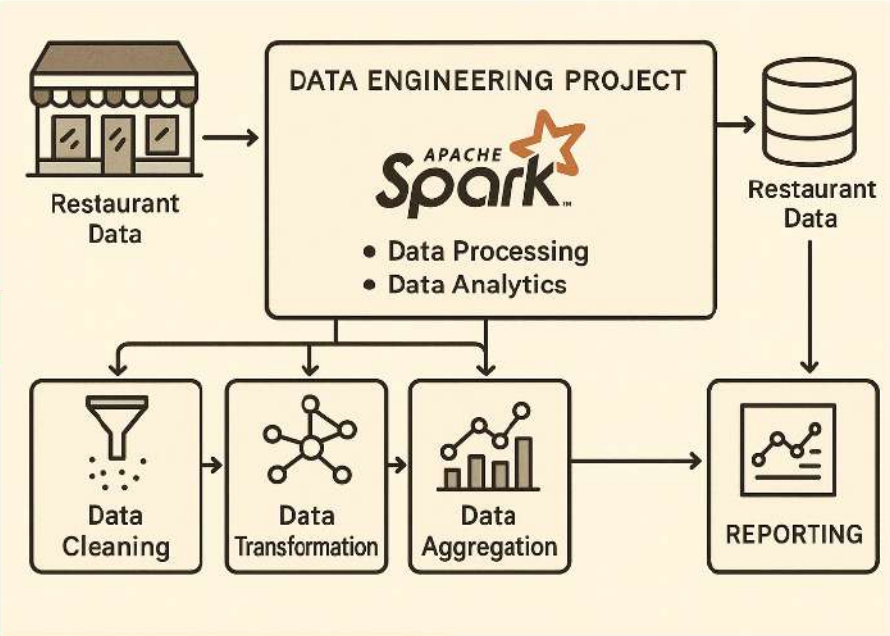
Comprehensive Data-Driven Analytics for Restaurant Operations

From Data Ingestion and Enrichment to PowerBI Reporting and Operational Insights



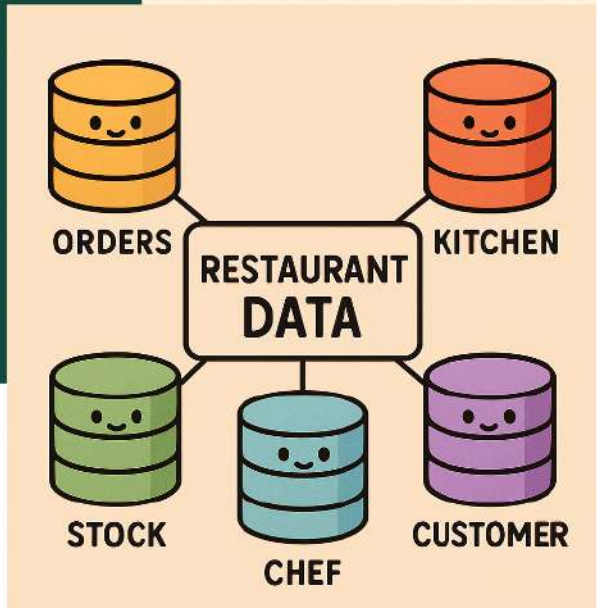


Table of Contents

Table of Contents

Introduction to the Restaurant Project

Project Objective

Delivering data-driven operational analytics and reporting for restaurant outlets.

Key Focus Areas

Enhancing visibility into sales, kitchen efficiency, inventory management, and customer behavior.

Technological Approach

Implementing integrated data processing pipelines for efficient data handling and analysis.

Scope of Operations

Includes data generation, ingestion, enrichment, and visualization to support business intelligence decisions.

Data Architecture and Domain Separation

Data Architecture and Domain Separation

The system uses a four-layer data lake architecture tailored for restaurant operations:

- Raw Layer: Stores ingested operational data (CSV, JSON, JDBC)
- Input Layer: Holds validated GeoJSON files
- Enrich Layer: Contains Delta Lake tables with enriched metrics
- System Layer: Hosts analytical results and aggregates

Strict domain separation ensures the restaurant data processing remains isolated from other business domains, maintaining focus and integrity.

Core Infrastructure Components

DataLakeIO Capabilities

Manages path resolution across layers and environments.

DataLoader & DataWriter Functionalities

Handle multi-format reading/writing, supporting JDBC, Delta, and JSON formats.

Spark Sessions Usage

Configured for specialized JDBC integration as well as large-memory transformations.

DeltaLakeOps Table Maintenance

Provides operations such as vacuuming, optimizing, and restoring tables.

Orchestration with Prefect and DaskTaskRunner

Utilizes Prefect flows with DaskTaskRunner to achieve concurrent pipeline execution and efficient error handling.

Sample Data Generation for Restaurant Operations

Target Outlet Locations

Koramangala
Indiranagar
Whitefield
MG Road
Electronic City

Customer Data

800 customers
Each with unique mobile numbers

Chef Assignments

3-6 chefs per outlet
Each chef has a unique ID

Menu Structure

10 menu items
Customizable toppings

Order Simulation

300 orders per outlet per day
Simulations span configured date ranges

Generated Outputs

Dimension and fact JSON files
Captures customers, chefs, sales, kitchen events, and stock snapshots

Source Facts Pipeline

Data Ingestion and Validation Process

The pipeline ingests operational data from SQL Server using JDBC.

It supports both full and incremental (delta) data loads.

Configuration-Driven Data Validation

Validation features include mandatory columns checking, type casting, and verifying allowed values.

The pipeline detects duplicates and anomalies in the ingested data.

Dynamic SQL Query Construction

Queries are dynamically constructed based on timestamp columns for delta filtering.

These queries enable precise control over data ingestion.

Output to Delta Lake Raw Layer

Validated and cleansed data is written to the Delta Lake raw layer.

Supports full overwrite and UPSERT merges for efficient data management.

Fact Enrichment Process

Overview of the Enrichment Process

The enrichment process transforms raw facts into analytical datasets.

This transformation involves calculating over 30 derived attributes.

Calculated Temporal Attributes

Enrichment includes deriving time dimensions such as order date, hour, week, and weekend flags.

These attributes support temporal-based analyses.

Customer Behavior Metrics

Derived customer behavior metrics include lifetime value and repeat status.

These enhance the understanding of customer engagement and loyalty.

Operational Performance Metrics

The process calculates kitchen SLA compliance and cooking durations.

These metrics evaluate operational performance efficiently.

Automation and Dispatch System

Enrichment employs a harmonizer factory pattern.

This pattern dispatches processes to specialized enricher classes per fact table.

Enrichment of Fact Tables

Sales Fact Table Enrichment

Adds detailed time attributes.

Aggregates item metrics.

Computes customer KPIs.

Kitchen Fact Table Enhancements

Calculates cook durations.

Flags orders delayed beyond the SLA threshold of 600 seconds.

Stock Fact Table Improvements

Categorizes inventory status into low, medium, and high.

Raises restock alerts when quantities fall below 10 units.

Enrichment of People (Dimension) Tables

Customer Data Enhancements

Incorporates ordering patterns.

Analyzes lifetime value.

Identifies repeat customer indicators.

Chef Performance Metrics

Aggregates cooking efficiency.

Monitors SLA compliance.

Evaluates chef-level KPIs.

Support Fine-Grained Analytics

Facilitates insights into customer preferences.

Improves operational staffing effectiveness.

Enrichment of Outlet-Level Data

Aggregate Key Metrics for Outlets

Outlet-level enrichment involves the integration of sales, kitchen, and stock metrics into a comprehensive dataset specific to each outlet.

Operational KPI Reporting

This process provides crucial operational indicators such as sales volume, adherence to kitchen SLA requirements, and the overall health of inventory levels.

Facilitated Cross-Outlet Analysis

The enriched data enables benchmarking across outlets and aids significantly in making strategic decisions regarding outlet management and inventory optimization.

PowerBI Integration and Reporting

Integration Pipeline Functionality

The integration pipeline publishes enriched datasets from Delta Lake to SQL Server.

Supports two schemas: enrich (fact and aggregated tables) and people (dimension tables).

Data Loading Mechanism

Uses JDBC-enabled Spark sessions to load 7 key tables.

Key tables include sales, kitchen, stock, customers, chefs, and outlet performance.

PowerBI Reporting Integration

Data is consumed by two PowerBI reports.

Reports include Outlet Operations Dashboard and Customer Analytics Dashboard.

Insights and Interactivity

Enables interactive visualization capabilities.

Covers aspects like sales, kitchen efficiency, inventory status, and customer behavior.

Detailed Data Model and Table Structure

Introduction to the Data Model

The restaurant project implements a star schema providing efficient analytical capabilities.

Dimension Tables

The dimension tables include dim_customer, dim_outlet, dim_chef, dim_item, and dim_stock_item, designed for attribute storage.

Fact Tables

The fact tables consist of fact_sales (incorporating nested order_menu_details JSON), fact_kitchen, and fact_stock for performance metrics.

Data Integrity Measures

The tables employ primary keys, foreign keys, and validation constraints to maintain data consistency and referential integrity.

Analytical Readiness

The structured star schema prepares the model for robust analytical operations and comprehensive data insights.

Schema Details and Relationships

Fact Sales Schema Overview

The 'fact_sales' table links to the 'customers' and 'outlets' tables. It contains order details stored as nested arrays.

Fact Kitchen Schema Description

The 'fact_kitchen' table manages chef assignments to cooking events, including timestamps and Service Level Agreement (SLA) flags.

Inventory Schema Relationship

The 'fact_stock' table captures data on inventory snapshots, associating them with 'stock items' and 'outlets'.

Dimension Tables Roles

Dimension tables act as master data holders, maintaining uniqueness and key constraints for data reliability.

Schema Relationships Functionality

Defined relationships enable complex joins, facilitating accurate KPI calculations for sales, kitchen, and inventory operations.

PowerBI Reports Overview

Outlet Operations Dashboard

Focus: Outlet performance, sales metrics, kitchen efficiency, inventory status.

File: outlet.pbix.

Connected to SQL Server enriched schema.

Provides actionable insights to operations teams.

Customer Analytics Dashboard

Focus: Customer behavior, ordering patterns, preferences, satisfaction metrics.

File: customer.pbix.

Connected to SQL Server enriched schema.

Provides actionable insights to marketing teams.

PowerBI Data Refresh and Integration Pipeline

Prefect-Orchestrated Refresh Pipeline

The automated refresh pipeline is orchestrated by Prefect utilizing DaskTaskRunner for parallel execution.

Data Extraction from Delta Lake

The pipeline extracts enriched data from Delta Lake tables using DataLakeIO and SourceObjectAssignment abstractions.

SQL Server Integration

Data is written to SQL Server using the DataWriter component via JDBC connections and schema mappings.

Configurable Load Modes

The system supports both full and delta load modes, allowing configuration-driven table selection and environment targeting (prod/dev).

Robust Logging and Validation

Comprehensive logging, error handling, and validation mechanisms are implemented to maintain data consistency.

ETL Orchestration with Prefect and DaskTaskRunner

ETL Orchestration with Prefect and DaskTaskRunner

Workflow orchestration employs:
- Prefect flows to define pipeline logic and task dependencies
- DaskTaskRunner for distributed parallel task execution enhancing performance
- Task metadata tagging for monitoring and filtering
- Graceful error handling, retries, and logging integrated with Prefect UI

This framework maintains clear separation of orchestration from business logic, enabling modular and testable pipeline components.

Summary and Next Steps

Understanding the Architecture

The Restaurant Project offers a structured layered architecture, processing data from ingestion to analytics.

Utilizing Robust Infrastructure

Our infrastructure supports various data formats and ensures seamless transformations.

Exploring PowerBI Dashboards

Integrated dashboards provide insights into outlet operations and customer behaviors, aiding analytics.

Enhancing Metrics with KPIs

Future plans include incorporating new Key Performance Indicators aligned with evolving business requirements.

Implementing Automation Processes

Scaling and refreshing data efficiently is a next step through automation strategies.