**Snowflake Migration – Strategy Document**

Version 1.0

Table of Contents

[**1.** **System Requirements** 2](#_Toc43368281)

[**2.** **Fraud Monitoring – POSTGRES** 2](#_Toc43368282)

[**3.** **Short-term (Max: last 24 months)** 3](#_Toc43368283)

[**4.** **History (24 months – 10 years)** 3](#_Toc43368284)

[**5.** **User Defined Tables** 3](#_Toc43368285)

[**6.** **ETL** 4](#_Toc43368286)

[**7.** **Access** 4](#_Toc43368287)

[**8.** **Additional Requirements** 4](#_Toc43368288)

# **System Requirements**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| Stability | Platform should show stable behavior throughout |
| High availability | * Ideally there should not be any production outages * Planned outages needs to be informed to users 3 days prior * Any random production outages should be addressed and fixed in less than 30 mins |
| Updates are made available without impact to production | Any system level updates should be made without impacting production environment |
| JDBC/ODBC connectivity | BI tool connected to Impala for showing dashboards. There should be a facility for connecting various tools to Snowflake (Tableau, Power BI, Looker) |
| Strong Data typing (dates, int, boolean, string) | System should support strong datatypes like dates, int, boolean, string and collection datatypes (array, map, key value pair etc.) |
| Indexed by mtcn, pcp, and transaction date when possible | Faster querying and index needed on few primary fields like MTCN, PCP and transaction date |
| Number of columns | Unlike Kudu there should not be any column limitations (currently Kudu supports only 300 columns in a table) |
| Records update/modify functionality | System should be capable of handling CDC-Change Data Capture (s-packet disposition modified data to be captured properly) |
| Specific data subject for Transaction XML | Make sure that data is available in new system and should be able to query it similarly like SQL |

# **Fraud Monitoring – POSTGRES**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| Table shortlist [FM tables] (6 months) | All the mentioned tables to be included-  <https://confluence.corpprod.awswuintranet.net/display/DATA/MRT-PostgresSQL> |
| Update Frequency – real-time | Updating should be consumed real time (data should be available immediately from OLTP) |
| Performance: < 10 seconds | Results should be fetched within 5-10 secs |
| Summarized customer tables (up to 10 years) | Below two tables which needs to be backed up with 10 years of data:  Devicetransaction, PCPHistory |

# **Short-term (Max: last 24 months)**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| Table shortlist:   * Vendors * Rules * Transaction Data * Centralized Fraud * Centralized Transaction * Matchedtransaction (impala) * Age Variables * Velocity variables | All the tables and fields should be available in Snowflake without any formatting and data issues (List of shortlisted tables and columns to be attached) |
| Update Frequency: Every 6 hours | Realtime |
| Query Performance: < 2-3 min | Results should be fetched in 2-3 minutes |

# **History (24 months – 10 years)**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| Table shortlist | All the tables and fields should be available in Snowflake without any formatting and data issues (List of shortlisted tables and columns to be attached) |
| Retention at a table level | Based on GDPR mandate. (Few are for 6 mnths, 1 yr, 10 yrs) |
| Update Frequency: once a day/week | Historical data should be backed up every week without impacting the production environment |
| Performance: < 30 min -within 2-3 mins | Data should be fetched within 2 mins |
| Define process for data deletion, similar to GDPR | Possibility of adding automated script to automatically delete additional data against the retention period |

# **User Defined Tables**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| User should have the ability to create temporary tables | Users should be able to create temporary tables for their analysis |
| Capability to upgrade view to production environment |  |

# **ETL**

|  |  |
| --- | --- |
| **Indicator** | **Acceptance Criteria** |
| Shortlist of Blaze lookup tables | Murthy/Baba to help on tables(lookup/dimension) which are currently loaded via ETL (FX tables, reported fraud files etc.) |
| Information from other systems (Apollo, UCD, …) | Ex.: Unisys S packet (all the tables loaded from all systems other than Blaze) |

# **Access**

* + Risk users
  + Non-risk users

Table shortlisted should include the following categorization

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Source | Risk Users Only (Y/N) | Retention (in years) |
|  |  |  |  |

# **Additional Requirements**

|  |  |
| --- | --- |
| **Problem** | **Solution/Requirement** |
| Data storage | Store only last event / Retention policy – CDC (ex.:- event ordering) |
| Migration checks are burden to business | Automatically verify migrated tables on engineering side |
| New tables pop-up out regularly | Governance process on table creation (clear owners, retention for every table) |
| Tables have same counts, metrics are different | Business driven audit (SQL v/s BDP v/s Snowflake should match) |
| Stable platform for history and short-term | Available (Same platform for both historical data and short-term data) |
| Storage shortages | Capacity Planning – storage (Platform architects to frequently gauge the storage/performance capabilities and address the shortcomings for next 3-5 years) |
| Automated queries not user friendly | Scheduled queries/jobs |
|  | Clean data marts (aggregate tables for easy business querying) |
|  | Analysis workbench (is it available?) (DS team might need R, Python or ML analysis engines on top of data lake) |
|  | Upload/Download user should be to upload/download their files and analysis |
| Different UAT and Production environments | There should be different environments –  UAT – For user testing  Production |