

WORD DOCUMENT

Capstone Project: Pharmaceutical Supply & Cold-Chain Compliance Dashboard

Group No. 5

Document Version / Details: Ver. 2.1/ 02-Jan-2023

Capstone Project: Pharmaceutical Supply & Cold-Chain Compliance Dashboard

Business Scenario: Building a Secure Data Pipeline with AWS, Snowflake, Informatica Intelligent Cloud Services, and Tableau for a Global Pharmaceutical Supply & Cold-Chain Compliance Dashboard

Project Overview: General

The goal of this capstone project is to design, develop, and deploy a secure data pipeline that extracts, transforms, and loads (ETL) shipment, temperature-sensor, and facility data from AWS S3 into Snowflake using Snowpipe and Snowpark (Python). The project also includes securing data using Data Masking, setting up AWS SQS for automated event notifications, and leveraging Informatica Intelligent Cloud Services (IICS) for advanced data integration. Additionally, Tableau will be used to create meaningful visualizations and insights from cold-chain and compliance data.

As a Snowflake Developer in a global pharmaceutical manufacturer and distributor, the objective is to analyze and generate reports for better decision-making using logistics and temperature-monitoring data. This involves extracting, transforming, and loading data into a centralized data warehouse using IICS Mapping Tasks with Data Integration Services.

Tools Utilized:

- **Snowflake:** For data loading and unloading and for Storage Integration using Snowpipe
- **IICS:** For creating Data Integration Tasks to move data from Snowflake to IICS
- **Tableau:** For Report Development and visualization for getting business Insights

Objective:

- Design and implement a robust data integration pipeline using Snowflake, IICS and Tableau to facilitate seamless data movement and transformation.
- Develop interactive dashboards and reports to visualize insights derived from shipment and sensor data.
- Ensure scalability, security, versioning, availability, and reliability in handling data integration tasks through Tableau.
- Optimize data models and configurations in Tableau for enhanced performance and resource utilization.
- Document the entire data integration and visualization process for future reference and analysis in a PowerPoint template shared by LTIMindtree SPOC on Day 3.

Assumptions:

- The project team consists of 4-5 associates proficient in Snowflake, IICS and Tableau and other required tools.
- Access to necessary Tableau environments and datasets is provided.
- Moderate to advanced knowledge of data integration concepts and data visualization techniques is assumed for all team members.

Expectations:

- Each use case must be error-free and bug-free.
- Design test cases wherever possible and prepare test-case passing reports.
- Implement appropriate validation wherever necessary.
- Adoption of Agile practice in the end-to-end lifecycle.
- Identify unique propositions for the project and highlight them in the presentation.
- Implementation of non-functional parameters like Security, Availability, Scalability is mandatory.
- Careful design of appropriate Data Models with Normalization must be done and presented.
- Design all the required Snowflake, IICS Tasks and Tableau Visuals error-free with robust design to handle data-pipeline errors.
- Identify the Unique Selling Proposition (USP) for Cold-Chain Compliance Analysis.
- Follow Agile practices and create Scrum Board and Burn-Down Chart etc.
- Project completion within 4 days, followed by 2 additional days for enhancements, presentations, and further extensions.
- Documentation of all project components, including code, configurations, and insights.
- Regular communication and collaboration among team members, SMEs, and stakeholders.
- Adherence to best practices for data integration, visualization, and optimization.

Use Cases:

1. **Data Integration:** Designing and implementing data integration workflows in Tableau to extract, transform, and load shipment and sensor data.
2. **Data Visualization:** Creating interactive dashboards and reports in Tableau to visualize temperature excursions and compliance metrics.
3. **Scalability and Reliability:** Ensuring scalability and reliability in handling data integration tasks through Tableau.
4. **Optimization:** Optimizing data models and configurations in Tableau for improved performance and resource utilization.
5. **Security Measures:** Implementing security measures in Tableau to protect sensitive product and shipment data and ensure compliance with regulations.

Submission Guidelines:

- Regular updates on project progress and challenges.
- Documentation of Tableau reports, dashboards, and insights.
- Presentation of key findings and recommendations.
- Demonstration of data integration workflows and visualization techniques.
- Collaborative review and feedback sessions.

Pharmaceutical Supply & Cold-Chain Compliance Dashboard:

Technical Specification

Case Study Context:

You are part of the **Global Quality Analytics Division at Mediflow Biotech**, a multinational producer of vaccines and biologics with operations across North America, Europe, and Asia. The company maintains thousands of temperature-controlled shipments each month through a network of distribution centres and 3PL carriers.

Business leadership wants to implement a centralized Cold-Chain Compliance Dashboard to monitor shipment integrity, track temperature excursions, and enhance GDP (Good Distribution Practice) compliance.

You are tasked with building this platform using:

- Snowflake for modern cloud data warehousing and automation
- IICS and Snowpipe for ingesting structured and semi-structured IoT data
- Tableau for interactive visual analytics

Business Objectives

- Unify shipment, sensor, and facility datasets from diverse systems.
- Create a robust, automated data pipeline with incremental updates.
- Enable secure sharing of compliance dashboards with regulators and vendors.
- Apply enterprise-grade governance (Row-Level Security, masking, tags).
- Build executive-ready Tableau dashboards for real-time monitoring and decision-making.

Input Datasets (All Provided as .csv):

Dataset Name	Description	Source
Shipment_Events.csv	Outbound and inbound shipments with status, carrier, and temperature profiles	Manual Upload
Sensor_Readings.csv	IoT temperature and humidity readings for each shipment interval	Streaming to S3 via Snowpipe
Facility_Master.csv	Cold storage warehouse and distribution centre metadata	MySQL (via IICS)
Product_Master.csv	Product name, category, and temperature sensitivity	Static Table
Carrier_Master.csv	Carrier, mode of transport, and SLA attributes	Static Table
Region_Hierarchy.csv	Geographic mapping of region, country, zone	CSV Upload

Project Phases and Detailed Tasks

PHASE 1: Data Acquisition and Loading

Task 1: Setup Snowflake Environment

- Create warehouse, database, schema.
- Setup users, roles, and assign privileges.
- Configure resource monitors for cost control.

Task 2: Load Flat Files via SnowSQL

- Use PUT and COPY INTO for:
 - Product_Master.csv
 - Region_Hierarchy.csv
 - Carrier_Master.csv
- Create matching staging tables with proper data types.

Task 3: MySQL → Snowflake via IICS

- Use IICS to extract from facility_master (MySQL).
- Transform and load into stg_facility_master in Snowflake.
- Schedule using a Taskflow.

Task 4: Real-Time Sensor Ingestion via Snowpipe

- Define External Stage pointing to S3 bucket.
- Create File Format (CSV, detect date types).
- Create Pipe and enable Snowpipe auto-ingestion from Sensor_Readings.csv.
- Test with 2–3 new file drops into S3.

PHASE 2: Data Modeling and Star Schema

Task 5: Build Dimension Tables

- DimProduct – ProductID, ProductName, Category, SensitivityClass, TempRange
- DimCarrier – CarrierID, CarrierName, Mode, SLAType
- DimFacility – FacilityID, FacilityName, City, Country, StorageType, RegionID
- DimRegion – Zone, Country, Region, RegionHierarchy

Task 6: Build Fact Tables

- FactShipment – joins DimProduct, DimCarrier, DimFacility, DimRegion.
- FactSensor – time-series temperature and humidity records by shipment.
- FactCompliance – derived SLA and temperature excursion summary.

Task 7: Apply Surrogate Keys and Constraints

- Implement date dimensions, referential integrity.
- Use MERGE statements for upserts.

PHASE 3: Incremental Automation Using Streams, Tasks, and MVs

Task 8: Stream Setup

- Create STREAM objects on:
 - stg_sensor_readings
 - stg_shipment_events
- Monitor changes using METADATA\$ACTION.

Task 9: Task Automation

- Create hourly TASK chains to:
 - Insert new sensor data into FactSensor.
 - Update latest SLA and excursion stats in FactCompliance.
 - Log changes into audit_log table.

Task 10: Materialized Views for Performance

- mv_excursion_summary – aggregated temperature excursions by product and region.
- mv_carrier_sla – on-time delivery vs breach percent.
- mv_facility_performance – facility-wise average temperature stability.

PHASE 4: Security, Governance and Sharing

Task 11: Data Masking and RLS

- Apply MASKING POLICY to:
 - DimProduct.ProductName
 - DimCarrier.CarrierName
- Implement ROW ACCESS POLICY to restrict FactCompliance based on user's region.

Task 12: Secure Data Sharing

- Create SECURE SHARE:
 - Includes selected fields from mv_excursion_summary and mv_carrier_sla.
- Test access from reader account.
- Document sharing process.

PHASE 5: Visualization Using Tableau

Task 13: Data Connection

- Connect Tableau to Snowflake using live/extract mode.
- Build calculated fields and data model within Tableau.

Task 14: Design Dashboards

Dashboard	Visuals & Features
Cold-Chain Integrity Overview	Gauge & trend charts for temperature stability and excursion count
Shipment SLA Performance	Bar chart for on-time vs delayed shipments by carrier
Facility Compliance Heatmap	Facility × Region heatmap for temperature deviation
Sensor Drift Analysis	Line graph for sensor calibration variance
Regulatory Compliance KPIs	Table for GDP violations and response time

Task 15: Interactivity Features

- Region/Product/Carrier filters.
- Drilldowns (Continent → Country → Facility).
- Alerts for temperature excursion breaches.
- Parameter-based views (e.g., daily vs monthly).

PHASE 6: Documentation and Submission

Task 16: Reporting

- Project architecture diagram (Snowflake + IICS + Tableau).
- ER Diagram of final schema.
- Data flow pipelines.
- Masking/RLS/Share configuration proofs.
- SnowSQL scripts, Tableau workbook, CSVs.

Deliverables

Type	Description
SQL Scripts	DDL, Stream, Task, Policies
Data Files	Cleaned .csv source files
Tableau Dashboards	.twbx or PDF exports
Documentation	Word or PDF project report
Screenshot Proofs	For shares, security, ingestion, etc.

**Let's get to the
future, faster.
Together.**

