Data Governance

Purpose and Scope

The purpose of this governance is to ensure that the data from each table transform correctly into OLAP model. The data itself is accurate and has no duplication on it where answer some question such as:

- The platform over dependent on a few tokens or not
- Trading volumes genuinely growing or inflated by some customer
- Do users start from p2p transfer eventually trade or just churning
- Retention differs based on region and token category
- Anomalies such as duplicate trades or suspiciously high value trades
- Need quality check for the data

The scope of the data will be:

- Source:
    - raw_kyc.raw_users
    - raw_tokens.raw_tokens
    - raw_transactions.raw_p2p_transfers
    - raw_transactions.raw_trades
- Target:
    - analytics.user_dim
    - analytics.token_dim
    - analytics.date_dim
    - analytics.time_dim
    - analytics.p2p_transfer_fact
    - analytics.trades_fact
- Process:
    - Moving the raw source into staging schema while changing the data type since every source data is STRING
    - On staging schema, the data will be queried and inserted into analytics schema
    - The deduplication data on fact table will be transformed using CTE and will be inserted into analytics schema.

Roles and Responsibilities

- Data Governance:
    - Oversee the entire process of data governance policies and procedures
    - Facilitating communication and approvals between Product Owner, Data Analyst, and Data Engineer

    - Monitor compliance and reports on data quality metrics
    - Manage documentation including data dictionary, data lineage, and Star Schema
- Product Owner:
    - Accountable for every business process for P2P Transfer and Trade
    - Approves the business rule, duplicate checking, and data quality
- Data Engineer:
    - Responsible for design, implement, and maintain ELT that transform data from OLTP into OLAP model
    - Make sure the deduplication, data transformation, and data quality is running on pipeline
- Data Analyst:
    - Collaborate with Product Owner to define the business rules for data transformation, unique, and quality
    - Investigate the data quality issues if its discovered
    - Develop report and analysis based on the designed OLAP

Data standards and definitions

Business Glossary

- Buy : Activity where user buy the token
- Sell : Activity where user sell the token
- Filled : Status for user where the token successfully bought or sold
- Failed : Status for user where the token failed to be bough or sold

Quality threshold

All the table should be not null except the first_transfer_date or first_trade_date from *user_dim*

Naming Conventions

- Every dimension table should be end with *_dim*
- Every fact table should be end with *_fact*
- All the table and column using *snake_case* naming conventions. Where the naming should always in lower case and the space should be changed to "_"

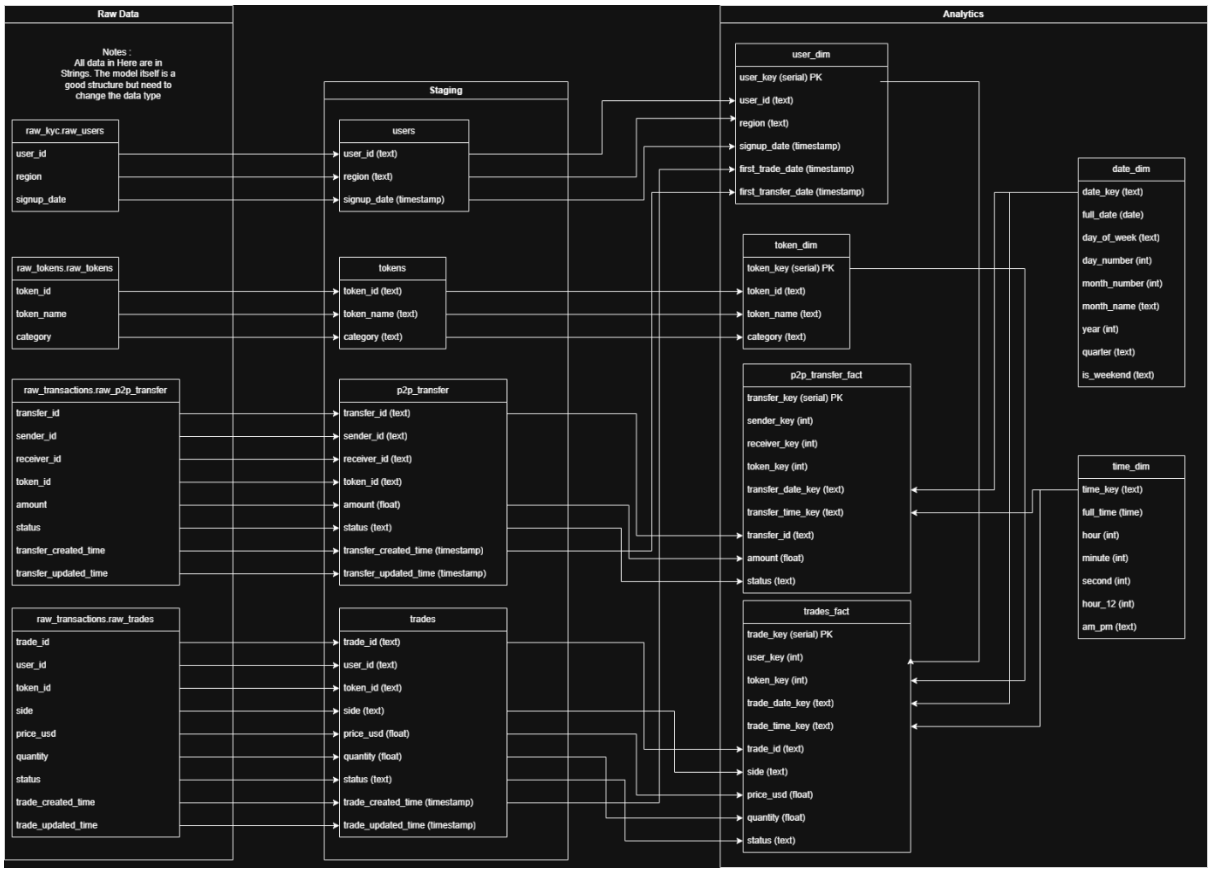
Classification tables

- analytics.user_dim (Restricted)
- analytics.token_dim (Internal)
- analytics.date_dim (Public)
- analytics.time_dim (Public)
- analytics.p2p_transfer_fact (Restricted)

- analytics.trades_fact (Restricted)

Master data hierarchies

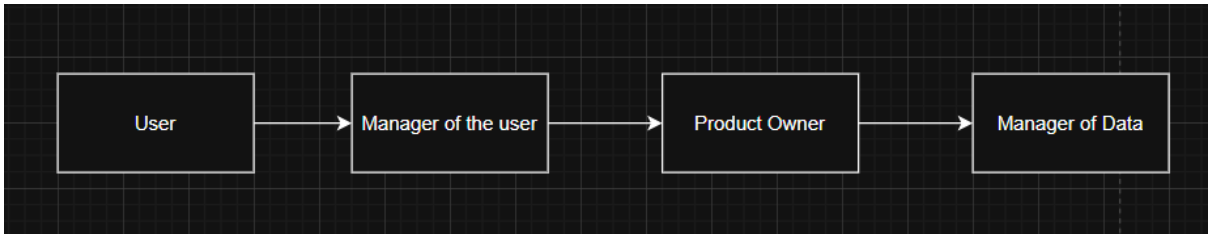
Please check the data lineage below.



Procedures and Workflows

Data Access Request

Data access request should be come with this flow. With the 48 hour SLA for the request to be approved with every stakeholder.



Quality Issue Escalation

When the data break, please inform the Product Owner. The product Owner will check it with teams. Please do issue the detailed specification of the error with the screenshots of it if possible. Here are the procedures:

- A data quality issue discover by the user

- The user will report to the respective product owner for the data quality issue including the impacted column, the problem, and the example of the issue.
- Data analyst will investigate the issue and gather more evidence.
- If the error comes from ETL/ELT pipeline, Data analyst will escalate it to Data Engineer. If the error comes from business rule, Data analyst will escalate it to Product Owner.
- The relevant parties will check the root cause together and implement the fix.
- Data Governance will ensure the stakeholders about the issue, the impact, and the resolution progress.
- If it is finished, then data analyst will close the ticket

New Dataset Onboarding

If there is new dataset onboarding, please set the meeting with the Product Owner and Data Analyst. Here are the steps:

- Product Owner and Data Analyst identify the need of new dataset. Aligning the proposal for the requirements
- Data Governance and Data Engineer will overview the proposal and check the availability with the current existing data
- If it is indeed needed, the Product Owner will welcome the onboarding
- Data Engineer will develop the ETL / ELT and documented the result
- Data Governance also will add the data into data glossary
- Data Analyst will check if the data indeed loaded into the environment
- If all the steps fulfil, the data will be released for analytics purposes

Change Management

If there is a need to modification the existing dataset. Please do follow this procedure:

- Users identify the need for changes and submit the change request for the proposed modification and potential impact.
- Data governance will review the change request and potential impact.
- Data governance will set a meeting with all impacted stakeholders and receive a feedback and concern for them.
- Based on the review and concern, data governance and product owner will approve the change.
- Data engineer will implement the change in ETL / ELT.
- Data engineer will test the pipeline to ensure the functionality didn’t break and the changes work as expected.
- Data engineer will make a plan for deployment, contacting every stakeholder for the potential downtime and the changes.
- Data engineer will deploy the changes.
- Data engineer and data analyst will check the changes work in production or not.

Exception Handling

The exception handling happens if there is an emergency case such as urgent hotfix. Should be for critical cases. Here are the procedures for it:

- Team members find something that needs an exception to a standard procedure
- The requestor documented the situation and proposed for urgent fix
- Product Owner and Data Governance will review the situation
- If the proposed deviation from standard procedure approved. Data governance will document the agreement and the duration of the exception.
- Data Engineer will implement the exception.
- Data Governance will document all the exceptions, justifications, approval, and implemented actions.
- After the workaround finished, all stakeholders will review for the better work around that goes by the standard solution.

Compliance and Enforcement

This section details how Data Governance policies are monitored, enforced, and measured, focusing on enabling compliance through automation and integration.

- Automated Monitoring:
    - ETL Checks: Real-time logging and alerting for ETL errors, data quality failures (NULLs, invalid formats, failed lookups), duplicate rejections, and volume anomalies.
    - Database Constraints: Use of UNIQUE and FOREIGN KEY constraints in the OLAP database to automatically prevent duplicates and referential integrity errors during loading.
- Audits:
    - ETL Logs: Monthly review of error/duplicate logs by Data Governance to spot trends.
    - Access Permissions: Quarterly review of user access to the OLAP warehouse by Data Governance.
    - Metrics Review: Monthly review of Data Quality Metrics by stakeholders.
- Training & Consequences
    - Training: Mandatory, role-specific initial training and annual refreshers on policies and procedures for Product Owner, Data Governance, Data Engineer, and Data Analyst roles.
    - Consequences (Progressive):

        - Automated: Data failing checks is quarantined/rejected with alerts.
        - Notification: Violations lead to notification of the individual and manager for correction.
        - Remediation: Persistent issues may require retraining or process fixes.
        - Access Review: Deliberate/negligent violations may lead to access review/revocation.

Measuring Success

- Key Metrics:
    - Reduction in quarantined records.
    - ETL job success rate.
    - Data Access Request SLA adherence.
    - Time-to-resolution for data quality issues.
- Enabling Approach
    - Prioritize Automation: Embed policy checks (quality, uniqueness) directly into ETL and database structures.
    - Integrate Workflows: Build checks into CI/CD pipelines and standard development processes.
    - Provide Support: Offer clear documentation, templates, and guidance to make compliance easy.
    - Use Feedback: Employ metrics for continuous improvement, not just punishment.