**Problem Statement**

Retail companies struggle with integrating **in-store POS transactions, online orders, and inventory systems** into a single analytics-ready platform.  
They need:

* **Batch processing** for daily sales reconciliation.
* **Real-time ingestion** for fraud detection & stock alerts.
* **Cloud-native scalability** to support peak season sales.
* **Unified warehouse (Snowflake / Synapse)** for advanced reporting

**Skill Towers Developed**

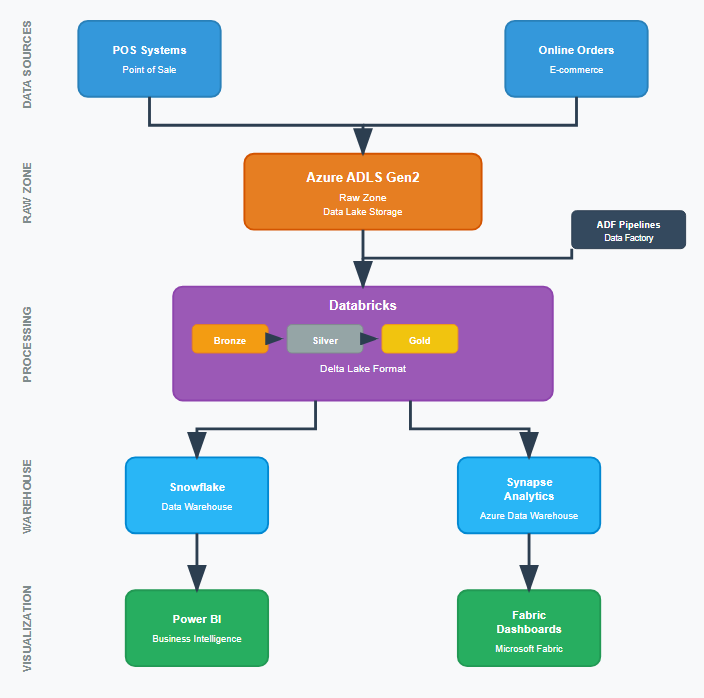
* **Data Engineering Fundamentals**: Batch & streaming design.
* **Azure Storage & ADF**: Ingestion pipelines for orders, customers, products.
* **Databricks & Delta Lake**: Cleansing, transformation, schema evolution.
* **Snowflake & Synapse**: Data warehousing & BI reporting.
* **Monitoring & Optimization**: Cost management, troubleshooting, scaling.
* **Advanced Topics**: Snowpark for ML-driven demand forecasting, Synapse for interactive dashboards.

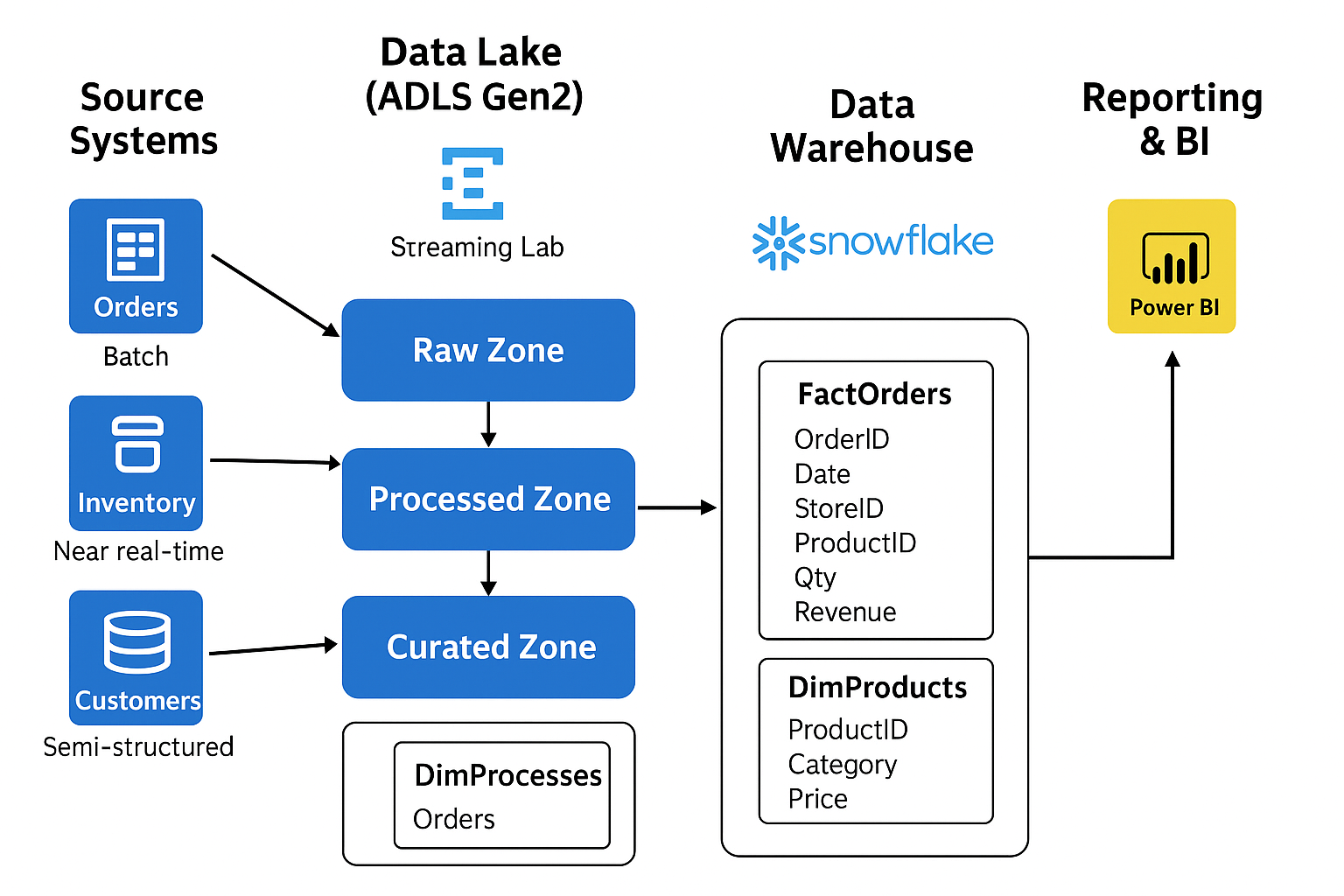
**Use Case / Architecture Diagram**

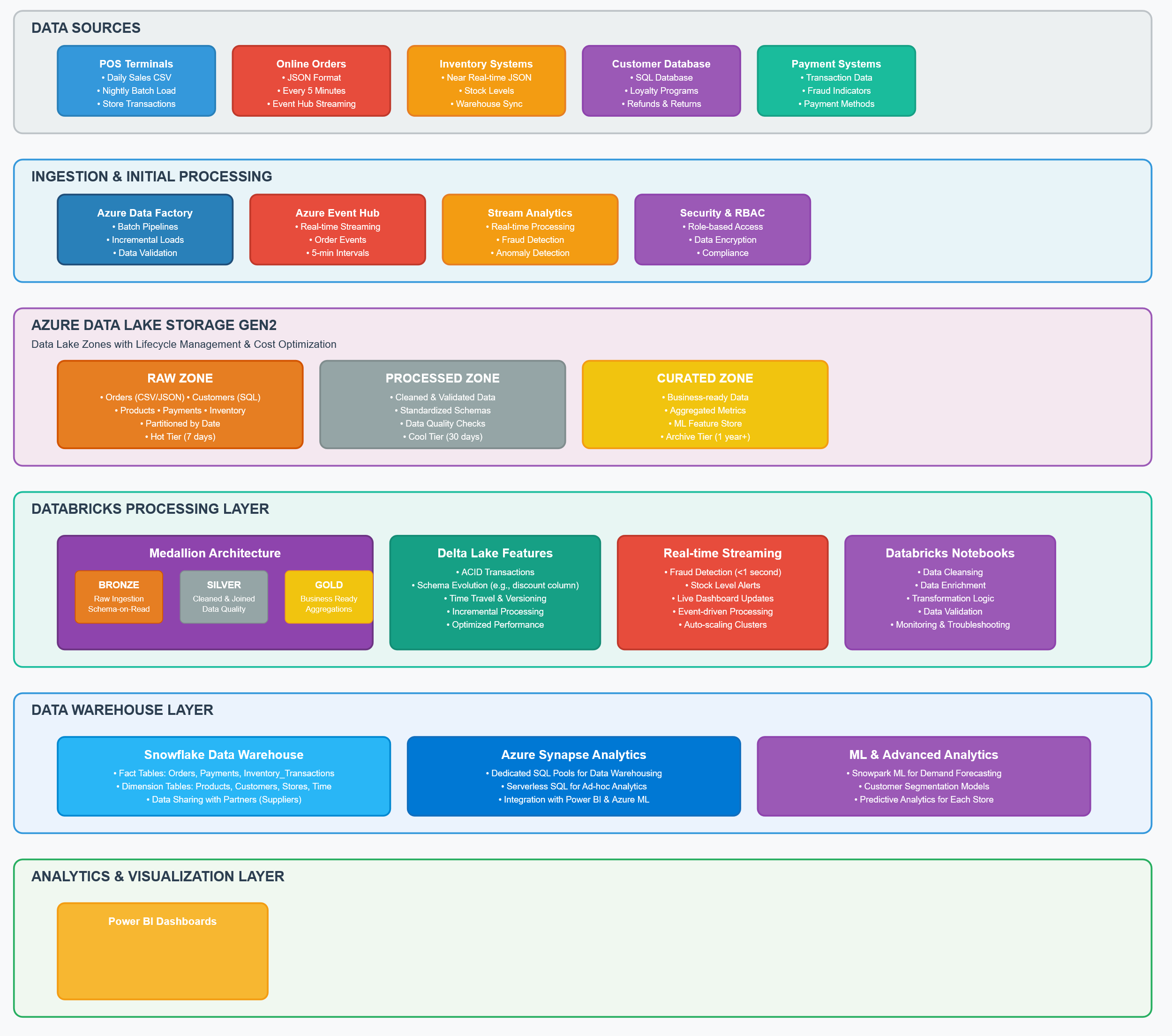
**Use Cases (Major for POS):**

1. **Order Ingestion** – Capture transactions from POS terminals (batch + streaming).
2. **Inventory Updates** – Sync stock in near-real-time across stores & warehouses.
3. **Customer Management** – Loyalty programs, refunds, personalization.
4. **Fraud Detection** – Flag abnormal orders instantly.
5. **Reporting & Analytics** – Sales by region, product, store, promotions.

**Architecture:**







**User Stories**

* As a **store manager**, I want a **daily sales dashboard** so I can track revenue across stores.
* As an **inventory manager**, I want **stock alerts** when product levels are low.
* As a **fraud analyst**, I want to detect **unusually high transactions in real-time**.
* As a **data engineer**, I want a **scalable pipeline** that can process both batch & streaming data.
* As a **data scientist**, I want to use **Snowpark ML** to predict demand for each store.

**Expected Deliverables**

* Data Lake zones: Raw / Processed / Curated for **orders, customers, products, payments**.
* ADF pipelines for batch ingestion of POS data.
* Databricks notebooks for cleansing, enrichment, and Delta Lake transformations.
* Incremental data ingestion (daily orders).
* Snowflake warehouse with **fact (Orders)** and **dimension tables (Products, Customers, Stores)**.
* Power BI / Synapse dashboards showing:
  + Daily/Monthly Sales
  + Top-selling products
  + Inventory shortages
  + Suspicious transactions
* Documentation: Architecture diagram + Troubleshooting Playbook.

**Milestones & Duration**

| **Milestone** | **Duration** | **Details** |
| --- | --- | --- |
| **M1 – Environment Setup** | 1 hr | Setup Azure resources, ADLS, Resource Groups |
| **M2 – Storage & Ingestion** | 1 hr | Create ADLS zones, ingest POS data (CSV, JSON) with ADF |
| **M3 – Transformation in Databricks** | 1 hr | Bronze → Silver → Gold Delta pipelines |
| **M4 – Warehouse (Snowflake/Synapse)** | 1 hr | Load curated POS data, create fact/dimension tables |
| **M5 – Advanced Features** | 1 hr | Incremental loads, real-time fraud detection, schema evolution |
| **M6 – Reporting & BI** | 1 hr | Power BI dashboards & Synapse integration |
| **M7 – Capstone Delivery** | 1 hr | Final demo + documentation submission |

**Implementation Notes**

* **Data Sources:**
  + Orders (CSV, daily batch)
  + Inventory (JSON, near real-time)
  + Customer details (SQL DB)
* **Batch Example:** Nightly load of all store sales.
* **Streaming Example:** Capture new orders every 5 minutes via Event Hub + Databricks streaming.
* **Delta Lake:** Use for **ACID transactions** & schema evolution (e.g., adding discount column).
* **Snowflake:** Use for warehouse & sharing data with partners (e.g., suppliers).
* **Synapse / Power BI:** Use for interactive dashboards.

**Evaluation Rubrics**

| **Criteria** | **Weightage** |
| --- | --- |
| Correct **Azure setup** (RG, ADLS, RBAC, Lifecycle) | 10% |
| ADF pipelines for batch & incremental ingestion | 20% |
| Databricks transformations & Delta Lake implementation | 20% |
| Warehouse schema design (Snowflake/Synapse) | 20% |
| BI Dashboard (sales, inventory, fraud detection) | 15% |
| Documentation (architecture, troubleshooting) | 10% |
| Bonus: Real-time fraud detection / ML demand prediction | 5% |