Achyuth Pilly (C0715089) & Pooja Nakar (C0727416)

Canadian Heritage

Hospital Data MigraTion

Project Report

***Index***

1. Project Overview

1.1 Project Summary

1.2 Milestones

* 1. Deliverables
  2. Project Cost Estimate and Source Funding
  3. Project Risk, Assumptions and Constraint

1. Project Organization

2.1 Project Governance

2.2 Roles and Responsibilities

* 1. Project Facilities and Resources

1. Project references
2. Glossary and Acronyms

Project Overview

Project Summary

*Migration of patient’s data of a hospital on to a relational database for analysis purpose:*

* + *The current state of the data is stored in flat files. Organization was facing difficulties in analyzing the data in the existing format.*
  + *This migration will benefit the analysis team and hospital management.*

### Project Goals, Business Outcomes and Objectives

| No. | Goals | Objectives | Business Outcomes |
| --- | --- | --- | --- |
| 1 | *Setting up the environment* | * *Gathering the physical resources necessary such as Processing Units, Hard disk.* * *Acquire the SAS enterprise version license.* | * *Resources gathered and configured for Installation.* |
| 2 | *Installation of OS, SAS tools* | * *Installing and configuring Linux OS and SAS tools like SAS enterprise minor, SAS SPDS, Base SAS, Enterprise Guide (EG)* | * *Ready environment for data migration.* |
| 3 | *Extraction of all flat files* | * *Provided Hard disk contains all the flat files which need to be extracted and upload on to the Linux server.* | * *All the files reside on server for transformation* |
| 4. | *Transforming the flat file* | * *Data wrangling is performed using the SAS tools.* | * *The data is ready for loading onto the SAS tables.* |
| 5. | *Loading the data on the tables* | * *Schema is designed based on the business requirement then Data is loaded onto the SAS SPDS tables.* | * *The data migration is done and is ready for analysis.* |

### Project Scope

*Migration of patient’s data of a hospital on to a relational database for analysis purpose*

Scope Definition

*The project scope is explained as below points:*

* *Setting up the environment: Gathering the physical resources necessary such as Processing Units, Hard disk. Acquire the SAS enterprise version license.*
* *Installation of OS, SAS tools: Installing and configuring Linux OS and SAS tools like SAS enterprise minor, SAS SPDS, Base SAS, Enterprise Guide (EG). Ready environment for data migration. Test environment, staging environment and production*
* *Extraction of all flat files: Provided Hard disk contains all the flat files which need to be extracted and upload on to the Linux server. All the files reside on server for transformation.*
* *Transforming the flat file: Data wrangling is performed using the SAS tools. Null values, duplicates in the data are removed. Data is indexed. The data is ready for loading onto the SAS tables. Data in different formats is converted into standard format.*
* *Loading the data on the tables: Schema is designed based on the business requirement then Data is loaded onto the SAS SPDS tables. Mapping of the data. Rules are defined. Relationships are defined between the tables. The data migration is done and is ready for analysis.*

### Boundaries

| Activities In Scope | Activities Out of Scope |
| --- | --- |
| 1. This will include documents such as Client Documents for project agreement – SRD, BRD, TIR(Technical Installation Report), TIP (Technical Installation Plan), Technical Support, Application knowledge repository | 1. The Application Server and Process Scheduler used must both be able to access the project files using the same path. This will require that both are running on Operating Systems that use compatible file access conventions |
| 2. Creation of ACL for managements users, end user developers, analysis team | 2. Data qualities inherited except some of the tune ups for the data during the transformation stage. |
| 3. Application training session and application technical support for a month, prior to final billing | 3. Addresses the possible obsolescence of the data carrier, but does not address the fact that certain technologies which run the data may be abandoned altogether, leaving migration useless. |
| 4. Cleansing the data by removing the corrupt files, missing data and duplicate data. | 4. Data analysis is not within the scope of migration. |

Milestones

| Project Milestone | Description | Expected Date |
| --- | --- | --- |
| 1. *Setting up the environment* | *Gathering the physical resources necessary such as Processing Units, Hard disk. Acquire the SAS enterprise version license.* | 3 week |
| 2. *Installation of OS, SAS tools* | *Installing and configuring Linux OS and SAS tools like SAS enterprise minor, SAS SPDS, Base SAS, and Enterprise Guide (EG). Ready environment for data migration.* | 3 week |
| 3. *Extraction of all flat files* | *Provided Hard disk contains all the flat files which need to be extracted and upload on to the Linux server. All the files reside on server for transformation.* | 2 weeks |
| 4. *Transforming the flat file* | *Data wrangling is performed using the SAS tools. Null values, duplicates in the data are removed. Data is indexed. The data is ready for loading onto the SAS tables.* | 4 weeks |
| 5. *Loading the data on the tables* | *Schema is designed based on the business requirement then Data is loaded onto the SAS SPDS tables.* | 4 weeks |

Deliverables

| Project Deliverable 1: *Installation of OS, SAS tools* | |
| --- | --- |
| **Stakeholder:** | DevOps team |
| **Description:** | *Installing and configuring Linux OS and SAS tools like SAS enterprise minor, SAS SPDS, Base SAS, and Enterprise Guide (EG). Ready environment for data migration* |
| **Acceptance Criteria:** | Tools should be platform independent. Environment should be successfully configured, high availability and scalable. |
| Due Date: | 3 weeks |
|  | |
| Project Deliverable 2: *Transforming the flat file* | |
| **Stakeholder:** | Developers |
| **Description:** | *Data wrangling is performed using the SAS tools. Null values, duplicates in the data are removed. Data is indexed. The data is ready for loading onto the SAS tables.* |
| **Acceptance Criteria:** | Data quality should increase not less than 75%. |
| Due Date: | 4 weeks |
|  |  |
| Project Deliverable 3: *Loading the data on the tables* | |
| **Stakeholder:** | End user |
| **Description:** | *Data wrangling is performed using the SAS tools. Null values, duplicates in the data are removed. Data is indexed. The data is ready for loading onto the SAS tables.* |
| **Acceptance Criteria:** | Data quality should increase not less than 75%. |
| Due Date: | 4 weeks |
|  |  |

Project Cost Estimate and Source of Funding

Project Cost Estimate



Project Risks, Assumptions, and Constraints

### Risks

| **No.** | **Risk Description** | **Probability (H/M/L)** | **Impact (H/M/L)** |
| --- | --- | --- | --- |
| 1 | There might be some data loss during the data migration | Medium | Medium |
| 2 | There might be downtime for the Service. | Low | High |

### Assumptions

The following table lists the items that cannot be proven or demonstrated when this project charter was prepared, but they are taken into account to stabilize the project approach or planning.

| No. | Assumptions |
| --- | --- |
| 1 | Raw data given in the form of hard disk is readable. |
| 2 | The given raw data volume is less than 20 PB. |

### Constraints

The following table lists the conditional factors within which the project must operate or fit.

| No. | Constraints |
| --- | --- |
| 1 | Cannot deal with streaming data. |
| 2 | To Load the final data will require at least 2-3 days of time. |

Project Organization

Project Governance

Roles and Responsibilities

| Project Role | Responsibilities | Assigned to |
| --- | --- | --- |
| Team Leader | To validate the user requirements. To manage the team and get the result. To review the outcome. | Pooja Nakar |
| Dev Ops team | Acquire the licence and do the necessary installation for setting up the environment for data migration | Achyuth Pilly |
| Developers | Perform data wrangling on the raw data. And migrate the data into schema specified tables. | Achyuth Pilly And Rizwan. |
| Tester | Test the data loaded in the table with valid values. Do negative testing and prepare test report. | Achyuth Pilly |

Project Facilities and Resources

*Dev Ops team is responsible for acquiring the licence of the SAS tools and installations of it. Resource manager will be providing the necessary hardware*.

* *Processing Unit, Hard disk, Storage*
* *Software tools Linux OS and SAS tools like SAS enterprise minor, SAS SPDS, Base SAS, and Enterprise Guide (EG).*

Project References

More information concerning this project can be found in the following documents:

| Document Title | Version # | Date | Author and Organization | Location (link or path) |
| --- | --- | --- | --- | --- |
| Project Charter Guide | 1D | 20-Mar-2019 | Chief Information Officer Branch (CIOB) | Y:\CIOB\Template |
| BRD | 2.1 | 20-Mar-2019 | CIOB Team | Y:\CIOB\BRD |
| Test Plan | 1.3 | 23-May-2019 | CIOB Team | Y:\CIOB\Test Plan |
| Test Report | 1.5 | 25-May-2019 | CIOB Team | Y:\CIOB\Test Report |

Glossary and Acronyms

| **Term** | **Definition** |
| --- | --- |
| SAS | software suite developed by [SAS Institute](https://en.wikipedia.org/wiki/SAS_Institute) for advanced analytics, [multivariate analysis](https://en.wikipedia.org/wiki/Multivariate_analysis), [business intelligence](https://en.wikipedia.org/wiki/Business_intelligence), [data management](https://en.wikipedia.org/wiki/Data_management), and [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics). |
| Data Wrangling | *Data wrangling* is the process of cleaning, structuring and enriching raw data into a desired format for better decision making in less time. |
| **Acronym** | **Name in Full** |
| SAS | Statistical Analysis System |
| ACL | Access Control level |