

Case Study

Microsoft abric

Transform Your Business with Microsoft Fabric

Datahub Consulting & Microsoft Fabric Helped an Airline Improve Reporting & Operational Efficiency by Connecting Multiple Data Systems.

Industry

The airline industry is highly competitive, with a focus on delivering safe, timely, and comfortable travel experiences while managing operational complexities. Airlines manage large volumes of data across reservations, operations, maintenance, and customer service. Efficient reporting and data-driven decision-making is critical for improving performance, customer satisfaction, and profitability in this competitive market.

Challenge

Within the airline industry third party systems handle the data for reservations, loyalty programme, and finance etc. The airline was using Power BI for reporting and was taking manual extracts from those third-party systems on a monthly basis to import into the power BI models. The Power BI reports used similar reservation, customer, operational, and financial data but each power bi report had its own model, creating siloed reporting.

Solution

As the existing internal reporting included manual extracts, individual semantic models for each report, and didn't have the ability to report across different systems then this was a challenge we needed to overcome.

The solution was to ingest all the data from the various source systems in a data lake and create one warehouse and semantic model to facilitate all reporting. This would then give the airline the ability to easily have a 360-degree view of their data.

We achieved this with the Microsoft Fabric platform with Power BI reports. The solution allowed for incremental loading on a daily basis and didn't involve manual extracts.

The diagrams below give a more detailed understanding of the solution.

Outcome

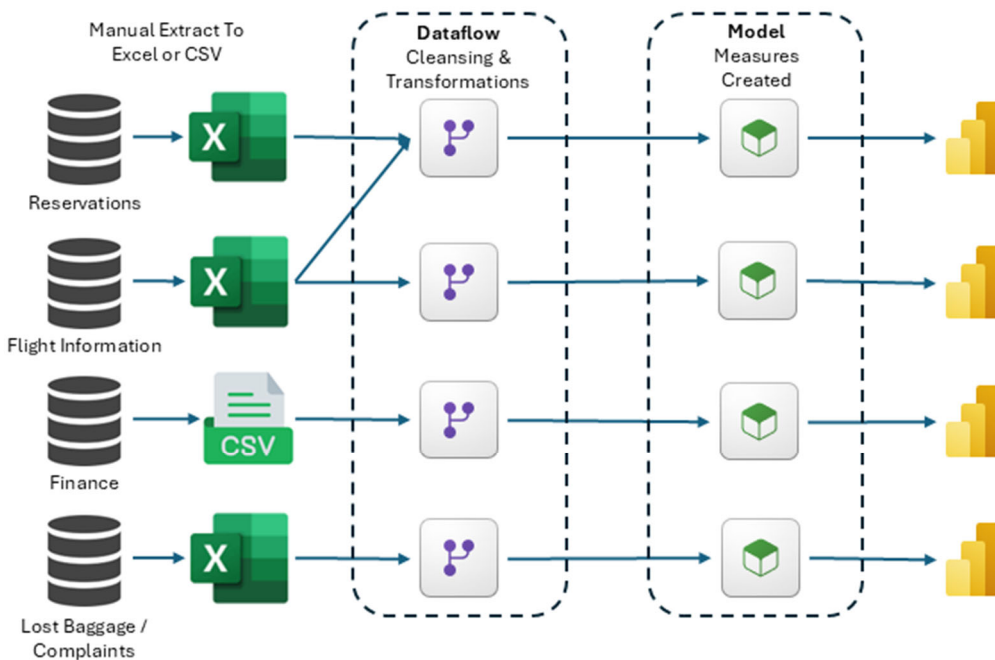
The airline for the first time could have a 360-degree view of the customer journey. Data was ingested into the model daily and then had accurate and timely reporting across systems. Utilizing their data from reservation through to landing at the destination, then performance and customer quality can be measured.

There are occasions where baggage goes missing that is beyond the control of the airline. The baggage logistics is the responsibility of the airport services. On these occasions the airline can monitor the performance of the airports regarding the airline's customers.

Previous Method (Built Internally)

Data hosted in separate databases for the different systems. Some of these are industry standard third-party systems used by most airlines. A manual extract is taken on a monthly basis from the system and this extract is either in excel or csv format. The excel / csv is imported into power BI. Cleansing and transformation are completed in the dataflow (power query) engine.

Each report has its own model and there is no scope for reporting across the different systems. This can lead to time delays in reporting and additional manual work.



New Solution

The solution ingested a copy of the raw data from the various systems and loaded the data into delta tables. From there a medallion architecture approach was used.

Dataflow Gen 2 was used to copy the data from the source system to OneLake and load into the delta tables.

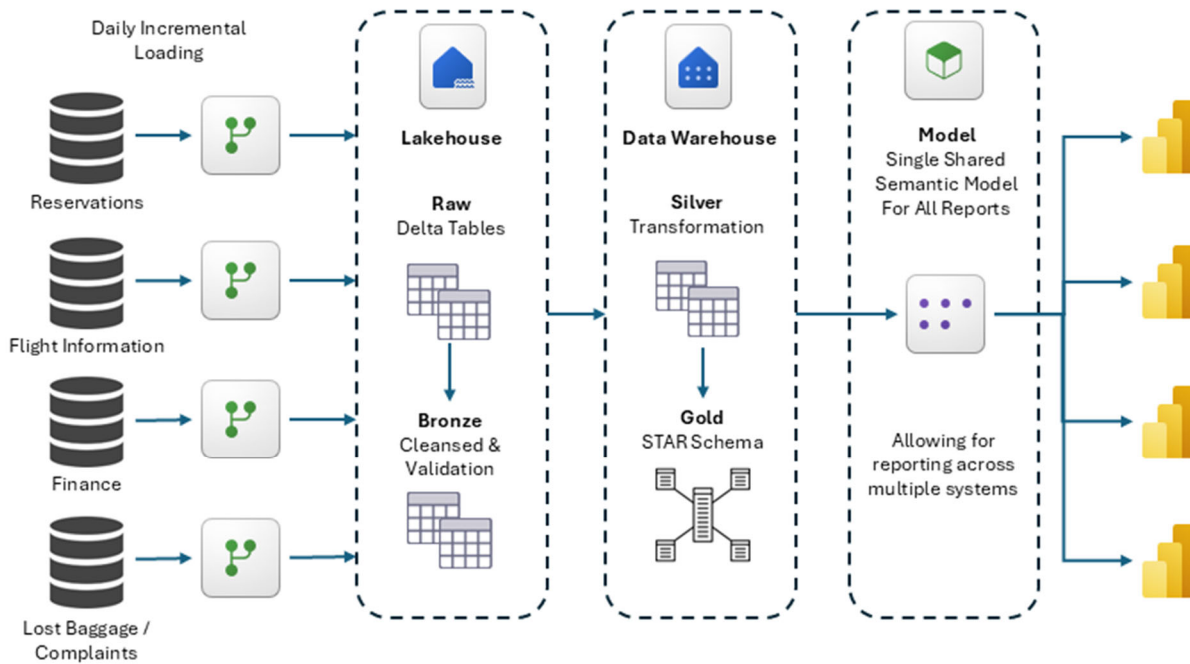
Fabric notebooks were used in the bronze layer as part of the cleansing and validation. This involved checking the data quality, identifying blanks, or nulls, and removing any unwanted columns.

In the data warehouse SQL stored procedures were used to format the data and to create the fact and dimension tables in the STAR schema.

Fabric Data Pipelines are used to trigger the activities including Dataflow Gen 2, Notebooks, and SQL Stored Procedures.

Finally, one single model was created utilizing all the data from the various sources to give a complete reporting solution. Users accessed the reports through a Power BI APP with multiple audiences created. This provided the security so only the authorized users could see the various reports. Data is partitioned in the model for performance and to allow for refreshing of the data using the XMLA endpoint, if needed in the future.

Data within aviation does include a lot of personal information of passengers and the solution took data protection laws into consideration and personal information was masked or removed at the bronze layer.



Technology We Used



OneLake

Central to any Microsoft Fabric project, this is where all the raw data is copied from the various systems. The raw data is loaded into delta tables where the data will be available in the Fabric platform.



Fabric – DataFlow Gen 2

Dataflow gen 2 is used to copy the data from source system to the Lakehouse delta table via a gateway. There are no transformations in the dataflow activities, all the cleansing, validation, & transformations are carried out in the bronze and silver layers.



Fabric – Notebooks

Used PySpark to do data quality checks and validation in the bronze layer. Also, the notebooks are used to remove the columns in the data frame of data that is not required in the solution. So only moving forward with the minimum number of columns necessary. Once that the data quality and validation is completed in the notebooks then data is available in the silver layer where transformations take place.



Fabric - Data Warehouse

A data warehouse was created using fact and dimensions tables to facilitate storing historic data for reporting. The data warehouse would be

incrementally loaded on a daily basis providing accurate and timely reports.



Fabric – Data Factory

Data pipelines are used to trigger the activities in the lakehouse and data warehouse. There is also logging of pipeline success / failure that is loaded into a log table. This log table is available to analyze the performance of the ELT process. Also, if there's a failure then this log will help with investigation.



Fabric – Semantic Model

There is one semantic model create that included operational reporting, financial, and customer service reporting. The financial reporting included ticket cost, Cost of refunds, and aircraft costs like landing fees etc. The models does not include any personal information allowing for compliance with data protection laws.



Power BI

A suit of reports was created from the models. Using shared models allowed for consistent reporting and prevented siloed reporting. Security of reports was integrated with the security groups set up by the IT team.

Integration with Machine Learning

The solution use will allow for the airline to use machine learning (predictive analytics) in the future to further enhance the solution. Using Microsoft Fabric that integrates multiple technologies and services allows for the system to easily be enhanced with additional features.



To see if Datahub Consulting could help your organisation with a Microsoft Fabric project. Please contact one of our office's where we'll be happy to arrange a call to discuss.

Email Address: info@datahubconsulting.co.uk

Leicester Office: +44 (0)116 223 0689

London Office: +44 (0) 2039188 799

European Office (Madrid): +34 911 11 58 11

