BCBSRI_CorpMaster_Logo

****

**Snowflake, Blade Bridge & Matillion - POC**

|  |  |
| --- | --- |
| **Version:** | 0.1 |
| **Version Date:** | 03/10/2022 |
| **Prepared By:** | Sunil Rangineni |

**Table of** **Contents**

[1. Executive Summary 3](#_Toc46755882)

[2. Problem Statement 3](#_Toc46755883)

[3. Solution Statement 4](#_Toc46755884)

[4. Associated Documents 6](#_Toc46755885)

[5. System Architecture Diagrams 6](#_Toc46755886)

[System Context Diagram 6](#_Toc46755887)

[Data Context- Cloud DW 7](#_Toc46755888)

[Conceptual Tasks - 8](#_Toc46755889)

[6. Support Responsibilities 9](#_Toc46755890)

[7. Infrastructure Architecture Risks 9](#_Toc46755891)

[8. Design Decisions 9](#_Toc46755892)

[9. Stakeholders/Approvers 9](#_Toc46755893)

[10. Document Revision History 10](#_Toc46755894)

[Appendix A – Design Decision Details 11](#_Toc46755895)

[General 11](#_Toc46755896)

[Security 11](#_Toc46755897)

[Appendix B 12](#_Toc46755898)

[Glossary 12](#_Toc46755899)

[Enterprise Architecture Diagramming Standards 12](#_Toc46755900)

# Executive Summary

As part of BCBSRI’s intent to migrate to nextgen data plaforms, we have to evaluate different technology platforms for replacing IBM Infosphere DataStage & IBM Netezza.

# Problem Statement

*What is the business or technical problem that this phase of the SA document is trying to solve?*

BCBSRI is looking for options to replace the existing warehouse and analytics platforms with modernaized data integration and analytical platforms on the cloud. The primary objective of this proof of value is to evaluate the technology platforms Blade Bridge, Matillion & Snowflake for viability to replace IBM Inforsphere DataStage & Netezza platforms.

# Solution Statement

*What is the technical solution outlined in this document to address the above mentioned business problem?*

* The primary technical objectives of this project are to:
* Perform proof of concept of the future state technology stack with BladeBridge, Matillion & Snowflake.

|  |  |  |
| --- | --- | --- |
| Category | Metric | Comments |
| Schema & Data Migration | Schema Conversion (DDL's) & Deployment efforts to Snowflake | Are there any automation tools that converts DDL's? |
| Initial Data Migration Efforts to Snowflake. | What's the ideal method to migrate data? ADF? |
| ETL Conversion, Migration & Execution | Blade Bridge Conversion Efforts & Automation Efficiency. | Efforts for Blade Bridge conversion of DataStage Code? Classification of DS code into simple, medium vs complex by Blade Bridge? Automatic Conversion efficiency of DataStage code? |
| Matillion code ETL migration efforts. | Efforts required to migrate code generated by Blade Bridge to Matillion? Manual Efforts required to the code to get the process into executable state? Configuration Efforts (if any)? Execution Times of Income Code within Matillion? |

Dimensional entities (out of our Income Data Mart) is the candidate for POC. There is currently no PHI information with the selected entities.

**Below are the current metrics of Income subject area:**

**Code Base Metrics:**

* 100+ IBM DataStage Jobs
* 10+ Netezza Tables (EDR) – Used reference keys.
* 35+ Netezza Tables (Income Data Mart)
  + 10+ Core tables
  + 25+ Audit/Stage/Reference.
* 45+ Netezza Tables (FDA – Facets Data Access)
* 5+ Oracle Tables for Reference data.

**Volumetrics (approx.):**

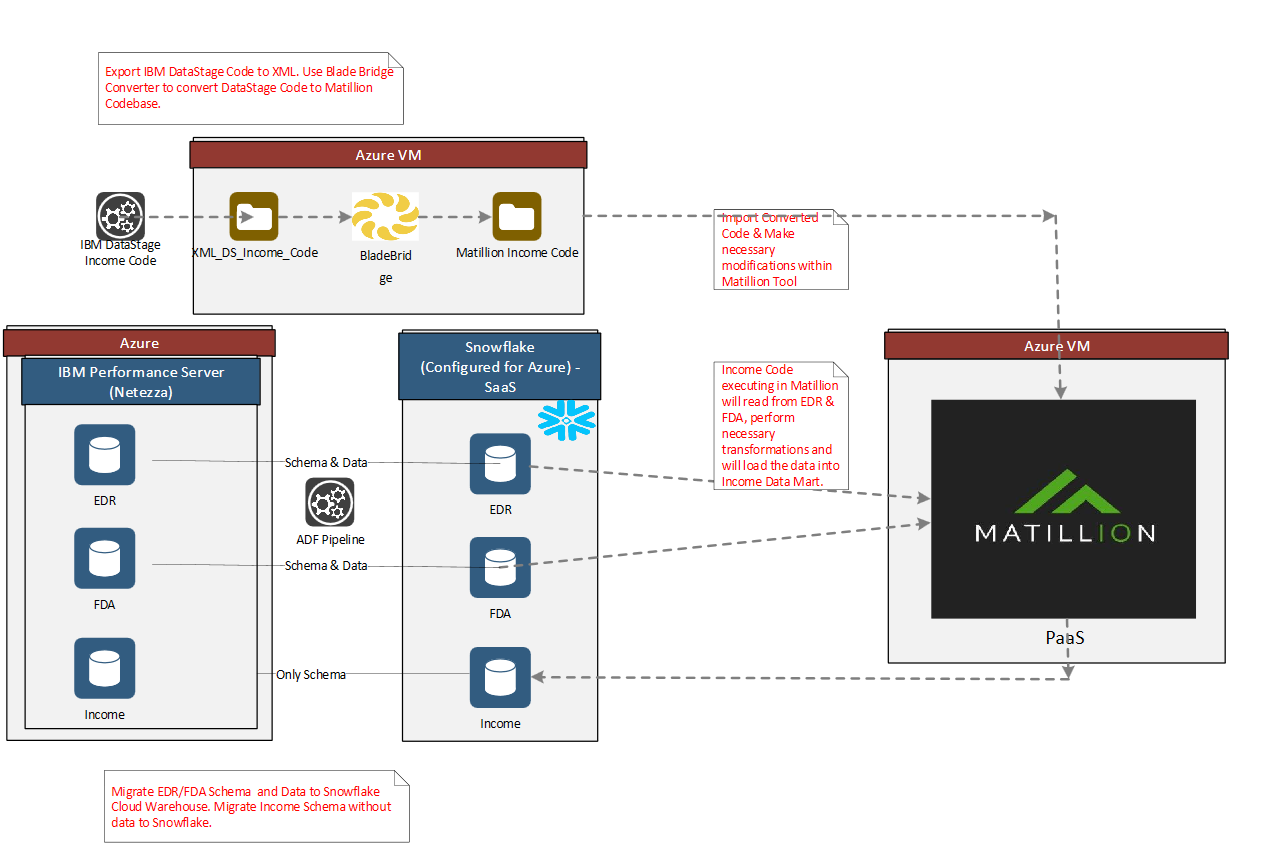
* 2+ GB -- Income Data Mart
* 40+ GB – FDA
* 10+ GB – Others (EDR/Oracle/Reference Data)

# Associated Documents

# System Architecture Diagrams

## System Context Diagram

This section contains a system context view of the architecture.



## Data Context- Cloud DW

Depicts this system as a “black box” and identifies data flowing in and out of this system.

Attached is the list of tables required to be



## Conceptual Tasks -

|  |  |  |
| --- | --- | --- |
| Category | Task | Assigned To |
| Snowflake | Acquire Snowflake Subscription & License |  |
| Setup necessary Snowflake configurations |  |
| Configure BCBSRI Azure subscriptions to Snowflake subscription ( Private Endpoint connections??) |  |
| Create Database, Schemas (if any) & Users in Snowflake. |  |
| Migrate DDL's for EDR, FDA & Income in Snowflake. |  |
| Data Migration  Prep-Activity | Extract DDLs for the Income objects from the following Netezza DBs  EDR, FDA & Income |  |
| Convert the DDLs to match to Snowflake Datatypes. |  |
| Create table and process to deidentify member\_id column from member table in EDR. |  |
| Data Migration | Create Tables using the scripts developed in Prep-Activity in Snowflake. |  |
| Migrate Data from EDR, FDA (Netezza) to Snowflake. |  |
| Azure | Identify an BCBSRI Azure Subscription (Dev Vs POC subscription) |  |
| Acquire new Azure VM's & Storage if required. |  |
| Identify network and connection requirements from Azure to other SaaS services. |  |
| ADF pipelines for pushing data from Netezza to Snowflake?? |  |
| Identify Firewall changes required (if any) and implement them (typically a 2 weeks turn around time for any firewall change in BCBSRI) |  |
| Data Stage | Identify DataStage Code for Income and create a xml/DataStage export file. |  |
| Transfer the file to Azure VM. |  |
| Blade Bridge | Acquire Blade Bridge Software and License. |  |
| Install Blade Bridge on the Azure VM. (Get Infosec Approvals as required). |  |
| Configure Blade Bridge to convert DataStage code to Matillion code. |  |
| Run Blade Bridge to create Matillion Code files. |  |
| Matillion. | Acquire Matillion subscription & license. |  |
| Configure BCBSRI Azure subscription to work with Matillion Subscription ( private endpoint connections if any) |  |
| Setup necessary Matillion Configurations. (source & target connections). |  |
| Import code from Blade Bridge conversion files into Matillion. |  |
| Make necessary Manual changes to code within Matillion |  |
| Execute Income Sequence within Matillion. |  |

**Infrastructure Pre-Requisites**

|  |  |  |
| --- | --- | --- |
| Infrastructure Pre-Requisite | Assigned To | Comments |
| Snowflake Subscription & License |  |  |
| Blade Bridge Converter |  |  |
| Matillion Subscription & License |  |  |
| Azure VM's |  |  |
| Azure Private Endpoints (if any) to connect to Snowflake & Matillion. |  |  |

# Support Responsibilities

This section identifies which teams are or will be responsible for supporting elements of this solution

|  |  |  |  |
| --- | --- | --- | --- |
| **Logical Architecture Node** | **Component** | **Support Team** | **New Type Of Support?** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Infrastructure Architecture Risks

These are any risks associated with this architecture. Only those that either directly impact or directly result from the architecture are covered here. They are operational risks presented that will be present in production; they are not project or delivery risks.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Risk** | **Likelihood** | **Impact** | **Mitigation** |
| 1 |  |  |  |  |

# Design Decisions

This “Table of Contents” summarizes the design decisions. For more information on these decisions, refer to Appendix A.

# Stakeholders/Approvers

|  |  |  |
| --- | --- | --- |
| **Name** | **Role** | **Context in Architecture** |
| Tracy McPhillips | Data & Analytics | Cloud DW |
| Tracey Heneghen | IT Enterprise Architecture | Cloud and Infrastructure services |
| Jeremy Williams | System operations | Cloud Network infrastructure,and Security and system operations |
| Jon Fredrickson | Information Security | Data and Information security |

# Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Modifications** | **Author** |
| 03/10/2022 | 0.1 | Document Creation | Sunil Rangineni |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Appendix A – Design Decision Details

This section catalogs the design decisions for this design. Security decisions are called out into a specific section.

## General

|  |  |
| --- | --- |
| 001: | |
| **Analysis** |  |
| **Desired End State** |  |
| **Factors contributing to not achieving End State** |  |
| **Associated Risks** |  |

## Security

|  |  |
| --- | --- |
| 020: | |
| **Analysis** |  |
| **Desired End State** | * This is the end state. |
| **Factors contributing to not achieving End State** |  |
| **Associated Risks** |  |

# Appendix B

## Glossary

|  |  |
| --- | --- |
| **Term** | **Definition** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## Enterprise Architecture Diagramming Standards