is inefficient to be able to easily run for a prolonged period of time. Thus, it is migrated to PySpark, which has direct connection or ease of access to the big data ecosystem. Altogether, the new system focuses on prioritizing of care, to achieve maximum efficiency and outcomes from care management teams and suggest an intervention approach to mitigate against undesired health outcomes, increased costs and also overcoming the drawbacks of earlier version of the system.

The process of migration includes thorough analysis of the existing system and the logic of computations of clinical metrics, in order to come up with an equivalent efficient way to attain the results. The input data is unstructured and are from different sources, namely, HRA, Claims, Pharmacy data etc. They are put together in a member level format to perform a member level analysis to derive the clinical metrics and evaluating their reasonability using pyspark.

The system works on the basis of ‘technology closed loop framework’, which works closely with the Hadoop ecosystem. The access and manipulation of data is performed in here. The computations performed using PySpark; are uploaded to the ecosystem, there by saving the vital checkpoints, which would be useful in the future.

PREFACE

Chapter I - **INTRODUCTION AND ORGANIZATION PROFILE** – Specifies the purpose, the objective and the environment in which the project works

Chapter II – **SYSTEM ANALYSIS** – Describes the existing tool along with its limitations, proposed updates and its features, and the software requirements specifications

Chapter III – **SYSTEM DESIGN** - Elucidates the architectural, structural and behavioural design of the dsystem

Chapter IV – **SYSTEM TESTING** – Specifies the testing techniques used, with test scenarios

Chapter V – **SYSTEM IMPLEMENTATION** - Lays out the implementation of how the system is being implemented

Chapter VI - **CONCLUSION** – Puts forth the special features and suggestions for future enhancements of this project

BIBLIOGRAPHY

CHAPTER I

INTRODUCTION

This chapter contains the introduction about the organization, about the population health monitoring report and system specifications, under which the project was developed and worked on.

1.1 ORGANIZATION PROFILE

SCIO for more than 14 years, provides its healthcare partners actionable insights and scalable solutions through deep expertise, flexible analytics, innovative services and technology that drive measurable outcomes improvement. It focuses on the areas of population health, payment integrity, risk adjustment, care management, opportunity analysis, incentive design, consumer segmentation and engagement, network performance and commercial effectiveness. Each year actionable insights reach more than 150 million covered lives, 406 million medical claims, 1.3 billion prescription claims, and over 60,000 pharma dashboards with 300 unique data sources, refreshed monthly.

EXL's acquisition of SCIO in 2018 expands the expertise as a strategic digital transformation partner and deepens analytics, service and technology capabilities across the healthcare continuum.

TEAM PROFILE

Data and Analytics Services provides rich data assets to support client data needs, delivers end-to-end data-enabled solutions and offers enterprise data and analytics. The team helps clients identify, optimize, create, capture and grow business value by leveraging data assets, analytic capabilities, and data management expertise. Care Optimization (CO) in data and analytics services is responsible for the following services.

- **Population Health**

Identify and stratify current and emerging risk, target highest area of impact and propensity to engage, intervene early, fuel engagement and outreach programs.

- **Risk adjustment and Quality**

Support quality outcomes measures, documentation accuracy, and optimized revenue

- **Real world insights**

Deliver persona and patient-centred insights and outcomes-based contracting

1.2 PROBLEM DEFINITION

To migrate and enhance the population health monitoring report which provides key insights about the population to assist the care management analytics team of the health care service provider organization located in Michigan; also aid them in comparing performance against past and upcoming scenarios so that the current or potential problems can be identified, analysed and prevented.

1.3 SYSTEM ENVIRONMENT

The following are the hardware and software specifications that are used in the development of the project.

1.3.1 HARDWARE SPECIFICATION

Processor: Intel® Core™ i5-8500M CPU

Clock Speed: 2.40 GHz

Hard Disk: 500 GB

RAM: 8.00 GB 1.2.2

1.3.2 SOFTWARE SPECIFICATION

Operating System: Windows 10 PRO 64-bit OS

Tools: Zeppelin, Ambari, Jupyter Notebook

Database: Hive

Language: PySpark, Python, SAS

Tools: Ambari, Zeppelin, SAS, Hive

1.3.3 TOOLS AND TECHNOLOGIES USED

The technologies and tools that are used in the project and a brief description about each of them are discussed in the following subsections.

- **Python Language**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's elegant syntax and dynamic typing, together with its interpreted nature, makes it an ideal language for scripting and rapid application development in many areas on most platforms.

- **Jupyter Notebook**

The Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebooks documents.

Jupyter is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media, usually ending with the “.ipynb” extension. Jupyter Notebook can connect to many kernels, (by default Jupyter Notebook ships with the IPython kernel) to allow programming in many languages.

- **PySpark**

Apache Spark is a lightning-fast cluster computing technology, designed for fast computation. It is based on Hadoop MapReduce and it extends the MapReduce model to efficiently use it for more types of computations, which includes interactive queries and stream processing. The main feature of Spark is its in-memory cluster computing that increases the processing speed of an application. Spark is designed to cover a wide range of workloads such as batch applications, iterative algorithms, interactive queries and streaming. Apart from supporting all these workloads in a respective system, it reduces the management burden of maintaining separate tools. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs.

- **Zepplin**

Apache Zeppelin is a web-based notebook which brings data exploration, visualization, sharing and collaboration features to Spark. It supports Python, but also a growing list of programming languages such as Scala, Hive, SparkSQL, shell and markdown. Zeppelin provides a “Modern Data Science Studio” that supports Spark and Hive out of the box. It supports multiple language backends which has support for a growing ecosystem of data sources.

- **Hive**

Apache Hive is an open-source data warehouse system built on top of Hadoop. It is used for querying and analyzing large datasets stored in Hadoop files. Hive gives a SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data.

Hive provides the necessary SQL abstraction to integrate SQL-like queries (HiveQL) into the underlying Java without the need to implement queries in the low-level Java API.

- **Ambari**

Apache Ambari is an open-source administration tool deployed on top of Hadoop clusters, and it is responsible for keeping track of the running applications and their status. Apache Ambari can be referred to as a web-based management tool that manages, monitors, and provisions the health of Hadoop clusters.

It provides a highly interactive dashboard that allows administrators to visualize the progress and status of every application running over the Hadoop cluster. Its flexible and scalable user interface allows a range of tools such as Pig, MapReduce, Hive, etc. to be installed on the cluster and administers their performances in a user-friendly fashion.

- **SAS**

SAS (Statistical Analysis System) is a software system for data inspection and report writing.

Statistical Analysis System is a batch of programs that work together to reclaim them and to reserve data values, reform data, evaluate complex and single exponential analyses and generate reports. SAS is a powerful and flexible statistical package that runs on many platforms, including Windows and Unix. The SAS Programming involves first creating/reading the data sets into the memory and then doing the analysis on this data.

CHAPTER II

SYSTEM ANALYSIS

This section deals with the overall description, which involves the dissection of the system into its component pieces, to investigate further upon how these components interact and work.

2.1 SYSTEM DESCRIPTION

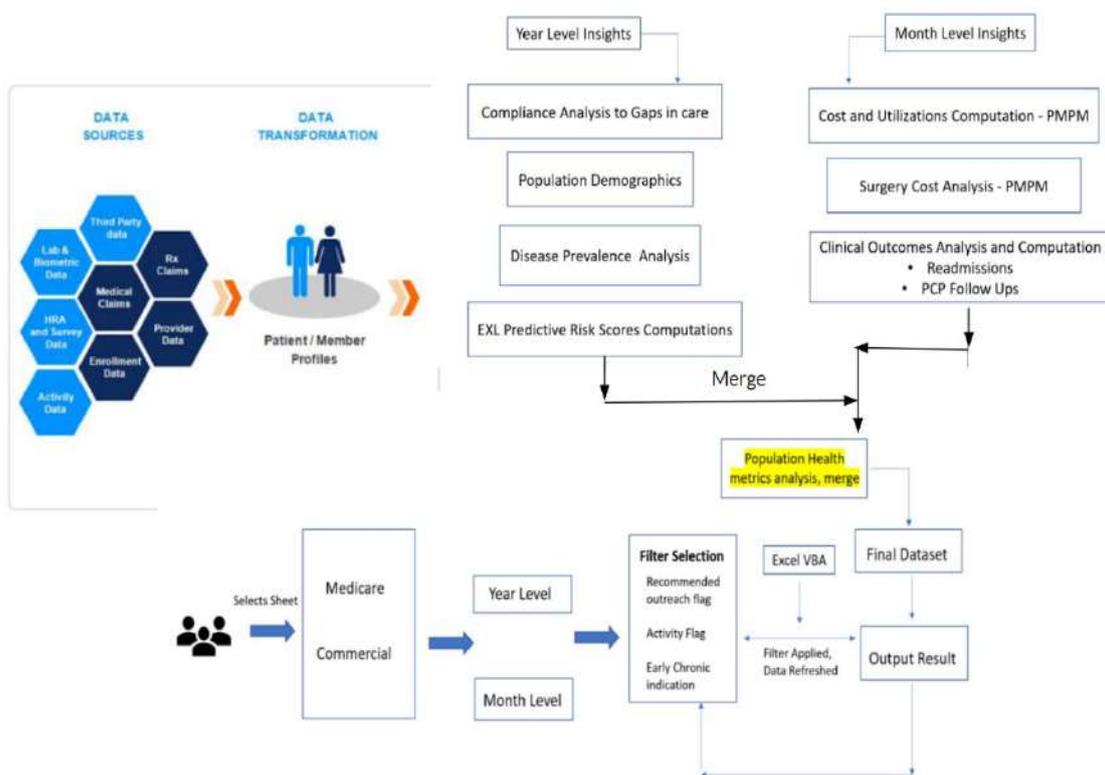


Figure 2.1 General Overview of Population Health Monitoring Report

As given in figure 2.1, the system mainly focuses on deriving the final input dataset for the Population health monitoring report. This dataset is the union of individual components which is eventually derived by member level analysis and computations.

Existing System Overview:

- The input data is fetched from the big data ecosystem / Hadoop
- The Individual analysis of metrics in the system is performed in SAS
- The user level view of the report as shown in figure 2.3 is followed earlier and currently as well
- The user makes a selection in the excel sheet, chooses the line of business
- Once the LOB (Line of Business) is chosen, the year or month level view sheet is selected
- The filters are applied in the filters section in the sheet
- Once the filters are selected, the data gets refreshed and applies to the filter of selection accordingly. This is achieved using the Excel VBA scripts, which when applied filters, gets triggered and fetches the output view

Drawbacks of the existing system:

- Drive Storage - Every computation made on the component of the system is stored locally, which takes up a lot of space
- Time Consuming - SAS by itself takes more time to perform computations
- Data Loss or Data tampering is more likely to take place

2.1.1 DATA SOURCES

In figure 2.1, the data to be used in the system is gathered together in the initial stage, which involves the pharmacy, lab, HRA (Health Risk Assessment) survey data, medical claims data, enrolment, provider and third-party data. All these data are used in a manner that the vital metrics from each are utilized for analysis and computations.

HRA data involves the Health Risk Assessment data, which is gathered together through a survey. This data allows clients to be proactive, rather than reactive. It identifies and predicts risk factors for preventable chronic conditions. Combined with readiness to change data, clients can pinpoint and prioritize the best programs and interventions for their populations.

Medical Claims Data gives an overview of the data of claims requested by patients. This data helps identify the patient's conditions and hospital visits. Similarly, the lab data and third party data is gathered for better analysis of the population health.

2.1.2 DATA TRANSFORMATION

The data is converted into member level profiles for a better patient centric approach of deliverable. It is known that almost 93 percent of the US population takes up claims to support their healthcare expenses. A member could take any number of claims and every claim is identified by a unique claim id. This member level information is used up for further analysis, having the primary key consideration as the member id.

2.1.3 ANALYTICS ENGINE

This section constitutes the Analysis performed on the data, involving evaluation of risk scores, such as the impactability, intervenability and clinical outcomes such as PCP follow ups and Readmissions.

The computations are done for both member and year level of insights. The existing system which is conventionally built using SAS, is converted into a system completely built using pyspark.

2.1.4 OVERVIEW ON THE DELIVERY

The report delivers the vital insights for the care management team, enabling to make decisions based on the same to prevent upcoming threats or scenarios with respect to patient health or engagements.

Care Management:

A Care Management team performs the set of activities intended to improve patient care and reduce the need for medical services by enhancing coordination of care, eliminate duplication and helping patients to effectively manage health conditions

Role of Care management in healthcare:

- Provide Quality Healthcare
- Patient Outreach and follow ups
- Providing Constant monitoring and support
- Patient-Centred Approach

Client Benefits from the Report:

Membership Insights

Informing what members truly need and how they need that care based on known social and clinical factors.

Care Management Efficiencies and Effectiveness

Focusing on the right members most likely to engage and improve health outcomes.

Risk and Cost Management

Moving down the risk funnel to mitigate potential future risk/costs.

Member Engagement & Health Outcomes

Creating far deeper member views inform effective intervention and engagement strategies.

Proactive Care Management

That addresses varying levels of member needs

2.2 USE CASE MODEL

The system use case model defines how the members engage and interact with the system. The usage of the system varies from the user to the developer. The way in which the user or client interacts with the system is different from the manner in which the developer or a member who builds the report interacts.

The use case model shown in figure 2.2, of the system gives an overview of vital tasks that is being performed by the members using the system. The data is prepared by the data management team and the vantage team, where in the data is prepared member level. The user selects or applies filter to the report, where the data gets refreshed according to the same. The testing is performed to the components of the system.

Ad Hoc testing is used to test the system, which doesn't require the system to be completely in order or to be testing ready. Since ad hoc testing is unstructured, there is no predefined order involved. Testing the component any point of development stage, gives more advantage over structured systematic testing.

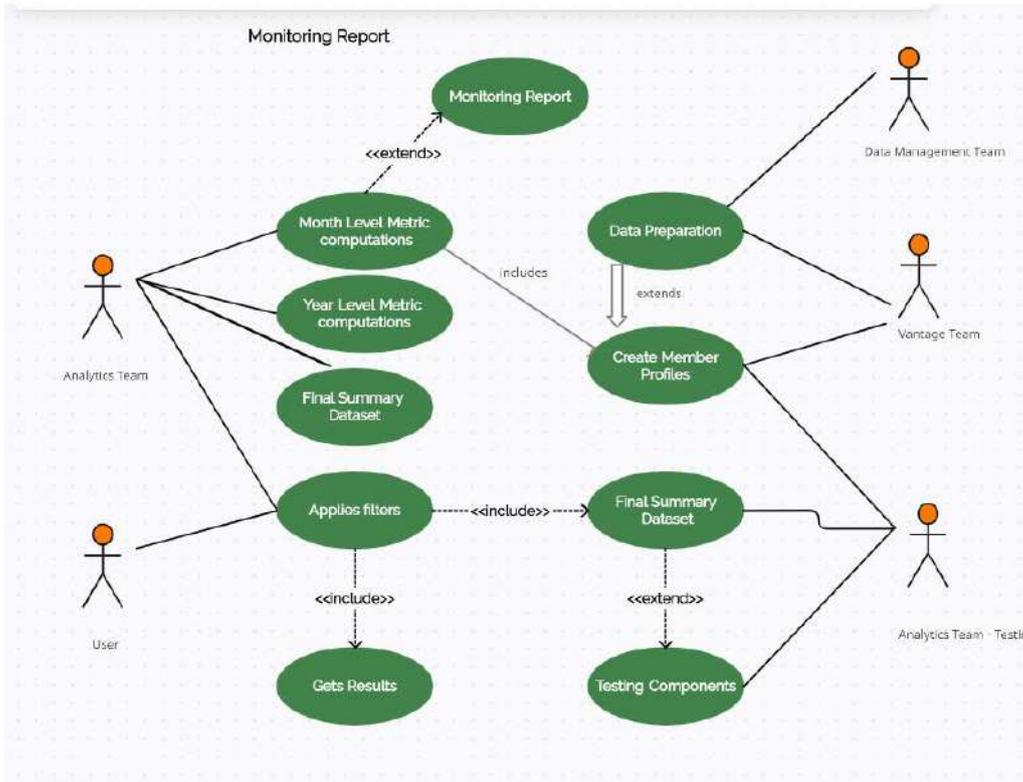


Figure 2.2 Use Case Diagram

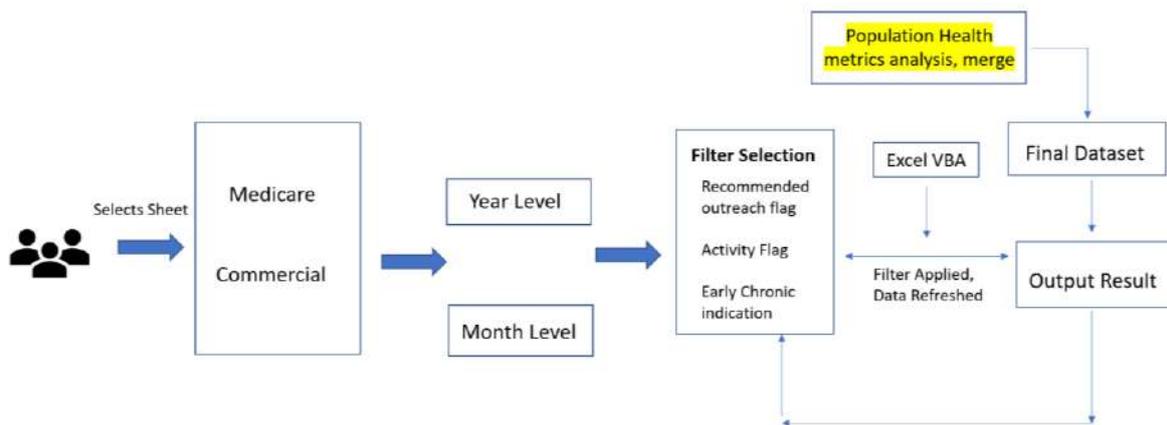


Figure 2.3 Client/User Interaction Flow

The above figure 2.3 shows the user interaction with the system. This user is the client who runs uses the monitoring report. There are two lines of business namely, Medicare and Commercial, which the user can select at the beginning stage. There are both year and month level of insights being provided for both the categories. The filters are applied to the report, which there by triggers the excel macro and calling the VBA function, to refresh and load the data according the filter applied.

2.3 SOFTWARE REQUIREMENT SPECIFICATION

This section focuses on the software requirement specifications, which covers the functional and non-functional requirements of the system

2.3.1 FUNCTIONAL REQUIREMENT SPECIFICATION

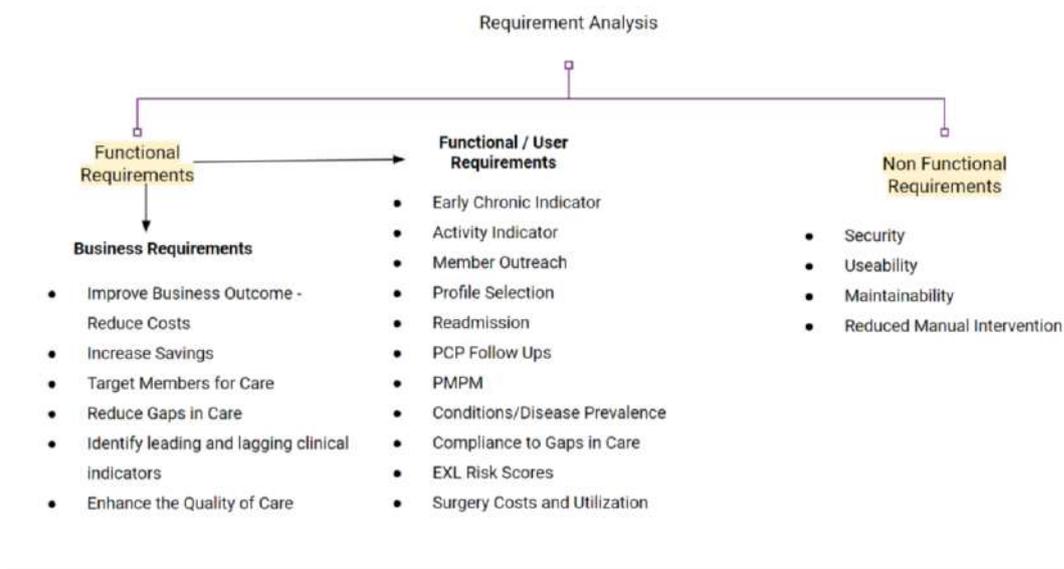


Figure 2.4 Requirement Analysis

The functional requirements denote the major working components of the system which constitutes the major part.

1. Early Chronic Indicator

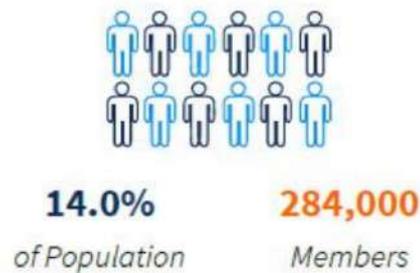


Figure 2.5 Early Chronic Indicator Percentage

Early Chronic Indicator is a flag which is being applied to the entire dataset which filters out the members having early chronic conditions. The excel VBA performs together with the Pivot to fetch the members under this category. Figure 2.5 shows how much percentage of the population shows an early chronic indication, since the analysis is performed in the member level.

Early Chronic indication is a condition or early symptom which depicts that the member is more likely to get sick in the near future or in the next 2-3 years. Members at both lines of business together comes up to around 8 million. Members in commercial or Medicare alone, comes up approximately around 4.5 million. The following metrics helps analyse and compute the chronic indication of members.

- Member Condition/Disease
- Member Claims
- Compliance to Gaps
- Barriers
- Activity information

2. Activity Indicator

Activity Indicator indicates the members who are been reached out by the primary care physicians. This helps or gives an idea of members who have been already reached out. The activity flag is used for further analysis like member outreach prediction and intervenability scores.

3. Member Outreach

Members who are more likely to interact or reply back to the primary care physicians are those who fall under this category. This metric or indication is a solution provided by EXL, using the following information.

- a. Member Activity Data/Indicator
- b. Demographics
- c. Health Literacy
- d. Barriers to Healthcare
- e. Lifestyle
- f. Income margin

4. Profile Selection

The profile selection is a feature in the monitoring report that acts like a filter. A profile generally in this case indicates the month in a year. Once the profile is selected, past 11 months from the selected profile is viewed in the report. This gives an idea of how the metrics have been changing over the period of time. The trend diagram is also shown for the same.

5. Readmission

Readmission is termed as an episode when a patient discharged from the hospital gets admitted again in a specific period of time interval. Readmission is considered as a vital metric in the report which gives the care management team an overview of percentage or rate of people who have been readmitted.

Readmission is computed based on the patient first date of service, last date of service, and several other vital metrics. This computation is performed for the member level of report for both lines of businesses namely, Medicare and commercial. Overall Population consists of members in both Medicare and commercial. The computations are performed on the whole data (8 million approximately) and then separated based on the lines of business, for the client.

Readmission could occur due to the following reasons:

- Early Discharge
- Poor Coordination
- Inadequate communication during discharge

The essential metrics considered while computing Readmission are as follows:

- Claim Date
- Last Date of Service
- Admission Date
- Claim ID

6. PCP Follow Ups

Follow ups are termed as the activity of PCP to reach out the patients, post to their discharge. Follow Ups and Readmission are both an important metric with respect to the care management strategy.

In order to provide preventive health care, the readmission and follow ups are monitored closely, to track how the care management team is performing. As the follow ups increase, the readmission is likely to drop. With this strategy, the care management team comes up with an innovative approach to handle any discrepancies in the care.

The clinical metrics considered for the computation of Readmission rate, are as follows:

- Claim Date
- PCP Date
- Discharge date
- Claim ID

7. PMPM – Costs and Utilizations

Costs incurred during the healthcare services offered are computed based on each category of expense. Surgery costs, contributes as a major source of costs among other costs. Healthcare in US, is expensive. The tracking of costs incurred during a visit to the hospital or during the episode of care, is monitored, which helps in reimbursement of the same. Per month per member deals with the cost for one person, during the month for the health insurance claims.



Figure 2.6 PMPM – Per Member Per Month

$$\text{PMPM} = \frac{\text{Sum (Total Amount)}}{\text{Sum (Total members)}}$$

The PMPM is calculated for one member based on the total amount incurred for claims, pooled together for a group of members in a certain category

8. Conditions/Disease Prevalence

Disease prevalence shows the percentage score of how much the condition is present in the population of that specific time period and selection. This gives an insight of the overall percentage of disease prevalence in the population. Guides the client to identify potential areas to focus on.

9. Compliance to Gaps in Care

Gaps in Care:

It is defined as the discrepancy between the care provided to patients and the recommended best practices in healthcare. Gaps in care are a vital issue for both patients and payers like self-insured employers. Closing these gaps can help support the goal of offering better health and wellness to patients.

Few Examples of Gaps in Care:

- A female patient age 50 with a family history of breast cancer who doesn't get annual mammograms
- A hypertension patient who doesn't refill their medication for blood thinners
- A patient's new diagnosis isn't shared with their PCP (Primary Care Physician) or specialist for another related condition
- A patient doesn't follow recommended prescription guidelines for a condition
- A patient isn't able to find the care they need in their network or area

Patient Compliance to Gaps:

Compliance rates to the gaps in care, denotes how much a condition is being complied against the gaps.

This information is vital for the care management team to better understand the gaps to make the prior arrangements to prevent or overcome the same, if the compliance rate is low for a condition. The report shows the compliance rates for the gaps in all conditions

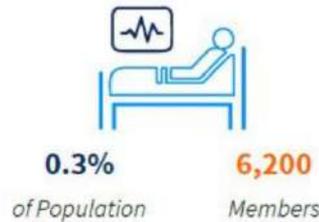


Figure 2.7 Members Compliance Rate

Members who take up claims and visit the hospital for treatment, fail to revisit or to follow the medications prescribed. For example, as given in figure 2.7, if 6,200 members take up claims for a medical condition, around 0.3% of the people tend to comply or fill the gaps in care.

10. EXL Risk Scores – Impactability and Intervenability



Figure 2.8 Risk Management Strategy

The risk scores serve an essential insight on the population health. The risk scores are a major resource for decision making in the care management team of a health care service provider organization.

As shown in figure 2.8, when there is an increased intervention and engagement with the patient, there is an improved health outcome. This is taken care by the primary care physicians, who are present for every patient as a personal care taker. Identifying the early chronic or risky patients in the cohort enables care management to take preventive care measures to mitigate any severity in the health and the overall costs.

Population at impactable stage, or having the impactability risk score. Impactability deals with identification of patients who are most likely to have a specific future event.

This score helps health care service providers to focus on an analytically driven approach, which is more patient-centered. Analyzing impactability can help organizations understand where to best focus on. Intervention is defined as the process of preventing and treating disease to improve health. Medical interventions could thereby be either preventive or therapeutic interventions.

2.3.2 BUSINESS REQUIREMENTS

- **Improve Business Outcome – Reduce Costs**

When illness is detected beforehand, the expense incurred could be less compared to the treatment taken place when the condition gets critical.

- **Increase Savings**

Great savings in cost could happen when the patient who are likely to be sick in near future are detected early. This concept brings in a lot of advantage in terms of savings, healthier patients and improved business outcomes.

- **Target Members for Care**

Using the report, the members who require care could be focused more. The focus could be streamlined and this saves up more cost and time.

- **Reduce Gaps in Care**

The gaps in care denotes the non-compliance to the medications or hospital visits suggested. The gaps have to be reduced so that the conditions could be prevented at an earlier stage. The reduction of gaps in care could be established by constant intervention by analysing the barriers for the patients with respect to access to care, demographics, health literacy and educational status

- **Identify leading and lagging clinical indicators**

Identifying the leading and lagging metrics in the population will help the business to better understand the population health and the performance of the business, in terms of profits, loss and savings during the period of time.

- **Enhancing the Quality of Care**

The overall objective for a care management team is to have a constant integration with the patients to be aware of their lifestyle, health conditions and health literacy. Enhancing the quality of care is focused to bring in maximum efficiency in the overall care management team and the healthcare insurance service provider, by improving cost savings and mitigating high expenses.

2.3.3 NON-FUNCTIONAL REQUIREMENT SPECIFICATION

- **Performance**

Performance of the system should meet the expectations

- **Security**

Data Security remains the most vital non-functional requirement since the data could be lost or replaced at any circumstance. In order to prevent this, it is vital for the system to have a safe data storage.

- **Reliability**

Data reliability should be ensured prior to any analysis. Proceeding with the analysis or computation with false data leads to incorrect results which could there by deteriorate the entire system result. A misinterpretation of a piece of information could lead to the whole system being incorrect.

- **Useability**

The usage of the system is a vital key. The system must be in a manner that it is easy to use and generate results as per required.

- **Maintainability**

The system must be easy to maintain for a long run. The components involved in the system must perform effectively in order to make sure that the system in properly maintained and working as expected.

- **Reduced Manual Intervention**

There should be reduced or minimal manual intervention as possible, since this could save up a lot of time and errors

2.4 TEST PLAN

System Testing:

- System Testing is testing conducted on a complete integrated system to evaluate the system's compliance with its specified requirements
- System testing takes, all of its components or individual components to validate the results
- The purpose of a system test is to evaluate the end-to-end system specifications

2.4.1 AD HOC TESTING

Ad hoc analysis is generally performed by business users on an as-needed basis to address data analysis needs not met by the business's established, recurring reporting that is already being produced on a daily, weekly, monthly or yearly basis.

Testing the current system on an ad hoc basis is more adequate and apt since the system consists of multiple individual components that ultimately contributes to the final result.

Thus, it is important for the components of the system to be more accurate and individual testing of the same could be done as and when needed.

Ad hoc testing is un-structured and convenient, since it doesn't require any pre-defined complete set of codes to test.

2.4.2 TESTING SCOPE

Ad hoc testing is widely practiced by many software houses both product and project-based organizations as it allows testers to test the product as per client's requirements. Can get more detailed insight on the software product resulting in high product quality and customer satisfaction. Main aim of this testing is to find defects by random checking

2.4.3 TOOLS USED FOR TESTING

Ad Hoc testing for this system requires the following tools:

- Microsoft SQL Server Management Studio
- Minerva Application – PySpark Script
- Ad Hoc testing requires connection to specific servers and databases to trigger the spark session to test
- The snippet or part of the system to be tested is scripted in the Minerva application, which serves as a platform to connect to specific databases used in the code snippet
- The PySpark engine is triggered using the SQL server connection which has the link specification to connect to the script

CHAPTER III

SYSTEM DESIGN

This chapter deals with the overview of designs which includes architectural design, Structural design, behavioural design, navigation design and code design of the system.

3.1 ARCHITECTURAL DESIGN

Architectural design focuses on the components or elements of the structure. This design gives an overview of the process outlook in brief.

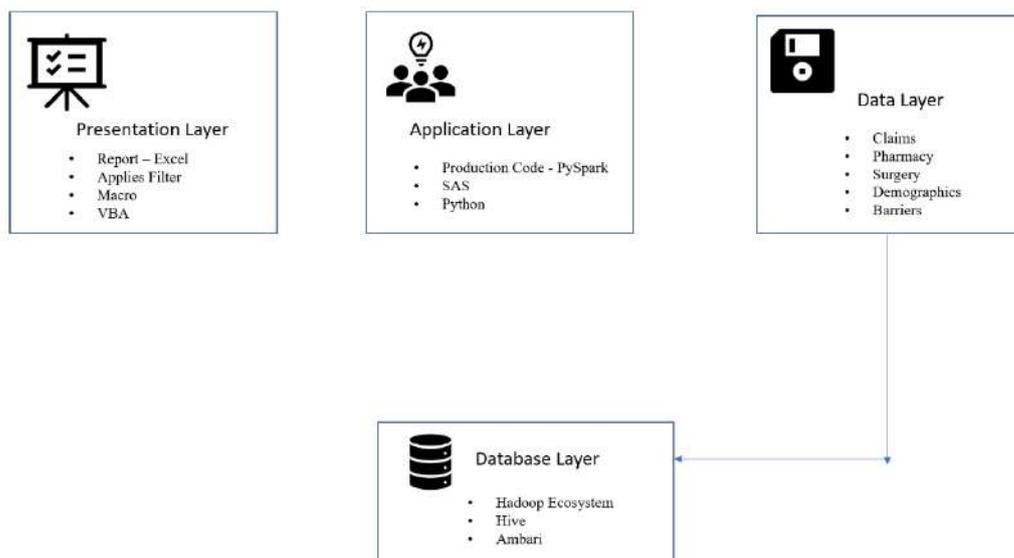


Figure 3.1 Three Tier Architecture Overview

3.1.1 ARCHITECTURAL OVERVIEW

As shown in figure 3.1, the database layer provides access to the client data and other databases required for analysis. The Hadoop ecosystem enables applications to work parallelly as containers.

The data layer gathers up together all the segments to be used for the monitoring report. Further, post to analysis and computations which is done using pyspark, in the application layer, is closely linked to the database layer where in the results are stored back into it. Earlier, this system was developed in SAS, which has results post to computations, being stored in local drives.

The Vantage and data management teams are responsible for creating profiles and organizing the data. The newer version of the system works closely with the EXL Hadoop ecosystem, which has more accessibility and saves a lot of drive space. The newer system is designed to be able to produce results which would thereby be loaded into the Hadoop ecosystem for future usage.

New System Overview:

- The new system involves computations performed and tested in PySpark
- Has a structured, organized code which generates the final resultant data set for the report
- This new system follows the principle of ‘Technology Closed Loop Framework’ which works closely with the existing technologies and complies to them
- Built in a manner that it is easier for data access and usage, also mitigating the drawbacks encountered in the earlier version of the system
- The intermediate outputs generated from the newer system is saved back into the big data ecosystem
- The workflow comparison of the earlier, or SAS version or system and the newer in the later part of this chapter

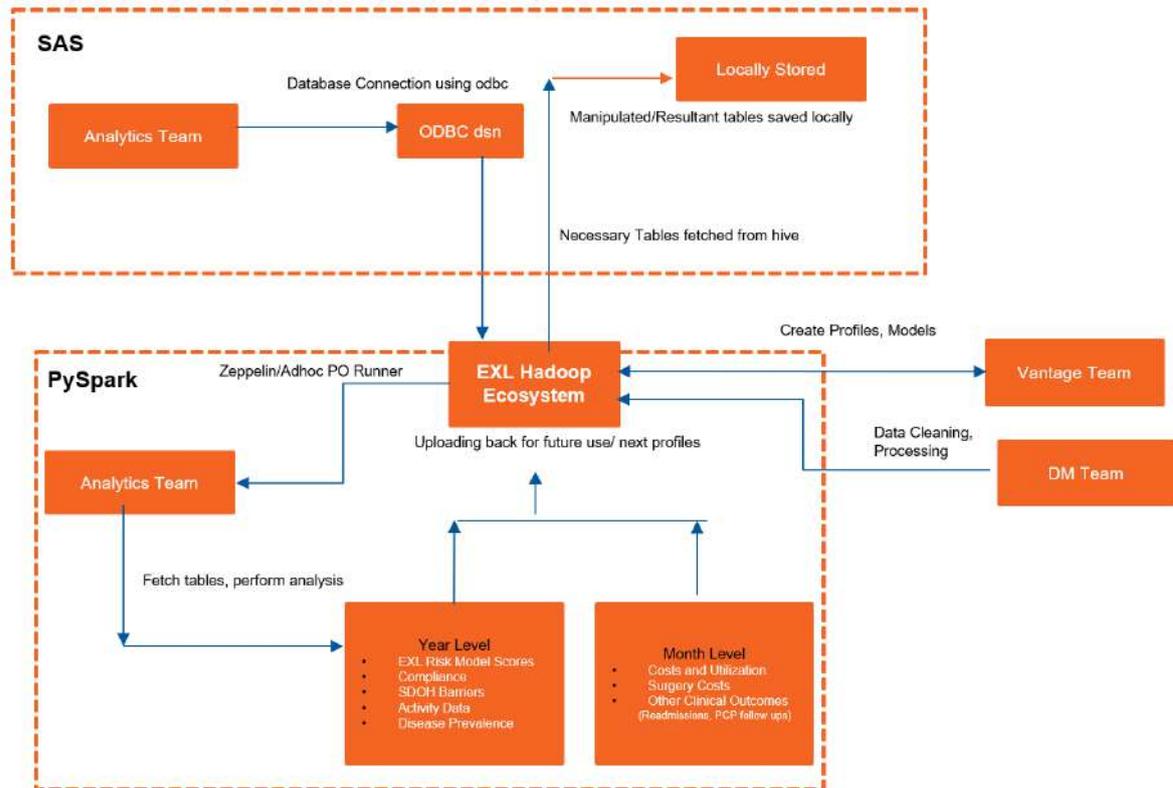


Figure 3.2 Workflow Comparison

3.1.2 BENEFITS OF MIGRATION FROM SAS TO PYSPARK

The main features that drive in migrating the SAS codes to apache spark are mentioned in the following.

- **In-Memory Computation in Spark**

With in-memory processing, we can increase the processing speed. Here the data is being cached so we need not fetch data from the disk every time thus the time is saved.

Spark has DAG execution engine which facilitates in-memory computation and acyclic data flow resulting in high speed

- **Swift Processing**

Using Apache Spark, we achieve a high data processing speed of about 100x faster in memory and 10x faster on the disk. This is made possible by reducing the number of read-write to disk.

- **Dynamic in Nature**

Parallel applications can be easily developed, as Spark provides 80 high-level operators Lazy Evaluation Outstanding feature of Spark which is call by need or memorization. It waits for instruction before providing final result which saves significant time.

- **Fault Tolerance in Spark**

Apache Spark provides fault tolerance through Spark abstraction-RDD. Spark RDDs are designed to handle the failure of any worker node in the cluster. Thus, it ensures that the loss of data is reduced to zero.

- **Real-Time Stream Processing**

Spark has a provision for real-time stream processing. Earlier the problem with Hadoop MapReduce was that it can handle and process data which is already present, but not the real-time data. but with Spark Streaming this problem can be solved.

3.2 STRUCTURAL DESIGN

Structural design deals with the solution design. This design focuses on picturizing the well-organized elements of the solution.

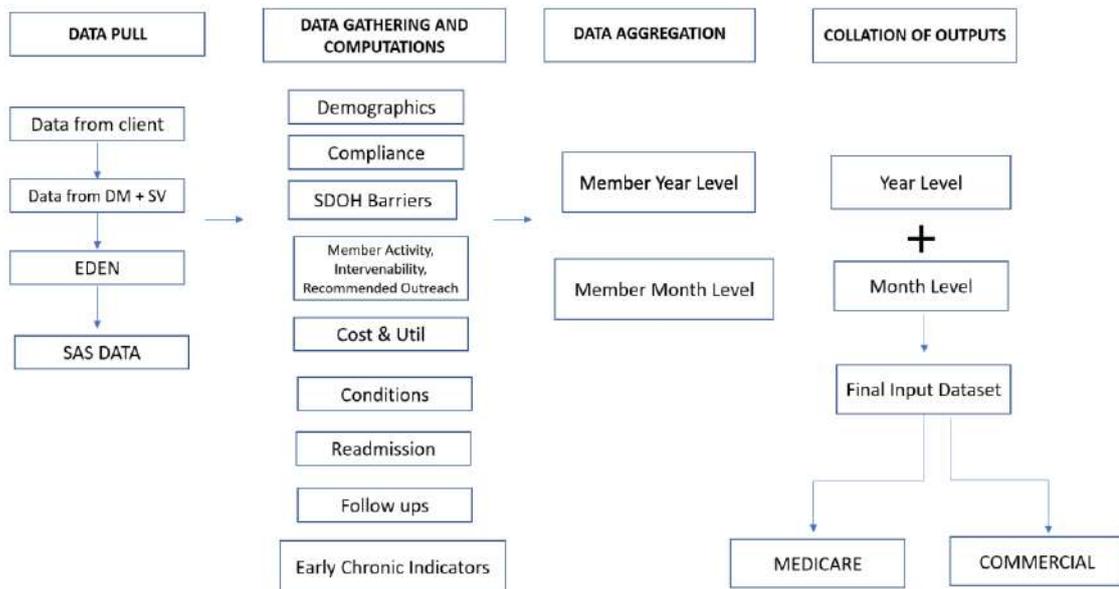


Figure 3.3 Structural Design

Structural Design Overview:

The computations and aggregations of data is categorized into both year and month level. There is a difference in both the levels, since the client/healthcare service provider gets separate insights in the individual levels as shown in figure 3.3. PMPM (Per Month per member) is a concept widely used in the healthcare insurance claims, which requires focusing population health over a period of month.

The clinical outcomes include information related to the primary care physician follow ups and readmissions which infers valuable insights to the care management team. Risk scores are better analysed with a period of 12 months and over, along with disease/condition prevalence and compliance rates.

3.3 BEHAVIOURAL DESIGN

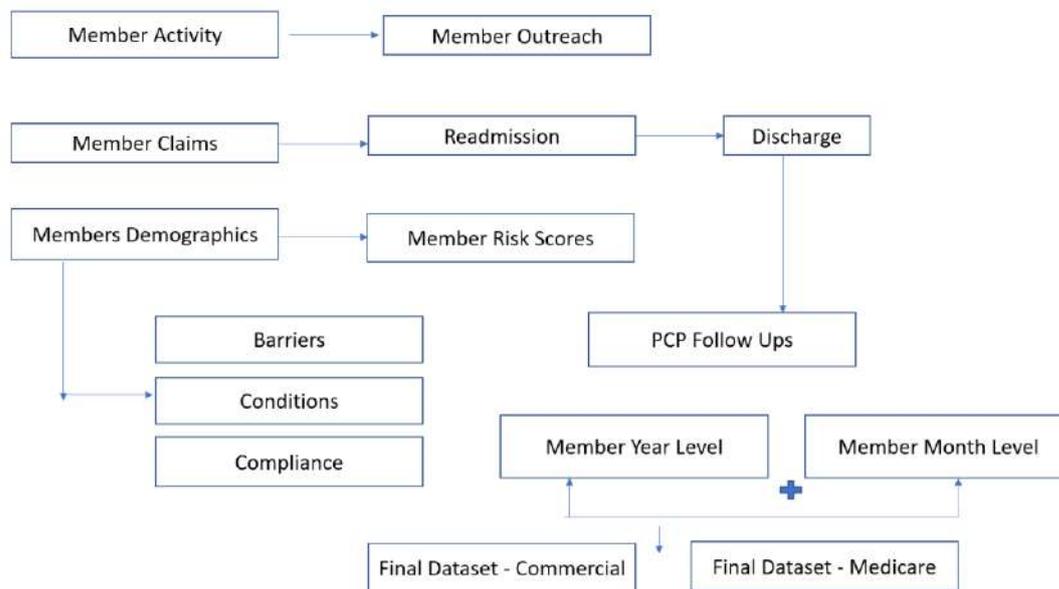


Figure 3.4 Behavioural Design

Discharge (Uses Readmission) + PCP → **Follow Ups**

Cost and Utilizations → Surgery Costs Analysis + Utilizations

Member Year Level → Member Activity + Conditions+ Risk Scores + Demographics + Barriers + Compliance

Member Month Level → Member year Level + Readmissions + Follow Ups + Costs and Utilizations

- The Model has components that has relationship within each other, which consists of deriving the later from the previous parts
- This relationship among the components and the way they interact with each other makes them reliable on each of them
- Incorrect or false computation of readmission could lead to a false value of discharge
- Which again, could affect the follow up rate which depends on the discharge data output

- Thus, the relation and interaction of the components in the system is important to recognize and to give all component the same priority
- The member month level data uses the year level data result to get the activity data insights from it. Thus, the year level data should be made sure that it denotes the right information, and should be testing for the accuracy of results generated by it
- The accuracy and adequacy of the whole system depends on its individual component's accuracy

Figure 3.3 shows how the source data is fetched for the analysis of individual components. The data is first pulled from the big data warehouse and gathered for computations for each level. Once the individual components are computed, they are tested for their accuracy. Later which, they are merged together and the output is collated in the form in which it is given as an input to the excel.

3.4 CODE DESIGN

The code generated for the system is made sure that it is efficient compared to the earlier version. Good coding practices and guidelines are followed for achieving the same

Object Oriented Designs followed:

- The system follows the object-oriented methodology in the code, which there by enables to structure and organize the individual components
- Enables reusing of code using inheritance
- Flexibility through polymorphism

- **Abstract Factory Method:**

It is a Creational design pattern which solves the problem of creating product objects without specifying their concrete classes. This technique is used in a part of the system. The abstract factory method enables interaction of methods from difference classes, which makes it easier to access a functionality or logic present in a component.

```
import pyspark
from pyspark import SparkContext
from pyspark.sql import SparkSession
sc = SparkContext.getOrCreate()

profileid = '202201'

class readmission:

    def __init__(self, id):
        self.id = profileid
        return self.id

    def compute(self):
        return 'readmission'

class followups:

    def __init__(self, id):
        self.id = profileid
        return self.id

    def compute(self):
        return "followups"

if __name__ == "__main__":

    rdms = readmission()
    flw = followups()

    print('{rdms.compute()}')
    print('{flw.compute()}')
```

Figure 3.5 Abstract factory method- Design Pattern Example

- **Adapter Design Pattern:**

This is a structural design pattern that allows objects with incompatible interfaces to collaborate.

This technique is being used in the system where the objects of classes of components which are not related, interacts. The adapter pattern has the advantage of code reusability and flexibility. Allows interface of existing class to be used as another interface

Coding practices followed:

- **Well Documented**

The code has to be well documented, so that it is well understood regarding the classes and methods used in it

- **Inline Comments**

Comments are very vital for the purpose of understanding the code, or while going through the code. Adding a line or two at each checkpoints helps users identify and understand the code flow better

- **Structured Well Organized using Object Oriented Techniques**

The code has to be organized and structured well, in order to avoid confusion. Since the current system comprises of various parts, it is vital for the code to be structured

- **Does not Produce Errors**

The code has to be error free and should generate the results as desired

- **Does not have duplication of steps/codes**

There should be no duplication of steps or functions in the code.

If so any, all of it that is similar has to be merged or made into one common function to access information from it. The existing system is made sure that it does not have any duplications

Chapter IV

SYSTEM TESTING

4.1 TESTING TECHNIQUES

Unit Testing

Unit Testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. It is performed by both developers and testers.

Integration Testing

Integration Testing is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units. Test drivers and test stubs are used to assist in Integration Testing. It is performed by testers.

Acceptance Testing

Acceptance testing, a testing technique performed to determine whether or not the software system has met the requirement specifications. The main purpose of this test is to evaluate the system's compliance with the business requirements and verify if it is having met the required criteria for delivery to end users

4.2 CURRENT SYSTEM TESTING:

Ad Hoc testing is used for the testing and validating the results generated from the system. This type of testing is a component level testing which enable better integration of components by ensuring its individual accuracy.

The flow of Ad Hoc testing for the system is as follows, in addition to the testing discussed in the earlier section (2.4)

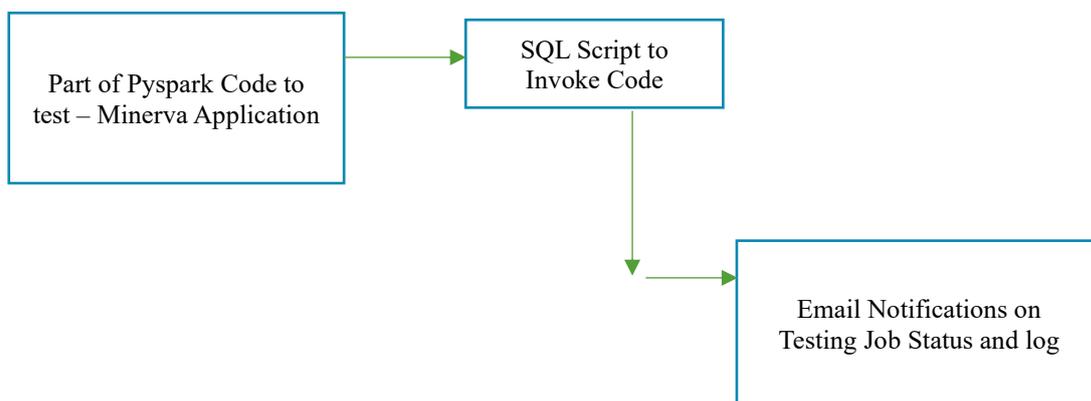


Figure 4.1 Testing Flow for the System

Purpose and Advantages of Ad Hoc Testing:

- This type of testing is performed when there is a limited amount of time
- This testing does not require pre planning, documentation or process to be followed
- Ad Hoc Testing enables detection of more defects than planned testing
- Since the system constitutes of individual components, thus the system is testing for each of its components individually for its accuracy

4.3 TEST CASES AND SCENARIOS

Test Scenario	Data Used	Expected Result	Result
Compliance	Member Claims	To Match SAS Output	Pass
Follow Ups	Member Claims	To Match SAS Output	Pass
Readmissions	Member IP claims	To Match SAS Output	Pass
Member year level	Merged - metrics	To Match SAS Output	Pass
Member month level	Merged - metrics	To Match SAS Output	Pass
Summary Dataset	Year+ Month	To Match SAS Output	Pass

TESTING CHECKPOINTS/ VALIDATIONS

- To validate the total members, count in Medicare. It should comply with the previous total Distinct Members trend
- To validate the total members, count in Commercial. It should comply with the previous total Distinct Members trend or results
- Rows Count must be validated in the summary dataset
- Presence of Clinical Metrics or Columns should be validated
- The Accuracy or reasonability of Clinical Measures are checked, Member Wise
- Total Average Costs Comparison should have minimum difference or similar to the past trends

CHAPTER V

SYSTEM IMPLEMENTATION

The System currently comprises of the Excel Report which is delivered to the client.

Client Perspective:

- The implementation of the system by the client is easier, and the pre requisite is to have Microsoft Excel
- The data or report gets refreshed manually by clicking the user created update button in the sheet prepared, which will trigger the Macro to call the VBA functions to populate the newer data, for every month of delivery
- This system could be implemented by anyone inside the organization, but implementing it outside the organization is not possible

Developer Perspective:

- The Pyspark code is executed, which has connection to the big data ecosystem
- The output datasets are uploaded to the ecosystem
- The final summary dataset for a profile is uploaded in Excel, using a macro auto update user defined button
- Using pivot, macro and VBA in Excel, the dataset is presented as a monitoring report

CHAPTER VI

CONCLUSION

Overview of System Developed:

The System thus created using pyspark generates the overall input summary dataset for the monitoring report, for both levels and lines of businesses. The System is made to test using Ad hoc for accurate generation of outputs matching with the expected ones. The newer system mitigates problems faced by the conventional system, like time taken and drive space constraints. The newer system works closely with the technology closed loop framework, to make ease of computations for further analysis. The data loss scenarios is taken care in the new system, where the data is safely stored in the big data ecosystem. The system is developed in a more organized way, to avoid confusions while generating the resultant input dataset for the report during every month of delivery. Clinical Outcome evaluation and Quality check is performed at the later stage to ensure that the system is complying to the regular value trend

Future Scope and Enhancements:

- The Excel Report could be converted into a Tableau/ Power BI Dashboard for a better interpretation and user experience for the care management team in the healthcare service provider organization/ Client – If the client wishes to switch
- The System can include another flag as a filter which would fetch the members who have engaged with the PCPs

- Introducing a separate analysis for instate and outstate members of Michigan
- Introducing and addition of new gaps in compliance rates, to make the care management team get a better overview of conditions

BIBLIOGRAPHY

1. American Academy of Family Physicians Government Relations. Government Affairs Weekly.
http://www.aafp.org/dam/AAFP/documents/advocacy/washington_update/2015/grweekly-jan-30.pdf. Accessed June 16, 2015
2. Alexandar JA, Markovitz AR, Paustian ML, et al. Implementation of patient-centered medical homes in adult primary care practices. *Med Care Res Rev.* 2015;72(4):438-467.
3. U.S. Department of Health and Human Services. Better, smarter, healthier: in historic announcement, HHS sets clear goals and timeline for shifting Medicare reimbursements from volume to value.
<http://www.hhs.gov/about/news/2015/01/26/better-smarter-healthier-in-historic-announcement-hhs-sets-clear-goals-and-timeline-for-shifting-medicarereimbursements-from-volume-to-value.html>. Accessed January 20, 2016.

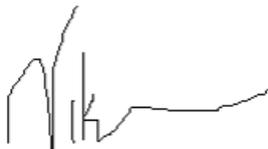
Website References:

- <https://www.exlservice.com/healthcare>
- <https://www.nlm.nih.gov/research/umls/>
- <https://machinelearningmastery.com/>
- <https://www.datacamp.com/tracks/big-data-with-pyspark/>
- <https://medium.com/>
- <https://support.sas.com/documentation/onlinedoc/>

Date: 12-05-2022

This is to certify that **Ms. Pavithra.S**, student of M.Sc. Decision and Computing Sciences (Integrated) from “**Coimbatore Institute of Technology**” has been working with **EXL Service.com (India) Private Limited** as an **Intern** for the period of 13th December 2021 to 13th June 2022. She has undergone academic training program at our organization as part of completion of requirement for the award of M.Sc. She is working on project work entitled “**Migration and Enhancement of Population Health Monitoring / Tracking Report**” during the said duration.

For, EXL Service.com (India) Private Limited



Vikram Pragadeesh

Assistant Vice President – Advanced Analytics / Data Science, EXL Health

TO WHOMSOEVER IT MAY CONCERN

20^h May, 2022

This is to certify that **Mr. Gowtham E**, a student of M.Sc. Decision and Computing Sciences from Coimbatore Institute of Technology is pursuing his final semester internship at OptiSol Business Solutions Private Limited, Chennai. The Internship commenced on 27th December, 2021 and shall be completed on 20th June, 2022.

For OptiSol Business Solutions Private Limited



Antony Justine D Julian
Senior Manager - Human Resources

TO WHOMSOEVER IT MAY CONCERN

20^h May, 2022

This is to certify that **Mr. Gowtham E**, a student of M.Sc. Decision and Computing Sciences from Coimbatore Institute of Technology is pursuing his final semester internship at OptiSol Business Solutions Private Limited, Chennai. The Internship commenced on 27th December, 2021 and shall be completed on 20th June, 2022.

For OptiSol Business Solutions Private Limited



Antony Justine D Julian
Senior Manager - Human Resources

INTERNSHIP EXPERIENCE LETTER

VG/HR/F&F/Intern/0522/003
16-May-2022

TO WHOMSOEVER IT MAY CONCERN

We are pleased to inform you that Mr Sridharan S Reg No: 1733030 Coimbatore Institute of Technology, is undergoing his Internship with Vanenburg Software (India) Private Limited from 06-January-2022 and will end on 31-May-2022.

During the period of his internship with us, Sridharan is working on the project **“Digest Email - a feature for Collabrr, a micro apps platform”**.

During his internship tenure we found his attendance and behaviour good

We wish him all success in his future endeavours.

Yours Sincerely

For Vanenburg Software (India) Private Limited



Saji Antony
Director

Certificate of Internship



Date of Issuance: 23rd May 2022

This is certify that

Ms. Swethaa Shri S

Has successfully completed 4 months (1st Jan 2022 – 1st May 2022) of internship at Satyukt Analytics Pvt Ltd as Database Designer Intern. She was intensively working on the Sat2Farm Data Management System. During the period of her internship, she was found to be punctual, hardworking and inquisitive.

A handwritten signature in blue ink, appearing to be "S Kumar", written over a horizontal line.

Dr. Sat Kumar
Chief Executive officer
Satyukt Analytics Pvt. Ltd.

A handwritten signature in blue ink, appearing to be "Yukti", written over a horizontal line.

Dr. Yukti Gill
Managing Director
Satyukt Analytics Pvt. Ltd.

May 16, 2022

To Whomever it may concern

This is to certify that **Miss. K R Sanjula** is pursuing her Internship in our organization from **13th December 2021**. She is hard working and her performance has been satisfactory.

For SCIOInspire Consulting Services (India) Private Limited



Payal N Chhabria

Assistant Vice President – Human Resource

**CLAIM ANALYSIS AUTOMATION AND
REPORTING IN HEALTHCARE INSURANCE**

SANJULA K R

Register No:1733024

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
M.Sc., Decision and Computing Sciences
OF ANNA UNIVERSITY



May 2022

**DEPARTMENT OF COMPUTING
COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution affiliated to Anna University)
COIMBATORE – 641014**

COIMBATORE INSTITUTE OF TECHNOLOGY
(Autonomous Institution affiliated to Anna University)
COIMBATORE 641014

(Bonafide Certificate)

Project Work - II
Tenth Semester

**CLAIM ANALYSIS AUTOMATION AND
REPORTING IN HEALTHCARE INSURANCE**

Bonafide record of work done by

SANJULA K R
(Register No:1733024)

Submitted in partial fulfilment of the
requirements for the degree of
M.Sc., Decision and Computing Sciences
of Anna University
May 2022



Faculty Guide



Head of the Department

Submitted for the viva-voce held on _____

Internal Examiner

External Examiner



May 16, 2022

To Whomever it may concern

This is to certify that **Miss. K R Sanjula** is pursuing her Internship in our organization from **13th December 2021**. She is hard working and her performance has been satisfactory.

For SCIOInspire Consulting Services (India) Private Limited

A handwritten signature in blue ink, appearing to read 'Payal'.

Payal N Chhabria

Assistant Vice President – Human Resource

2

SCIOInspire Consulting Services (India) Private Limited

Registered Office: TVH Belicia Towers, Tower 1, 6th Floor, Block No.94, MRC Nagar, Chennai, Tamilnadu 600028

Phone: 044 30994800

CIN: U74140TN2007PTC065322

CONTENTS

CHAPTER	PAGE NO
ACKNOWLEDGEMENT	i
SYNOPSIS	ii
PREFACE	iii
I INTRODUCTION	1
1.1 ORGANIZATION PROFILE	1
1.2 PROBLEM DEFINITION	3
1.3 DESCRIPTIVE STATISTICAL SUMMARY	6
1.4 OVERVIEW OF ANALYSIS	8
1.5 INFERENCES SUMMARY	13
II ANALYSIS PROCESS	14
2.1 DATASET	10
2.2 FUNCTIONAL REQUIREMENTS	18
2.3 NON-FUNCTIONAL REQUIREMENTS	20
2.4 US HEALTHCARE SYSTEM	21
III ANALYSIS PROCESS AUTOMATION	27
3.1 AUTOMATION MODEL	27
3.2 BEHAVIROUAL DESIGN	30
3.3 SELECTION DESIGN	33
3.4 WORK FLOW MODEL	34
IV SYSTEM TESTING	37
4.1 TEST CASES AND TEST REPORTS	37
V ANALYSIS REPORTS AND INFERENCES	41
5.1 REPORT / VISUAL FORMAT	41
5.2 DASHBOARD DESCRIPTION	42
VI CONCLUSION	49
BIBLIOGRAPHY	

ACKNOWLEDGEMENT

During the course of my internship and in the process of preparing this project, I was helped and guided by many respected persons, who deserve my deepest gratitude.

I would first like to thank **Dr. A. Rajeswari, Principal, Coimbatore Institute of Technology** for her kind patronage.

I would like to express my sincere gratitude and appreciation to **Dr. A. Kannammal, Head, Department of Computing (DCS), Coimbatore Institute of Technology** for giving me an opportunity to take up this project.

I'm also grateful to my guide **Dr. R. Umarani, Professor, Department of Computing, Coimbatore Institute of Technology** for supporting me throughout the execution of the project. I am thankful for her guidance, her invaluable constructive criticism and friendly advice during the project work.

I would thank EXL Services for hiring me as an intern and giving me the opportunity for doing this project. As the completion of this internship gives me much pleasure, I would like to thank **Mr. Karthick Anandaraman**, Associate Vice President-Analytics, EXL Service Private Limited, Chennai for sharing their expertise and more over for their timely support and invaluable guidance and suggestion till the completion the internship. I am also extremely thankful to my industry guide, **Ms. Soffiaa Anglo Charles, Senior Business Analyst**, EXL Service Private Limited, Chennai for her encouragement, support and guidance throughout the internship period.

I would also like to thank the almighty and my family, my colleagues at EXL Service and my peers at Coimbatore Institute of Technology for their constant support and encouragement.

SYNOPSIS

In today's complex health plan environment, there is a need to analyze huge data loads, coordinate payments to multiple carriers and manage a wide range of health plans across government and commercial clients. This makes payment integrity issues including fraud, waste and abuse (FWA) more common. FWA leads to higher costs for you and higher copayments and premiums for members.

In an effort to contain costs, health care organizations (HCO) are changing the way they pursue claims overpayments. Traditionally, the focus was mainly on fraud. But today, in a process more broadly labeled as payment integrity, health plans are seeking to uncover a wider range of abuse, waste and errors in claims processing and data-driven analytics is making that possible. Payment Integrity covers a broad spectrum of behavior from organized crime to simple mistakes in filling out a form to eligibility fraud.

It also covers subrogation, where another company is responsible for a piece of the claim. The objective of Payment Integrity Management is to ensure whether the claims are paid correctly. SCIO's Payment Integrity analytics and services assist health payers in controlling costs associated with incorrectly billed and overpaid claims and reducing overall adjudication and reimbursement expenditures.

Therefore, this comprehensive approach to payment integrity can help health plans to reduce administrative and medical costs and improve relationships. Learn how your health plan can move the lever and activate savings.

PREFACE

CHAPTER I

Gives a detailed description of the organization for which the system is developed along with the overview of the system, goals, and the scope of the proposed system. It gives the summary statistics and the system configuration used for this project.

CHAPTER II

Describes the nature of the dataset, functional requirements and non-functional requirements of the end-to-end system. It also gives the introduction to US Healthcare system,

CHAPTER III

Dives into the system building process i.e., the automation model along with the work flow is described and implementation process with the tool.

CHAPTER IV

Discusses about the test cases and reports made for the end-to-end system. It gives the different type of testing plan made in each unit of project.

CHAPTER V

Describes the features of the Report built and in-depth knowledge of the dashboards and how it is used across the organization for the betterment.

CHAPTER VI

Describes the conclusion of the end-to-end project built.

CHAPTER - I

INTRODUCTON

This section gives a detailed description of the organization for which the system is developed along with the overview of the project, goals and the scope of the proposed system. It also specifies the system environment used in the development of the proposed system and gives brief introduction about the various technologies used in the development of the system.

1.1. ORGANISATION PROFILE

- ✚ SCIO Health Analytics, an EXL company, is a leading health analytics solution and services company. It serves over 100 healthcare organizations representing over 130 million covered lives across the continuum including providers, health plans, Pharmacy benefit management (PBM's), employers, health services and global life sciences companies.
- ✚ SCIO provides predictive analytic solutions and services that transform data into actionable insights, helping healthcare organizations create the understanding that drives change and proven outcomes. The organization profile, work domain and various tools & technologies used in the development of the system are explained in this chapter. It also specifies the system environment used during the development phase.
- ✚ SCIO Health Analytics, an EXL company, founded in 2007, provides its healthcare partners, actionable insights and scalable solutions through deep expertise, flexible analytics, innovative services and

technology that drive measurable outcomes improvement. It focuses on the areas of population health, payment integrity, risk adjustment, care management, opportunity.

- ✚ analysis, incentive design, consumer segmentation and engagement, network performance and commercial effectiveness.
- ✚ Each year actionable insights reach more than 150 million covered lives, 406 million medical claims, 1.3 billion prescription claims, and over 60,000 pharma dashboards with 300 unique data sources refreshed monthly.
- ✚ EXL's acquisition of SCIO in 2018 expands the expertise as a strategic digital transformation partner and deepens analytics, service and technology capabilities across the healthcare continuum.

1.1. 1 Team Profile

Data and Analytics Service provides rich data assets to support client data needs, delivers end-to-end data-enabled solutions, and offers enterprise data and analytics. Our team helps clients identify, optimize, create, capture and grow business value by leveraging our data assets, analytic capabilities, and data management expertise. Risk Optimization (RO) in data and analytics services is responsible for the following modules.

✚ **Population Health**

The objective of this service is to identify the health outcomes of a group by monitoring and identifying individual patients within that group. Typically, in Population Health programs, we use a business intelligence (BI) tool to aggregate data and provide a comprehensive

clinical picture of each patient. Using that data, providers can track and hopefully improve clinical outcomes while lowering the costs.

Risk Adjustment and Quality

The identification and appropriate documentation of a disease condition and the management of care necessary to address a current disease condition or prevent a potential disease condition.

Real World Insights

Using the deep industry knowledge and advanced analytics methodologies we deliver persona and patient-centered insights and outcomes-based contracting.

1.2 PROBLEM DEFINITION

1.2.1 OBJECTIVE

Analytics is focused on two categories, the Care Analytics and the Health Analytics. Under Health Analytics comes the Payment Integrity Process which deals with Health Insurance Claims. SCIO's Payment Integrity analytics and services assist health payors in controlling costs associated with incorrectly billed and overpaid claims and reducing overall adjudication and reimbursement expenditures. SCIO acts as a vendor for Insurance companies. The main objective of the vendor is to find whether the provider has billed and claimed correctly for a subscriber bill. If it's incorrectly billed the vendor will get paid for the finding.

Initially, the provider i.e., the hospital sends the hospital bills or expenses of the subscriber i.e., the patient to the insurance companies as claims data. The insurance companies have a network of providers. The insurance companies collect those claims data from various providers. Then the data is sent to different vendors as 1st pass, 2nd pass and so on. The vendors review the data, analytics is done on the data and finally the overpaid claims are sent back to the insurance companies as findings and the vendors get paid. The workflow representation is depicted in figure 1.1.

and the vendors get paid. The workflow representation is depicted in figure 1.1

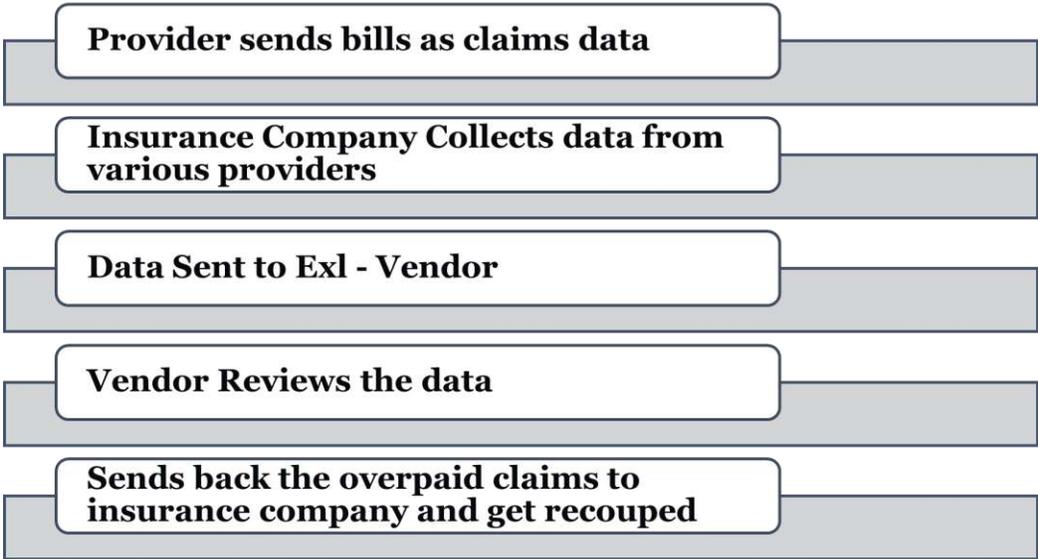


Figure 1.1 Workflow of Business

1.2.2 AIM OF PROJECT

The Aim of the project mainly focuses on creating and building the following line items in the time frame,

+ Payment Integrity Service:

Creating an operational cycle end to end system

+ Build Selections:

To ensure the claims have been correctly paid.

+ Provide Insights

Assist health payers in controlling costs and reimbursement expenditures.

1.2.3 SCOPE OF PROJECT

The scope of the project is to implement the following line items,

- + Optimize care quality while improving the medical cost trend and reducing the PMPM**
- + Track program outcomes and focus and effective plan management.**
- + Identify incorrect coding quality issue.**
- + To consolidate all the business requirements into a single system.**
- + To establish relationships between the given sources.**
- + Make the workflow dynamic for future loads.**
- + Achieve better understanding of the data and give in depth insights about the business of the organization.**
- + Integrate data from various source and load it to a data mart after performing the necessary Extract-Transform-Load operations**
- + Generate an interactive report.**

1.2.4 USERS

The users of the system are,

- ✚ Administrator
- ✚ Planning Team
- ✚ POD Leaders
- ✚ Spoke Person
- ✚ Team Leaders
- ✚ Team Members

1.3. DESCRIPTIVE STATISTICAL SUMMARY

With dataset having 400+ attributes in the table, certain key attributes such as claim details, provider details, concept details, amount spent, finding and audited are selected based on heuristic knowledge. The finding and audited attributes together derive a column named hit rate, this on further grouping based on provider forms provider-based hit rate and grouping based on concept details forms concepts-based hit rate. The descriptive statistics of certain important attributes are as follows,

Table 1.3.1 Descriptive Statistics of Provider performance

Provider performance	Values
Mean	58.02
Standard Error	0.7847
Median	66.66
Mode	100
Standard Deviation	40.79
Sample Variance	1664.14
Kurtosis	-1.478
Skewness	-0.354
Range	100
minimum	0
Maximum	100
Sum	156779.8
Count	2702

The table 1.3.1. shows the descriptive statistics of the provider performance. With the range being 0 – 100, median being 66.66% says among 2702 providers the central tendency of the performance score seems to be approximately 67%.

Table 1.3.2 Descriptive Statistics of Concept performance

Concept performance	Values
Mean	59.26
Standard Error	2.92
Median	63.157
Mode	100
Standard Deviation	36.22514
Sample Variance	1312.261
Kurtosis	-1.2705
Skewness	-0.38538
Range	100
Minimum	0
Maximum	100
Sum	9066.993
Count	153

The table 1.3.2. depicts the descriptive statistics of concept performance. With the range being 0 – 100, the median pointing to 63.157% says among 153 concept the central tendency of the performance seems to be having an approximate 63%.

1.4. OVERVIEW OF ANALYSIS

An overview of system environment, tool and the technologies used for the same are detailed in this section.

1.4.1 SYSTEM CONFIGURATION

The following are the hardware and software specifications that are used in the development of the project.

1.3.1 Hardware Specification

Processor : Intel® Core™ i5-8500M CPU
Clock Speed : 2.40 GHz
Hard Disk : 500 GB
RAM : 8.00 GB

1.3.2 Software Specification

Operating System : Windows 10 PRO 64-bit OS
Tools : Ambari, Jupiter Notebook, Tableau
Database : Hive, SQL
Language : PySpark, Python, SAS

1.4.2. TOOLS AND TECHNOLOGIES USED

A programming tool is a computer program that software developers use to create, debug, maintain, or otherwise support other programs and applications. Technology is the collection of techniques, skills, methods and processes used in the production of goods or services or in the accomplishment of objectives. The tools and technologies used in the concerned projects are detailed in this section.

1.4.2.1 Python language

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, makes it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's elegant syntax and dynamic typing, together with its interpreted nature, makes it an ideal language for scripting and rapid application development in many areas on most platforms. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. It has a wide range of applications from web development like Django and Bottle, scientific and mathematical

computing (Orange, SymPy, NumPy) to desktop graphical user interfaces such as Pygame and Panda3D.

1.4.2.2 Jupyter Notebook

The Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebooks documents. Jupyter is a JSON document, following a versioned schema and containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media, usually ending with the “. ipynb” extension. Jupyter Notebook can connect to many kernels, (by default Jupyter Notebook ships with the IPython kernel) to allow programming in many languages

1.4.2.3 PySparkA

Apache Spark is a lightning-fast cluster computing technology, designed for fast computation. It is based on Hadoop MapReduce and it extends the MapReduce model to efficiently use it for more types of computations, which includes interactive queries and stream processing. The main feature of Spark is its in-memory cluster computing that increases the processing speed of an application. Spark is designed to cover a wide range of workloads such as batch applications, iterative algorithms, interactive queries and streaming. Apart from supporting all these workloads in a respective system, it reduces the management burden of maintaining separate tools. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs. It also supports a rich set of higher-level tools including Spark SQL for SQL

and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming.

1.4.2.4 Hive

Apache Hive is an open-source data warehouse system built on top of Hadoop. It is used for querying and analyzing large datasets stored in Hadoop files. Hive gives a SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data. Hive provides the necessary SQL abstraction to integrate SQL-like queries (HiveQL) into the underlying Java without the need to implement queries in the low-level Java API.

1.4.2.5 Ambari

Apache Ambari is an open-source administration tool deployed on top of Hadoop clusters, and it is responsible for keeping track of the running applications and their status. Apache Ambari can be referred to as a web-based management tool that manages, monitors, and provisions the health of Hadoop clusters. It provides a highly interactive dashboard that allows administrators to visualize the progress and status of every application running over the Hadoop cluster. It's flexible and scalable user interface allows a range of tools such as Pig, MapReduce, Hive, etc. to be installed on the cluster and administers their performances in a user-friendly fashion.

1.4.2.6 Statistical Analysis System

SAS (Statistical Analysis System) is a software system for data inspection and report writing. Statistical Analysis System is a batch of programs that work together to reclaim them and to reserve data values, reform data, evaluate complex and single exponential analyses and generate reports. SAS is a powerful and flexible statistical package that runs on many platforms, including Windows and Unix. The SAS Programming involves first creating/reading the data sets into the memory and then doing the analysis on this data.

1.4.2.7 Tableau

A data visualization tool developed in Salesforce to connect with any database, be it SQL or MongoDB, and interact freely is called Tableau. It is mostly used in the Business Intelligence industry, and raw data is simplified easily to any format understandable by the users. Visualizations are made in the form of dashboards, and data should be represented in tabular format. It is also used in reporting and is mostly called a reporting tool. It helps to explore data, visualize and prepare reports for the same data. It helps users create different charts, graphs, maps, dashboards, and stories for visualizing and analyzing data, to help in making business decisions. It can connect to several data sources that other BI tools do not support. Tableau enables users to create reports by joining and blending different datasets. Tableau Server supports a centralized location to manage all published data sources within an organization.

1.5 INFERENCES SUMMARY

The proposed system starts by streamlining the pipeline as shown in Figure 1.1. once streamlined the claims move smoothly along the pipeline and enters the final report generating phase. The proposed system is to lend a helping hand to the auditors who manually review thousands of claims in a short-stipulated time. This system not only reduces the manual effort but also increases the selection rate and increase revenue. Hence, the proposed system optimizes the time spent by the operations team by increasing the selection rate by tapping the untapped opportunity with available human resource.

CHAPTER II

ANALYSIS PROCESS

This section describes the nature of the dataset, and techniques studied for building the model. This chapter also justifies the reason behind the use of the technique.

2.1 DATASET

There are five main data sources available for running the business. The data source is very confidential as it involves patients details along with their health information. Our client loads data to us in a periodic manner such as per month basis. We at SCIO process the data given under different programs such as DRG, RUG, Outpatient, Data mining etc. To intercept the claims data given to us, we have methods namely Clinical and Non-Clinical under which programs can be classified. Programs can be classified into two types based on their review and turnaround times we receive the file.

- ✚ **Pre pay:** The client sends the data before paying the provider. Hence the pre pay programs must have a TAT of within 3 days and reviews are held on a daily basis.
- ✚ **Post pay:** The client sends the data after paying the provider. Hence the post pay programs have a much greater TAT ranging from 2 week to 3 months.

To understand this interception better, we need to know about medical records and Contracts. We initially receive data called as Electronic Health record (EHR) although it might give a lot of information about that particular claim it doesn't record information right from the

beginning of the subscriber’s visit to the hospital. Contracts are the agreement between the Payor and Subscriber or Provider regarding the insurance plan. For example, if a person has an insurance that covers only 80% of his medical expenses and if the payor has paid more than 80% percent of his bill, we can identify that by just reviewing the contract. A claim can be reviewed from both a Clinical and Non-Clinical perspective as depicted in figure 2.1

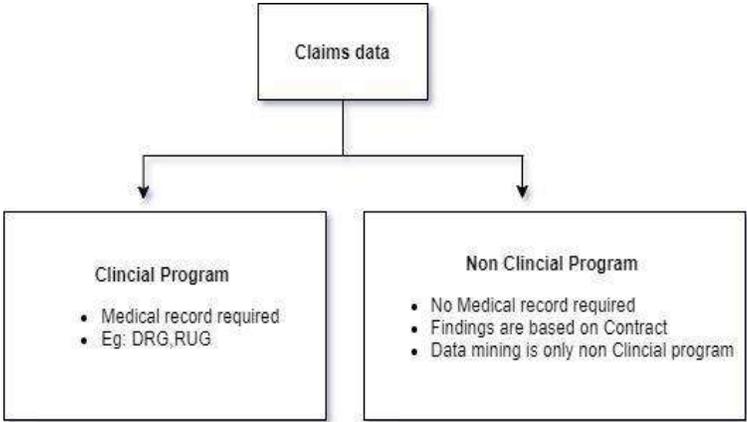
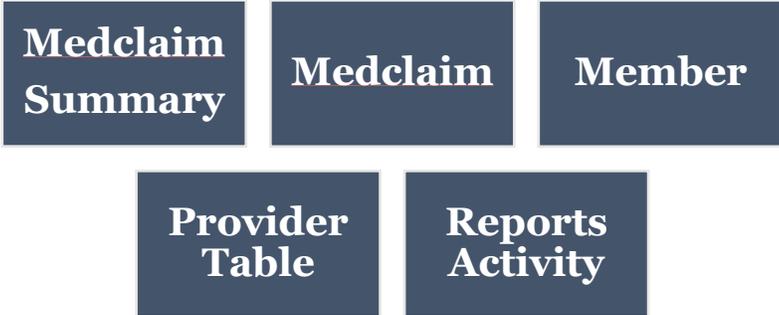


Figure 2.1 Clinical and Non-Clinical Program

2.1.1 The Five Major Data sources are



Medclaim Summary Dataset

Medclaim Summary table has the summary level details regarding the claim. It has details like claim number, patient, who enabled claim, the provider, Total paid amount etc. One claim number may contain several services, so these will be group together claimsummaryid (Primarykey), medcalimnumber, clientnumber, platformcode (system to use code), Line of business- Medicare and Medicare or commercial.

Medclaim Dataset

Medclaim table has the line level information like the amount paid, what kind of service given, what procedure they have undergone, different types of codes to identify the treatment, dates related to treatment, recent version indicator, Point of admission indicator etc.

Member Dataset

Member tables has the details of members. The main attributes are Memberid (primary key) and Demography details of member stored in another table like address, age, phone number etc.

Provider Table

Provider Table has details of all providers (hospitals, clinics, doctors etc.) and attributes like the address, Taxid, contact number, providerassigneid. Here the providerassigneid and providertaxid together act as primary key for the table.

Reports activity

This table contains other vendor details and their activities related on the claims. The activities related to the findings and no findings of the claims will be given along with the order and the process date.

2.1.2 DATA INTERPRETATION AND PIPELINE

The objective of this section is to explain briefly regarding the data flow in our organization. The flow chart below gives an idea about that. After our operations team reports a particular claim as a finding, they will update it in the client system. The Client reviews it and if found as a finding a particular percentage is paid as a commission to us.

The Data Pipeline is represented in figure 4.3

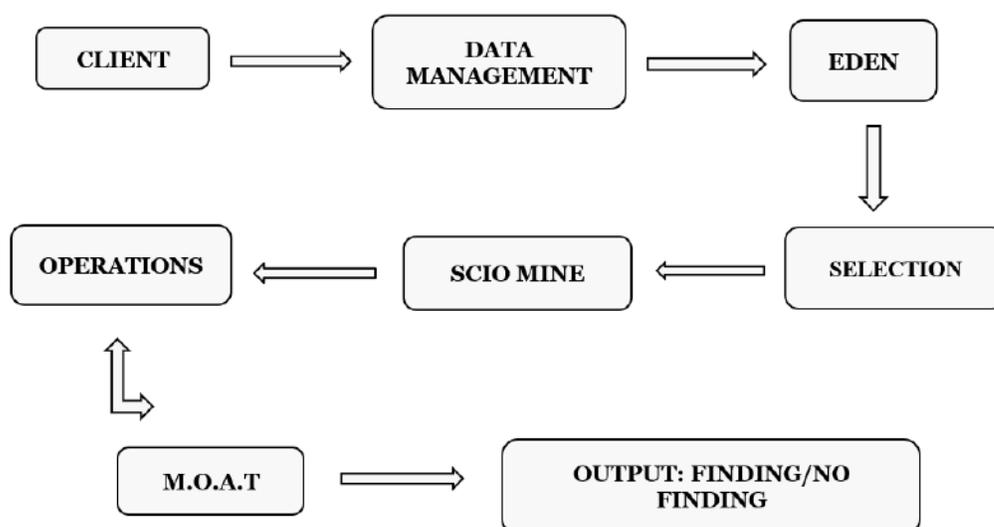


Fig 2.2 The Data Pipeline

Health plans are seeking to uncover wider range of abuse, waste, and fraud which are happening during claim processing. Therefore, we EXL (health care service company) team leverage our deep domain and health care knowledge to help our customers by providing them analytical and technology enabled services to the clients by providing them valuable insights.

2.2 FUNCTIONAL REQUIREMENTS

The system should satisfy the functional requirements as depicted in figure. Design is essentially the bridge between requirement specification and the final solution for satisfying the requirements. Figure 2.3 depicts the proposed design and functional requirements of the member list outreach. Along with that, the Data repository, Analysis service and Reporting functionalities for the end-to-end system is maintained and configured.

2.2.1 DATA REPOSITORY

The repository database is an intermediate database stage and mart where the cleansed data is placed in the repository and further flattening hierarchy and constraints inclusion happens in the repository database. It also serves as a replica of the mart database structure. The only difference between the repository and the mart database is that manipulations and further transformations can be done in the repository database whereas the mart database is just the final structure upon which no further manipulation is done.

2.2.3 ANALYSIS SERVICE

The Analysis Services should hold aggregated data in the structures necessary for the reporting requirements of the system. The tables should be created for each logical dimension required for the system. The analysis is done upon the views, to avoid the unnecessary chaos to the actual data, also adding new calculated column in a view would do no harm compared to that of adding it to actual one. It is made sure that the cube should have all calculations needed for the report.

2.3.4 REPORTING

The front end for the analysis services is the Excel 2019 Pivot tables. Any user who has a reporting need should be able to build a report as they wish. The user has the flexibility to view the profits and ranges across many dimensions. Some of the reports produced are trend report, the gross profit drivers report, missed opportunity Reports.

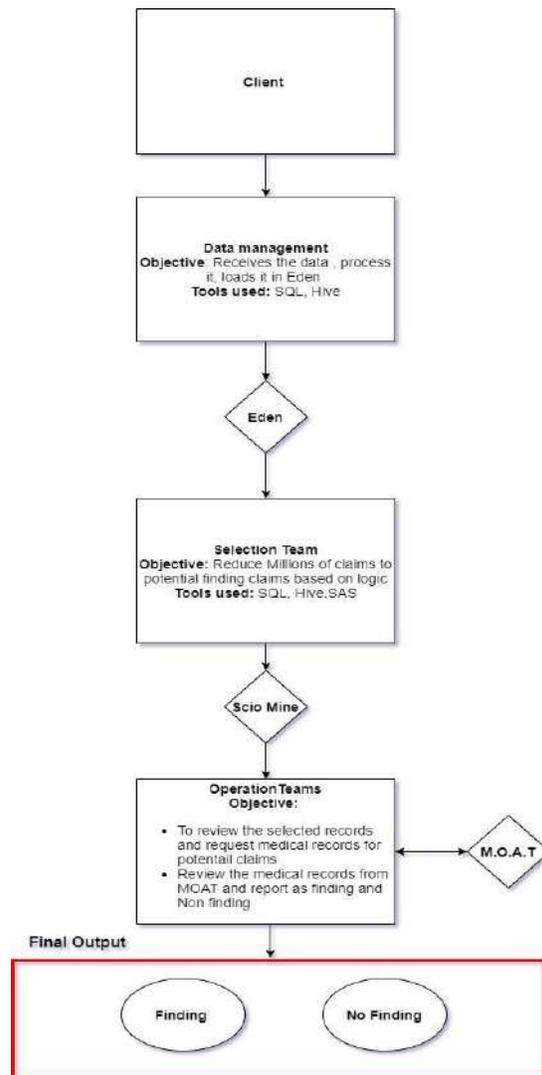


Figure 2.3 Functional Requirement Design

2.3 NON-FUNCTIONAL REQUIREMENTS

SCALABILITY

The database will hold the current dimension data and the fact data with a year of margin and transactional data until the system is stabilized within 6 months of go live.

AUDIT AND CONTROL

Auditing and error handling must be done throughout. Errors have to be captured in an error logging table which the users should be able to view in the administration module and support teams should be accessible to view and trouble shoot.

TRACEABILITY

All the tables in the data mart should have identifiers to ensure that all data can be traced back through the system and maintenance and support is as straight forward as possible. The data in the repository should be easily traced to the data and time of load using load references.

ACCURACY

The accuracy of the data is of highest importance as this data will impact the decisions to be taken in the client organization and business driven solutions are taken from that. The Accuracy of data is high mandatory.

SECURITY

Security will be set by the client and the internal team as data is confidential because it has member details of the patients.

FLEXIBILITY

The system will include flexibility to move the system between servers and change database names as required by the client and the program.

2.4 US HEALTHCARE SYSTEM

Healthcare is delivered by practitioners in medicine, chiropractic, dentistry, nursing, pharmacy and allied health. The delivery of modern healthcare depends on an expanding group of trained professionals coming together as an interdisciplinary team. Healthcare in US involves the following and the pictorial representation of the same is shown in figure 2.4.

- ✚ Patients and Doctors
- ✚ Nurses
- ✚ Dentists
- ✚ Hospitals and Nursing Homes
- ✚ Manufacturers of Healthcare Equipment and Devices
- ✚ Pharmaceutical and Life Sciences Industry



Figure 2.4 Healthcare in US

2.4.1 Participants in US Healthcare

The main participants of US Healthcare are patient, provider, payor, supplier and researcher.

Patient:

A patient is a person who receives medical attention, care or treatment. One who is visiting a physician for a routine check-up may also be viewed as a patient.

Provider:

A provider is a person who is qualified to deliver proper health care in a systematic and professional manner to any individual in need of health care services. Providers include two categories, Physicians and Facilities.

Physician is of two types,

- **Primary Care Physician** – Provides general medical care and treats routine problems.
- **Specialty Care Physician** – Has advanced medical training and is certified to practice in a specific field.

There are different types of facilities as follows,

- **Hospitals** – Provides comprehensive and intensive medical services for acute and chronic care.
- **Ambulatory Care Centers** – Freestanding facility other than a physician's office that operates exclusively to provide surgical services to patients who do not require hospitalization.

- **Home Health Care** – Care given to a person in their own home to allow a person suffering from an illness/injury to remain at home without being moved to a nursing care or rehabilitation facility.
- **Hospice** – Palliative and supportive care for terminally ill people to spend the final stages of their lives at home or in home-like settings.
- **Skilled Nursing Facility (SNF)** – care of individuals in a residential facility, usually there on a long-term basis.

Payor:

A payor is someone who assumes the risk of paying for medical treatments. They can be entities such as Third-Party Administrators (TPA), Self-Funded/Self-Administered Employer Groups and Insurance carriers. Suppliers are organizations that sell healthcare products to providers to be used in the delivery of healthcare. Researchers are people or organizations that use scientific methods to discover the cause of illness, methods of treatment or ways to avert illness. This research leads to technological advances in healthcare. Examples of Providers are Physicians, Nurses, Hospitals, Nursing homes and examples of Payors are Medicare, Medicaid, Blue Cross Blue Shield, Insurance Companies. The participants of US Healthcare are depicted in figure 2.2.

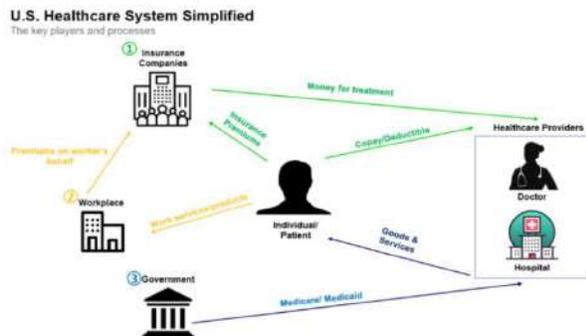


Figure 2.5 Participants in US Healthcare

2.4.2 Government Based Health Insurance Plans

Government Based Health Insurance Plans mainly fall into 2 categories –

- ✚ Medicare
- ✚ Medicaid.

The pictorial representation of the same is shown in figure 2.6.

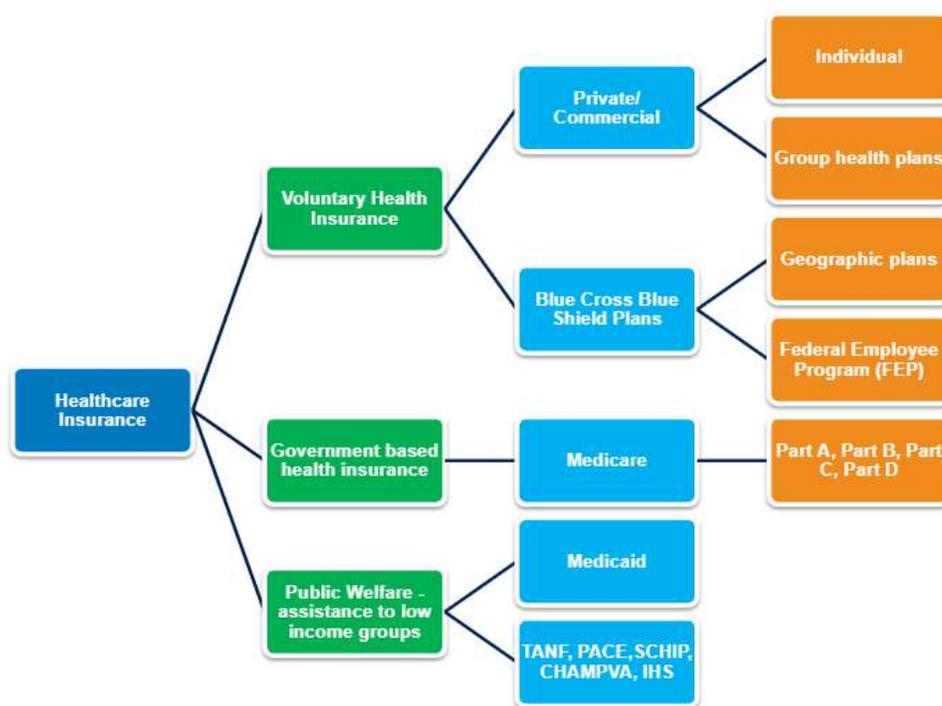


Figure 2.6 Types of Health Insurance Plans

Medicare

The Health Insurance for the Aged and Disabled Act (title XVIII of the Social Security Act), known as "Medicare," is a broad program of health insurance designed to assist the nation's elderly to meet hospital, medical, and other health costs. Medicare is health insurance for people

65 or older, for people under 65 with certain disabilities and for people of any age with End-Stage Renal Disease.

Medicaid

Medicaid was created on July 30, 1965 (title XIX of the Social Security Act). Medicaid is a joint federal-state program that provides health coverage or nursing home coverage to certain categories of low-asset people, including children, pregnant women, parents of eligible children, people with disabilities and elderly needing nursing home care. Medicaid was created to help low-asset people who fall into one of these eligibility categories.

The Centers for Medicare and Medicaid Services (CMS) set out the general rules under which Medicaid operates, each state runs its own program.

Under certain circumstances, an applicant may be denied coverage. As a result, the eligibility rules differ significantly from state to state, although all states must follow the same basic framework. Each state administers its own Medicaid program while the federal Centers for Medicare and Medicaid Services (CMS) monitors the state-run programs and establishes requirements for service delivery, quality, funding, and eligibility standards.

Insurance is a way of guarding against the financial risks that are always present in life. A health insurance policy is a contract between an insurer and an individual or group against expenses incurred through illness of the insured. The flow between the payor, provider and the subscriber is represented in figure 2.7.

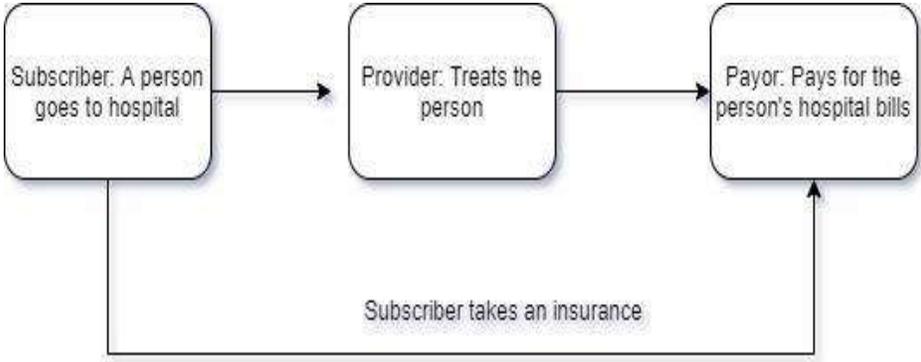


Figure 2.7 Flow Between Participants

CHAPTER III

ANALYSIS PROCESS AUTOMATION

This section dives into the model building process, the tools used for the same. Architecture and its layers are looked deeper and how the problem has been approached. It identifies the software components, specifies relationships among components, properties of both components and relations, defines program structure and provides a blue print for implementation.

The following chapter deals with the various design issues that guides the interface development of the application. System design is also been explained, Systems design is a set of procedures performed to convert the needs and requirements of a business or organization into a design that can be implemented on the organization's computer system.

3.1 AUTOMATION MODEL

The analytics model used to develop the integrated system is **Cross Industry Standard Process for Data Mining (CRISP-DM)**.

CRISP-DM breaks the process of data mining into six major phases:

-  Business Understanding
-  Data Understanding
-  Data Preparation
-  Modeling
-  Evaluation
-  Deployment

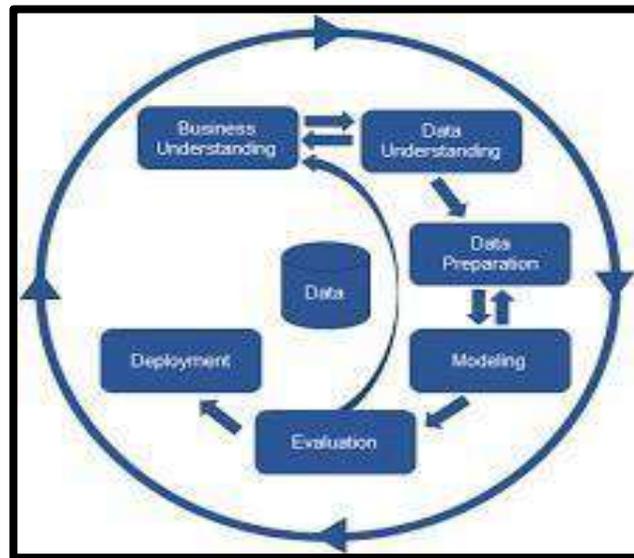


Figure 3.1 CRISP DM

The sequence of the phases is not strict and moving back and forth between different phases as it is always required. Figure 3.1 gives the process diagram of the approach. The arrows in the process diagram indicate the most important and frequent dependencies between phases.

The outer circle in the diagram symbolizes the cyclic nature of data mining itself. A data mining process continues after a solution has been deployed. The lessons learned during the process can trigger new, often more focused business questions, and subsequent data mining processes will benefit from the experiences of previous ones.

The CRISP-DM approach is used in the developing the system. The model was chosen because of both its focus on both technical and business sense.



Figure 3.2 Architectural Design of end-to-end system

The figure depicts the architecture design for the end-to-end system. The architecture has the following layers.

- **Operational Database layer:** The source data for the data warehouse - an organization's enterprise resource planning systems and the data management team fall into this layer.
- **Data Access layer:** The interface between the operational and informational access layer -Tools to extract, transform, load data into the warehouse fall into this layer after satisfying the requirements.
- **Metadata layer:** The data dictionary -this is usually more detailed than an operational system data dictionary. There are dictionaries for the entire warehouse.
- **Informational Access layer:** The data accessed for reporting and analyzing and the tools for reporting and analyzing data -this is also called the data mart. Business intelligence tools fall into this layer.

3.2 BEHAVIORAL DESIGN

Behavioral models show the dynamic behavior of the system and the process flow as it's executing. They show what happens or what is supposed to happen when a system responds to a sequence flow.

3.2.1. CLAIMS MANAGEMENT

The Claims Management Process involves a sequence of steps as shown in figure 3.3.

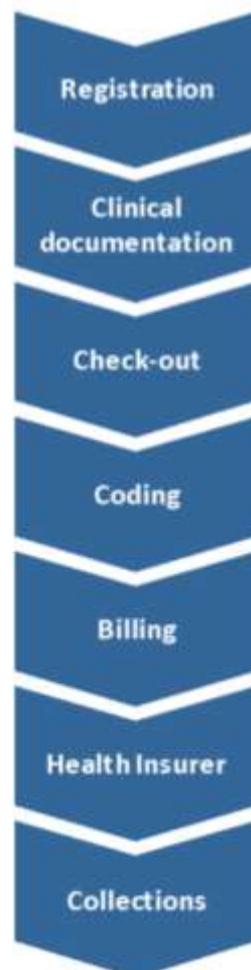


Figure 3.3 Claims Management

The steps are explained in detail below.

Step 1: Patient Registration

The first step is patient registration. The registration staff should collect the patient's demographic information and health insurance information and accurately enter it into the practice management database. The registration staff should confirm the patient's benefits, applicable deductibles and/or co-payments by calling the health insurer or through online verification. The registration staff should make a copy of the patient's health insurance card to obtain his or her health insurance information.

Registration staff can enter this information in their health insurer reference log and health insurer follow-up log for future reference. The registration staff should also verify whether a returning or established patient has had any change in his or her health insurance information. The registration staff should also give the patient a copy of the physician practice's payment and privacy policies during check-in.

Step 2: Clinical Documentation

The treating physician and/or clinical staff should document the patient's history, symptoms, diagnosis and treatment plan including appropriate tests that may be ordered in the medical record. The treating physician and/or clinical staff should assign the appropriate International Classification of Disease – 9th Edition – Clinical Modification (ICD-9-CM) code(s) and CPT code(s), document these codes in the medical record, and record code information on the physician practice's super bill.

Step 3: Check Out

The registration staff should collect the patient's balance (e.g., deductible, co-payment) and schedule the next appointment in an order and record it.

Step 4: Coding

The coding professional should verify and review the codes the physician and/or other clinical staff provide, based on the documentation in the medical record. The coding professional should contact the health insurer for pre-authorization, pre-certification or pre-determination of the patient's benefit coverage prior to a procedure or service. The coding professional should document the health insurer's authorization number and supporting documentation and forward this information to the staff responsible for billing.

Step 5: Billing

The billing staff should enter the codes and fees accurately as they appear on the physician practice's super bill or patient encounter form and then generate a paper or electronic claim. The billing staff should review each claim for completeness and accuracy before submitting it to the health insurer.

Step 6: Health Insurer

The health insurer should process the claim, and if they approve it, they should route a payment to the physician practice along with a copy of the EOB. The health insurer should route the original EOB to the patient.

Step 7: Collections

The collections staff should follow up with the health insurer after submitting the claim to verify that the health insurer received the claim and ensure that they are processing it. The collections staff should verify the payment and post it in the physician practice's accounts receivable. If the collections staff deems the payment inappropriate, they should investigate why the health insurer did not pay the claim appropriately and determine whether they should appeal it.

3.3 STRUCTURAL DESIGN

Structural design is the methodical investigation of the stability, strength and rigidity of structures. The basic objective in structural analysis and design is to produce a structure capable of resisting all applied loads without failure during its intended life.

3.3.1 DATA FLOW DIAGRAM

A data flow diagram (DFD) is a significant modeling technique for analyzing and constructing information processes. DFD literally means an illustration that explains the course or movement of information in a process. DFD illustrates this flow of information in a process based on the inputs and outputs. A DFD can be referred to as a Process Model.

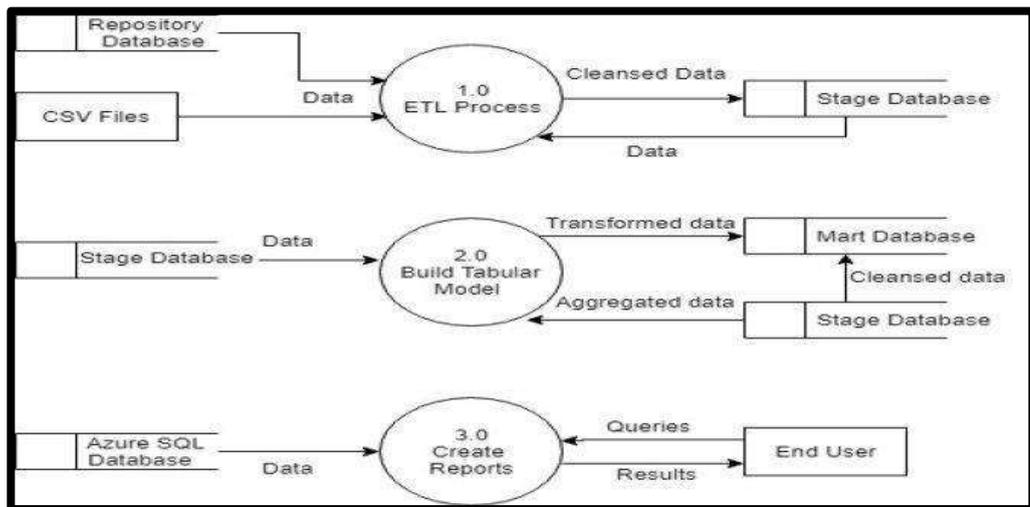


Figure 3.4 Level 1 Data Flow Diagram

It shows how the data flows into the main subprocess and datastores.

3.4 WORK FLOW MODEL DESGIN

The work flow model of projected is depicted in figure 3.5.

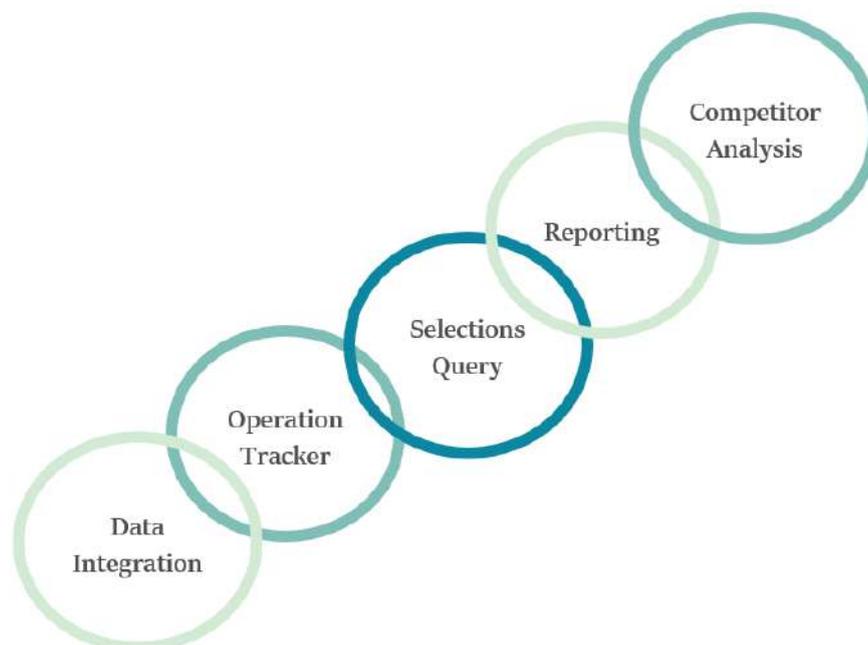


Figure 3.5 WORK FLOW MODEL

3.4.1 Data Integration:

Integrating multiple data sources and analyzing the requirement columns for the whole process of building a system is the initial step. This data integration step involves

- ✚ Requirement gathering according to different programs
- ✚ Mapping the attributes as per the idealization.
- ✚ Generating coding script to integrate the data

3.4.2 Operation Tracker:

Operation tracker comes into the process once after the data is integrated according to the requirements and the programs. This Operation tracker tracks the Selection rate, Rejection rate, received rate, Moat accepted rate, the file tracker. This tracker is received for both posts pay and prepay. Rejected reasons for each program is a Separate Dashboard sheet to know about the rejected reasons for the selected program

3.4.3 Selection Query Model

The process of selections plays a vital role in bucketing a claim as a finding and a no findings. This particular process involves several stages. It is depicted in figure 3.6.

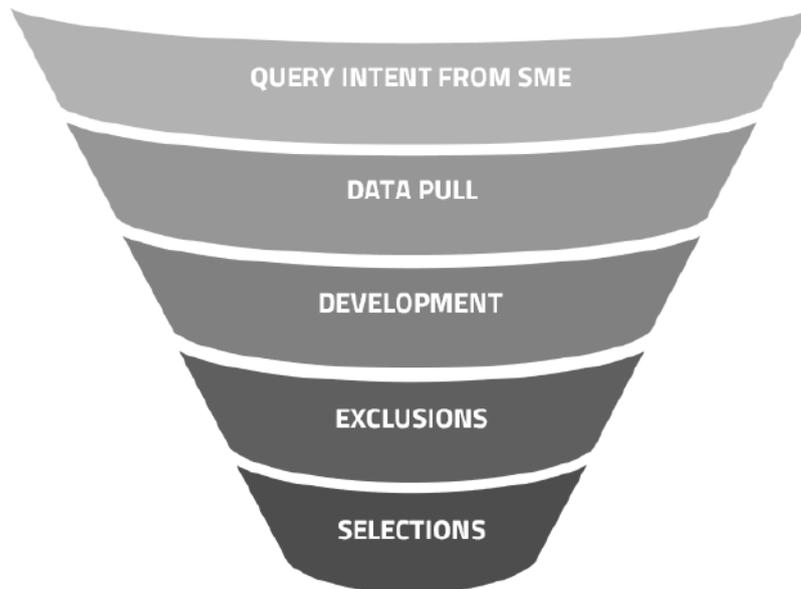


Figure 3.6 Selection query model

+ Query Intent from SME:

The subject matter experts will send the query intent via mail and tracker using QAT application. The intent of a query will be sent after an in-depth analysis of claims according to the set of codes like procedural code, HCPC code, diagnosis code, CPT code which is verified in terms of medical standards.

+ Data Pull:

Once the intent is understood, the original data pull happens in which the claims are in line level detail as well as summary level detail.

+ Development:

From the intent and data, the query is developed after an analysis. The query which is developed will be sent for a review to SME's. One the review is done the query will be pushed into production after testing.

Exclusions:

A set conditions and Instructions will be sent by the client mentioning that few set of claims should come under selection criteria. For those respective instructions query is developed and the procedure is run in which the corresponding claims will be excluded and do not come under selection criteria.

Selections:

The final queries after the development phase are moved into production. The production script is run for the selections process. Once the selection is completed it is uploaded into client system, which is further process for auditing.

3.4.4. Reporting

Reporting is the process of Analyzing the Model on different phases

 **Inventory analysis:** To track the inventory (selections reasons) present in the particular program.

 **Audit Analysis:** To track different audit status happening right from claim received till recoupment.

 **Outpatient Analysis:** To track performance of different outpatient programs across the clients.

The detailed description of reporting is described in chapter 5 (Analysis reports and Inferences)

3.4.5 Competitor Analysis

Competitor Analysis is the process of analyzing the other vendor activities for individual claim. It keeps track of findings marked by other vendor which will be a missed opportunity for EXL. This is done to increase the revenue and the EXL missed count. Therefore, the main objective is to track the number of no findings claims marked by findings by other vendors.

CHAPTER IV

SYSTEM TESTING

System testing is a generic metric of performing a variety of tests on a system to explore functionality or to identify problems. A series of systematic procedures are referred to test how the system should perform and where common mistakes may be found by entering data that may cause the system to malfunction or return incorrect information. The purpose of testing is quality assurance, verification and validation, or reliability estimation.

A Test Plan documents the strategy that will be used to verify and ensure that a product meets its design specifications. The three major elements that constitute the test plan are Test Coverage, Test Methods, and Test Responsibilities.

4.1 TEST CASES AND TEST REPORTS

Test cases are built around the requirements and specifications i.e., what the system is supposed to do. Test cases are generally driven from the external descriptions of the framework and design parameters.

Each test case contains item criteria such as:

PASS

All expected results are achieved and/or all unexpected events are resolved.

FAIL

Testing process response does not confirm the expected result.

4.1.1 UNIT TESTING

Developers write individual module's test cases to test various functionalities of a module. Manual testing was done.

Table 4.1 Test Cases for Stored Procedures for all the Dimension Loads

No.	Test case	Actual Output	Pass/Fail
1	Implementing start log at the start of the package	A new record is generated in the LoadLog_log table and the status is sent to 'Running'	Pass
2	Implementing end log with status as 'Success'	After the package runs successfully, the end time is updated in the dbLoadlog table and status is set to 'Success'	Pass
3	Implementing end log with status as 'Failure' in the 'On Error' event handler	If the package fails, the end time is updated in the dbLoadlog table and status is set to 'Failure'	Pass
4	Implementing write log after every task using 'On Post Execute' error handler	The dbLoadDetailLog table is updated for every task that runs in the package, for both Success and Failure cases	Pass

4.1.2 INTEGRATION TESTING

For use integration test bottom-up approach was used where the lowest level components were tested first, and then used to facilitate the testing of higher-level components. For use integration tests are basically the same as the unit tests just without using the stubs and drivers. Integration testing is performed on functional units which are integrated at same level or at different Levels. It basically checks the coupling within the unity Functionalities

Table 4.2 Test Cases for Pipeline overall load

No.	Test case	Actual Output	Pass/Fail
1	Running a successful end to end load	and the logs are as expected, row count is logged and stage table is truncated before each file is loaded to it	Pass
2	Check If any other Test case is Running before starting the package.	The package runs for only when there are no running procedure.	Pass

4.1.3 SYSTEM TESTING

4.1.3.1 FUNCTIONAL TESTING:

Under this the whole system is tested under the development team. Each of the functionalities as per requirements is tested here. This will be according to business requirements.

4.1.3.2 NON - FUNCTIONAL TESTING:

Under this the whole system is tested by the development team and all the quality related aspects are tested to confirm with the non-functional requirements of the system.

CHAPTE V

ANALYSIS REPORTS AND INFERENCES

Report analysis and inference is the key work of data scientists. The insights achieved by the analysis model can easily be depicted with the help of dashboards. This section details the same.

5.1 REPORT / VISUAL FORMAT

The dashboard is built using Tableau. The data sources used are on SQL Server and Hive environment.

The data sources used for reporting are,

-  Medclaim Summary Table
-  Medclaim Table
-  Member Table
-  Provider Table:
-  Reports activity Table

The objective of building the dashboards and monitoring them is to gain insights that can increase the revenue for the company and provide insights to clients upon which business driven solutions re made.

5.1.1 TABLEAU DASHBOARDS

There are three target groups for whom the dashboard is created.

-  Board Members of the company who would like to see high level details (Audit Dashboard)
-  Program Category Managers who are interested in their portfolio (Detailed Report suggesting the missed opportunity details)

- ✚ Data Management Team in who wants to tracks the file received rate from the clients to ensure the threshold is met each time (Prepay Dashboard)
- ✚ Developer who wants to see the performance of selections (Outpatient Dashboard)

5.2 DASHBOARD AND REPORTS DESCRIPTIONS:

The Dashboards and Reports generated for the end-to-end system is discussed in this selection along with the scop and objective of each dashboard report. The sample snapshot dashboard is shown in figure 5.1

5.2.1 OPERATION TRACKER:

This Dashboard is built to monitor Humana prepay selections by

- ✚ Rejection number, Rejection percentage, Rejection reason
- ✚ Claims Received count
- ✚ Claims selected count and the percentage
- ✚ Claims Accepted count and the percentage
- ✚ Blank count and the percentage

Layers of Dashboard

✚ Rejection Rate (Layer 1)

Shows the MOAT rejection percentage on daily basis for all the programs

✚ MOAT Summary (Layer 2)

Shows the overall counts of claims received, loaded to MOAT, Accepted & rejected

✚ Prepay File Calendar (Layer 3)

Helps to find the files received, not received of the days for the selected month.

✚ Rejected Reasons (Layer 4)

Rejection reasons for each program will be displayed and the number of claims under each rejection reason will be mentioned. Files which have 80% and above rejected claims will be highlighted in red)

Users:

- ✚ SPOC and POD leads review the Selection/Exclusion numbers
- ✚ Reach out to Client - on files not Received and if threshold is not received.
- ✚ Track the rejection rate and can reach the POD SMEs to improve logics.
- ✚ Rejection/Selection trend or the pattern can be identified

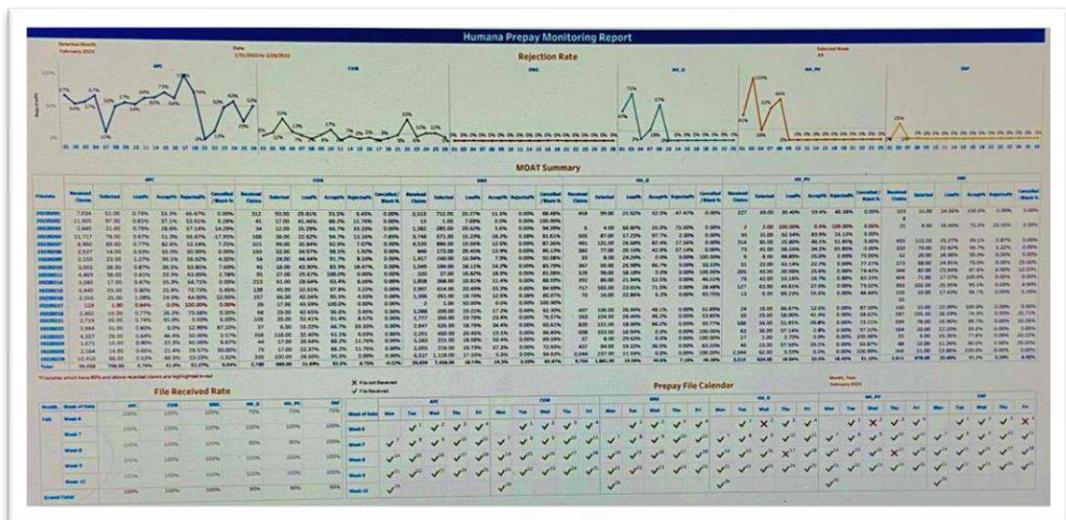


Figure 5.1 Operation Dashboard Sample Snapshot

5.2.2 OUTPATIENT DASHBOARD

The main objective of this dashboard is to track performance of different outpatient programs across the clients. The dashboard is classified into 4 analyses they are,

- + **Main dashboard:** The main overall dashboard gives us the overview view across programs and clients along with important KPI.
- + **Query dashboard:** Query dashboard gives the performance of queries that are in production along with new selections and old selections time frame.
- + **Provider dashboard:** Provider dashboard gives the provider wise performance along the key attributes and the overall view.
- + **Major call outs:** Major call out sheet gives Analysis is which are computed according to the requirements.

The sample snapshots of outpatient dashboard are given in figure 5.2 and figure 5.3

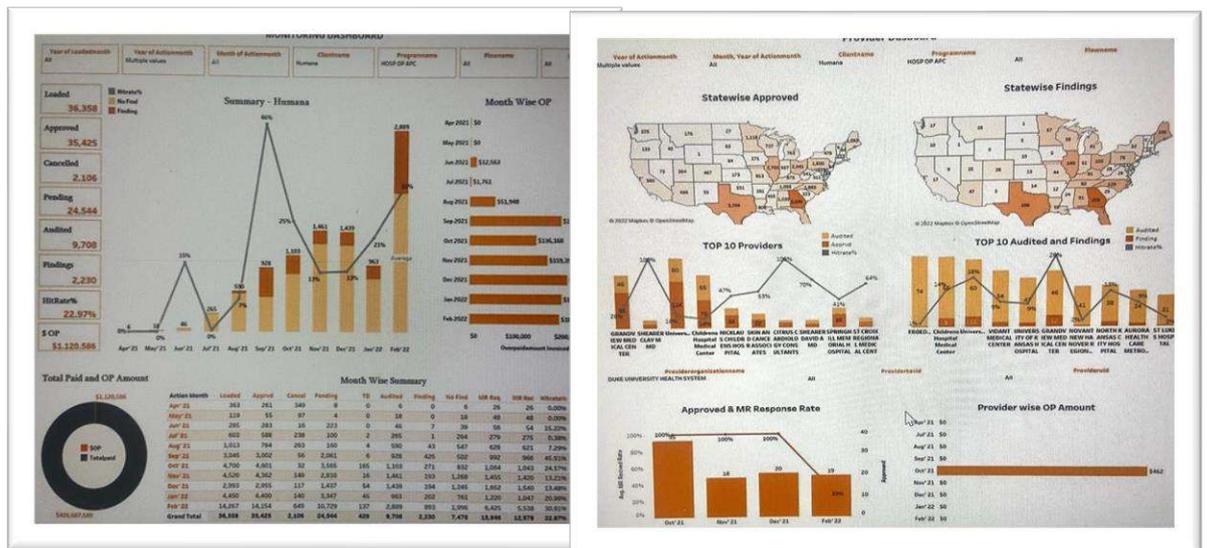


Figure 5.2 Outpatient Dashboard Sample Snapshots 1

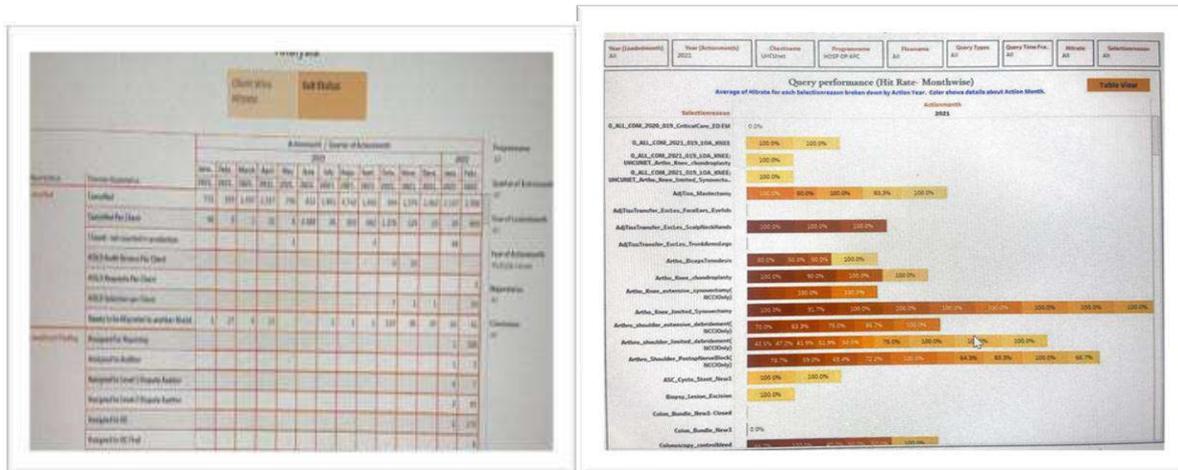


Figure 5.3 Outpatient Dashboard Sample Snapshots 2

5.2.3 Inventory Dashboard and Audit Dashboard

The main objective of inventory and audit dashboards are to track the inventory (selections reasons) present in the particular program and to track different audit status happening right from claim received till recoupment. It gives selection reason wise performance metrics like Hit Rate, OP, OP/finding etc., and the analysis of the number of claims in different statuses based on selection reasons. It has two parts which is explained below,

Part 1: Inventory Optimization

This is for a concept wise summary of different metrics like hit rate, OP\$, and number of claims in different statuses, like findings, no findings, InProgress etc., so as to know what the output is from each selection reason. This is done with selection month in the columns as it gives the summary over a required period of time. Filters used for the dashboard are Selection Month, Quarter and Year, Priority Buckets for Concepts based on Hit Rate and OPS. Sample snapshots of Audit dashboard is depicted in figure 5.4.

Queries	QIP Priority	Selections	Findings	InProgress	QP\$	QP\$/Find	Hit Rate
304	High Med Low	14,304	642	2,548	68,68M	62,028	5.42%

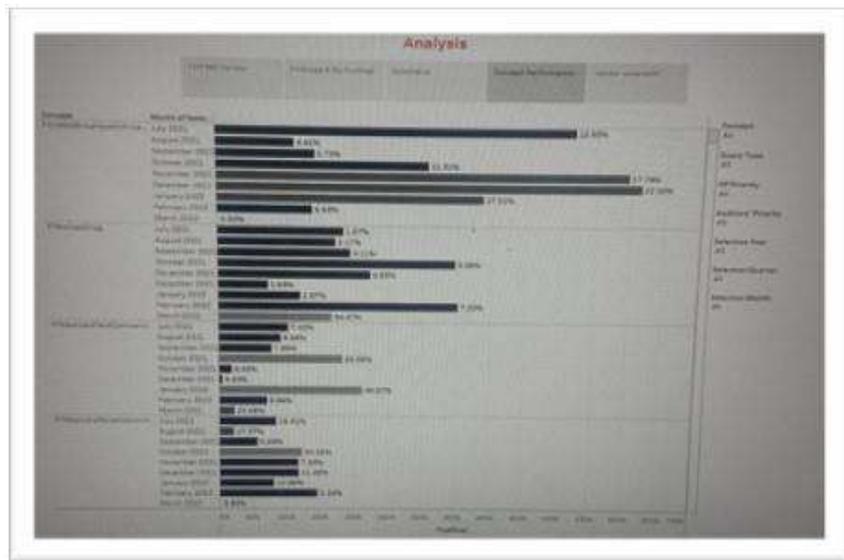


Figure 5.4 Inventory Dashboard Sample Snapshots

Part 2: Audit Dashboard

Audit dashboard gives the high-level summary of Audits for the particular selections.

Contents: Audit Summary - Selection Month wise split of different metrics (Hit Rate, OP\$, Find, No Find etc.), Selections vs Audit Rate - Number of Selections and how many have been audited and the percentage, Trend of Find, No Find and Pending,

along with Trend of OP\$ and Hit Rate. Sample snapshots of Audit dashboard is depicted in figure 5.5.

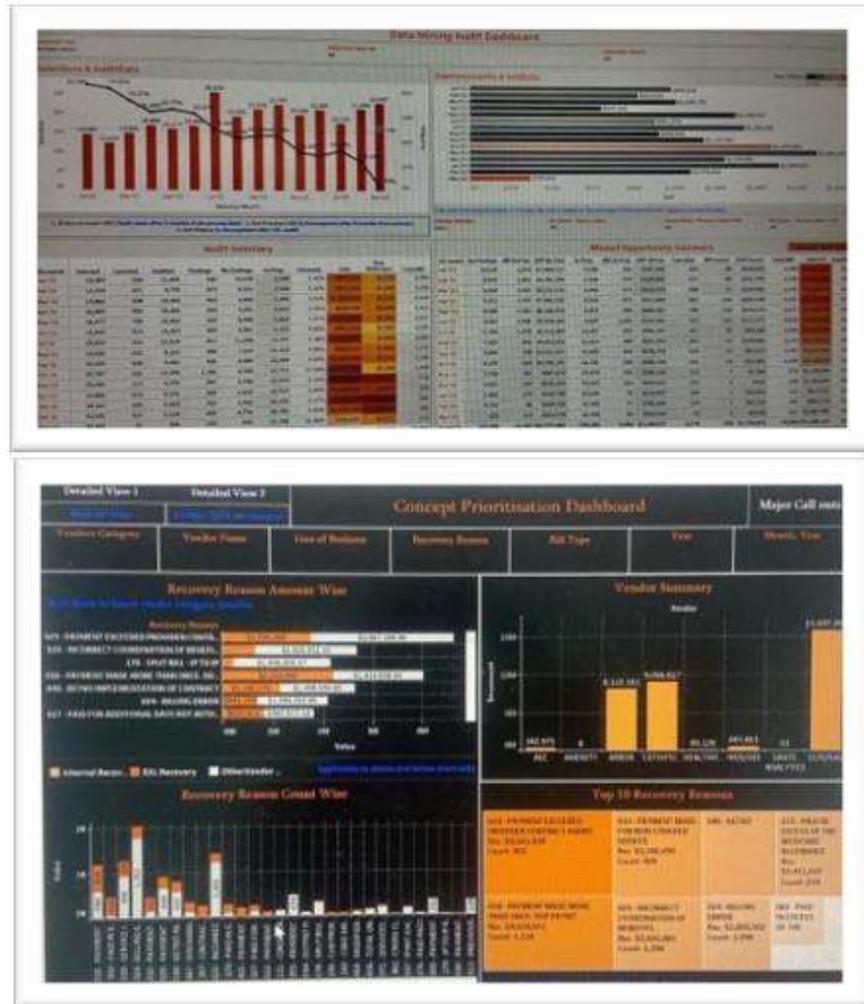


Figure 5.5 Audit Dashboard Sample Snapshots

5.2.3 Competitor analysis

The idea of this analysis is to identify any missed opportunities in Data Mining from the set of claims that EXL didn't mark as findings - includes No Findings, Cancelled, In Progress. This analysis deep dive into the different statuses of claims marked by us, in comparison to the recoveries made by different vendors also the reason for which the claims have been selected. The sample snapshots of competitor analysis is given in figure 5.6.

Key Metrics:

- ✚ Our Final Status vs other vendor recovery Sub status wise list of vendor recovery of claims.
- ✚ Selection reason vs recovery reason. (Selection reason is reason selected by EXL for a particular claim where in recovery reason is reason selected by another vendor for particular claim)
- ✚ To analyze the difference in selection reason to improve the findings count.

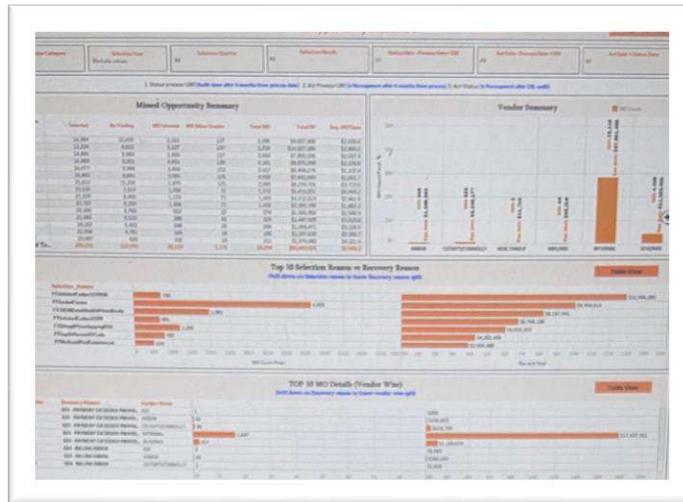


Figure 5.6 Competitor Analysis Snapshots

CHAPTER VI

CONCLUSION

Payment Integrity Project helped to understand how the US Healthcare System works and how the US Health Insurance Claims are handled through the HIPAA Privacy Rule (Health Insurance Portability and Accountability Act) which establishes national standards to protect individual's medical records and other personal health information and applies to health plans, health care clearinghouses and those health care providers that conduct certain health care transactions electronically.

It provided a complete understanding of every concept and helped to adapt into the corporate world. It gave opportunities to not only learn new tools and technologies like Ambari, SAS, Hive etc but also implement them according to the business problem. It also proved to be a learning to face critical issues and resolve them immediately. The project paved the way to interact with business directly, which improved communication skills. Overall, the modules gave a complete view of the business and the technical aspects of providing data and analytics services. Also, the development of the system has given a complete understanding of an end to-end Business Intelligence systems lifecycle.

The technical advantages of the applications are:

- Flexible
- Highly reliable
- Simple and easy to understand by business users.
- Fully automated
- Provide powerful insights

The project helped to experience the life cycle of an end – to – end business intelligence system in health care industry. It provides a complete understanding of every concept and helped to adapt into the corporate world. It also proved to be learning to face critical issues and resolve them immediately for a change request. The project paved a way to interact with clients directly, which improved communication skills also helped me to outward my ideas to them. Overall, the project gave a complete view on the business and the technical aspects of building a Business Intelligence (BI) system and Analytics in US Healthcare System.

BIBLIOGRAPHY

- [1] Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender systems in e-commerce. In Proceedings of the 1st ACM conference on electronic commerce (pp. 158- 166). ACM.
- [2] <https://towardsdatascience.com/evaluation-metrics-for-recommender-systems-df56c6611093> , Evaluation Metrics for Recommender Systems, November 23 2018
- [3] Ferrari Alberto (2015), “The Definitive Guide to DAX: Business intelligence with Microsoft Excel, SQL Server Analysis Services, and Power BI (Business Skills)” by Microsoft
- [4] Knight, Brian, Mitchell, Allan, Green, Darren (January 2007). “Professional SQL Server 2005 Integration Services (Programmer to Programmer)”, Wiley, John & Sons, Incorporated.
- [5] Delaney, K. (2000). Inside Microsoft SQL Server 2000. Microsoft Press.
- [6] Larson, B. (2006). Delivering business intelligence with Microsoft SQL server 2005. McGraw-Hill, Inc.
- [7] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (pp. 29-39). Citeseer.
- [8] Nazneen Fatema Rajani, Mihaela Bornea and Ken Barker.2017. Stacking with Auxiliary Features for Entity Linking in the Medical Domain, In Proceedings of the BioNLP, pages 39-47.

Emulya Technologies Private Limited

(Formerly known as Intain AI Private Limited)



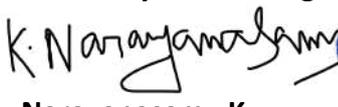
16 May 2022

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Sri Hari K V - Reg No : 1733028**, Student of M.Sc Decision and computing Science, Coimbatore institute of Technology, is undergoing internship with Emulya Technologies Private limited from the period of **09 Dec 2021**.

During the Period of internship programme with us he had completed all the task assigned to him. `

For Emulya Technologies

Narayanasamy K
Chief Technology Officer



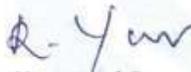
Certificate of Internship

Sussma. S (1733034), M.Sc. Decision and computing sciences, student of Coimbatore Institute of Technology, has completed her Internship Program in our company, VJH Software Private Limited on the title, 'The Timetracker for Workforce'. The internship period is from January 3, 2022, to May 31, 2022.

We wish her future success.

Sincerely,

For VJH Software Pvt.Ltd


Yuvarani R
Hr & Admin



VJH Software Pvt. Ltd.

"Race View Tower" No.126, Anna Salai, Guindy, Chennai - 600 032. Phone : 044-49500788
E-mail : admin@vjhsoftware.in Website : www.vjhsoftware.in

May 16, 2022

To Whomever it may concern

This is to certify that **Mr. Christopher Immanuel Prince P J** is pursuing his Internship in our organization from **13th December 2021**. He is hard working and his performance has been satisfactory.

For SCIOInspire Consulting Services (India) Private Limited



Payal N Chhabria

Assistant Vice President – Human Resource

www.thorogood.com

info@thorogood.com

Bangalore

ACR Towers

2nd Floor

32 Residency Road

Bangalore 560 025

Tel: +91 (0) 80 4123 3700

Fax: +91 (0) 80 4123 3800

London

Tel: +44 (0) 20 0231 0800

New York

Tel: +1 077 905 5010

Philadelphia

Tel: +1 877 905 5010

Singapore

Tel: +65 9450 5329

18 May 2022

TO WHOMSOEVER IT MAY CONCERN

THOROGOOD®

This is to certify that Ms Sowmya D has done the internship with M/s Thorogood Associates India Pvt.Ltd., Bangalore.

As per our internal records, Ms Sowmya D 's internship details are as given below :

Intern ID: 800

Intern Name: Sowmya D

Internship Start Date: 9th December 2021

Internship End Date: 30th June 2022

This letter is being used as an internship proof.

For Thorogood Associates India Pvt.Ltd.,


Authorized Signatory

Thorogood Associates India Pvt Ltd

Registered Number: U72900KA2004PTC070740

Registered Address: Unit 201, ACR Towers, No 32, Residency Road

Bangalore, Karnataka, India 560025

Country of registration: India



Internship Completion Certificate

16-May-2022

To,
Head of Department
Coimbatore Institute of Technology
Coimbatore, Tamilnadu

This is to certify that **Jeevika. G**, student of M.Sc., Decision and Computing Sciences with Registration No: **1733010**, has completed her internship with us from **12,Jan 2022 to 31,May 2022**.

As part of her internship, she has done a study on the topic "**CUSTOMER SENTIMENT ANALYSIS**" using Machine Learning Techniques. She has also contributed towards different classification models to provide more reliable results.

During the tenure with us, we found her sincere, inquisitive and result oriented. She was always ready to put her best efforts and get the task completed on time.

We thank her for all the efforts and wish her the best in all future endeavors.

For **STRATFORGE INDIA PRIVATE LIMITED**

Mohan Yaaraasi
Head of Data Science & Operations Research

Stratforge (India) Private Limited.

No.14, 1st Floor, Haddows Road 1st Street, Nungambakkam, Chennai 600 006, India. Tel: +91 44 42125826.
Regd. Office: No. Door No.8153, TVH Lumbini Square, 27, Bricklin Road, Purasaiwalkam, Chennai – 600 007
Web: www.stratforge.com CIN – U72900TN2016PTC112928