Data Warehousing – Log Analytics with Metadata Enrichment

📂 Project Files

1. **`logs.py`**
    - **Purpose**: First step in the pipeline.
    - Downloads and extracts raw log data.
    - Creates `logs_*` tables in SQLite (`logs.db`).
    - Populates these tables with raw simulation logs.
2. **`metadata_init.py`**
    - Initializes **metadata dimension tables** (chip family, design block, team, test condition, etc.).
    - Prepares mapping rules for enrichment.
    - Runs after `logs.py` to ensure raw log data is available.
3. **`metadata_enrich.py`**
    - Reads from both log tables (`logs_*`) and metadata tables.
    - Enriches log data with attributes:
        - Chip family
        - Design block
        - Team
        - Test condition

        - Impact score
    - Produces **enriched log tables** that are structured and ready for analysis.
4. `dash.py` **(Dashboard)**
    - Streamlit-based analytics dashboard.
    - Key features:
        - **Filters** (chip, block, team, impact score, test condition).
        - **Metrics/KPIs**: total logs, unique chips, impacted blocks, impacted teams.
        - **Visuals**:
            - Error distribution by block, team, score.
            - Error trends over time.
            - Heatmap (team vs block).
            - Categorization trend (failure mode).
            - Anomaly detection with rolling averages.
        - **Raw log viewer + CSV export**.

🚀 How to Run

Step 1 – Ingest Logs

python logs.py

- Downloads & extracts raw logs.
- Creates `logs_*` tables in SQLite DB.

Step 2 – Initialize Metadata

python metadata_init.py

- Creates metadata mapping tables (chip, block, team, test condition).

Step 3 – Enrich Logs

python metadata_enrich.py

- Joins raw logs with metadata.
- Produces **enriched log tables** with searchable attributes.

Step 4 – Launch Dashboard

python3 -m streamlit run dash.py

- Opens interactive dashboard in browser.
- Apply filters to explore enriched logs and visual insights.

🔑 Key Outcomes

- **Centralized SQLite log warehouse** with enriched metadata.
- Engineers can quickly **filter by chip, block, team, or condition**.
- Dashboard provides **fast insights**:
    - Error trends & anomalies
    - Workload distribution across teams
    - Pareto & correlation analyses
    - Heatmaps for problem localization