HealthCare Insurance

Project Report

"ENABLING BIG DATA WITH CI/CD"

Version: 1.2.3

Document ID: 18506

Date: 25th November 2020

Submitted to: Montreal College of IT

Table of Contents

Table of Figures

Author Introduction

TRISHA SOLANKI

Business Analyst

ABSTRACT

An experienced entrepreneur with a demonstrated history of working in the Information Technology, Civic, Education, and Healthcare Industries. Skilled in Business Analytics, Data Analytics, and Project Management with an excellent background in DevOps, Web Development, and Digital Marketing. Proficient with all six Business Analysis Knowledge Areas. Hands-on experience with analytical tools and technologies such as Tableau, Power BI, MySQL, Python, R, MS SQL Server, MS Excel, MS Access, and MS Project.

Email

trisha@healthcarein.ca

Executive Summary

Company Description

The company is one of the largest insurance companies in the Canada and offers a wide range of primary and health insurance products to businesses (employees) and individuals (private). This publicly operated company currently has over 10,000 employees in total and has revenue of over $20 billion over its various business units and 18 branches throughout Canada.

Business Challenge

This insurance company has Big Data computational requirements for actuarial designs which are executed against a fine range of databases habitually on daily basis. These databases run on nearly 150 Virtual Machines (VMs), and need to frequently be managed, monitored, and upgraded all while enabling Continuous Integration and Continuous Development (CI/CD) in testing and production environments.

We are looking for a tool that would enable support of our existing tool chain, and customized cloud environment - Chef and VMware with a custom-built distro (distribution - deployment of software - is nearly always used in a Linux context.), while providing strong custom workflows to support the variety of tooling.

Technical Aspects

- HealthCare Insurance needs a Cloud-based Big Data service – Cloudify Orchestration Platform – to deploy databases on demand, manage them, and upgrade them as needed through a simple application blueprint while tying all of the pieces of the infrastructure and application together.
- We are looking for application-level orchestration solutions that can support our existing stack, communicate, and process data coming from our infrastructure orchestrator, as well as support a multiple existing applications.
- The complex architecture (solution architecture) contains a management portal to request environment provisioning, an existing IaaS orchestrator that sends JSON requests to Cloudify, a DNS/DHCP that receives the IPs, Cloud, Docker containers, Chef for application deployment, along with more than 3 backup and monitoring tools.
- We selected Cloudify based on the open and highly modular plugin architecture that supports any technology, as well as the support for custom, complex workflows, that enables us to innovate and leverage Cloudify for less orthodox scenarios with our latest technology based environments.

Goal Chart

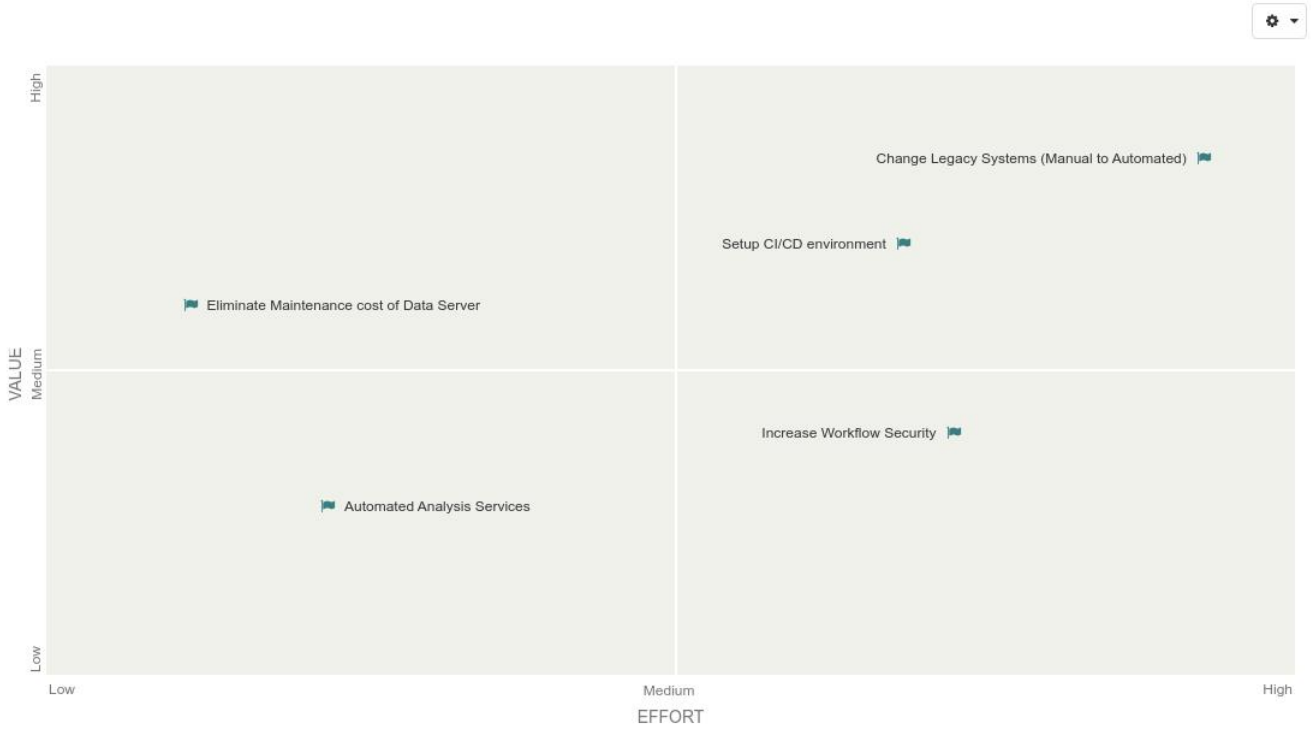


Figure 1 Goal Chart

Project Scope

In Scope

- A new work environment to handle Data operations
- Deployment of latest legacy systems
- An automated custom CI/CD workflow
- Cloud based services and tools
- Big Data to handle complex Data requirements
- Latest Backup and Monitoring tools which can support the new systems
- Eliminate Hardware Data Storage (on premises storage)
- Transfer all the data to Cloud services
- Hiring Data Engineers (if necessary)
- Training for Data Team and other related employees to use the new technologies and tools
- Account setup, Permission assignments, and other security tasks

Out of Scope

- Support for any other team apart from Technical and Operations Team
- Allowance of more than one Cloud services
- Usage of manual (old) procedures, systems, or tools
- New Hardware implementation
- Customer Support (Engagement)

Resource List

Sr. No.	Resource Description
01	AWS Cloud Services
02	Docker
03	Jenkins
04	MS Office Subscription
05	Jira
06	Database Administrators
07	Cloud Practitioners
08	Developers
09	Testers
10	Business Analysts
11	Project Managers
12	Backup Tools
13	Project Management Tools
14	Data Analysis Tools

SWOT Analysis

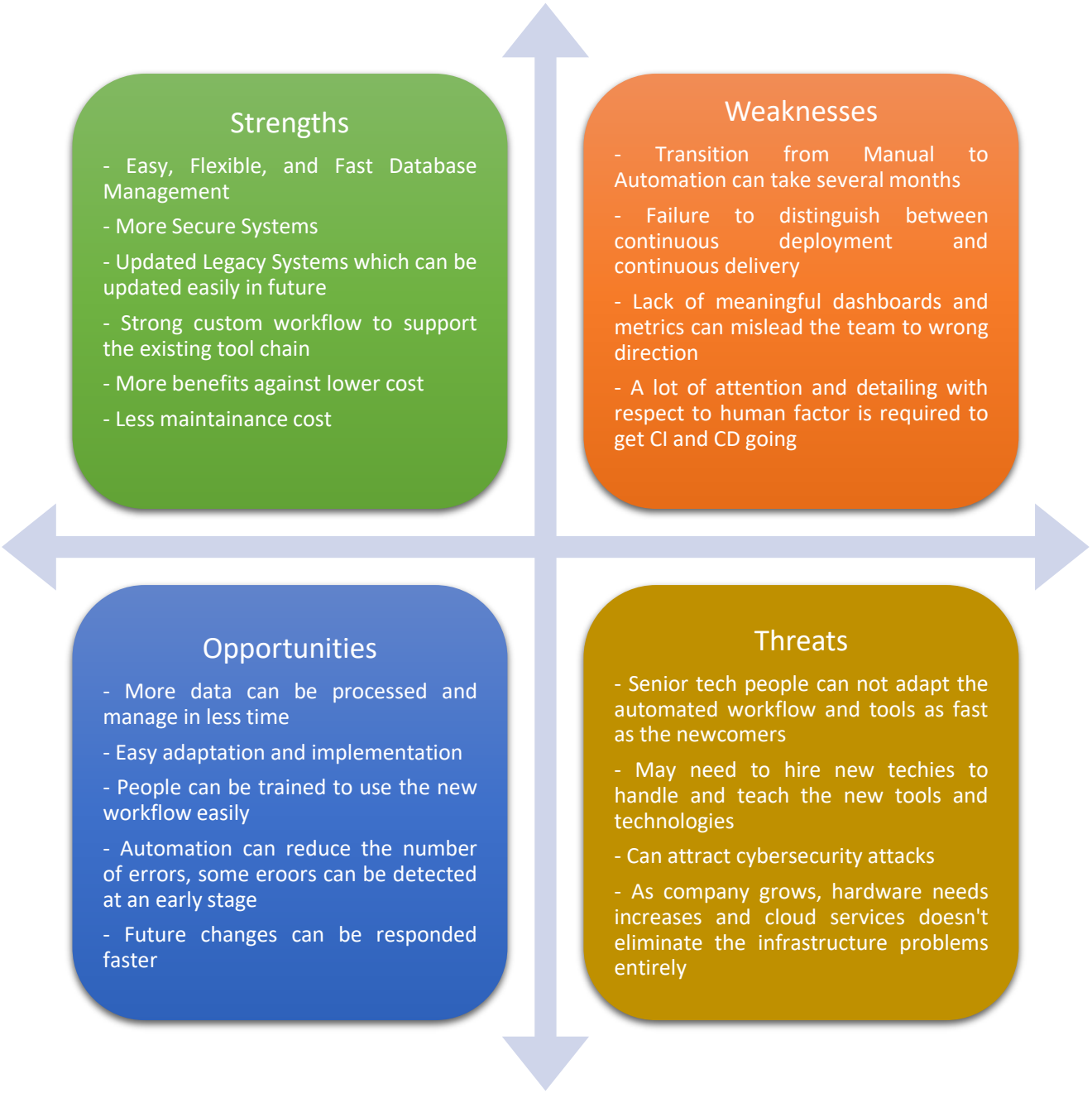


Figure 2 SWOT Analysis

Change Management

GAP Analysis

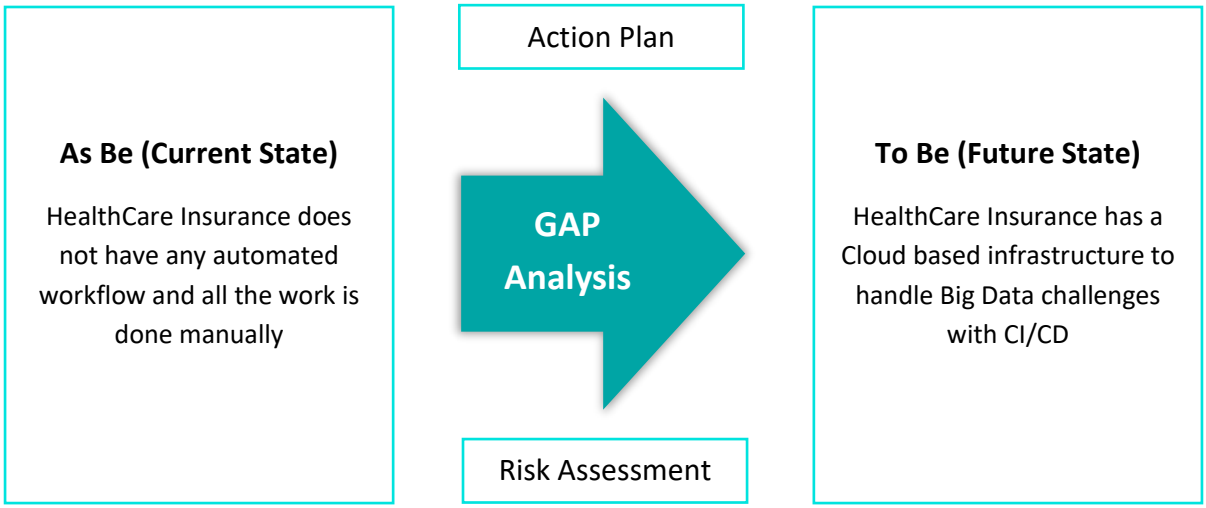


Figure 3 GAP Analysis

Business Analysis Core Concept Model (BACCM)

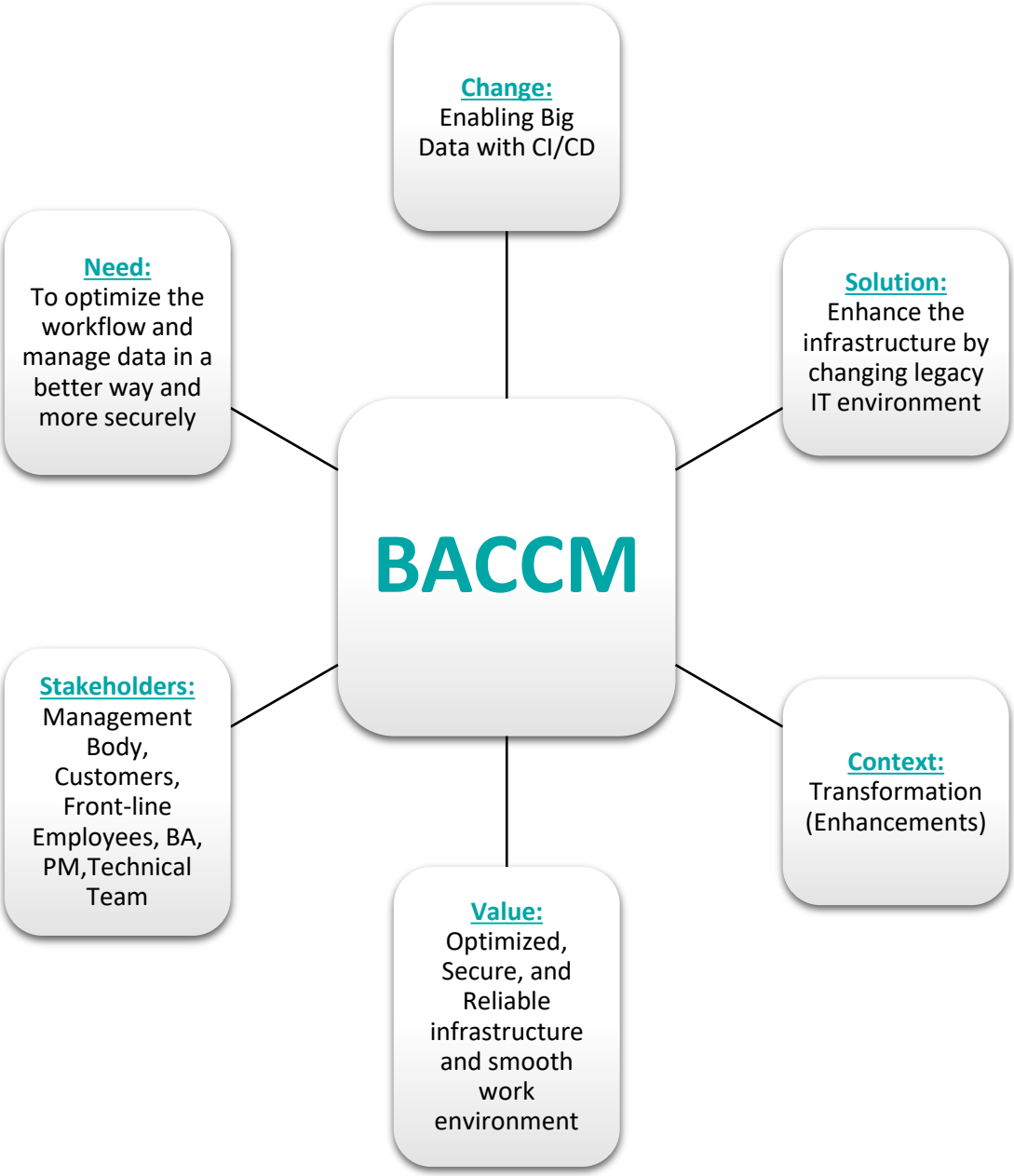


Figure 4 BACCM

Stakeholder Analysis

Stakeholder Mapping (Onion Diagram)

Figure 5 Stakeholder Mapping

RACI Matrix

	Planning & Analysis	Requirement Gathering	Designing	Deployment	Testing & Bug Fixing
Developers	**I**	**I**	**R**	**R/A**	**C**
Testers	**I**	**I**	**I**	**C/I**	**R/A**
Cloud Practitioners	**C**	**C**	**R/C**	**R**	**C**
Database Administrators	**C**	**C**	**R/C**	**R**	**C**
Project Managers	**R**	**R/C**	**A/C**	**A/C**	**A**
Business Analysts	**R/A**	**R/A**	**A/C**	**C/I**	**A**
Sponsors	**A**	**A**	**I**	**I**	**I**
Investors	**I**	**I**	**I**	**I**	**I**

Communication Plan

Communication	Frequency	Goal	Owner	Audience
Synchronous Communication				
Meetings				
Team meeting	Daily	Discuss what each team member did yesterday, what they'll do today, and any blockers	Project manager	Project team
Project review	At milestones	Present project deliverables, gather feedback, and discuss next steps	PM / BA	Project team + Project sponsor
Post-mortem meeting	At end of project	Assess what worked and what did not work and discuss actionable takeaways	BA	Project team
Team stand-up				
Task progress updates	Daily	Share daily progress made on project tasks	PM, BA	Project team
Video Conference Calls				
Major work updates	Weekly	Share weekly progress made on project tasks and give major updates on project	PM, BA	Project team + Client + Senior Management
Asynchronous Communication				
Email				
Project status report	Weekly	Review project status and discuss potential issues or delays	Project manager	Project team + Project sponsor + Senior Managers

Business Requirements

In Scope

- A new work environment to handle Data operations
- Deployment of latest legacy systems
- An automated custom CI/CD workflow
- Cloud based services and tools
- Big Data to handle complex Data requirements
- Latest Backup and Monitoring tools which can support the new systems
- Eliminate Hardware Data Storage (on premises storage)
- Transfer all the data to Cloud services
- Hiring Data Engineers (if necessary)
- Training for Data Team and other related employees to use the new technologies and tools
- Account setup, Permission assignments, and other security tasks

Out of Scope

- Support for any other team apart from Technical and Operations Team
- Allowance of more than one Cloud services
- Usage of manual (old) procedures, systems, or tools
- New Hardware implementation
- Customer Support (Engagement)

Assumptions

- New legacy systems and workflow will be easy to adapt
- Anyone will be able to use it once the training is given
- Fast & Secure operations will be performed compared to old systems

Constraints / Risks

- Workflow should be as simple as possible, Cloud services should be selected from which are available in the market, and Resource utilization should be done properly for existing and new tools
- Cybersecurity attacks, May be difficult to adapt for senior employees, Automating the wrong processes first, Lack of coordination between continuous integration and continuous delivery

Process Flow Diagrams

Swimlane Diagram

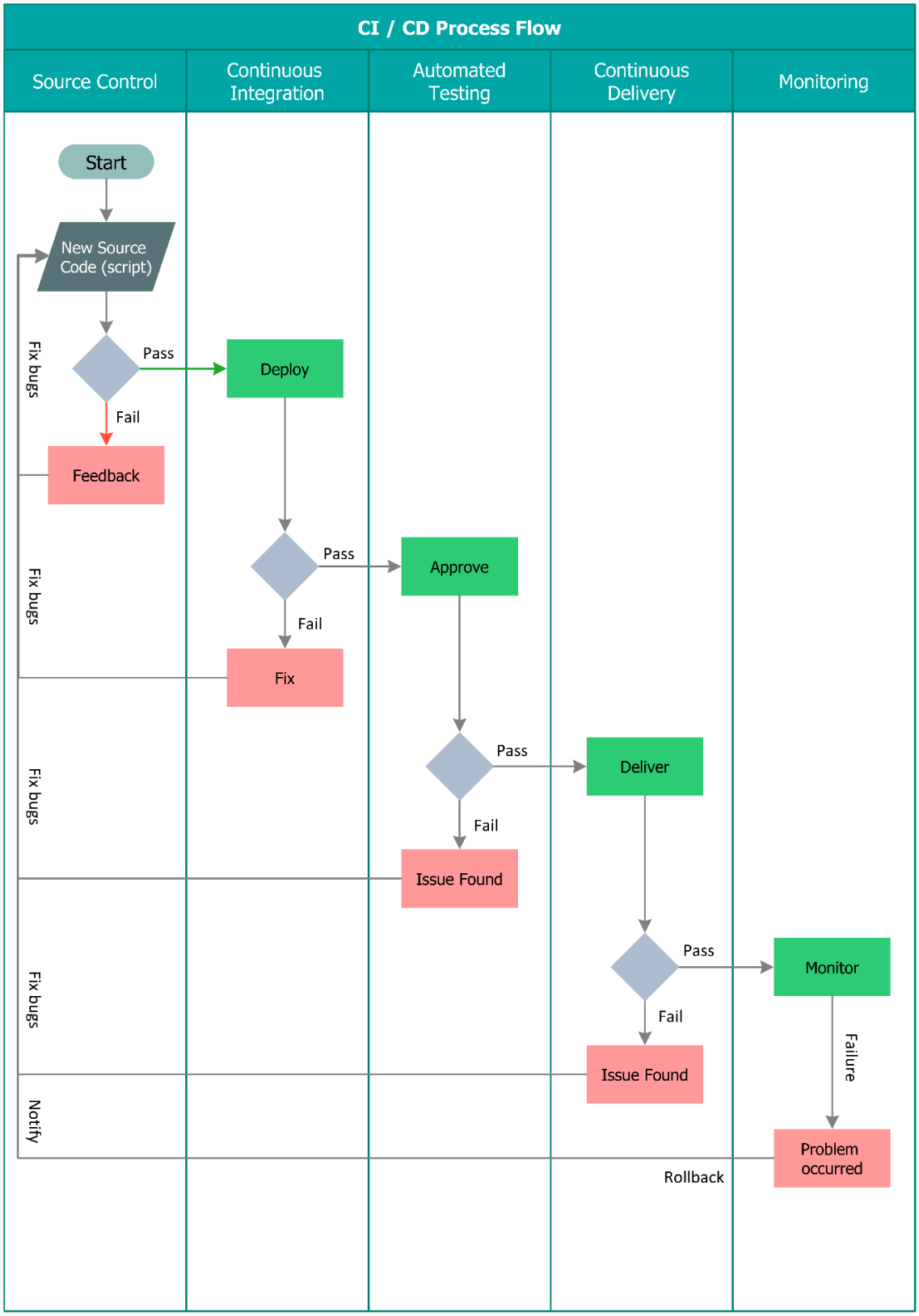


Figure 6 Swimlane Diagram

Use Case Diagram

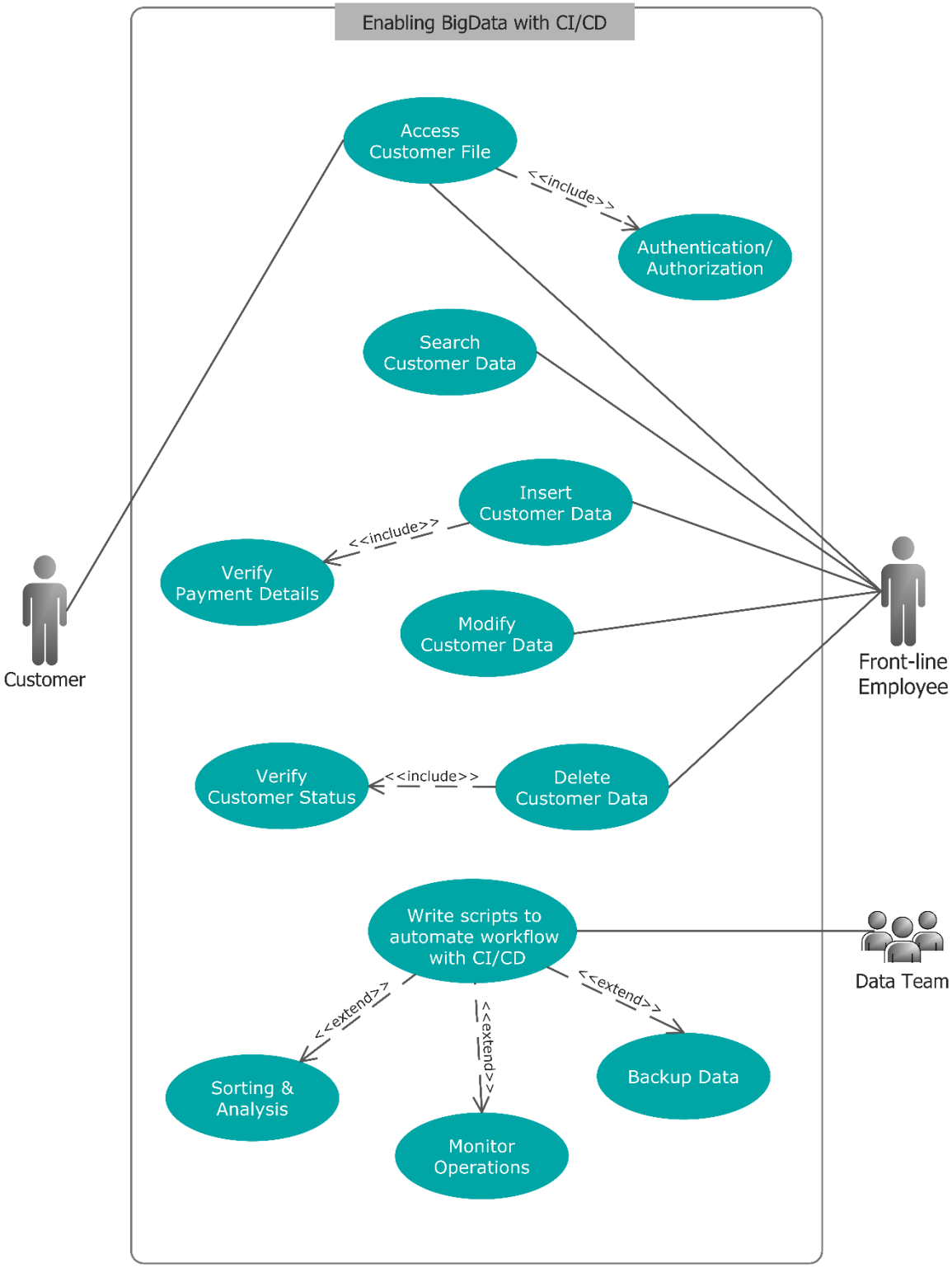


Figure 7 Use Case Diagram

Requirement Analysis & Grooming

List of Requirements

Business Requirements

Code	Requirements
BA01	Data migration to cloud services
BA02	Eliminate Data Servers on premises
BA03	Clean and Setup migrated data on Cloud
BA04	Create an automated workflow with cloud services
BA05	Change the legacy system of data workflow
BA06	Limit the access to a few number of people
BA07	Setup automated analytical services
BA08	Add security to data

Functional Requirements

Code	Requirements
FR01	User should be able to find data on Cloud databases
FR02	Data should be cleaned and organized automatically
FR03	Data should be sorted automatically
FR04	Workflow (CI/CD) should involve data security
FR05	User or Customer should be authorized before accessing any data

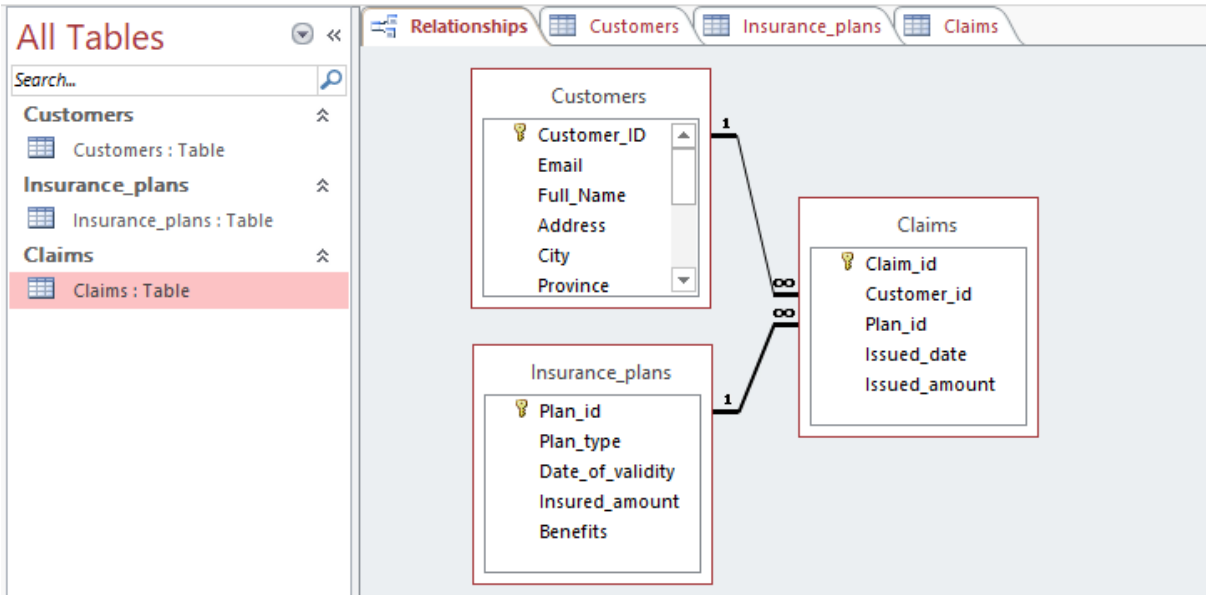
FR06	Only Managerial level employees have the right to view revenue data
FR07	Data Analysis should be done without any manual help (automatic)
FR08	Data is continuously refreshed in case of any updates

Non-functional Requirements

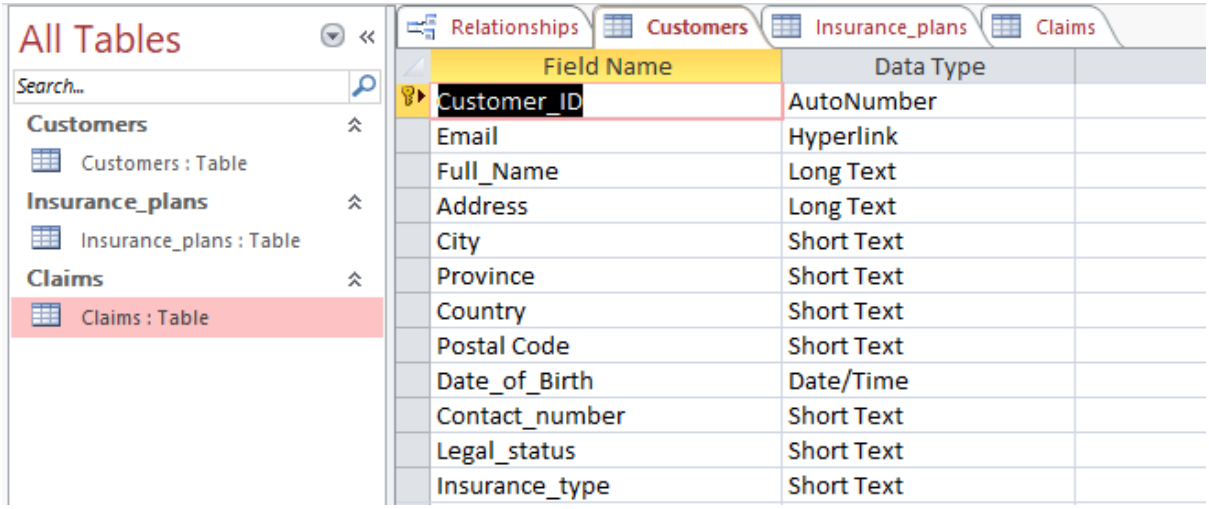
Code	Requirements
NR01	Scalability feature should be added in case of too many parallel customers try to access the data
NR02	Data should updated within 5 seconds
NR03	Error message should be displayed in case of wrong data access
NR04	Every unsuccessful attempt by a user/employee to access an item of data shall be recorded
NR05	The software should be portable
NR06	Privacy of information, the export of restricted technologies, intellectual property rights, etc. should be audited
NR07	Automated data operations should run 100x faster than manual ones
NR08	Execution of queries should be 80% faster and 95% reliable

Database Design Requirements

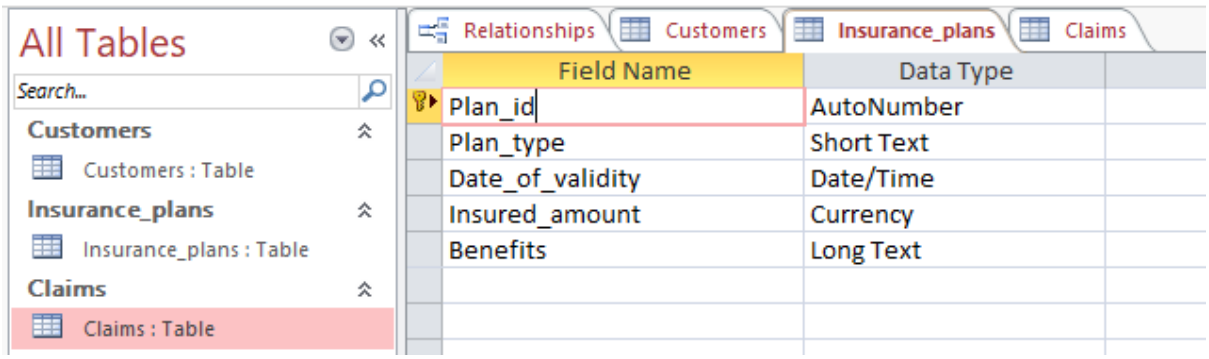
1. Tables and their relationships



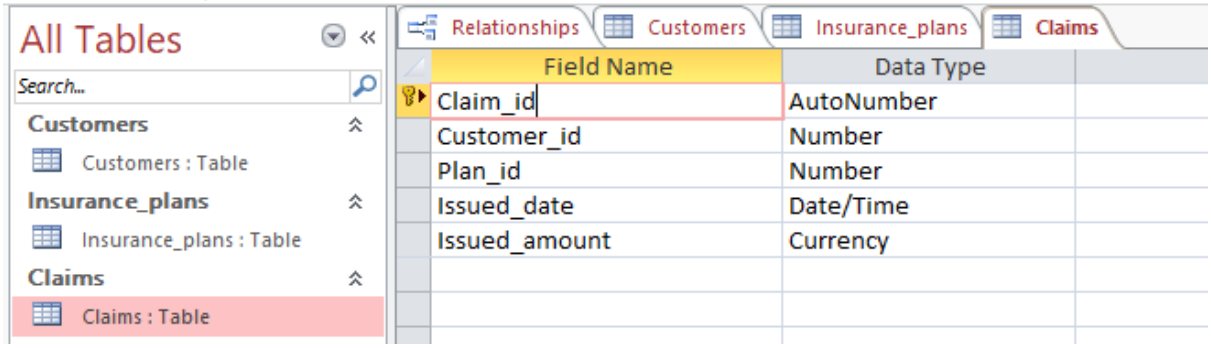
2. Table 1: Customers



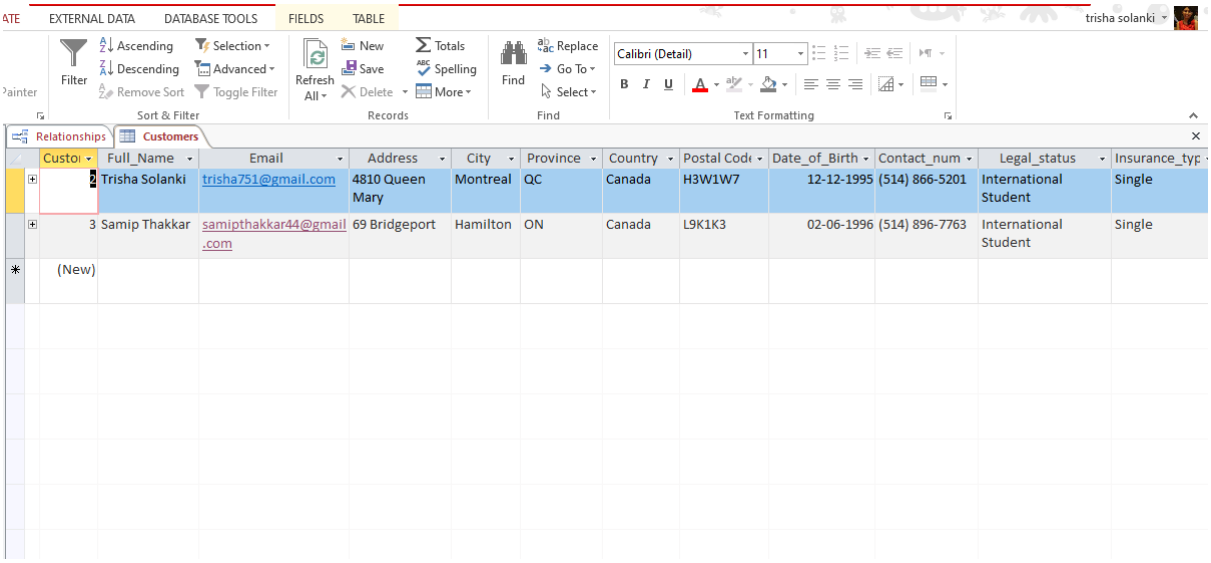
3. Table 2: Insurance_plans



4. Table 3: Claims



5. Records insertion in “Customers” Table



User Stories

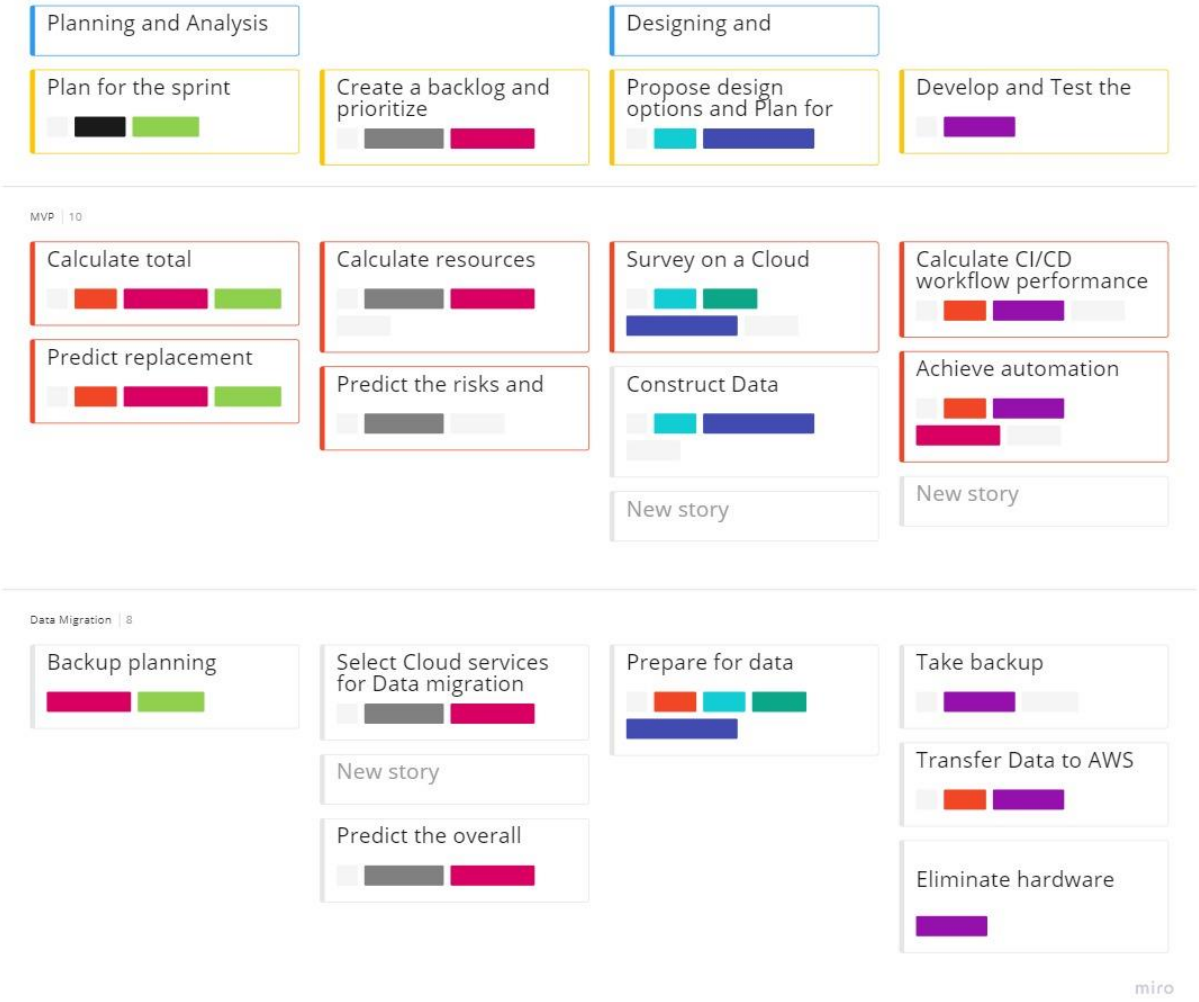


Figure 8 User Stories

Feature Description

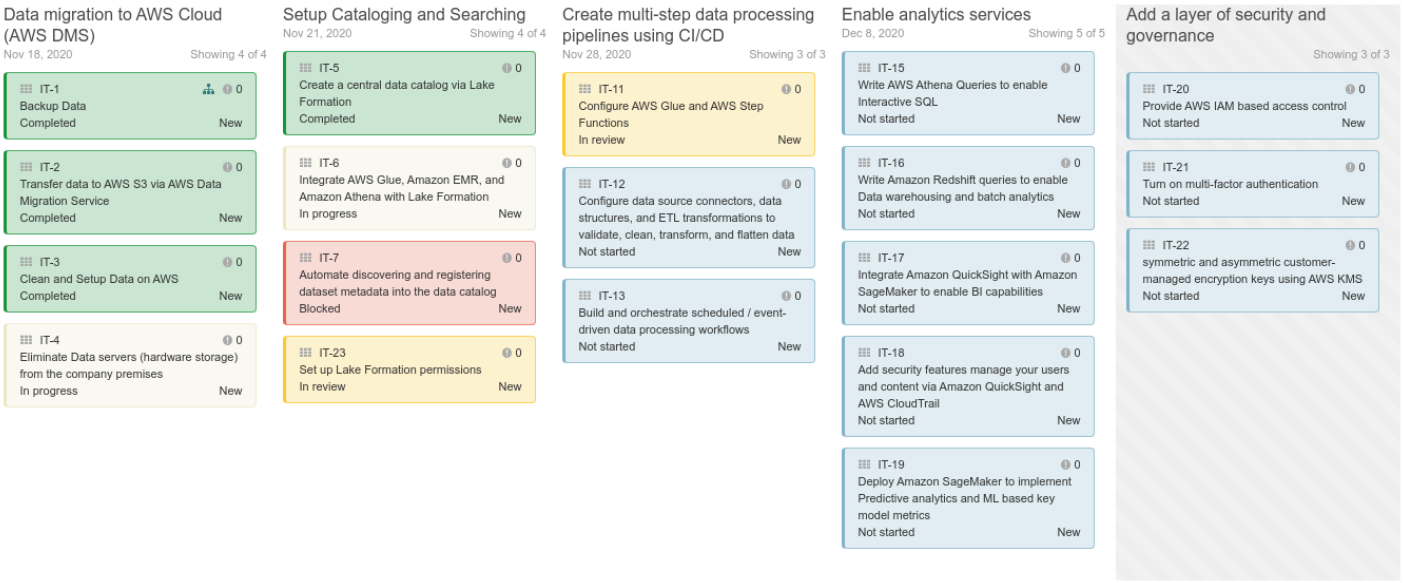


Figure 9 Feature Description

Entity Relationship (ER) Diagram

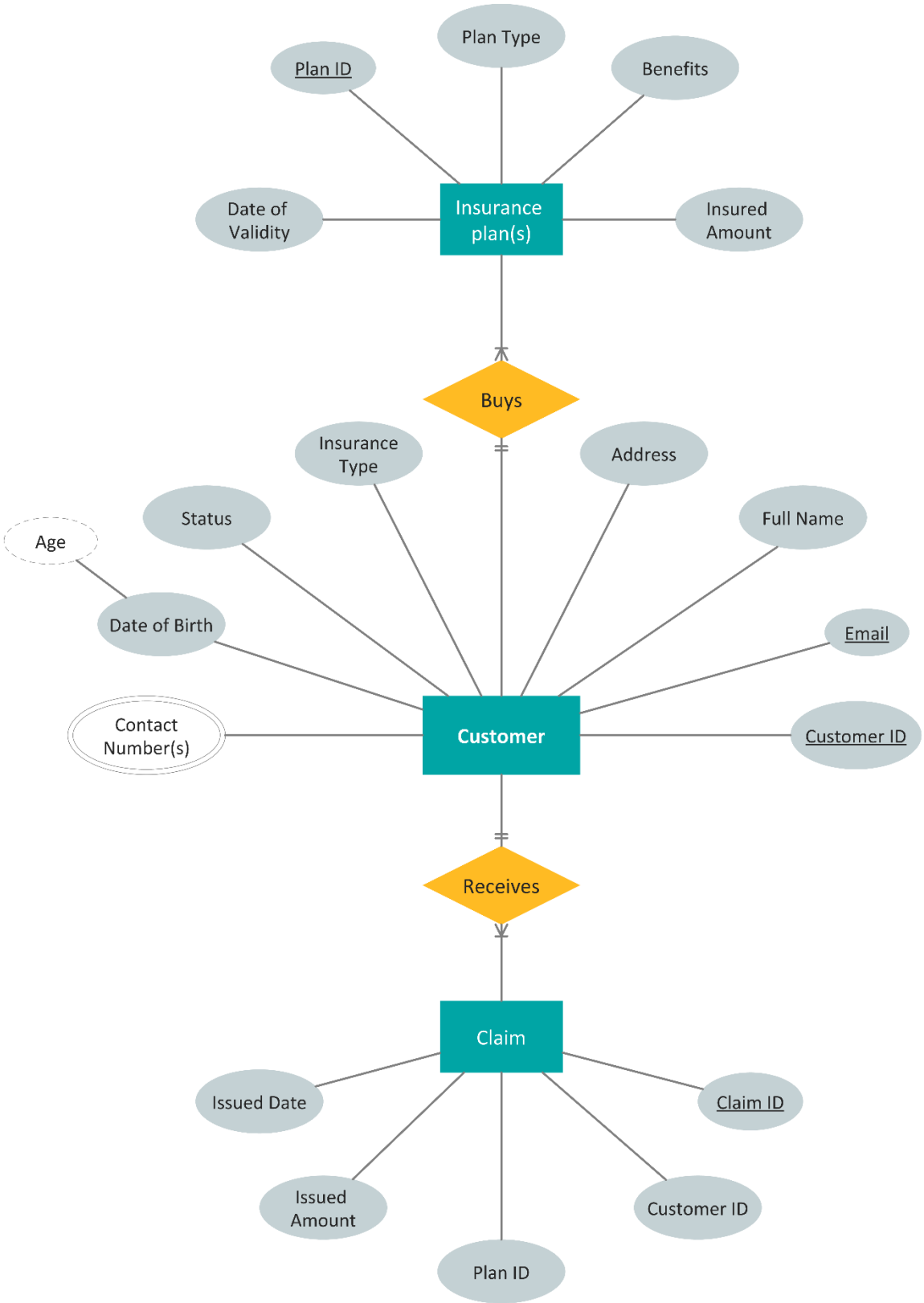


Figure 10 ER Diagram

Data Flow Diagrams

Automating Analytical workflow

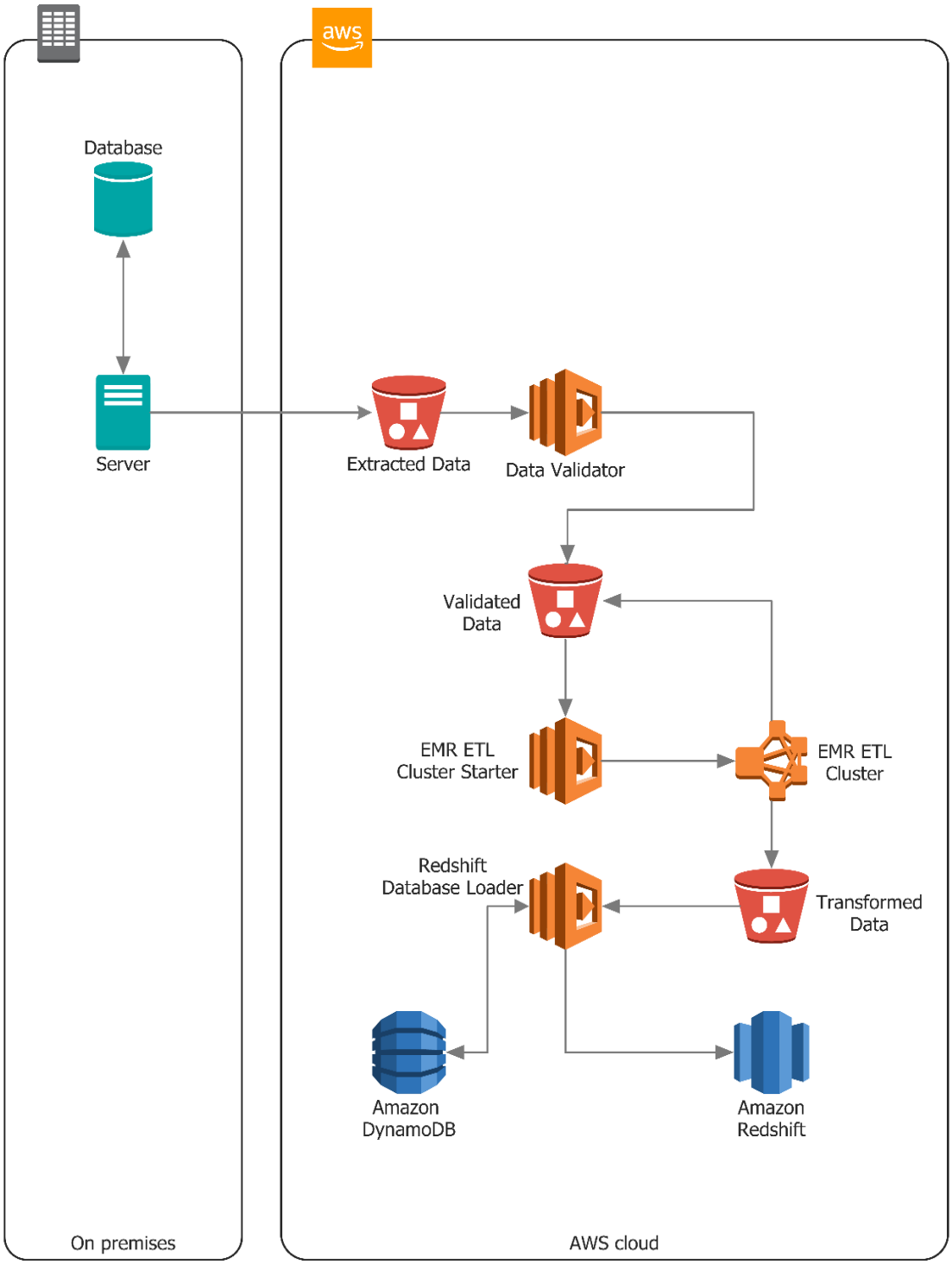


Figure 11 Automatic Analytical Workflow

DFD I

This Data Flow Diagram shows how a new customer's data interacts databases and how data flows in the system in which order

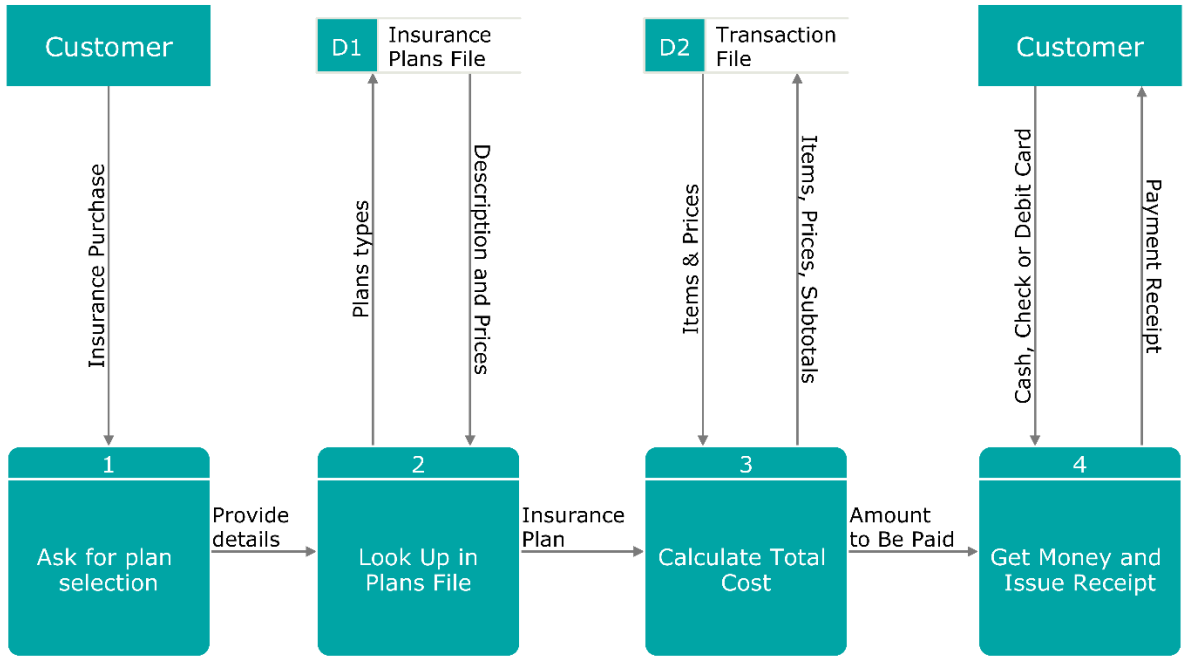


Figure 12 DFD I

DFD II

This Data Flow Diagram shows how an existing customer's data interacts databases and how data flows in the system in which order

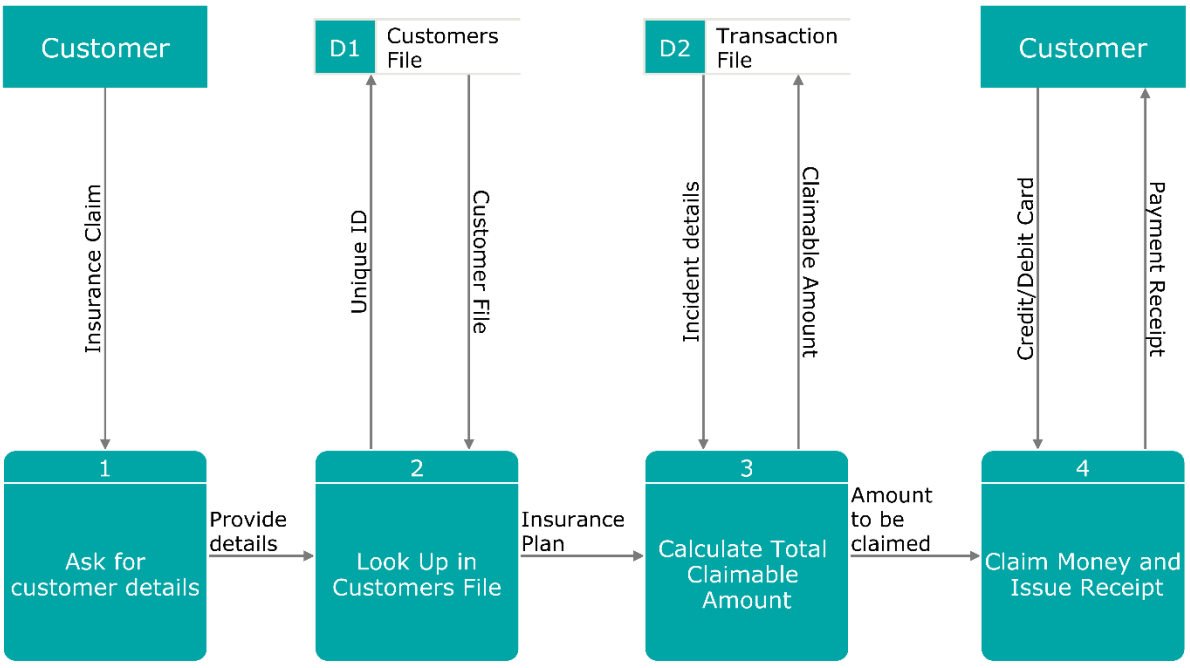


Figure 13 DFD II

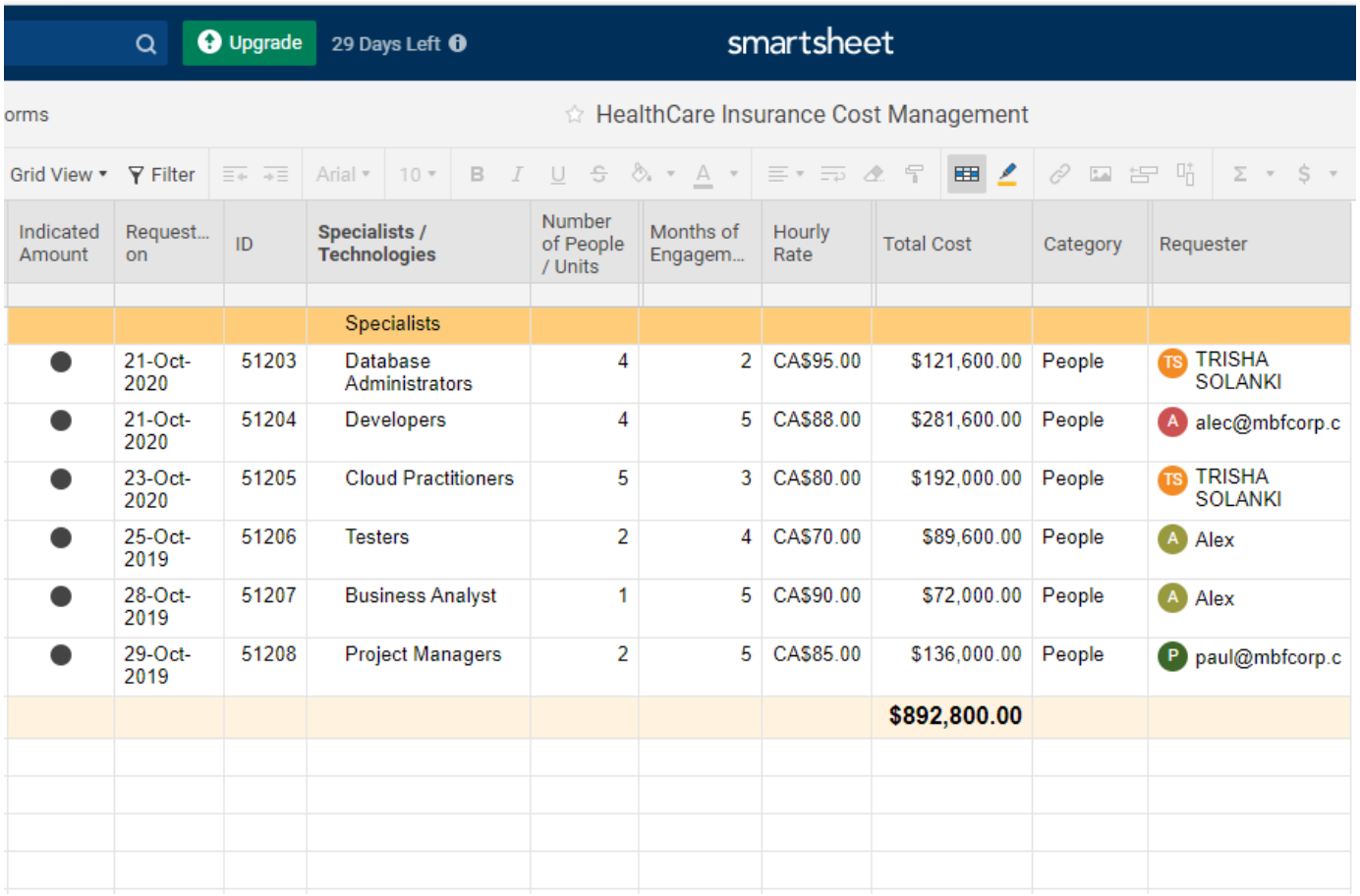
Work Breakdown Structure (WBS)

ID	Task Name	Duration	Start	Finish	Predecessors
1	**1 Data migration to AWS Cloud (AWS DMS)**	**14 days**	**Mon 02-11-20**	**Fri 20-11-20**	
2	1.1 Backup Data	4 days	Mon 02-11-20	Thu 05-11-20	
3	1.2 Transfer data to AWS S3 via AWS Data Migration Service	3 days	Thu 05-11-20	Sun 08-11-20	2
4	1.3 Clean and Setup Data on AWS	7 days	Mon 09-11-20	Tue 17-11-20	3
5	1.4 Eliminate Data servers (hardware storage) from the company premises	2 days	Wed 18-11-20	Thu 19-11-20	4
6	1.5 Data Migration Done	0 days	Fri 20-11-20	Fri 20-11-20	
7	**2 Setup Cataloging and Searching**	**5 days**	**Mon 16-11-20**	**Mon 23-11-20**	
8	2.1 Create a central data catalog via Lake Formation	2 days	Mon 16-11-20	Tue 17-11-20	
9	2.2 Integrate AWS Glue, Amazon EMR, and Amazon Athena with Lake Formation	2 days	Tue 17-11-20	Wed 18-11-20	8
10	2.3 Automate discovering and registering dataset metadata into the data catalog	2 days	Thu 19-11-20	Fri 20-11-20	9
11	2.4 Set up Lake Formation permissions	2 days	Sat 21-11-20	Sun 22-11-20	9
12	2.5 Catalog and Search Setup Done	0 days	Mon 23-11-20	Mon 23-11-20	
13	**3 Create multi-step data processing pipelines using CI/CD**	**6 days**	**Mon 23-11-20**	**Sun 29-11-20**	
14	3.1 Configure AWS Glue and AWS Step Functions	2 days	Mon 23-11-20	Tue 24-11-20	11
15	3.2 Configure data source connectors, data structures, and ETL transformations to validate, clean, transform, and flatten data	2 days	Wed 25-11-20	Thu 26-11-20	14
16	3.3 Build and orchestrate scheduled / event-driven data processing workflows	2 days	Fri 27-11-20	Sat 28-11-20	15
17	3.4 CI/CD Pipelines Creation Done	0 days	Sun 29-11-20	Sun 29-11-20	
18	**4 Enable analytics services**	**0 days?**	**Wed 18-11-20**	**Wed 18-11-20**	
19	4.1 Write AWS Athena Queries to enable Interactive SQL				
20	4.2 Write Amazon Redshift queries to enable Data warehousing and batch analytics				
21	4.3 Integrate Amazon QuickSight with Amazon SageMaker to enable BI capabilities				
22	4.4 Add security features manage your users and content via Amazon QuickSight and AWS CloudTrail				
23	4.5 Deploy Amazon SageMaker to implement Predictive analytics and ML based key model metrics				
24	4.6 Provide AWS IAM based access control				
25	4.7 Turn on multi-factor authentication				
26	4.8 symmetric and asymmetric customer-managed encryption keys using AWS KMS				

Page 1

Figure 14 Work Breakdown Structure (WBS)

Cost Management Plan



Indicated Amount	Request... on	ID	Specialists / Technologies	Number of People / Units	Months of Engagem...	Hourly Rate	Total Cost	Category	Requester
			Specialists						
●	21-Oct-2020	51203	Database Administrators	4	2	CA$95.00	$121,600.00	People	TS TRISHA SOLANKI
●	21-Oct-2020	51204	Developers	4	5	CA$88.00	$281,600.00	People	A alec@mbfcorp.c
●	23-Oct-2020	51205	Cloud Practitioners	5	3	CA$80.00	$192,000.00	People	TS TRISHA SOLANKI
●	25-Oct-2019	51206	Testers	2	4	CA$70.00	$89,600.00	People	A Alex
●	28-Oct-2019	51207	Business Analyst	1	5	CA$90.00	$72,000.00	People	A Alex
●	29-Oct-2019	51208	Project Managers	2	5	CA$85.00	$136,000.00	People	P paul@mbfcorp.c
							$892,800.00		

Figure 15 Cost Management Plan

Solution Evaluation

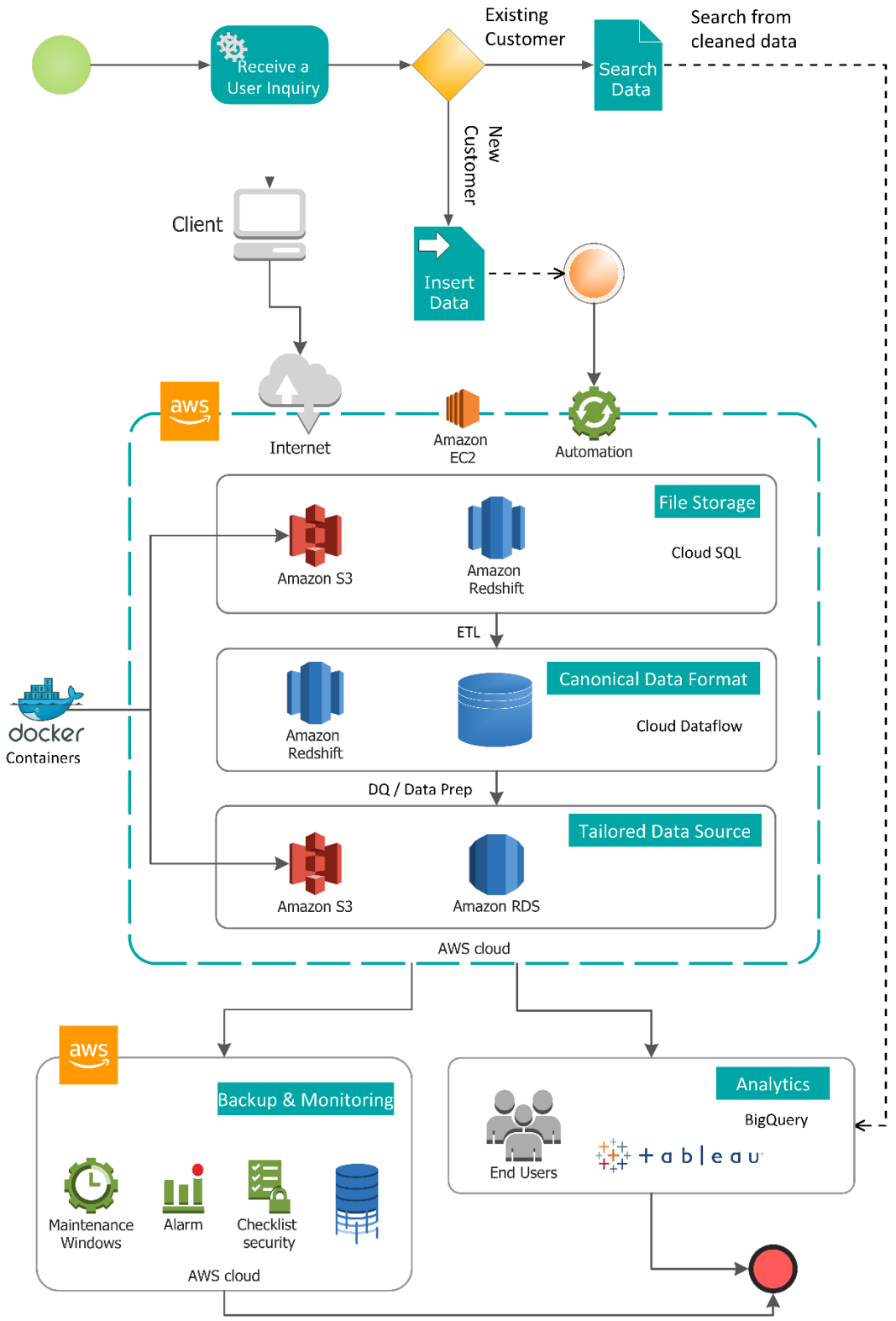


Figure 16 Solution Evaluation

Feature Lists

Workspace name	Feature reference #	Feature name	Release name	Feature status
HealthCare Insurance	IT-1	Backup Data	Data migration to AWS Cloud (AWS DMS)	Completed
HealthCare Insurance	IT-2	Transfer data to AWS S3 via AWS Data Migration Service	Data migration to AWS Cloud (AWS DMS)	Completed
HealthCare Insurance	IT-3	Clean and Setup Data on AWS	Data migration to AWS Cloud (AWS DMS)	Completed
HealthCare Insurance	IT-4	Eliminate Data servers (hardware storage) from the company premises	Data migration to AWS Cloud (AWS DMS)	In progress
HealthCare Insurance	IT-5	Create a central data catalog via Lake Formation	Setup Cataloging and Searching	Completed
HealthCare Insurance	IT-6	Integrate AWS Glue, Amazon EMR, and Amazon Athena with Lake Formation	Setup Cataloging and Searching	In progress
HealthCare Insurance	IT-7	Automate discovering and registering dataset metadata into the data catalog	Setup Cataloging and Searching	Blocked
HealthCare Insurance	IT-11	Configure AWS Glue and AWS Step Functions	Create multi-step data processing pipelines using CI/CD	In review
HealthCare Insurance	IT-12	Configure data source connectors, data structures, and ETL transformations to validate, clean, transform, and flatten data	Create multi-step data processing pipelines using CI/CD	Not started
HealthCare Insurance	IT-13	Build and orchestrate scheduled / event-driven data processing workflows	Create multi-step data processing pipelines using CI/CD	Not started
HealthCare Insurance	IT-15	Write AWS Athena Queries to enable Interactive SQL	Enable analytics services	Not started
HealthCare Insurance	IT-16	Write Amazon Redshift queries to enable Data warehousing and batch analytics	Enable analytics services	Not started
HealthCare Insurance	IT-17	Integrate Amazon QuickSight with Amazon SageMaker to enable BI capabilities	Enable analytics services	Not started
HealthCare Insurance	IT-18	Add security features manage your users and content via Amazon QuickSight and AWS CloudTrail	Enable analytics services	Not started
HealthCare Insurance	IT-19	Deploy Amazon SageMaker to implement Predictive analytics and ML based key model metrics	Enable analytics services	Not started
HealthCare Insurance	IT-20	Provide AWS IAM based access control	Add a layer of security and governance	Not started
HealthCare Insurance	IT-21	Turn on multi-factor authentication	Add a layer of security and governance	Not started
HealthCare Insurance	IT-22	symmetric and asymmetric customer-managed encryption keys using AWS KMS	Add a layer of security and governance	Not started
HealthCare Insurance	IT-23	Set up Lake Formation permissions	Setup Cataloging and Searching	In review

Figure 17 Feature List

Pivot Report

Workspace name	Release quarter	Feature status	Not started	In progress	In review	Completed	Blocked
HealthCare Insurance	2020 Q4		Add security features manage your users and content via Amazon QuickSight and AWS CloudTrail Build and orchestrate scheduled / event-driven data processing workflows Configure data source connectors, data structures, and ETL transformations to validate, clean, transform, and flatten data Deploy Amazon SageMaker to implement Predictive analytics and ML based key model metrics Integrate Amazon QuickSight with Amazon SageMaker to enable BI capabilities Write Amazon Redshift queries to enable Data warehousing and batch analytics Write AWS Athena Queries to enable Interactive SQL	Eliminate Data servers (hardware storage) from the company premises Integrate AWS Glue, Amazon EMR, and Amazon Athena with Lake Formation	Configure AWS Glue and AWS Step Functions Set up Lake Formation permissions	Backup Data Clean and Setup Data on AWS Create a central data catalog via Lake Formation Transfer data to AWS S3 via AWS Data Migration Service	Automate discovering and registering dataset metadata into the data catalog
			Provide AWS IAM based access control symmetric and asymmetric customer-managed encryption keys using AWS KMS Turn on multi-factor authentication				

Figure 18 Pivot Report

Project Schedule and Timeline

Roadmap

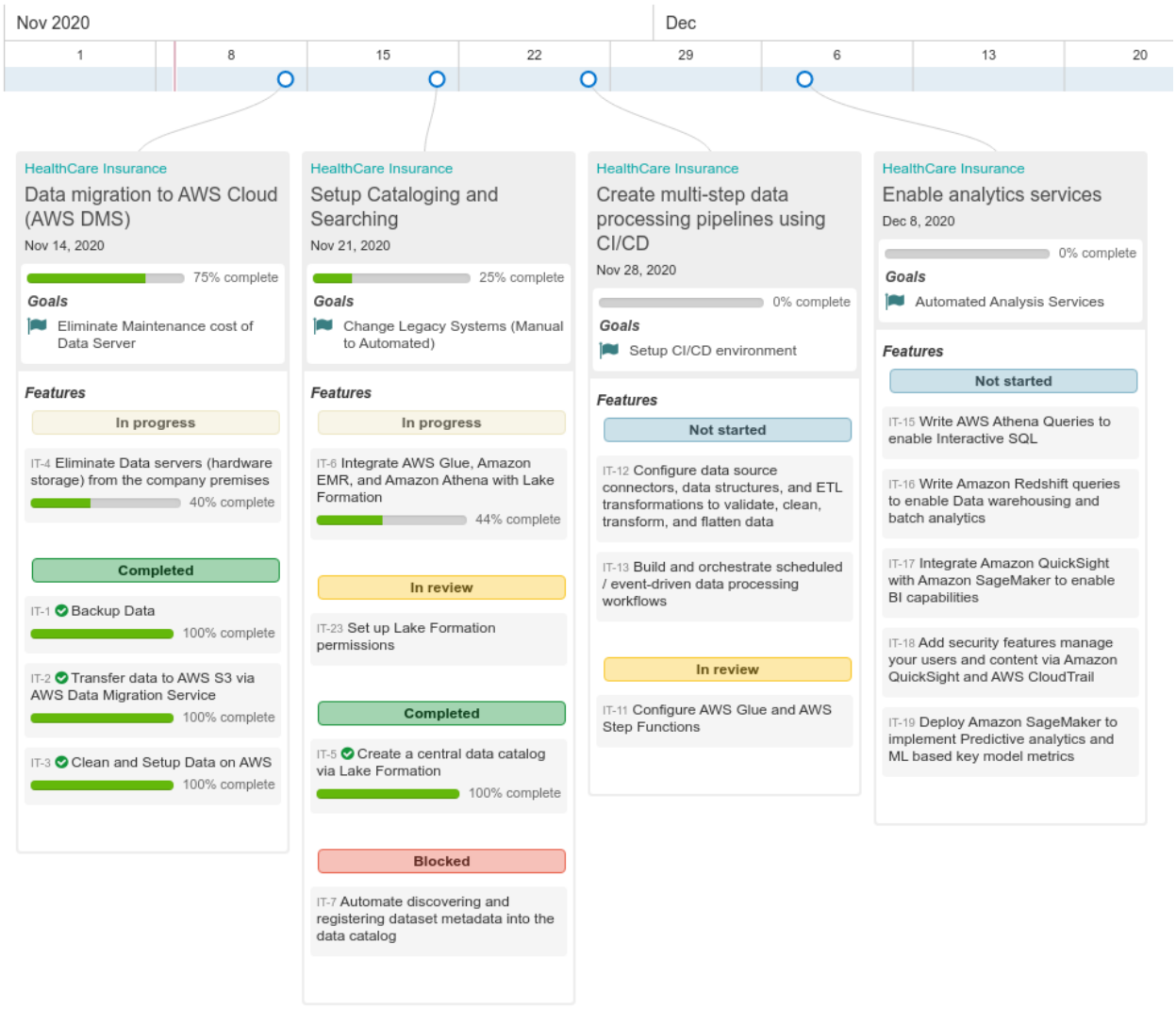


Figure 19 Roadmap

Gantt chart

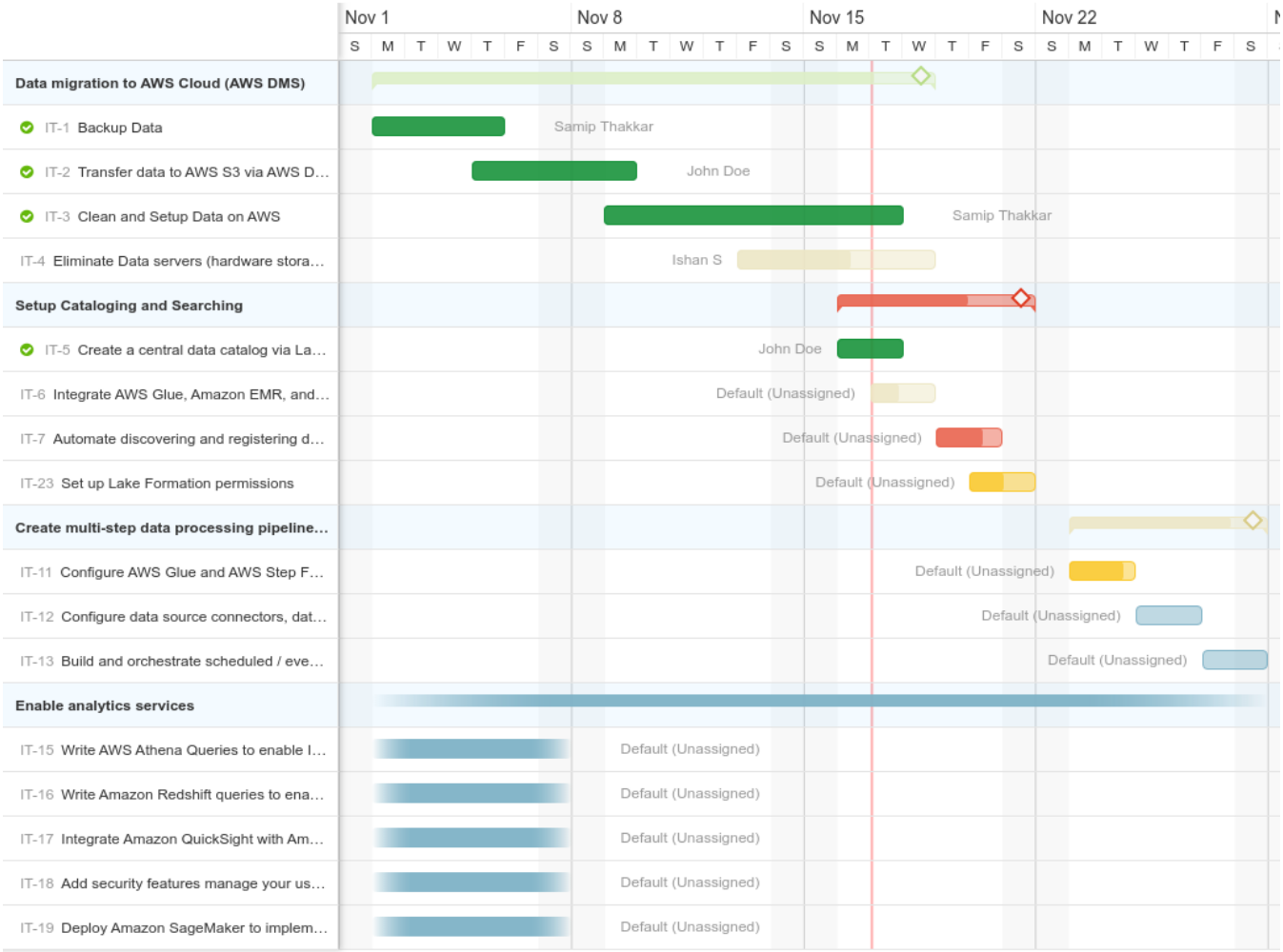
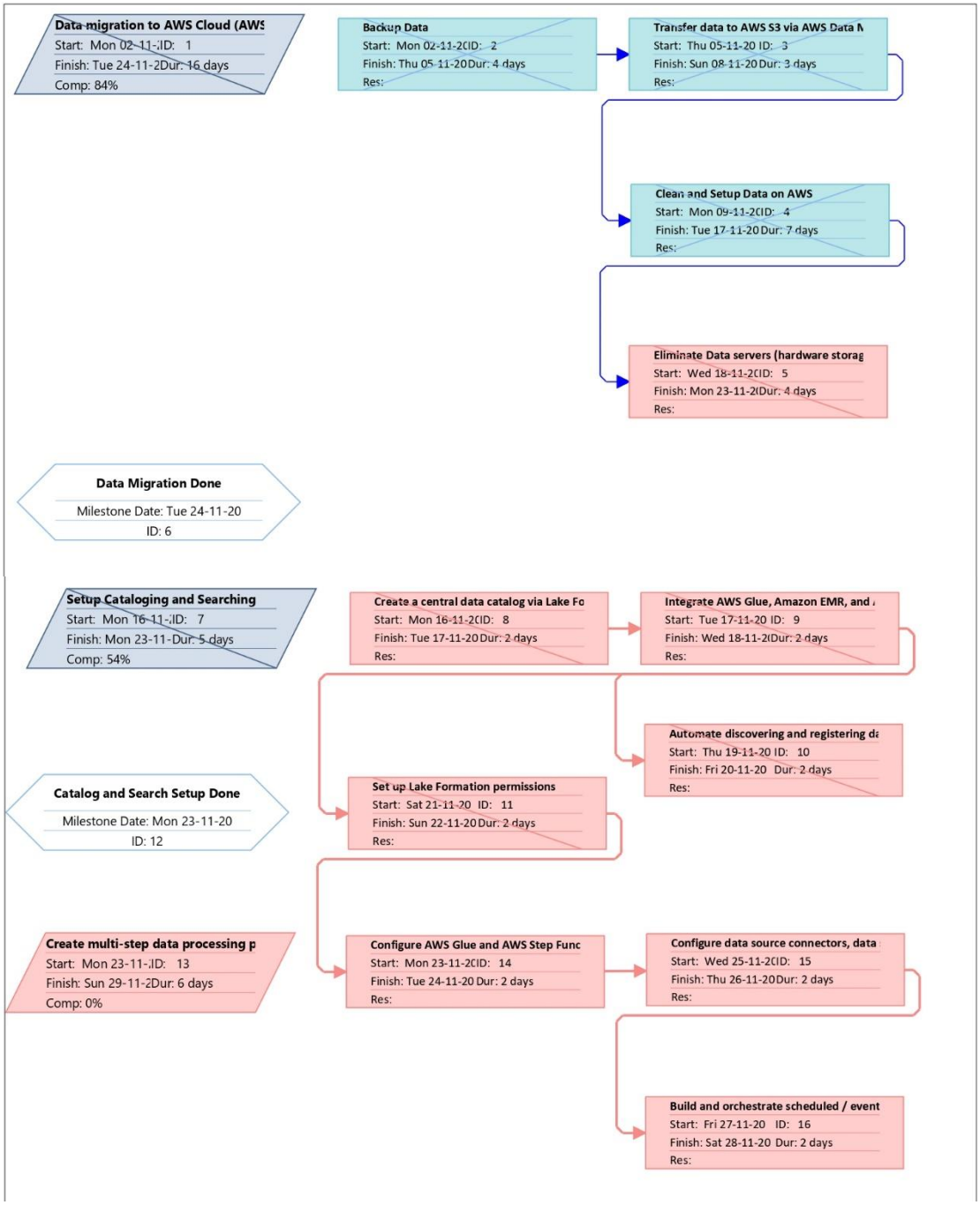
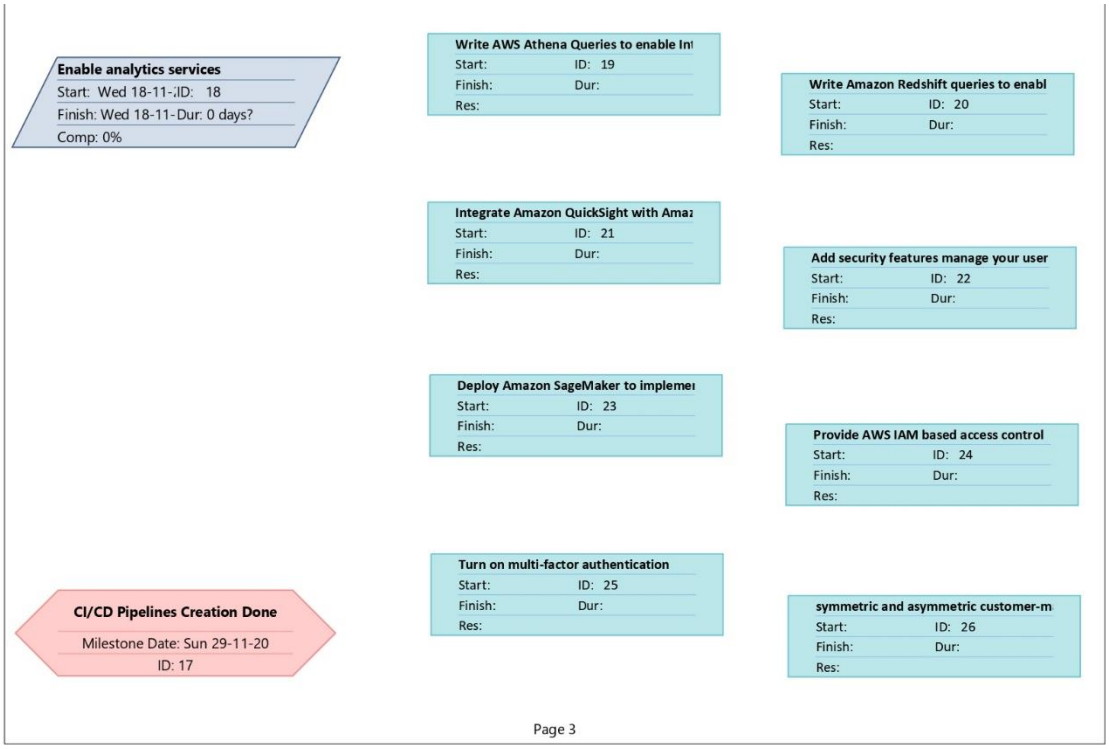


Figure 20 Gantt chart

Network Diagram





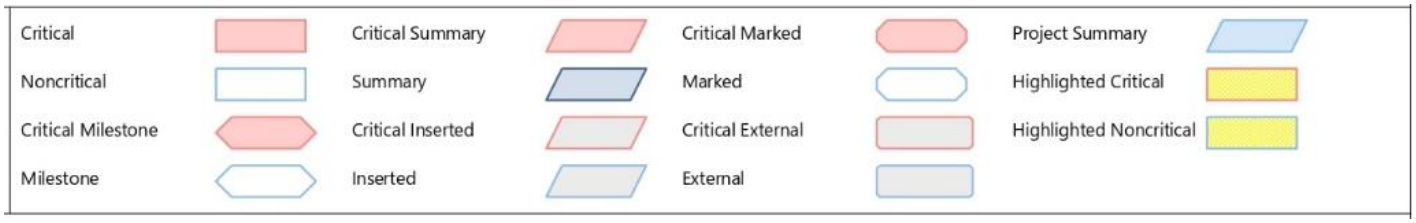


Figure 21 Network Diagram

Appendix

CI/CD	It refers to the combined practices of continuous integration and continuous delivery or continuous deployment. CI/CD bridges the gaps between development and operation activities and teams by enforcing automation in building, testing and deployment of applications.
Big Data	Big data is a field that treats ways to analyse, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.
AWS	Amazon Web Services (AWS) is a subsidiary of Amazon providing on-demand cloud computing platforms on a metered pay-as-you-go basis.
VM	Virtual machine (VM) is an emulation of a computer system. Virtual machines are based on computer architectures and provide functionality of a physical computer. Their implementations may involve specialized hardware, software, or a combination.
DNS	The Domain Name System (DNS) is a hierarchical and decentralized naming system for computers, services, or other resources connected to the Internet or a private network.
DHCP	The Dynamic Host Configuration Protocol (DHCP) is a network management protocol used on Internet Protocol (IP) networks, whereby a DHCP server dynamically assigns an IP address and other network configuration parameters to each device on the network, so they can communicate with other IP networks.
DML	A Data Manipulation Language (DML) is a family of computer languages including commands permitting users to manipulate data in a database.
Orchestration	In system administration, orchestration is the automated configuration, coordination, and management of computer systems and software.
Cloud Computing	Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.
Legacy Systems	In computing, a legacy system is an old method, technology, computer system, or application program, "of, relating to, or being a previous or outdated computer system," yet still in use. This can also imply that the system is out of date or in need of replacement.
Cyber-attacks	In computers and computer networks an attack is any attempt to expose, alter, disable, destroy, steal or gain unauthorized access to or make unauthorized use of an asset. A cyberattack is any type of offensive manoeuvre that targets computer information systems, infrastructures, computer networks, or personal computer devices.
Script	In computer programming, a script is a program or sequence of instructions that is interpreted or carried out by another program rather than by the computer processor (as a compiled program is).
SQL	SQL is a domain-specific language used in programming and designed for managing data held in a relational database management system, or for stream processing in a relational data stream management system.
ETL	In computing, extract, transform, load is the general procedure of copying data from one or more sources into a destination system which represents the data differently from the source(s) or in a different context than the source(s).

References

Aha! (2020). *Project Management Services*. Retrieved from Aha!: https://www.aha.io/roadmap/project

Amazon Web Services. (2020). *AWS*. Retrieved from Cloud Computing Platform: https://aws.amazon.com/

ConceptDraw. (2020). *Business and Technical Diagramming Package*. Retrieved from ConceptDraw Diagramming Tool: https://www.conceptdraw.com/

IIBA BABOK Guide. (2020). *BABOK Guide*. Retrieved from International Institute of Business Analysis™: https://www.iiba.org/

Microsoft Access. (2020). *MS Access*. Retrieved from Microsoft: https://www.microsoft.com/en-CA/microsoft-365/access

Microsoft Project. (2020). *MS Project*. Retrieved from Microsoft: https://www.microsoft.com/en-ca/microsoft-365/project/project-management-software

Miro . (2020). *User Stories*. Retrieved from Online Whiteboard: https://miro.com/

Smartsheet.com. (2020). *Work Management Tool*. Retrieved from Smartsheet: https://www.smartsheet.com/