**CONFIDENTIAL**

****

|  |
| --- |
| **Proof of Concept (POC)**  **Requirements & Design Document**  **For**  **Data Platform & Governance** |

This document is owned by Bursa Malaysia Berhad and/or the Bursa Malaysia group of companies (“Bursa Malaysia”). No part of the document is to be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission in writing from Bursa Malaysia.

Prepared By: **[Bursa Data Business/ Enterprise Data Management] &   
[Group Technology/Data Platform]**

Table of Contents

[1 INTRODUCTION 8](#_Toc82441023)

[1.1 Instructions to Vendors 8](#_Toc82441024)

[1.2 Objective 8](#_Toc82441025)

[1.3 Scope 9](#_Toc82441026)

[2 TECHNICAL REQUIREMENTS 13](#_Toc82441027)

[2.1 Cloud Architecture Diagram 13](#_Toc82441028)

[2.2 Talend Modules and Server Architecture Diagram 14](#_Toc82441029)

[2.2.1 Hardware Requirements Specification 14](#_Toc82441030)

[2.3 Talend Data Flow Diagram 15](#_Toc82441031)

[2.3.1 Option 1 16](#_Toc82441032)

[2.3.2 Option 2 16](#_Toc82441033)

[2.3.3 Option 3 17](#_Toc82441034)

[2.4 Network & Security Requirements -Outbound Rules 14](#_Toc82441035)

[2.4.1 Outbound rules for EC2 Engine hosted in Bursa Cloud 14](#_Toc82441036)

[2.4.2 Inbound rules for EC2 Engine hosted in Bursa Cloud 15](#_Toc82441037)

[2.4.3 Inbound rules for Redshift Database - Bursa Cloud 16](#_Toc82441038)

[2.4.4 Inbound rules for SFTP Server - Bursa on-Prem Data Centre 17](#_Toc82441039)

[3 USER REQUIREMENTS 18](#_Toc82441040)

[3.1 An Overview of Use Cases (Data Governance) 18](#_Toc82441041)

[3.1.1 Use Case 1 - Self Service search and access to data 19](#_Toc82441042)

[3.1.2 Use Case 2 - Management view on data management matrix/KPI 20](#_Toc82441043)

[3.1.3 Use Case 3 - Business glossary 21](#_Toc82441044)

[3.1.4 Use Case 4 -Data profiling based on pre-define data quality rules 22](#_Toc82441045)

[3.1.5 Use Case 5 - Demonstrate impact to the downstream systems (i.e. data lineage) when there is change to metadata at the source system 23](#_Toc82441046)

[3.1.6 Use Case 6 - Merging of master data and creation of golden source master data 24](#_Toc82441047)

[3.1.7 Use Case 7 – Scan Metadata 26](#_Toc82441048)

[3.1.8 Use Case 8 - Centralize access to common statistics 27](#_Toc82441049)

[3.2 Data Platform 28](#_Toc82441050)

[3.2.1 Data Anonymization 30](#_Toc82441051)

[3.2.2 Dashboard Development (Use Case 9) 31](#_Toc82441052)

[3.2.3 Entity Relationship Diagram – BursaLink, BMGS & CDS 32](#_Toc82441053)

[3.2.4 Business Transformation Rules 33](#_Toc82441054)

[3.2.5 Business KPI Measurements – from BDB-DA Team 35](#_Toc82441055)

[4 APPENDICES 36](#_Toc82441056)

[4.1 Schema Overview for the Nine CSV Files 36](#_Toc82441057)

[4.2 An Overview of BursaLink Metadata 36](#_Toc82441058)

[4.3 An Overview of CDS Metadata 36](#_Toc82441059)

[4.4 An Overview of BMGS Metadata 37](#_Toc82441060)

List of Figures

[Figure 1: High level Data Platform Architecture Diagram 9](#_Toc82432227)

[Figure 2: Entity Relationship Diagram for 3 common Tables identified in the POC project 10](#_Toc82432228)

[Figure 3: AWS Architecture Diagram 13](#_Toc82432229)

[Figure 4: Talend Modules and Server Architecture Diagram 14](#_Toc82432230)

[Figure 5:Talend Data Flow Diagram 15](#_Toc82432231)

[Figure 6: Data Lineage Option 1 16](#_Toc82432232)

[Figure 7: Data Lineage Option 2 16](#_Toc82432233)

[Figure 8: Data Lineage Option 3 17](#_Toc82432234)

[Figure 9: Use Case Diagram 18](#_Toc82432235)

[Figure 10: Use Case 1 – Self Service search and access to data 19](#_Toc82432236)

[Figure 11: Use Case 1a - Details 19](#_Toc82432237)

[Figure 12: Use Case 2 - Management view on data management matrix/KPI 20](#_Toc82432238)

[Figure 13: Use Case 2a - Details 20](#_Toc82432239)

[Figure 14: Use Case 3 – Business Glossary 21](#_Toc82432240)

[Figure 15: Use Case 3a – Details 21](#_Toc82432241)

[Figure 16: Use Case 4 – Data Profiling based on pre-define data quality rules 22](#_Toc82432242)

[Figure 17: Use Case 4a – Details 22](#_Toc82432243)

[Figure 18: Use Case 5 – Demonstrate impact to the downstream systems (i.e. data lineage) when there is change to metadata at the source system. 23](#_Toc82432244)

[Figure 19: Use Case 5a – Details 24](#_Toc82432245)

[Figure 20: Use Case 6 – Merging of master data and creation of golden source master data 25](#_Toc82432246)

[Figure 21: Use Case 6a - Details 25](#_Toc82432247)

[Figure 22: Use Case 7 – Scan Metadata 26](#_Toc82432248)

[Figure 23: Use Case 7a - Details 26](#_Toc82432249)

[Figure 24: Use Case 8 – Centralize access to common statistics 27](#_Toc82432250)

[Figure 25: Use Case 8a - Details 27](#_Toc82432251)

[Figure 26: Current Setup for AWS Tactical Solution 28](#_Toc82432252)

[Figure 27: Proposal for Setup for Data Platform 28](#_Toc82432253)

[Figure 28: Proposal for Data Lineage for Azure Data Platform 29](#_Toc82432254)

[Figure 29: Attributes/Fields to be Anonymized 30](#_Toc82432255)

[Figure 30: Report 1 (R1) 31](#_Toc82432256)

[Figure 31: Report 2 (R2) 31](#_Toc82432257)

[Figure 32: Report 3 (R3) 31](#_Toc82432258)

[Figure 33: Report 4 (R4) 31](#_Toc82432259)

[Figure 34: Report 5 (R5) 32](#_Toc82432260)

[Figure 35: Dataset “BursaLink\_CDS\_BMGS ” 32](#_Toc82432261)

List of Tables

[Table 1: User Role and Respective Functions 8](#_Toc82432262)

[Table 2: Three sources to be ingested into POC Project 10](#_Toc82432263)

[Table 3: Definitions, Acronyms and Abbreviations 11](#_Toc82432264)

[Table 4: Hardware Requirements Spec for VPC 13](#_Toc82432265)

[Table 5: Outbound rules for EC2 Engine hosted in Bursa Cloud 14](#_Toc82432266)

[Table 6: Inbound rules for EC2 Engine hosted in Bursa Cloud 16](#_Toc82432267)

[Table 7: Inbound rules for Redshift Database - Bursa Cloud 16](#_Toc82432268)

[Table 8: Inbound rules for SFTP Server - Bursa on-Prem Data Centre 17](#_Toc82432269)

[Table 9: DQ Rules to Apply for Respective Attributes 23](#_Toc82432270)

[Table 10: Profiles to create Master Table 25](#_Toc82432271)

[Table 11: Rules for Transformation from Source 34](#_Toc82432272)

[Table 12: Transformation for Dashboard 35](#_Toc82432273)

Requirements Acceptance

|  |  |  |  |
| --- | --- | --- | --- |
| Requestor | | Date | |
| …………………………..  Goh Leng Leng  VP – Enterprise Data Management  Market Operations  Bursa Malaysia |  | |
| …………………………..  Tan Sieow Yeek  VP – Data Analytics & Products  Biz. Analytics & Info. Services  Bursa Malaysia |  | |
| …………………………..  Sadesh Manikam  SVP - Data Platform  Group Technology  Bursa Malaysia |  | |

|  |  |
| --- | --- |
| Vendor | Date |
| …………………………..  Akshant Trivedi  Solution Engineer  Celebal |  |

# INTRODUCTION

Bursa Malaysia is one and the only national Stock Exchange of Malaysia. It operates a fully integrated exchange, offering the complete range of exchange-related services including listing, trading, clearing, settlement and depository services.

Aspire to become data driven organisation, Bursa Malaysia have adopted an Enterprise Data Management framework in May 2020 based on DAMA BOK, ISO8000 and CMMI. The framework layout the key building blocks and knowledge areas that are to be followed by Bursa Malaysia in order to elevate data management maturity level and become data driven organisation. Bursa Malaysia has embarked on the foundational activities to support the data lifecycle processes which include establishment of Data Governance Steering Committee, Data Governance Working Group, standard operating procedures for all the data management knowledge areas as per the framework. Since the establishment of the framework, we have operationalized Data Governance, Business Glossary, Metadata Management and Data Classification SOP manually. The next step is to operationalize data quality management.

Bursa Malaysia is looking for Data Governance, Metadata Management and Data Quality Management tools that could facilitate the following:

1. Efficient discovery and tracking of metadata

2. Detection of data quality issues to build trusted data

3. Enforcement of data governance policies and procedures

4. Enable self-service access to data

## Instructions to Vendors

The POC belongs to Bursa Malaysia. The vendor is required to hold all proprietary and confidential information related to the Bursa Malaysia in the strictest form complying to Non-Disclosure Agreement agreed by both Bursa Malaysia and the participating vendor. Any code snippets or intellectual property within the POC scope is strictly prohibited to be duplicated, distributed or disclosed to any third party without written consent from Bursa Malaysia.

## Objective

The objective of this POC is to explore features and functions in relation to:

* Data Governance
* Metadata Management
* Data Quality Management
* Master Data Management

## Scope

The scope of this POC involves the following user roles and exploration of the following functions based on use cases defined in the section 3.

|  |  |  |
| --- | --- | --- |
| **No** | **User Role** | **What they can do / Function** |
| 1 | Data User | a) Search for data b) Self Service Access to data c) Access to common statistics |
| 2 | Management User | View dashboard on data governance matrix |
| 3 | Enterprise Data Steward (EDS) | a) Validate and approve business glossary b) Set Data Governance policies c) Merge Master Data |
| 4 | Business Data Steward  (BDS) | a) Tag data sensitivity level b) map business glossary to metadata c) create and update business glossary d) define data quality rules |
| 5 | Technical Data Steward (TDS) | a) Access to data lineage b) Perform impact analysis due to changes to metadata |
| 6 | Data / Business Owner | a) Approve access to data |

Table 1: User Role and Respective Functions

To facilitate the above POC, connectivity to the following source system and the relevant data in the development environment need to be setup as per the following diagram:

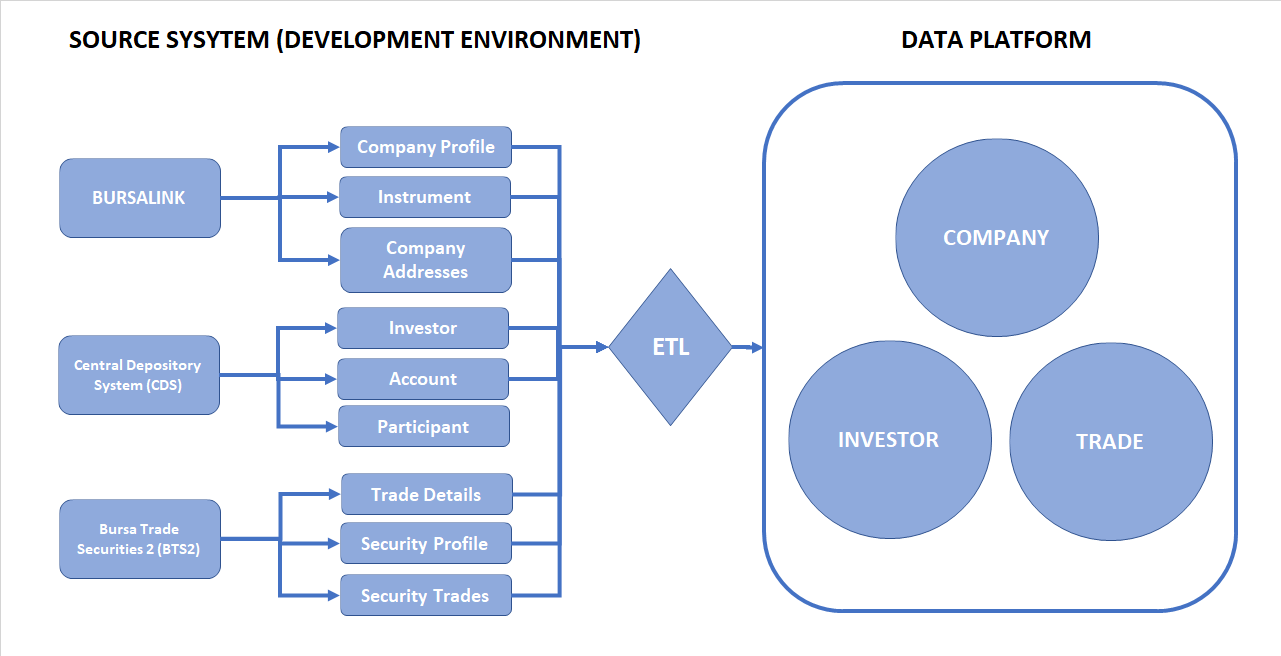


Figure 1: High level Data Platform Architecture Diagram

Figure 2: Entity Relationship Diagram for 3 common Tables identified in the POC project

The POC use cases will be based on the data ingested from the above source systems. ETL is to be developed by vendor and dashboard on the data platform will be developed by Bursa Malaysia internal Data Analytics team*. Data Platform will be provided by Bursa Malaysia. Vendor is to integrate their tools with Bursa Malaysia existing data platform.*

As minimum requirement, the following data is to be ingested from the source system for this POC (latest 3 months of data, size to be provided by TDS):

| **System** | **Data** | **Table / Object** | **Attributes** |
| --- | --- | --- | --- |
| **BursaLink** | **Company Profile** |  |  |
|  | - Master Profile | PR\_COMPANIES | Name, Regno, Board, Sector, FYE |
|  | - Address | PR\_COMPANY\_ADDRESSES | Address, Contacts |
|  | - Instrument Profile | PR\_INSTRUMENT\_PROFILES | Stock Code, ISIN Code, Listing Date, Date Delisted |
| **CDS** | **Investor Profile** |  |  |
|  | - CDS Accounts Details | ACCOUNT | Account Type, Status, Open Date |
|  | - Investor /Depositor | INVESTOR | Investor Type and Demography |
|  | - Nationality | NATIONALITY | Nationality |
| **BTS2 (BMGS)** | **Market Data** |  |  |
|  | - Securities Detail | SECURITY | Name, Issued Qty, Sector, Listed Date |
|  | - Market Data | SECBOARD | Open, High, Low Close, Trade Vol / Val |
|  | - Trade Detail | TRADE\_CONSUMPTION\_DATA | Buy/Sell Type, Broker, Price, Quantity |
|  |  |  |  |

Table 2: Three sources to be ingested into POC Project

*Please refer to Appendix for the metadata of the systems.*Definitions, Acronyms and Abbreviations

|  |  |  |
| --- | --- | --- |
| **No** | **Acronym/ Abbreviation** | **Descriptions** |
| 1 | BDS | Business Data Steward |
| 2 | Data Owner | Individual who owns the data |
| 3 | DG | Data Governance |
| 4 | DGSC | Data Governance Steering Committee |
| 5 | DGWG | Data Governance Working Group |
| 6 | DQ | Data Quality Management |
| 7 | EDM | Enterprise data management |
| 8 | EDS | Enterprise Data Steward |
| 9 | POC | Proof of Concept |
| 10 | MM | Metadata Management |
| 11 | System Owner | Individual who owns the system |
| 12 | TDS | Technology Data Steward |
| 13 | BMD | Bursa Malaysia Derivatives |
| 14 | ERM | Enterprise Risk Management |

Table 3: Definitions, Acronyms and Abbreviations

# TECHNICAL REQUIREMENTS

Refer latest task list document

## Cloud Architecture Diagram

Figure 3 below describes the security and network configuration of EC2 via Virtual Machine (VM) for Talend Apps and Bastion Jump Host whereby whitelisted users from Talend are provided access to both Talend Apps and Bastion Jump Host. User from Talend is requested to access RedShift via Bastion Jump Host.

Both Talend Apps and Bastion Jump Host will be hosted in public subnet while RedShift will be hosted in protected subnet.

Bursa development applications servers such as CDS, BursaLink and BTS2(BMGS) will push the CSV file to SFTP server. Meanwhile, the Talend Apps server will pull the CSV file hosted by the SFTP server.

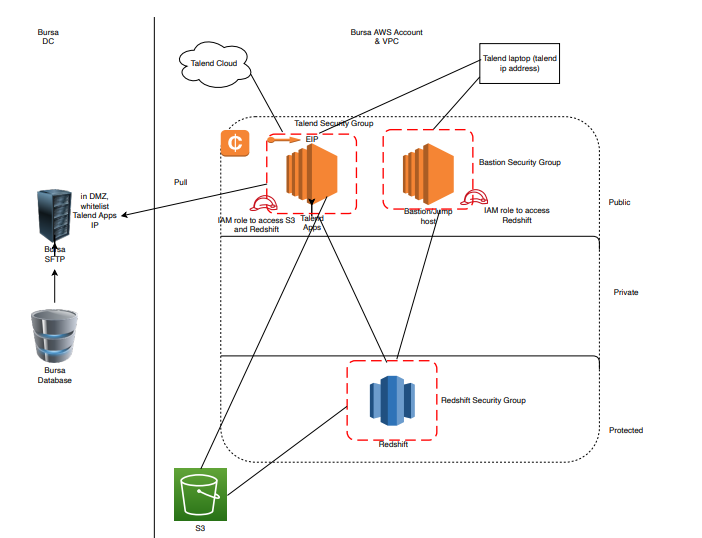


Figure 3: AWS Architecture Diagram

## Talend Modules and Server Architecture Diagram

Figure 4 represents Talend Modules which consists of Talend Data Catalog (installed on EC2 via AWS Cloud) and Talend Data Fabric is posted in Talend Virtual Private Cloud (VPC).

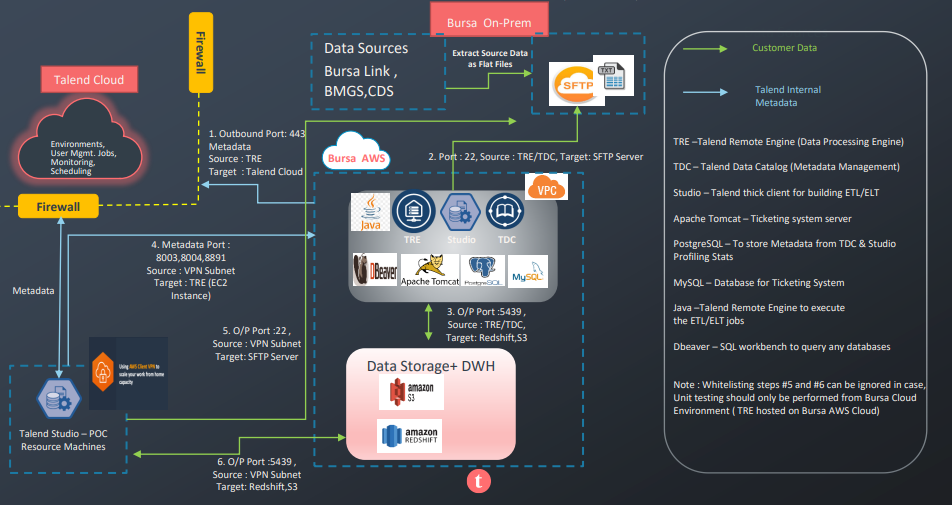


Figure 4: Talend Modules and Server Architecture Diagram

The Talend Data Catalog (TDC) will communicate with Talend Data Fabric (TDF) via public IP which the range of IP from the TDF environment has been whitelisted. The thick client via Talend vendors has been also included in the whitelisting table in AWS.

### Hardware Requirements Specification

|  |  |  |
| --- | --- | --- |
| **Node Desc** | **Node Spec.** | **Size** |
| Talend Data Catalog (TDC) App. Server + Remote Engine | Nodes # | 1 |
| Server Cores | 8 x Intel Xeon 2.1GHz |
| Server RAM | 64 GB |
| Storage | 200 GB |
| Operating System | Windows Server 2019/2016 |
| Server Type | AWS EC2—R5.2xlarge |
| Database | PostgreSQL opensource |

Table 4: Hardware Requirements Spec for VPC

## Talend Data Flow Diagram

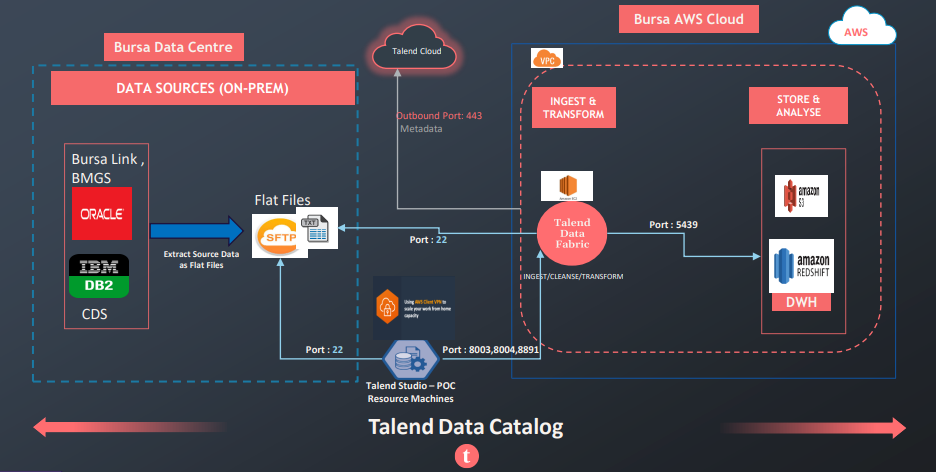
Figure 5 shows the data flow between data source (BursaLink, BMGS, CDS) and Talend Data Catalog (TDC). Due to project timeline and scope, site to site VPN was not able to be established due to constraint in scope and timeline. Thus, an intermediate medium of SFTP was recommended for exchange of data and metadata from data source to TDC using Comma Separated Value (CSV) formatted file.

Figure 5:Talend Data Flow Diagram

Unfortunately, metadata from the data source will not be captured in the CSV file. Instead, metadata management will only start from SFTP onwards. After a lengthy discussion, the POC project team agreed to use Option 2 as it will manage the scope creeping within the timeline. Below are the three options discussed.

### Option 1

Bursa provides Oracle data into flat files & based on the figure 6 below using SFTP port: 22. So, in this case data lineage start from flat file > Target (S3 & Redshift).

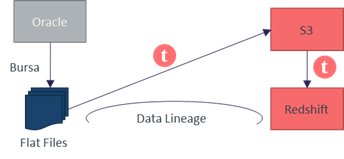


Figure 6: Data Lineage Option 1

### Option 2

Direct connection to Oracle where Talend job will extract the data, if Bursa allows TDC to allow to access Oracle metadata using read only id. Then data lineage will come from Oracle > Target (S3 & Redshift).

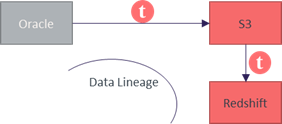


Figure 7: Data Lineage Option 2

### Option 3

Talend is also able to generate flat file from Oracle & then SFTP it on S3 & load into redshift. In this case, also TDC requires read access on Oracle Metadata, File system in between & Target system. So, data lineage will come from Oracle > Flat file >Target (S3 & Redshift).

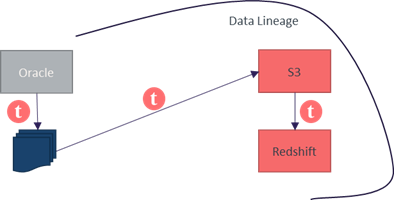


Figure 8: Data Lineage Option 3

In summary, Bursa has agreed for Option 1 to be used for POC and Option 2 for production implementation.

## Network & Security Requirements -Outbound Rules

### Outbound rules for EC2 Engine hosted in Bursa Cloud

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Port range** | **Protocol** | **Source** | **Source Infrastructure** | **Direction** | **Security groups** | **Remarks** |
| 22 | SFTP(SSH) | SFTP or FTP Server IP | Bursa On-Prem Data Centre | OUT | Talend Secutity Group | Ignore |
| 443 | TCP | https://accounts-iam.ap.cloud.talend.com https://ap.cloud.talend.com https://apid.ap.cloud.talend.com https://cloud.talend.com https://dts.ap.cloud.talend.com https://engine.ap.cloud.talend.com https://iam.ap.cloud.talend.com https://ipaas.ap.cloud.talend.com https://log.ap.cloud.talend.com https://lts.ap.cloud.talend.com https://msg.ap.cloud.talend.com https://pair.ap.cloud.talend.com https://portal.ap.cloud.talend.com https://tdp.ap.cloud.talend.com https://tds.ap.cloud.talend.com https://tmc.ap.cloud.talend.com https://webhooks.ap.cloud.talend.com https://repo.ap.cloud.talend.com | Talend Cloud | OUT | Talend Secutity Group |  |
| 5439 | TCP | Redshift Cluster | Bursa AWS Cloud | OUT | Talend Secutity Group |  |

Table 5: Outbound rules for EC2 Engine hosted in Bursa Cloud

### Inbound rules for EC2 Engine hosted in Bursa Cloud

| **Port range** | **Protocol** | **Source** | **Source Infrastructure** | **Direction** | **Security groups** |  |
| --- | --- | --- | --- | --- | --- | --- |
| 8003 | TCP | XXX. XXX. XXX.XXX | POC resource systems via  AWS Client VPN Connectivity | IN | Talend Secutity Group |  |
| 8004 | TCP | XXX. XXX. XXX.XXX | Talend/Partner Resource laptop | IN | Talend Secutity Group |  |
| 8891 | TCP | XXX. XXX. XXX.XXX | Talend/Partner Resource laptop | IN | Talend Secutity Group |  |
| 3389 (RDP) | TCP | XXX. XXX. XXX.XXX | Talend/Partner Resource laptop | IN | Talend Secutity Group |  |
| 5432 (PostegreSQL Port) | TCP | XXX. XXX. XXX.XXX | Talend/Partner Resource laptop | IN | Talend Secutity Group |  |

Table 6: Inbound rules for EC2 Engine hosted in Bursa Cloud

### Inbound rules for Redshift Database - Bursa Cloud

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Port range** | **Protocol** | **Source** | **Source Infrastructere** | **Direction** | **Security groups** | **Remarks** |
| 5439 | TCP | EC2 Engine hosted in Bursa Cloud | AWS Bursa Cloud | IN | Redshift Secutity Group |  |
| 5439 | TCP | Bastion Server | AWS Bursa Cloud | IN | Redshift Secutity Group |  |

Table 7: Inbound rules for Redshift Database - Bursa Cloud

### Inbound rules for SFTP Server - Bursa on-Prem Data Centre

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Port range** | **Protocol** | **Source** | **Source Infrastructure** | **Direction** | **Security groups** | **Remarks** |
| 22 | SFTP | EC2 Engine hosted in Bursa Cloud | AWS Bursa Cloud | IN | XXXXX-XXXXXX-21 | ignore |
| 22 | SFTP | AWS Client VPN Subnet | POC resource systems via AWS Client VPN Connectivity | IN | XXXXX-XXXXXX-21 | ignore |

Table 8: Inbound rules for SFTP Server - Bursa on-Prem Data Centre

# USER REQUIREMENTS

## An Overview of Use Cases (Data Governance)

This POC will cover eight use cases as per the Use Case Diagram below.



Figure 9: Use Case Diagram

The subsequent sub-sections elaborate each use case above and the possible scenarios.

### Use Case 1 - Self Service search and access to data

Allow data user to search for data via the data catalogue and facilitate access to the data. While facilitating the access, ensure that data is protected based on the sensitivity level. User will then be allowed to store the data in data lake for reuse later.

**Figure 10**: Use Case 1 – Self Service search and access to data



Figure 11: Use Case 1a - Details

**Scenarios Example (Search by data sensitivity level):**

1. Public information - User search for Bursa Malaysia, details for Bursa Malaysia which include stock code, Registered address, ISIN code should be displayed
2. Confidential information - User search for Rakuten, approval from data owner will be required, upon approval, system should display CDS account number (masked), New IC Number (masked), and any other account information available in the data lake or source system.

### Use Case 2 - Management view on data management matrix/KPI

Show case on the data governance dashboard providing the following KPIs:

a) Number of data access with status (approved, rejected, etc.)

b) Number of business glossary changes (created, updated, etc.)

c) Number of metadata changes

d) Number of data quality issues and status (whether it has been rectified)



Figure 12: Use Case 2 - Management view on data management matrix/KPI



Figure 13: Use Case 2a - Details

**Scenarios Example:**

1. User performed following transactions and view the statistics (able to drill down to the details of the transactions):
2. Data access requests (UC1)
3. Business glossary changes (UC3)
4. Metadata changes (UC5)
5. Raised data quality issues based on the DQ profile (UC4)

### Use Case 3 - Business glossary

Show case on how the BDS can create and update the business glossary and tag to the metadata

**Figure 14:** Use Case 3 – Business Glossary



**Figure 15**: Use Case 3a – Details

**Scenarios example:**

1. New Business Glossary - BDS create new business glossary for Board, system allow user to tag the terms to BOARD\_CODE in BursaLink
2. Existing Business Glossary - BDS attempt to create the same business glossary in scenario 1 again for Board, system will reject due to duplication

### Use Case 4 -Data profiling based on pre-define data quality rules

Show case on how the BDS can define the data quality rules and how the system assesses the data quality based on the rules set and after rectification, populate into the data lake, the statistics to be shown on the DQ dashboard



Figure 16: Use Case 4 – Data Profiling based on pre-define data quality rules

**Figure 17**: Use Case 4a – Details

**Scenario Examples:**

System to profile the data based on the following DQ rules and list out a list of data which does not meet the criteria.



Table 9: DQ Rules to Apply for Respective Attributes

### Use Case 5 - Demonstrate impact to the downstream systems (i.e. data lineage) when there is change to metadata at the source system

Show case on how the system provide visibilities on the impact to the downstream systems when there is changes to metadata of the source system.



Figure 18: Use Case 5 – Demonstrate impact to the downstream systems (i.e. data lineage) when there is change to metadata at the source system.



Figure 19: Use Case 5a – Details

***Note:***

* *Ownership is by division*
* *Sensitivity Level include the following*

*a) Secret – Access not permitted;*

*b) Confidential – Data to be masked; Approval from Data Owner required*

*c) Restricted – Approval from business owner required;*

*d) Public – No restriction.*

**Scenarios Example:**

1. Change the Investor Type (INVESTOR.TY\_INVESTOR) from C3 to G1 / G2 / G3 in the source system (CDS) and observe the change to the lineage and impact to the downstream system.

### Use Case 6 - Merging of master data and creation of golden source master data

Show case how the system can merge master data from multiple data sources and allow the multiple parties to validate and confirm the data and create a golden source



Figure 20: Use Case 6 – Merging of master data and creation of golden source master data



Figure 21: Use Case 6a - Details

**Scenario Example:**

1. Merge the Company Profile, Instrument profile from BursaLink and Security profile in BMGS to create the Master Table for Instrument / Company.



Table 10: Profiles to create Master Table

Table 5 shows the merge of Company Profile, Instrument profile from BursaLink and Security profile in BMGS to create the Master Table for Instrument / Company.

### Use Case 7 – Scan Metadata

Showcase how the system can automatically scan and capture the metadata from the source system and detect any changes to the metadata.



TDS

Figure 22: Use Case 7 – Scan Metadata



Figure 23: Use Case 7a - Details

### Use Case 8 - Centralize access to common statistics

Allow data user to access to commonly access statistics; Aggregated data will be provided. Dashboard will be developed by Bursa Malaysia.



Figure 24: Use Case 8 – Centralize access to common statistics



Figure 25: Use Case 8a - Details

## Data Platform

To facilitate the above POC, connectivity to the following source system and the relevant data in the development environment need to be setup as per the following diagram indicate in Figure 26.

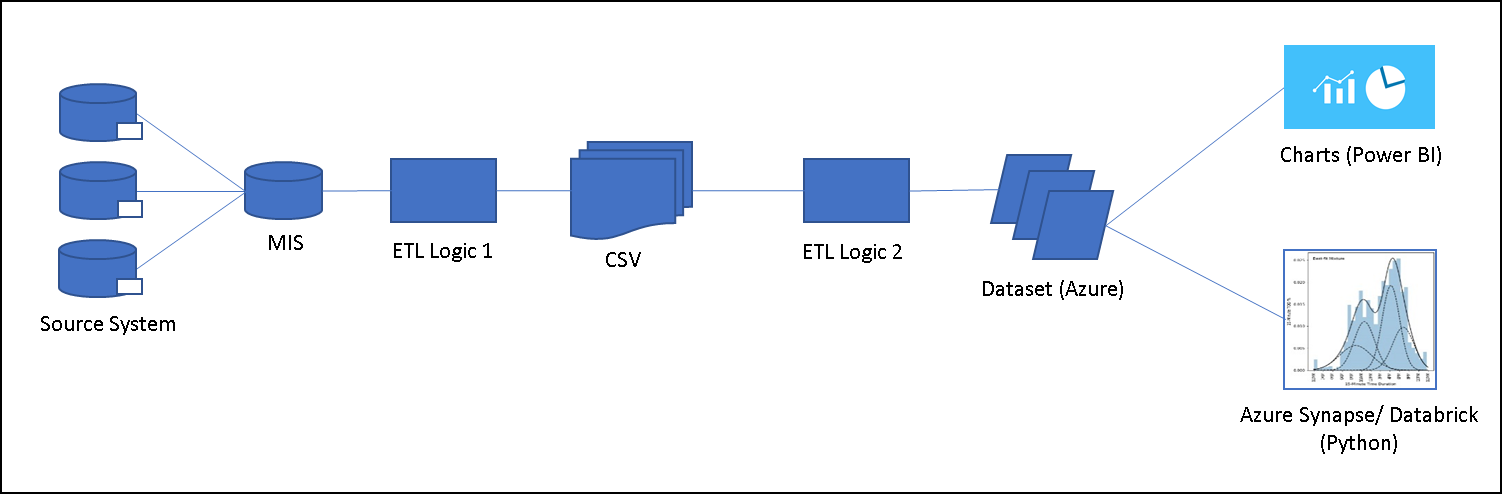


Figure 26: Current Setup for AWS Tactical Solution

The POC use cases will be based on the data ingested from the above source systems. ETL is to be developed by vendor and dashboard on the data platform will be developed by Bursa Malaysia internal Data Analytics team*. Data Platform will be provided by Bursa Malaysia. Vendor is to integrate their tools with Bursa Malaysia existing data platform.*

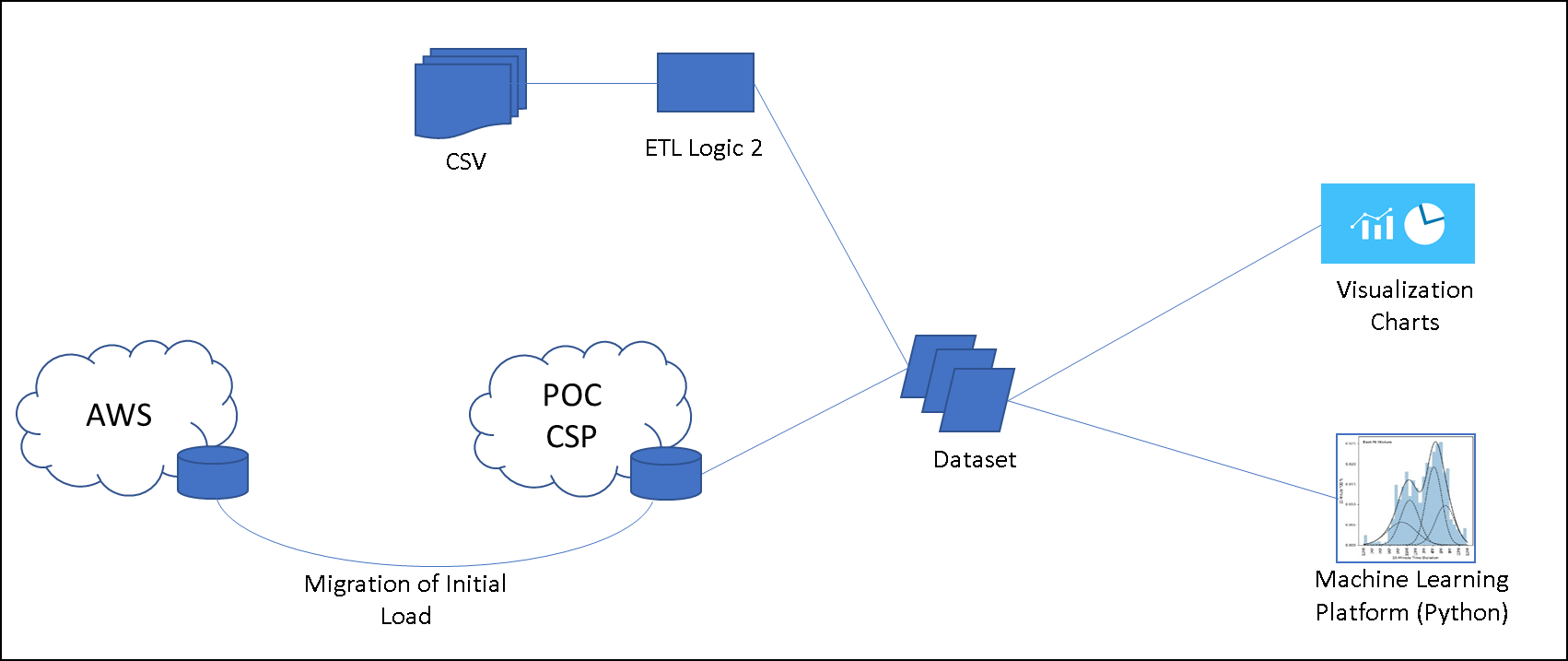
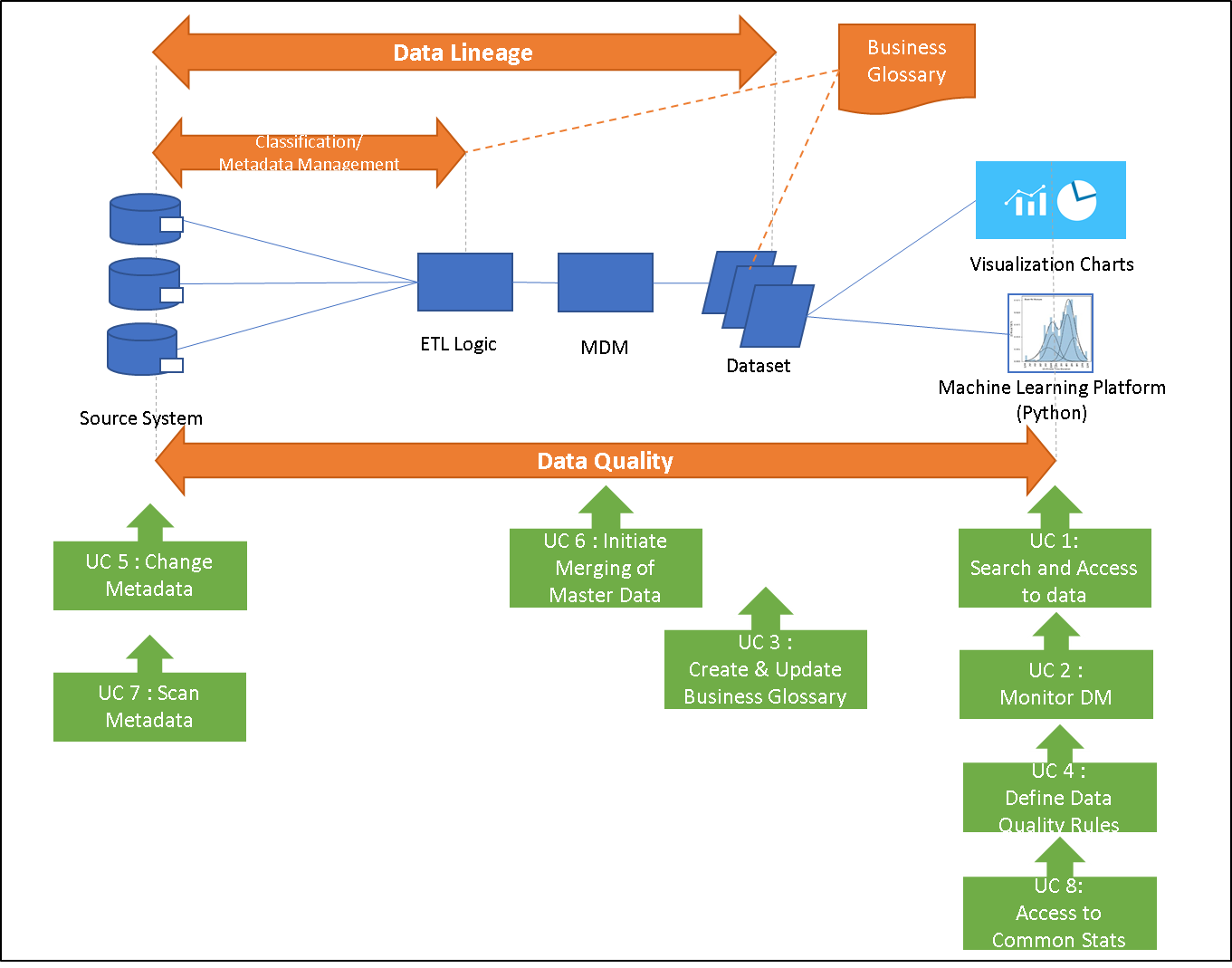


Figure 27: Proposal for Setup for Data Platform

Presently we have data in cloud via AWS tactical solution, although it’s a good option to migrate selected data to POC cloud which helps to optimize ETL process. However due to sensitive data this option can be discussed of further before scrambling data prior exposing for visualization (an option to discuss with participating cloud vendors).

Figure 28: Proposal for Data Lineage for Azure Data Platform

This section describes the configuration of data platform and the usage. There will be two main functionalities tested:

1. Data Anonymization for confidential classified data
2. Dashboard Development

### Data Anonymization

Figure 29: Attributes/Fields to be Anonymized

Figure 29 above indicates confidential attributes/fields which requires to be anonymized. In this case, we are using hashing algorithm to scramble the data and retain the uniqueness. The anonymized data will be used as the primary key to merge between tables to provide a flattened query.

The anonymization will be developed by the Bursa Application team during the transformation process via Extraction, Transformation, Load (ETL).

### Dashboard Development (Use Case 9)

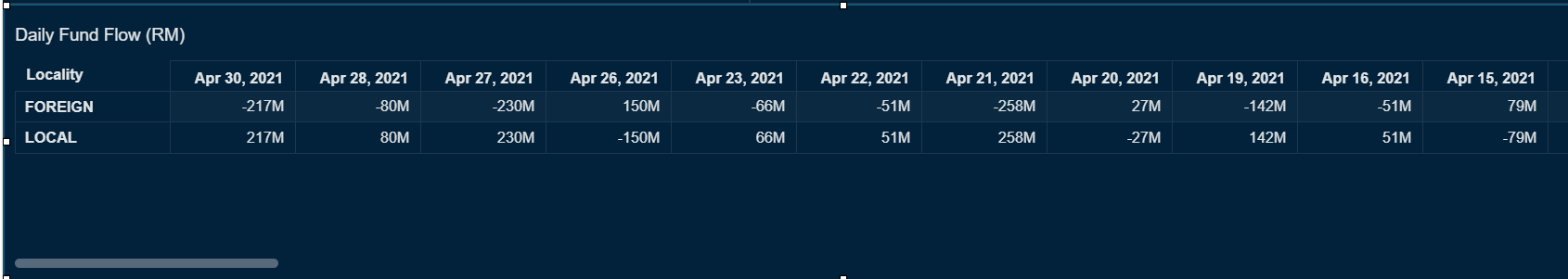
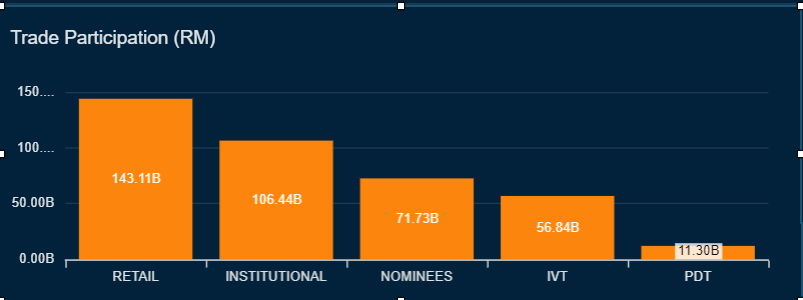


Figure 30: Report 1 (R1)

Figure 31: Report 2 (R2)

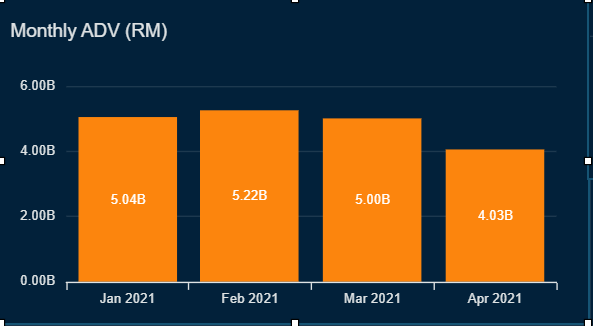


Figure 32: Report 3 (R3)



Figure 33: Report 4 (R4)

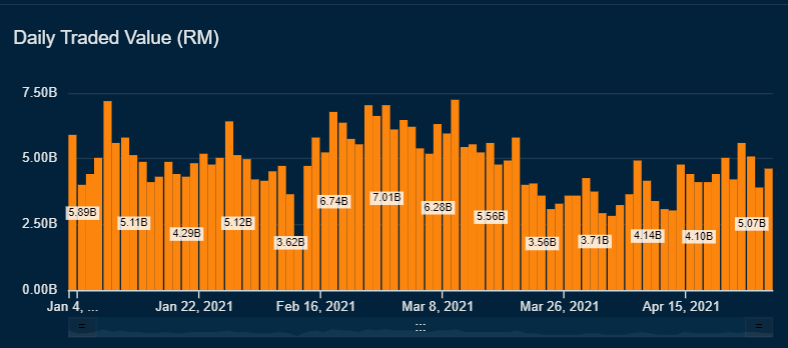


Figure 34: Report 5 (R5)

### Entity Relationship Diagram – BursaLink, BMGS & CDS

Based on Figure 35, nine CSV files mainly BursaLink (PR\_Companies.csv, PR\_Instrument\_Profiles.csv, PR\_Company\_Addresses.csv), BMGS (Security.csv, Secboard.csv, Trade\_Consumption\_Data.csv) and CDS (Account.csv, Investor.csv, Nationality.csv) will be merged using the defined primary keys and foreign keys.



Figure 35: Dataset “BursaLink\_CDS\_BMGS ”

A dataset called “BursaLink\_CDS\_BMGS ” has to be created to build the above dashboard listed in **Figure 30: Report 1 (R1)** to **Figure 34: Report 5 (R5)**.

### Business Transformation Rules

| **#** | **R1** | **R2** | **R3** | **R4** | **R5** | **Computed Fields** | **Source** | **Business Rule / Algorithm** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Y |  | Y |  |  | gross traded value buy | BTS2 (BMGS) | **GrossTradedValueBuy** = [Trade\_Consumption\_Data].[Price] x [Trade\_Consumption\_Data].[Quantity]  Only for applied for  [Trade\_Consumption\_Data].[Buyer\_Account\_Id] |
| 2 | Y |  | Y |  |  | gross traded value sell | BTS2 (BMGS) | **GrossTradedValueSell** = [Trade\_Consumption\_Data].[Price] x [Trade\_Consumption\_Data].[Quantity]  Only for applied for  [Trade\_Consumption\_Data].[Seller\_Account\_Id] |
| 3 | Y |  |  |  |  | locality | CDS | IF [Nationality].[NM\_Nationality] = ‘Malaysia’ THEN **Locality =** ‘Local’  ELSE  **Locality =** ‘Foreign’ |
|  |  |  |  |  |  |  |  |  |
| 4 | Y | Y | Y | Y | Y | trading date | BTS2 (BMGS) | **TradingDate** = [Trade\_Consumption\_Data].[Trade\_Date] |
| 5 | Y | Y | Y | Y | Y | market transaction type | BTS2 (BMGS) | IF [SecBoard].[Board\_ID]= ‘NM’ or ‘OD’ THEN **MarketTransactionType** = ‘OMT’  ELSEIF [SecBoard].[Board\_ID] = ‘DB’ THEN **MarketTransactionType** = ‘DB’  ELSEIF [SecBoard].[Board\_ID]= ‘BI’ THEN **MarketTransactionType** = ‘BI’  ELSE  **MarketTransactionType** = ‘Undefined’ |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| 6 | Y | Y | Y | Y | Y | stock code | BTS2 (BMGS) | **StockCode** = [Trade\_Consumption\_Data].[Stock Code] |
| 7 |  |  | Y |  |  | investor type | CDS | IF [Investor].[TY\_Investor] = ‘C5’OR ‘C6’ OR ‘C7’,OR ‘O’ OR ‘I’ THEN  **InvestorType = ‘**Retail’  ELSEIF [Investor].[TY\_Investor] = ‘C1’ OR ‘C2’ OR ‘C3’ OR ‘C9’ OR ‘T’ THEN **Investor\_Type = ‘**Institution’   ELSEIF [Investor].[TY\_Investor] = ‘C8’ THEN **Investor\_Type = ‘**Nominees’  ELSE **Investor\_Type**  = ‘Undefined’ |
| 8 |  |  | Y |  |  | intraday account type | CDS | IF [Investort].[NM\_Investor] contains ‘%IVT%’ THEN **IntradayAccountType** = ‘IVT’  ELSEIF [Investort].[NM\_Investor] contains ‘%PDT%’ THEN **IntradayAccountType** = ‘PDT’  ELSE **IntradayAccountType** = ‘Undefined’ |
| 9 |  |  |  | Y |  | market cap | BTS2 (BMGS) | **MarketCap** = [Security].[Issued\_Quantity] x [Secboard].[Close\_Price] |
| 10 |  | Y |  | Y | Y | Trade Value | BTS2 (BMGS) | **TradeValue** = [Secboard].[Value\_Today] for each Security each trading day |

Table 11: Rules for Transformation from Source

### Business KPI Measurements – from BDB-DA Team

|  |  |  |
| --- | --- | --- |
| **Object/Dataset Name** | **Computed Fields** | **KPI/Measurements** |
| cds\_trade\_demography | gross traded value buy gross traded value sell locality trading date market transaction type | **Measurement**: **Net Fund Flow**  **Formula**: ( sum[gross traded value buy] – sum[gross traded value sell]),  filter by market transaction type = 'OMT'  **Dimension**: locality, trading date |
|
|
|
| listed stocks and cds\_trade\_demography | gross traded value buy gross traded value sell trading date market transaction type stock code | **Measurement**: **ADV** **Formula**: (sum[gross traded value buy] + sum[gross traded value sell])/2)/distinct count (trading\_date), filter by market transaction type = 'OMT'  **Dimension**: trading\_date by month  **ETL**: a.listed stocks table left join with b.cds trade demography table ON ((a.stock\_code = b.stock\_code) AND (a.trading\_date = b.trading\_date)) |
|
|
|
| cds\_trade\_demography | gross traded value buy gross traded value sell trading date market transaction type investor type intraday account type | **Measurement**: **Traded Value OMT Formula**: (sum(gross traded value buy) + sum(gross traded value sell))/2), filter by market transaction type = 'OMT'  **Dimension**: investor type ? **intraday account type**  Formula for investor type:  if intraday account type = ivt then ivt, if intraday account type = pdt then pdt, else investor type |
|
|
|
| listed stocks and cds\_trade\_demography | gross traded value buy gross traded value sell trading date market transaction type stock code market cap | **Measurement**: **Market Cap** **Formula**: ((sum**(**[gross traded value buy]**)** + sum**(**[gross traded value sell]**)**) /2)/distinct count ([trading\_date])/ [average market cap] \* 248, , filter by market transaction type = 'OMT'  **Dimension**: trading date  **ETL**: a.listed stocks table left join with b.cds trade demography table ON ((a.stock\_code = b.stock\_code) AND (a.trading\_date = b.trading\_date)) |
| cds\_trade\_demography | gross traded value buy gross traded value sell trading date market transaction type | **Measurement**: **Traded Value OMT Formula**: (sum([gross traded value buy]) + sum([gross traded value sell])/2), filter by market transaction type = 'OMT'  **Dimension**: trading date |

Table 12: Transformation for Dashboard

# APPENDICES

## Schema Overview for the Nine CSV Files

Highlights the schema of nine CSV files mainly (refer the respective worksheets): -

1. **BursaLink** (PR\_Companies.csv, PR\_Instrument\_Profiles.csv, PR\_Company\_Addresses.csv),
2. **BMGS** (Security.csv, Secboard.csv, Trade\_Consumption\_Data.csv)
3. **CDS** (Account.csv, Investor.csv, Nationality.csv)

The fields/attributes with confidential classification will be anonymized using hash algorithm.



The hashing algorithm with retain the uniqueness of the key while scrambling the values with alphanumeric characters.

## An Overview of BursaLink Metadata



## An Overview of CDS Metadata



## An Overview of BMGS Metadata

