Cloud Architecture for Airports “AEROCLOUD”

AEROCLOUD

Written by: Ray Varghese

Professor: Dr. Junaid Qazi

Table of contents

1. Introduction

Aerocloud is an integrated and sophisticated cloud-based system hosted on the Azure ecosystem that can serve as the technological spearhead to automate global airport operations. Integrating state-of-the-art digital cloud technology on Azure, Artificial Intelligence, and the Internet of Things will enhance operational efficiency, improve security, and elevate customer experiences. This digital landscape powered by Azure cloud storage and analysis can enable data-driven decision-making. Leveraging these concepts in the Aerocloud architecture helps to optimize functions like maintaining security protocols, baggage handling, aircraft turnaround, streamlining operations, tracking inventory, and financial KPIs Integration of technology, data, and sensors helps to empower and scale up the operations of an airport.

2. Mission

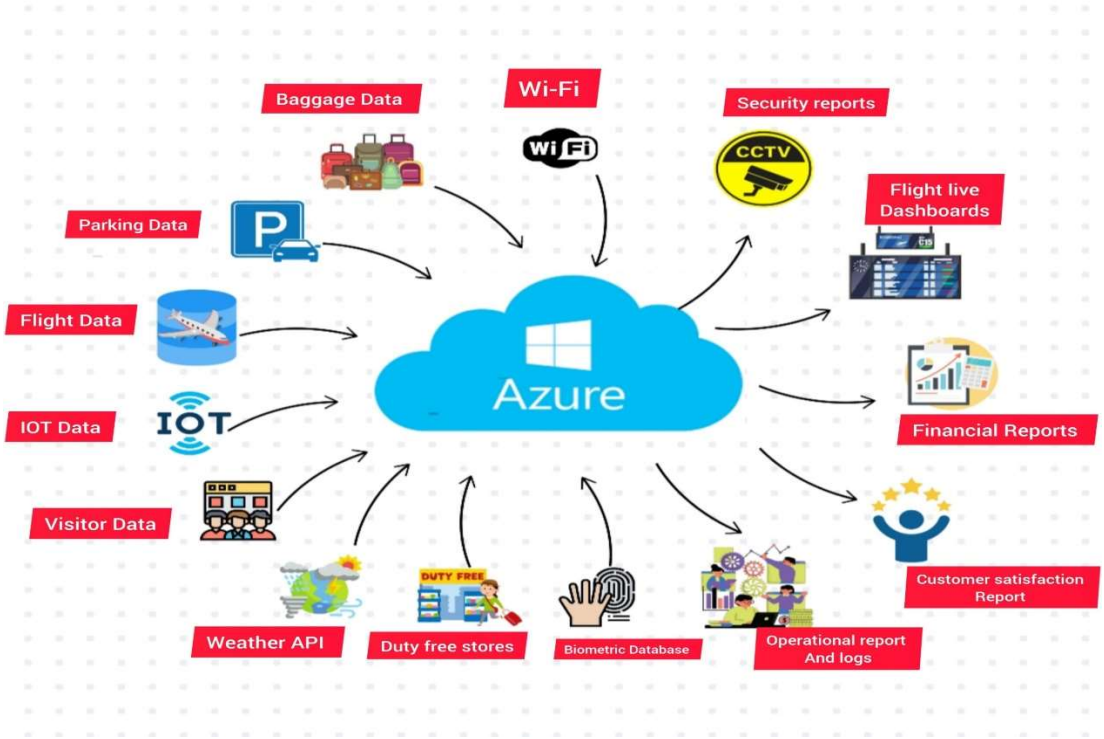
AeroCoud aims to make air travel smoother, smarter, and more secure. By bringing together the power of cloud technology, real-time data, and intelligent automation, we help airports run more efficiently—

making the journey better for travelers, safer for everyone, and easier for airport teams to manage. AeroCloud helps airports deliver superior customer experiences, achieve epitome safety compliance and standards, and manage resources efficiently.

3. Objectives

- Operational Efficiency- Improve efficiency, performance, and cost-effective methods to reduce redundancy. Optimize key airport functions like baggage handling, reduce aircraft delays, and streamline visitor traffic.
- Customer Experience- Focusing on delivering exceptional service and building long-lasting relationships
- Safety and Security- Strengthen Airport security using AI and automated systems that enable live feeds to make real-time decisions.
- Revenue Growth- Implementing strategies that drive financial performance and market expansion. Ensuring sustainability and social commitment as their long-term vision.

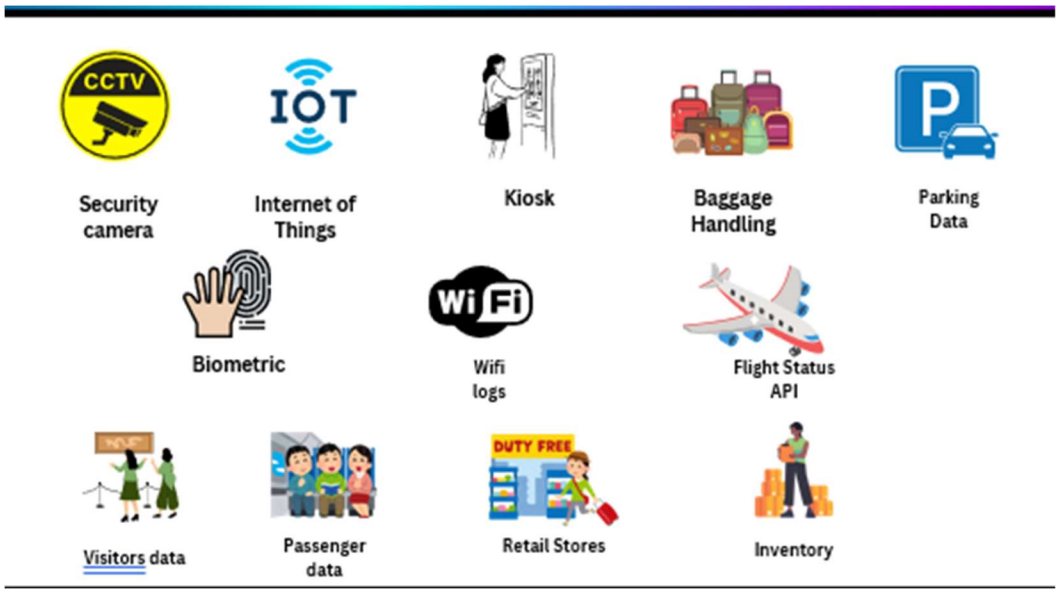
4. Visual diagram



5. Data Sources

As shown in the visual diagram, AeroCloud is powered by data sources from different spheres, being able to capture data from Kiosks, RFID tags, parking sensors, cameras, website APIs, Wi-Fi logs, biometrics, and the Internet of Things (IoT sensors). They can be broadly classified as batch and streaming data sources. The data gathered from these sources are ingested into the cloud

storage which is further transformed and aggregated to generate reports that enable data-driven decision making.



5.1 Streaming data sources (Real-time)

- Internet of Things (IoT)- Sensors mounted on Kiosks, baggage carousels, airport gates, passenger boarding areas, and aero bridges.
- Security Cameras for live surveillance of the airport premises.
- Kiosks which are used for tracking boarding passes, baggage weight, and dimensions.
- Biometric gates used at immigration counters to verify facial recognition and hold the biometric database of passengers.

- Baggage carousels and conveyors which track the passenger baggage RFID tags which streamline the baggage handling process.
- Flight status APIs which are connected to the airline database and update the status of the aircraft periodically. It helps to track the arrival, departure, gate status, etc.

5.2 Batch data sources (data ingestion at a fixed time)

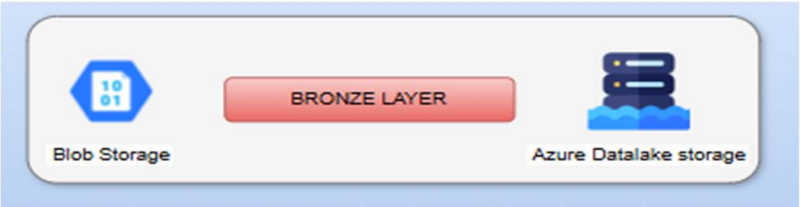
- Parking data which is collected when passengers or visitors book the parking spots from their mobile app or website.
- WiFi logs which can track the information of passengers and their connection trends which can track the footfall of visitors.
- Visitor and passenger data which can be correlated and identify the demographics and track the trends and passenger movements.
- Retail stores and their data can be correlated to rentals and financial analysis.
- Inventory data that tracks supplies such as fuel, catering, and fleet management.

6. Lakehouse Architecture

AeroCloud is managed in the Azure ecosystem and utilizes the Lakehouse architecture or the medallion architecture where the storage in Azure Data Lake storage Generation 2. The medallion architecture employs the bronze layer which stores the raw data in native format. The Silver layer contains curated and cleansed data. At this stage, Delta Lake ensures schema enforcement, and version control and stores data in parquet and delta file formats. The Gold layer contains aggregated or summarized data which can be transformed into financial reports, inventory reports, and other live dashboards.

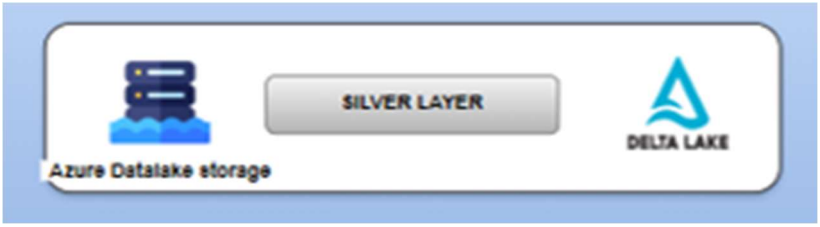
6.1 Bronze layer (Raw data)

All data- streaming and batch data get ingested or stored in this layer normally in the native format. Has access to hot and cold storage solutions. Azure Blob storage enables cost-effective cold or archive data. Azure data lake storage Gen 2 (ADLS) enables the storage of transactional and analytical data in the raw form. At this stage, all the data ingested has to be stored to enable a master copy.



6.2 Silver layer (Curated data)

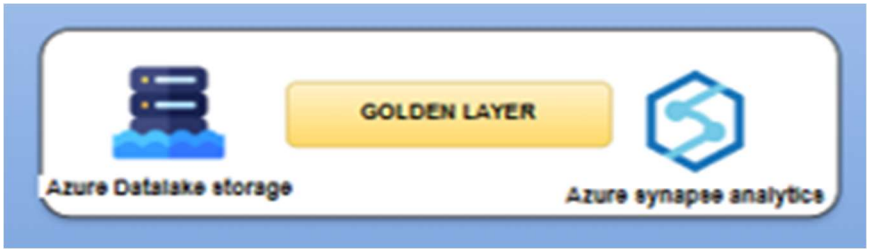
The data from the bronze layer undergoes ETL/ELT-Extraction, Transformation, and loading removing null values, duplicates, and errors. Standardization and normalization of data happen at this stage. Data is cleansed and curated to form a structured data format. Delta Lake enables schema enforcement and ACID transactions (Atomicity, Consistency, Isolation, and Durability). It uses parquet or Delta file formats to store the curated data with data versioning and time travel provisioning.



6.3 Gold layer (Aggregated data)

Highly structured and aggregated data is available at this staging area with derived insights and trends used for data-driven decision-making. This is the stage where reports are compiled and KPIs are evaluated. It combines cross-functional data from different business units or departments like Human resources, finance, and operations. The datasets are

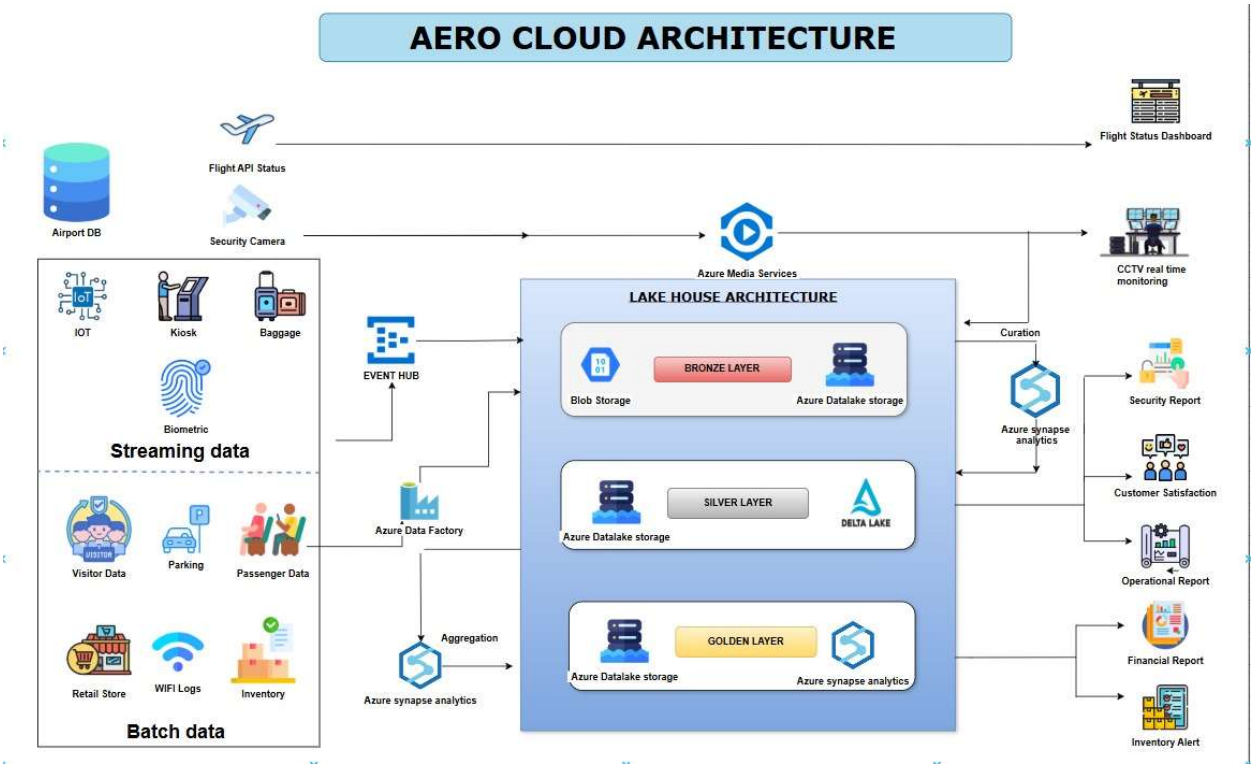
highly structured and relational. It enables live Power BI dashboards and KPI reporting. In this layer, Azure Synapse Analytics works as a data warehouse that stores the SQL queries on the dedicated pool to store reporting queries.



6.4 Consumer layer (Reports and destination)

This is the destination layer containing summarized and detailed reports for data consumption. It enables reports, dashboards, and alerts. AeroCloud has a live flight status dashboard that collects data from airline APIs, similarly, all the security footage is stored and summarized to store incident or security reports. Operational reports, financial reports, customer satisfaction reports, and inventory alerts are the outcomes of this layer. The audiences are data analysts, CEOs, and executive members.

7. Cloud Architecture- AEROCLOUD

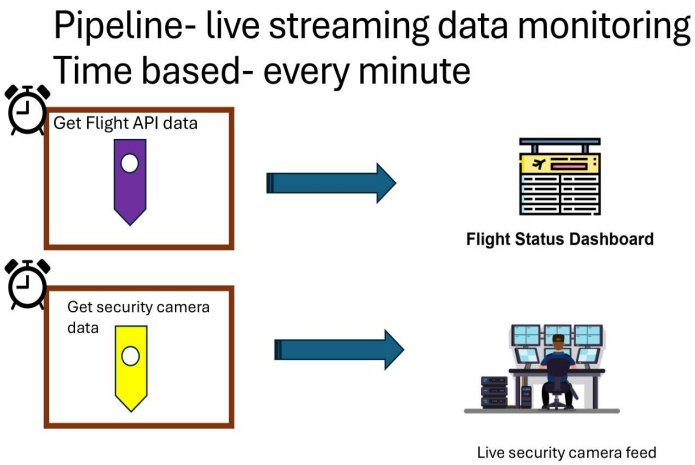


The AeroCloud architecture in the Azure ecosystem has an airline API that updates every minute and ingests the airline status arrival, departure, and boarding status into the Azure data factory which orchestrates the data into the flight status dashboard. The live security monitoring cameras ingest audio & video to the Event Hubs for streaming data and drive the Azure Media services for live streaming and CCTV monitoring center. This is essential for security protocols and compliance. The data is stored in the Azure data lake storage bronze layer where raw data is ingested. Streaming

data from other sensors, IoT, kiosks, baggage RFID tags, and biometric-enabled sources generates data that is ingested into the event hub or IoT hub to store in the bronze layer. The bronze layer has blob storage for cold data and ADLS for hot storage. Batch data- Passenger & visitor data, Wifi logs, parking, retail stores, and inventory data which is collected every 4 hours are ingested through the Azure data factory in its raw format in the bronze layer. The data in the bronze layer undergoes transformation or curation where data cleansing, standardization, removing duplicates, normalization, and error removals using Azure Synapse Analytics. This curated data flows into the silver layer where delta lake enforces schema formation and ACID transactions. Data versioning and storage optimization happen at this stage. The curated data is further aggregated and summarized using Azure Synapse analytics where transact structured query language T-SQL logical queries are stored on the dedicated SQL pool to form the KPIs, dashboards, and reports at the consumer layer. The data in the gold layer can be used to generate financial and operational reports along with inventory alerts and stock levels. It also generates security reports and customer satisfaction reports.

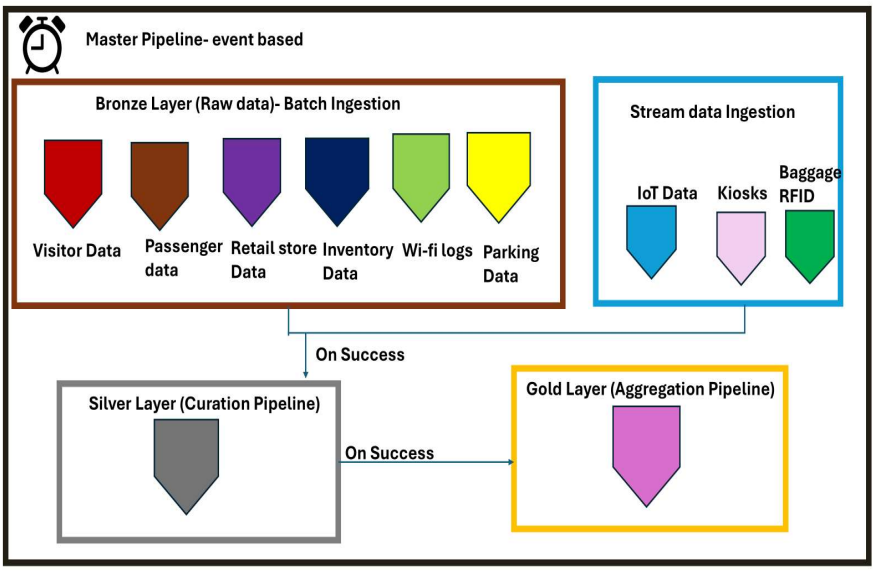
8. Pipeline Approach

- Time-based pipelines- Airline API data and security camera audio& video data are fetched every minute triggering live streaming and updating the live flight status dashboards to ensure less latency and real-time data.



- Event-based pipelines- The master pipeline is triggered every 30 minutes using a hybrid time and event-based approach. Following this, data is fetched from batch and streaming data sources causing ingestion into Azure data factory or Event Hub that causes dataflow into the bronze layer. On the success of ingestion from both batch and stream data, the curation or transformation

process begins where data cleansing and standardization occur. On the success of the curation event, the aggregation pipeline begins with the gold layer where summarized reports and alerts are generated.



9. Pipeline fail strategy

Failure of pipelines can occur due to unforeseen circumstances such as an unresponsive server, network connectivity problems, faulty sensors, or API connection timeout. As data engineers, it's essential to ensure the data and pipelines are always operational to ensure operational status. Using a retry mechanism ensures the number of retries and the time delay between each retry attempt. Azure data factory which handles batch ingestion has a default 30-second retry delay time and a

maximum of 3 retries. Azure Synapse Analytics can be configured to change the retry values. Azure Synapse environment has a monitor module that can verify pipeline runs and trigger runs.

Azure Synapse Analytics has a diagnostic setting that captures pipeline activity logs and stores them. Furthermore, the Azure ecosystem supports alerts when pipeline failure occurs which triggers email or text alerts using the Azure Monitor service. Azure logic apps can also automate the recovery or restart the pipeline activity.

10. Conclusion

AeroCloud is the cloud solution that ensures scalability in airport operations and a cost-effective storage & computing strategy. It's an upgrade that enables Smart airports to leverage tools like Artificial intelligence and Internet of Things (IoT) sensors. This paves the way for automation of various systems to ensure efficiency and less prone to human errors & negligence. This cloud upgrade ensures data security, and redundancy and enhances the ability of airport authorities to make data-driven decisions. Aerocloud is the answer to safety, security, extraordinary customer experiences, and sustainability.