**Aviation Data Platform Modernization Program**

Contents

[**Problem Statement** 3](#_Toc204133347)

[**Skill Tower Involved** 3](#_Toc204133348)

[**Use Case / Architecture Overview** 4](#_Toc204133349)

[**User Stories** 5](#_Toc204133350)

[**Expected Deliverables** 5](#_Toc204133351)

[**Milestones and Duration** 6](#_Toc204133352)

[**Implementation Notes** 7](#_Toc204133353)

[**Evaluation Rubrics** 7](#_Toc204133354)

# **Problem Statement**

An international airline operates across multiple airports, flight routes, and partner systems. However, its data is fragmented across flight operations, maintenance logs, passenger systems, and third-party sources (e.g., weather, airport data). The airline faces challenges in real-time performance monitoring, regulatory compliance (e.g., EASA, FAA), predictive maintenance, and passenger satisfaction insights.

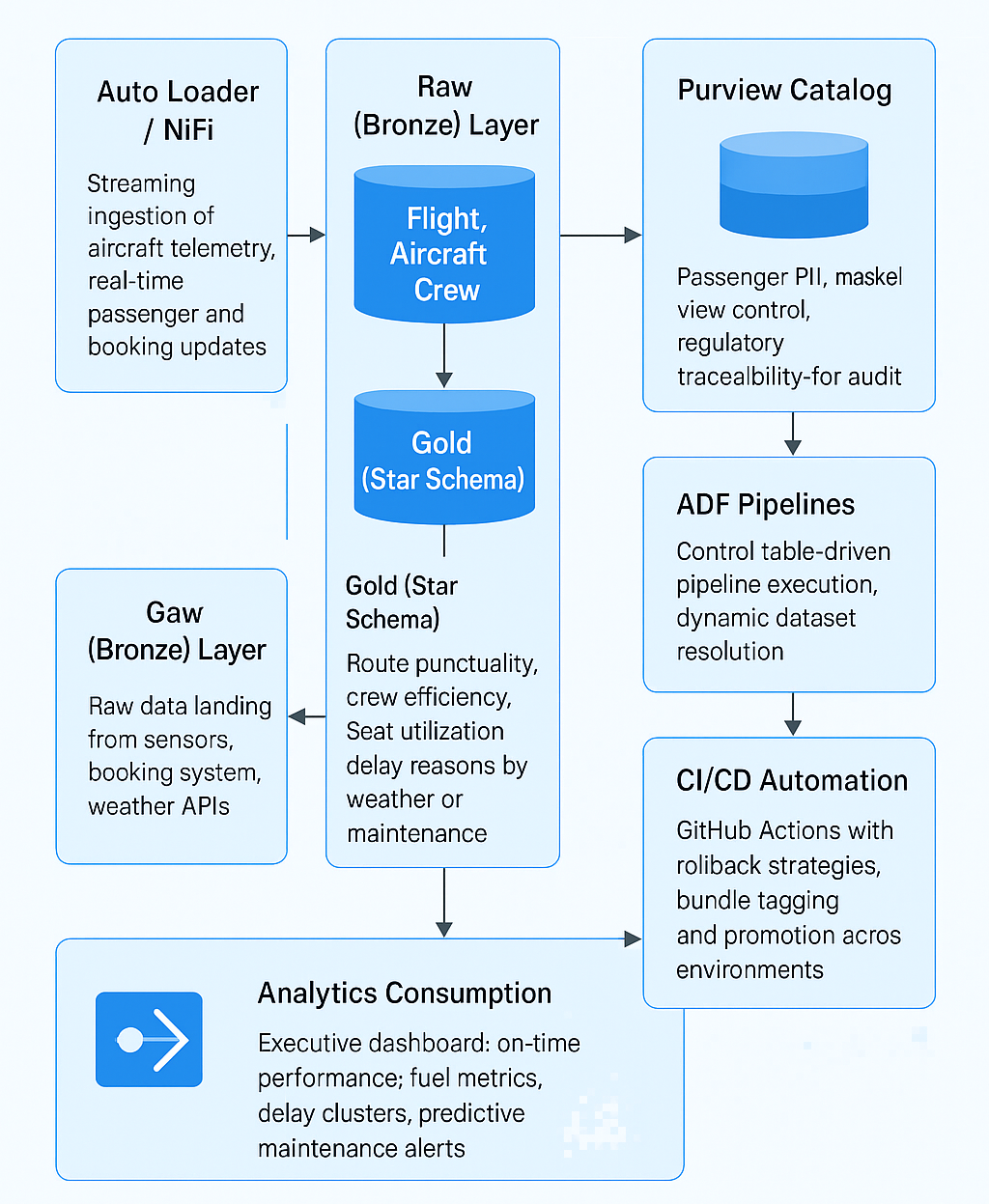
This program aims to build a **unified aviation data platform** to support streaming and batch analytics, data governance, predictive models, and real-time dashboards — built on Azure, Databricks, Delta Lake, and open standards.

# **Skill Tower Involved**

| **Skill Area** | **Description** |
| --- | --- |
| Dev Environment & Git | Signed-commit enforcement, pre-commit hooks, VS Code productivity, Azure RG setup |
| Python & CLI Tools | Data structure operations, CSV/Parquet handling, decorators, argparse, pytest, wheel packaging |
| Spark & Delta Lake | Streaming ingestion (telemetry), DataFrame API, caching, partitioning strategies |
| SQL Engineering | JOINs, CTEs, window functions for delay trends, MERGE UPSERT for SCD flight records |
| Data Modeling | Hub/Link/Satellite design for flights, aircrafts, crew, weather, passengers |
| Governance & Cataloging | PII/PIA tagging (e.g., passenger data), column masking, audit lineage via Purview |
| Orchestration | ADF pipelines with control tables, dynamic datasets, parameterized deployments |
| CI/CD | GitHub Actions for YAML-based deployments, rollback, test automation, secure bundle promotion |
| Domain Expertise | Aviation-specific modeling: aircraft telemetry, route delay analysis, crew assignment logs |

# **Use Case / Architecture Overview**

| **Component** | **Description** |
| --- | --- |
| **Auto Loader / NiFi** | Streaming ingestion of aircraft telemetry, real-time passenger and booking updates |
| **Raw (Bronze) Layer** | Raw data landing from sensors, booking system, weather APIs |
| **Business Vault (Silver)** | Modeled Hubs: Flight, Aircraft, Crew — Satellites for maintenance, telemetry, passenger behavior |
| **Gold (Star Schema)** | Route punctuality, crew efficiency, seat utilization, delay reasons by weather or maintenance |
| **Purview Catalog** | Passenger PII, masked view control, regulatory traceability for audit |
| **ADF Pipelines** | Control table-driven pipeline execution, dynamic dataset resolution |
| **CI/CD Automation** | GitHub Actions with rollback strategies, bundle tagging, and promotion across environments |
| **Analytics Consumption** | Executive dashboard: on-time performance, fuel metrics, delay clusters, predictive maintenance alerts |



# **User Stories**

| **ID** | **Title** | **Description** |
| --- | --- | --- |
| US\_01 | Ingest Telemetry | As a data engineer, I want to stream aircraft sensor data into the raw layer in near real time |
| US\_02 | Crew Assignment Vault | As a modeler, I want to model the relationship between flights and assigned crew across shifts |
| US\_03 | Maintenance Forecasting | As a data scientist, I want to build predictive models based on historical telemetry and weather |
| US\_04 | Passenger Privacy Compliance | As a compliance officer, I want Purview to mask and track PII access in reports |
| US\_05 | Route Performance Dashboard | As an operations manager, I want to analyze route-wise punctuality and delay breakdowns |
| US\_06 | Automated Deployment | As a DevOps engineer, I want to automate validation and release to staging and prod environments |
| US\_07 | Quality Monitoring | As a QA lead, I want to run Great Expectations checks on booking data completeness and accuracy |

# **Expected Deliverables**

| **ID** | **Deliverable** | **Description** |
| --- | --- | --- |
| D\_01 | Streaming Ingestion Layer | Aircraft telemetry and booking data ingested into Delta Bronze tables |
| D\_02 | Business Vault + Star Schema | Crew, Aircraft, Flight modeled in Vault; metrics exported in star schema |
| D\_03 | Governance & Purview Config | Passenger and crew data tagged, lineage visible, audit ready |
| D\_04 | Python Tooling & CLI | Tools for chunked CSV read, logging, argparse CLIs, packaged in private PyPI |
| D\_05 | CI/CD Workflows | Auto-deploy via GitHub Actions with secure rollback |
| D\_06 | Data Quality Framework | Great Expectations suite with checkpoints and webhook alerts |
| D\_07 | Executive Dashboard | Visual KPIs: delay analytics, crew performance, route heatmap, fuel efficiency metrics |

# **Milestones and Duration**

| **Milestone** | **Estimated Duration** |
| --- | --- |
| Git Setup, Azure RG Bootstrap | 1 days |
| Telemetry Ingestion & Bronze Tables | 1 days |
| Business Vault & Star Schema | 1 days |
| Purview + PII Masking + Audit Logging | 1 days |
| ADF Orchestration Pipelines | 1 days |
| CI/CD Integration & Test Automation | 1 days |
| Dashboard Build + Final Demo | 1 day |

# **Implementation Notes**

| **ID** | **Details** |
| --- | --- |
| IN\_01 | Telemetry should use micro-batch streaming with watermarks for late data |
| IN\_02 | Aircraft and Crew must be modeled with Satellite tables for effective tracking |
| IN\_03 | Use OPTIMIZE ZORDER BY (flight\_id, flight\_date) to speed up analytics |
| IN\_04 | All passenger-related columns (email, passport) tagged and masked in Purview |
| IN\_05 | Control tables should manage orchestration of data pipelines with timestamp bookmarks |
| IN\_06 | CI/CD pipelines use bundle.yml for validation, tagging, and environment promotion |
| IN\_07 | Route delay prediction models should log inference metadata to Delta for audit |

# **Evaluation Rubrics**

| **Criteria** | **Weight** |
| --- | --- |
| Real-Time Telemetry Integration | 20% |
| Data Modeling for Aviation Entities | 20% |
| Governance, Masking, and GDPR Compliance | 15% |
| CI/CD Pipeline with Rollback Support | 15% |
| Code & Data Quality Assurance | 15% |
| Executive Dashboard Clarity | 10% |
| Innovation (Predictive Maintenance, Delay Clustering) | 5% |