

Key Technologies Used:	2
Why They Are Used:.....	2
Demo Screenshots	2

Design Document

Key Technologies Used:

1. **Streamlit**: A framework to build interactive web applications easily, allowing users to upload data, input parameters, and view results without needing complex front-end code.
2. **Pandas**: A powerful library for data manipulation, used to load, clean, and process the data (both historical and real-time) in tabular format.
3. **Machine Learning (Isolation Forest)**: Anomaly detection model used to identify outliers in the data, helping flag unusual patterns or deviations from the norm.
4. **Python's OS Library**: Used to handle file operations like creating directories and saving output files.

Why They Are Used:

- **Streamlit** simplifies creating web apps and allows easy interaction with data.
- **Pandas** is essential for efficient data handling and analysis.
- **Isolation Forest** detects anomalies in data automatically.
- **OS** ensures the proper handling of files during processing.

These technologies together enable the creation of an easy-to-use web app for data reconciliation and anomaly detection.

Demo Screenshots

Step 1. Once we have executed `streamlit run app.py` command, below ui will pop-up. For full steps on how to run, please refer `readme.md` file inside code folder in repo.

UI URL – <http://localhost:8501>

The screenshot shows the application interface at localhost:8501. The sidebar on the left contains the following sections:

- Upload Data**
- Upload Historical Data (CSV)**: A drag-and-drop area with a limit of 200MB per file. A 'Browse files' button is present.
- Upload Real-time Data (CSV)**: A drag-and-drop area with a limit of 200MB per file. A 'Browse files' button is present.
- Enter Composite Keys (comma-separated)**: An empty text input field.
- Enter Value Keys (comma-separated)**: An empty text input field.
- Enter Date/Time Field Name**: An empty text input field.

The main area on the right displays the title 'Data Reconciliation & Anomaly Detection' with a colorful icon.

Step2: We can upload historical_data.csv and real_time_data.csv file, and enter composite keys, value and date/time field name as highlighted below. Click on Process data to get anomaly result with anomaly reason.

The screenshot shows the application interface after processing. The sidebar on the left shows the upload of historical_data.csv (501.0B) and real_time_data.csv (195.0B). The main area on the right displays the title 'Data Reconciliation & Anomaly Detection' and a 'Processing Complete!' message. Below the message is a 'Download Reconciled Data' button and a table of reconciled data.

	id	category	timestamp	price	status	Status	Anomaly Reason
0	101	A	2024-03-04 15:00:00	130	Active	Anomaly	Deviation of 23.81% from baseline.
1	102	B	2024-03-04 16:30:00	250	Inactive	Anomaly	Deviation of 19.05% from baseline.
2	103	A	2024-03-04 09:00:00	140	Inactive	Anomaly	Deviation of 3.45% from baseline.
3	104	C	2024-03-04 11:00:00	315	Active	Match	

