Tools and Frameworks:

- Databricks
- Azure Data Lake
- Azure Monitor
- Spark
- PostgreSQL

Scalable solution

- **Autoloader**:
    - Continuous file ingestion: Reads files from Azure Blob Storage as soon as they are added.
    - Schema inference and evolution: Automatically adapts to new or modified schema in incoming files.

- **Databricks Workflow**:
    - Schedule for continuous execution
    - The workflow triggers the Databricks Job that processes new files using Autoloader.

- **Job Cluster:**
    - Databricks Job Clusters are created to run the job. These clusters can be auto scaled to handle varying data loads.

- **Data Validation & Transformation**
    - Data cleaning: Ensures that missing or corrupted data is handled appropriately (e.g., quarantining invalid data).
    - Transformation: Includes any necessary adjustments such as normalizing units, converting timestamp formats, and handling missing values.

- **PostgreSQL Database**
    - Stores both raw data and aggregated data.

- **Error Logging and Monitoring**
    - Logging: Logs every step of the pipeline (ingestion, validation, transformation, aggregation, and storage).
    - Monitoring: Use Databricks' built-in monitoring or Azure Monitor to track job status, performance metrics, and errors.

- **Fault Tolerance and Retry**
    - Retry mechanism for failed data processing tasks.
    - Alerting system to notify about pipeline failures (via Databricks Jobs or Azure Monitor).