

## Case Study

### For a Cosmetics Retail Company on ETL and Reporting Implementation:

#### Problem Statement

The cosmetics retail company faced the following challenges, limiting their ability to make data-driven decisions:

1. **Lack of Proper ETL Mechanism:**

- Data was siloed across multiple systems (SQL Server, Oracle, SAP HANA, and SharePoint Excel), making it difficult to consolidate and analyze.
- Manual processes for data extraction and transformation led to delays and errors.
- Inconsistent data formats (e.g., varying product codes, date formats) made integration tedious.
- Increasing data volumes (~3 million rows daily) overwhelmed existing infrastructure.

2. **Lack of Centralized Reporting:**

- Insights were delayed due to fragmented reporting tools and spreadsheets managed independently by teams.
- Inability to get a unified view of sales, inventory, and campaign performance.
- Existing reporting tools could not handle large data volumes or provide real-time insights.
- Managers lacked self-service reporting capabilities, leading to dependency on IT teams.

---

## Proposed Solution

To address these challenges, the company implemented a modern data architecture using **Azure Data Factory (ADF)** for ETL, **Azure Data Lake Storage (ADLS)** as a centralized data repository, and **Power BI** for unified reporting and insights.

---

## Scenario Details

#### Data Sources

1. **SQL Server:** Transactional data from POS systems (1.5M rows/day).
  2. **Oracle:** Product catalog and supplier details (500K rows/day).
  3. **SAP HANA:** Inventory and logistics data (800K rows/day).
  4. **SharePoint (Excel):** Regional sales forecasts and campaigns (200K rows/day).
-

## Solution Architecture



### 1. Azure Data Factory (ADF) as ETL Tool

- Automates the extraction, transformation, and loading (ETL) of data from all sources.
- Supports incremental data loads and ensures data consistency across systems.
- Pipelines are scheduled to run every 4 hours for near-real-time data updates.

### 2. Azure Data Lake Storage (ADLS) as Centralized Repository

- Serves as a scalable and secure data lake for raw, processed, and aggregated data.

Data is structured in a hierarchical format:

```
/raw/           # Raw data from all sources
/processed/     # Cleansed and transformed data
/aggregated/    # Data optimized for reporting
```

### 3. Power BI for Reporting

- Connects directly to the **aggregated** layer of ADLS for analytics.
- Dashboards provide unified views of sales, inventory, and campaign metrics.
- Scheduled refresh every 4 hours ensures up-to-date insights.

---

## Reporting Capabilities

#### 1. Sales Dashboard:

- Track sales performance by region, store, and product category.
- Highlight best-selling products and regions contributing to revenue growth.

#### 2. Inventory Dashboard:

- Monitor stock levels and identify potential stockouts.
- Optimize inventory turnover and restocking schedules.

#### 3. Marketing Campaign Dashboard:

- Measure campaign effectiveness through ROI, sales uplift, and regional performance.

- Compare forecasted vs. actual sales for better future planning.

---

## Outcomes and Benefits

### 1. Improved Data Integration and Transformation:

- Automated ETL pipelines ensure timely and accurate data consolidation.
- Data standardization reduces integration challenges and improves reporting accuracy.

### 2. Real-Time Insights:

- Near-real-time data refresh enables quick decision-making.
- Managers and executives can access self-service dashboards in Power BI.

### 3. Scalability and Performance:

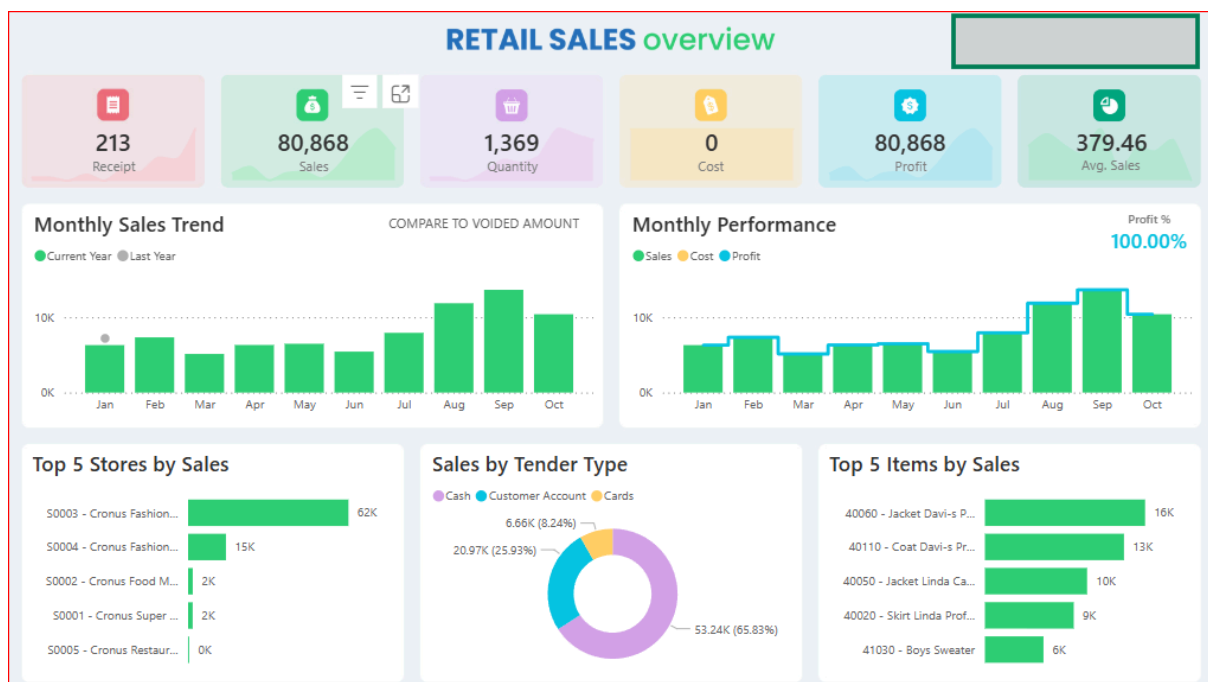
- ADLS and ADF handle growing data volumes (projected 20% annual growth).
- Reporting latency reduced to under 5 seconds for 90% of Power BI queries.

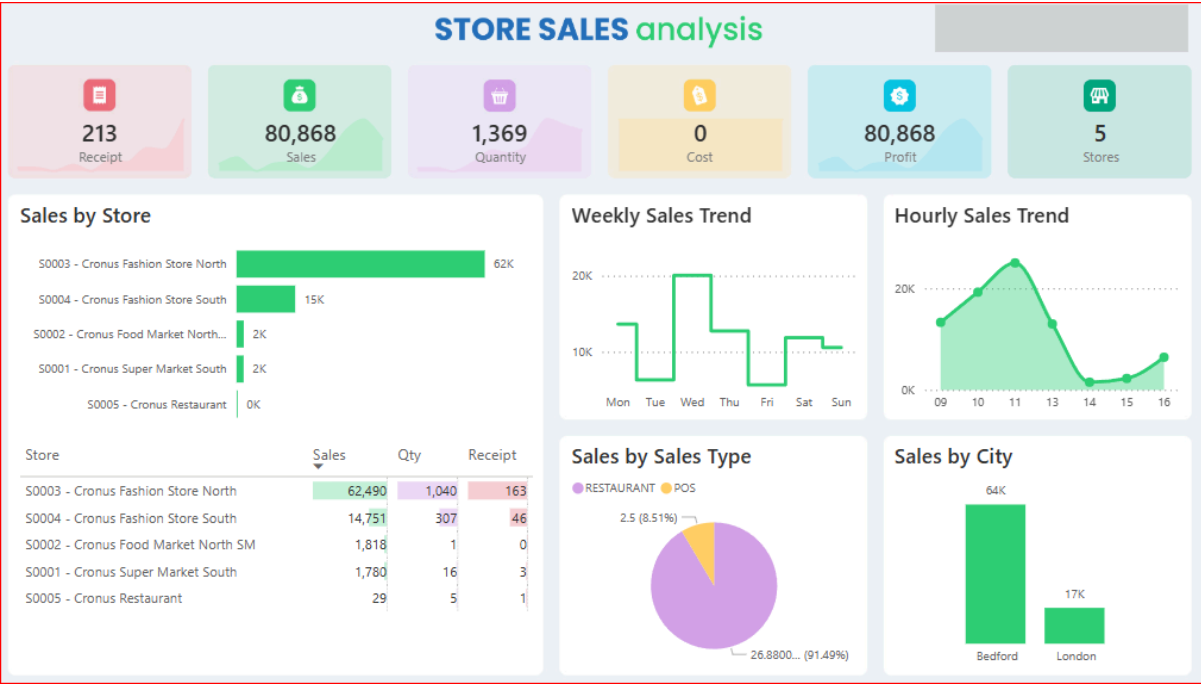
### 4. Enhanced Decision-Making:

- Unified reporting empowers teams to analyze sales trends, manage inventory, and evaluate campaign ROI effectively.
  - Insights drive better customer engagement and operational efficiency.
- 

## KPIs Post Implementation

1. **Sales Growth:** Monthly revenue growth improved by 10%.
2. **Inventory Efficiency:** Stockouts reduced to less than 1% of SKUs.
3. **Campaign Effectiveness:** Achieved a 5x ROI on marketing campaigns.
4. **ETL Performance:** Pipelines complete within 30 minutes for daily batches.





By leveraging Azure Data Factory, Data Lake Storage, and Power BI, the company transformed its data landscape, resolving its ETL and reporting challenges while enabling agile decision-making.