

## **Capstone Project: Insurance Claims Fraud Detection & Settlement Analytics**

**Business Scenario:** Building a Secure Data Pipeline with AWS, Snowflake, Informatica Intelligent Cloud Services, and Tableau for Insurance Claims Fraud Detection & Settlement Analytics

### **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) policy, claim, and FNOL (First Notice of Loss) data from AWS S3 into Snowflake using Snowpipe and Snowpark (Python). The project also includes securing sensitive PII 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 the data.

As a Snowflake Developer in a multi-line insurance carrier (Auto, Property, Health), the objective is to analyze and generate reports for better decision-making using Claims 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 multi-line Insurance Claims 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 advance 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-end lifecycle.
- Identify unique prepositions for the project and highlight it in the presentation.
- Implementation of non-functional parameters like Security, Availability, Scalability is mandatory.
- Very careful design of appropriate Data models with Normalization must be done and presented.
- Design all the required Snowflake, IICS Tasks and Tableau Visuals with Error-Free with Robust design to handle Errors in Data Pipeline
- Identify the Unique Selling Proposition (USP) for Claims Analysis
- To 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 multi-line Insurance Claims data.
2. Data Visualization: Creating interactive dashboards and reports in Tableau to visualize Claims and Fraud indicators to analyze trends and patterns.
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 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.

## Insurance Claims Fraud Detection & Settlement Analytics:

### Technical Specification

#### Case Study Context:

You are part of the Data Analytics & Risk Control Division at Guardian Insurance Group, a multi-line insurer covering Auto, Health, and Property segments. With operations across multiple regions and partner networks (garages, hospitals, third-party administrators), leadership is seeking to implement a centralized Claims Intelligence Platform to monitor fraud risk signals, track settlement timelines, evaluate provider anomalies, and improve reserve accuracy. 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 data
- Tableau for powerful, interactive visual analytics

#### Business Objectives

- Unify policy, claims, FNOL, and provider datasets from diverse systems
- Create a robust, automated data pipeline with incremental updates
- Enable secure sharing of reports with external audit and SIU teams
- Apply enterprise-grade governance (RLS, masking, tags)
- Build executive-ready Tableau dashboards for monitoring and decision-making

#### Input Datasets (All Provided as .csv):

Dataset Name	Description	Source
Policy_Master.csv	Active policy register with product, tenure, sum assured	Manual Upload
Claim_Transactions.csv	Claim lifecycle events (FNOL, survey, approval, payment)	S3 Bucket
Fraud_Alerts.csv	SIU red-flags, model scores, and investigation outcomes	Streaming to S3
Customer_Profile.csv	Customer demographics and contact (PII masked)	Static Table
Provider_Network.csv	Provider directory (hospitals/garages) with categories	Manual Upload
Region_Hierarchy.csv	Geographic mapping of country, state, branch	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:
  - Policy\_Master.csv
  - Customer\_Profile.csv
  - Region\_Hierarchy.csv
- Create matching staging tables with proper data types

#### Task 3: AWS → Snowflake via IICS

- Use IICS to extract from provider\_network (S3)
- Transform and load into stg\_provider\_network in Snowflake
- Schedule using a Taskflow

#### Task 4: Real-Time Fraud 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 Fraud\_Alerts.csv
- Test with 2–3 new fraud alert drops into S3

## PHASE 2: Data Modeling and Star Schema

### Task 5: Build Dimension Tables

- DimCustomer – CustomerID, CustomerName, Region, RiskBand
- DimPolicy – PolicyID, ProductType, Tenure, SumAssured
- DimProvider – ProviderID, ProviderName, Category, Region
- DimRegion – Zone, Country, State, Branch, RegionHierarchy

### Task 6: Build Fact Tables

- FactClaims: joins DimPolicy, DimCustomer, DimProvider, DimRegion
- FactSettlement: settlement approvals, TAT, leakage indicators
- FactFraud: SIU flags, model scores, investigation status

### 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:
  - o stg\_claim\_transactions
  - o stg\_fraud\_alerts
- Monitor changes using METADATA\$ACTION

### Task 9: Task Automation

- Create hourly TASK chains to:
  - o Insert new claims to FactClaims
  - o Update latest settlement status in FactSettlement
  - o Append investigation outcomes to FactFraud and log to audit\_log

### Task 10: Materialized Views for Performance

- mv\_fraud\_risk\_trends: aggregated fraud rate by product and region
- mv\_provider\_anomalies: provider outlier index vs peer benchmarks
- mv\_settlement\_tat: average settlement days and leakage by branch

## PHASE 4: 🛡️ Security, Governance and Sharing

### Task 11: Data Masking and RLS

- Apply MASKING POLICY to:
  - DimCustomer.email
  - DimCustomer.phone
- Implement ROW ACCESS POLICY to restrict access to FactClaims based on user's region and product line

### Task 12: Secure Data Sharing

- Create SECURE SHARE:
  - Includes only selected fields from mv\_fraud\_risk\_trends and mv\_settlement\_tat
- 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

Fraud Detection Overview KPI tiles (Fraud%), fraud score distribution, alerts timeline

Settlement Efficiency Box/line charts of TAT by branch and product

Provider Quality Scatter/heatmap of provider anomaly index and claim ratios

Regional Risk Heatmap Map of fraud intensity and leakage by geography

Claim Lifecycle Tracker Gantt/timeline from FNOL → survey → approval → payment

### Task 15: Interactivity Features

- Product/Region filters
- Drilldowns (Country → State → Branch)
- Alerts for KPI deviations
- Parameter-based views (e.g., 30/60/90-day windows)

## **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.