



## TAKE HOME TEST

### **Business Issue**

Your company, **CoinX**, is a fast-growing cryptocurrency exchange in Southeast Asia offering spot trading, peer-to-peer (P2P) transfers, and fiat on/off ramps.

Recently, the leadership team raised several concerns:

1. Trading concentration risk:
  - a. Is the platform overly dependent on a few tokens?
  - b. Are trading volumes genuinely growing or just inflated by a small set of traders?
2. User retention & cross-product usage:
  - a. Do users who start with P2P transfers eventually trade, or churn?
  - b. How does retention differ by region and token category?
3. Data reliability & compliance:
  - a. There are anomalies (e.g., duplicate trades, suspiciously high-value transfers).
  - b. Management needs trusted dashboards but is not sure the current data enforces enough quality checks.

### **Instructions**

Your tasks are to:

1. Create a proper and scalable data modeling design that can answer the leadership team questions.
2. Create a data governance plan to ensure data quality and reliability.
3. Explain and demonstrate how the insight will be presented to the management.
4. Please be mindful that:
  - a. All data types are in **STRING** format.
  - b. All timestamp data are in UTC.
  - c. You need to explicitly describe any definition that you come up with.

### **Expected Deliverables**

1. Please submit a public repository that contains:
  - a. SQL or any transformation query for all data models proposed.
  - b. Data lineage and ERD of the solution.
  - c. Data governance
2. (Optional) Report or dashboard layout and description for how it could answer the questions.

## **Data**

### raw\_trades

- source schema:: raw\_transaction
- PK: trade\_id
- FK: user\_id, token\_id

### raw\_p2p\_transfers

- source schema: raw\_transaction
- PK: transfer\_id
- FK: sender\_id, receiver\_id, token\_id

### raw\_users

- source schema: raw\_kyc
- PK: user\_id
- FK: -

### raw\_tokens

- source schema: raw\_config
- PK: token\_id
- FK: -