DATA+AI SUMMIT
BY databricks

Jet Streaming Data and Predictive Analytics:

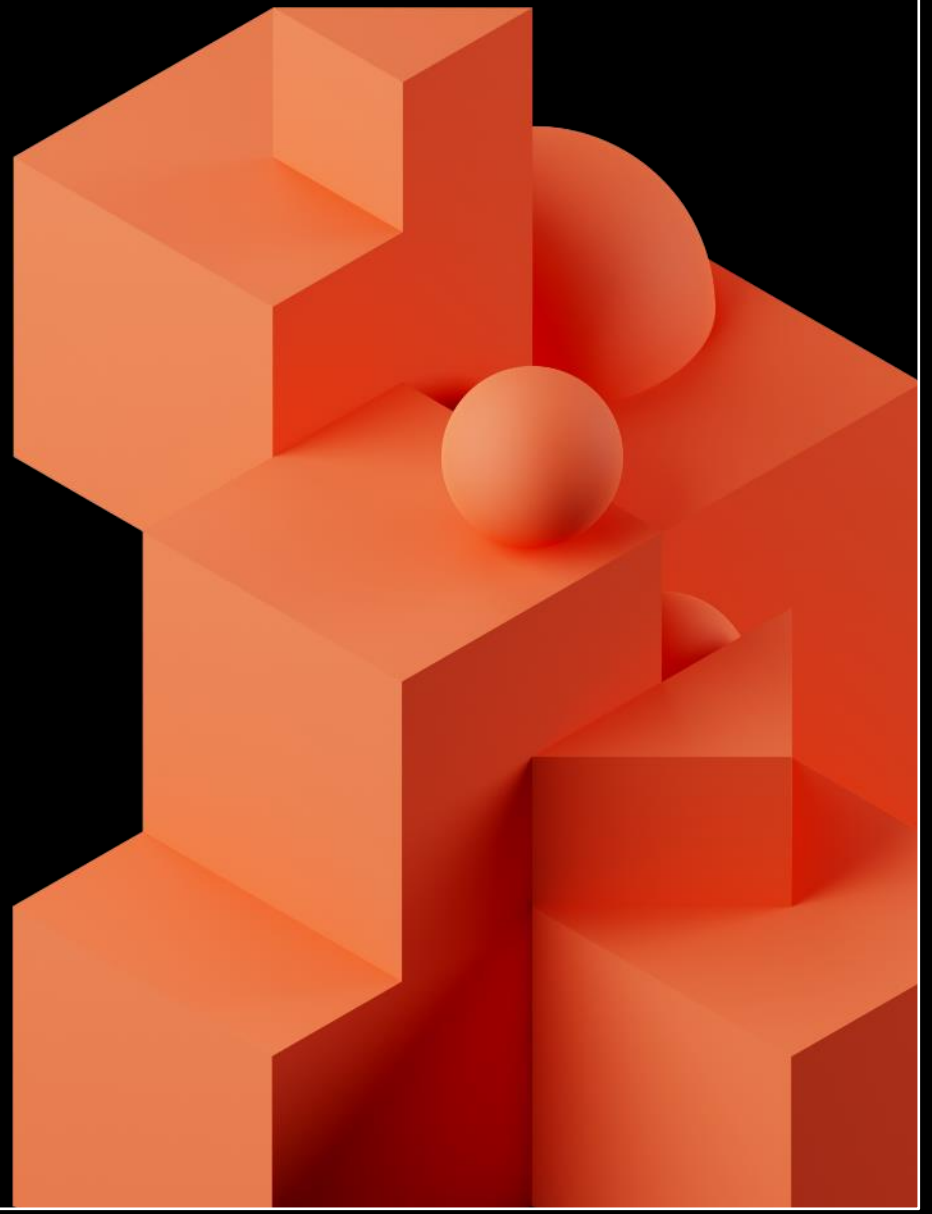
How the Lakehouse and Apache Spark™ Enable Collins Aerospace to Keep Aircraft Flying

Sanket Amin

Senior Manager, CAS Data Science and Analytics

Databricks 2023

Collins Aerospace



© 2023 Collins Aerospace | This document contains no export controlled technical data.

My Flight to a Bucket List Destination Was Delayed....

My Flight to a Bucket List Destination Was Delayed.... and delayed

My Flight to a Bucket List Destination Was Delayed.... and delayed, and delayed

My Flight to a Bucket List Destination Was Delayed....

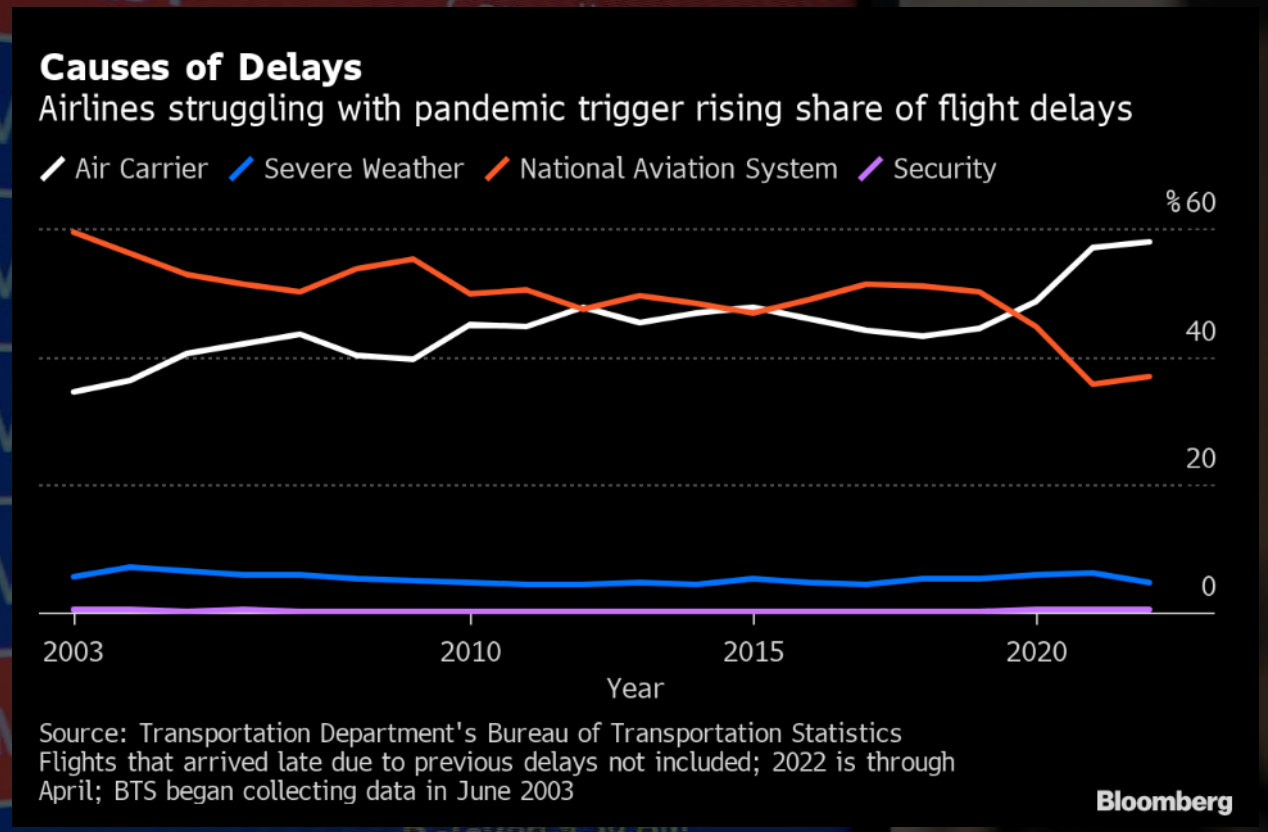
5 Times in a Single Day

Why did my Flight Get Delayed?

- Air Carrier Delay
- National Aviation System
- Extreme Weather
- Security Issue

Why did my Flight Get Delayed?

- Air Carrier Delay
- National Aviation System
- Extreme Weather
- Security Issue



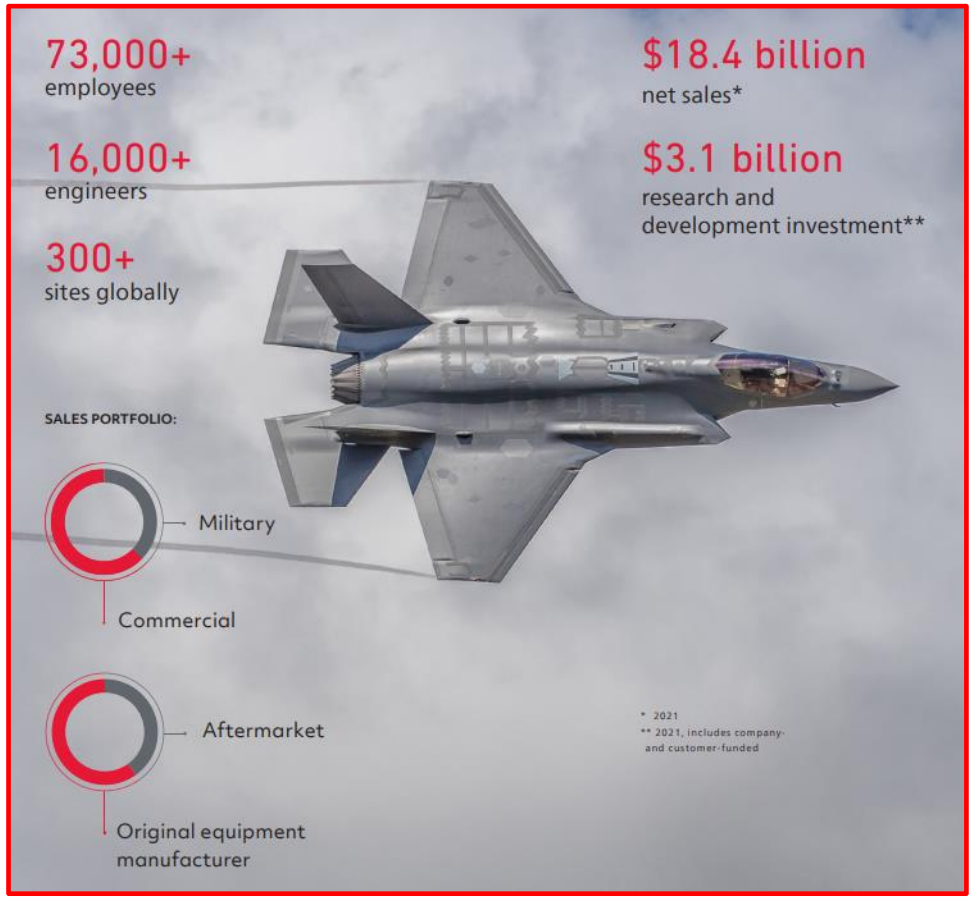
Cost to Large US Airline Operator

~$8M/Month

Who is Collins Aerospace?

RTX

By the numbers….



Look Left Look Right Look Up Look Down
Collins Aerospace is There

CAS.. is Leading the Digital Transformation in Aerospace

2000+ *employees in* **33** *countries*

22K *aircraft*

75M *messages / day**

250+ *airports around the globe*

5000+ *unique customers*

13M *Web & app users*

99.999 *network reliability*

Aviation connects the world. We connect aviation.

CONNECTED AVIATION SOLUTIONS

Based in Annapolis, Maryland

- Airport systems
- Applications, analytics & data products
- Business aviation flight support services
- Connectivity & network services

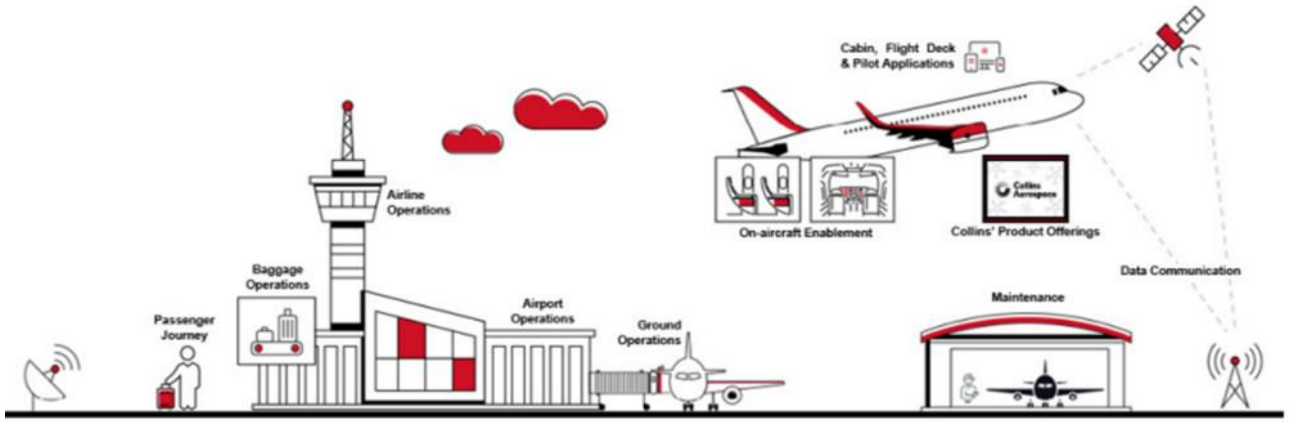
CAS is Leading the Digital Transformation in Aerospace

CONNECTED AVIATION SOLUTIONS

Based in Annapolis, Maryland

- Airport systems
- Applications, analytics & data products
- Business aviation flight support services
- Connectivity & network services

End to End Digital Technology Landscape



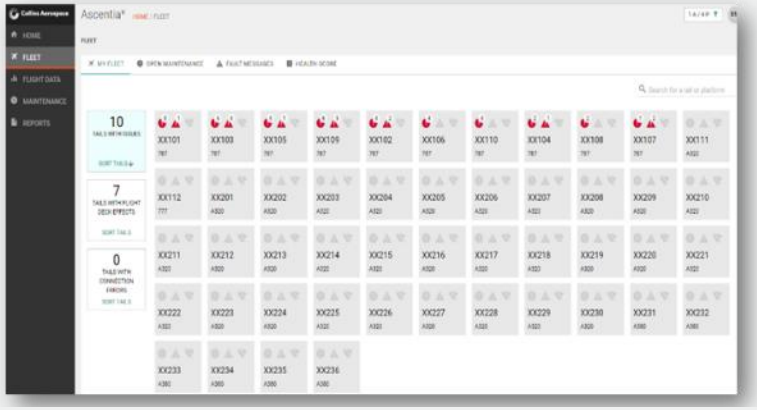
Aviation connects the world. We connect aviation.

Ascentia®

DATA DRIVEN. PREDICTIVE. PROACTIVE.

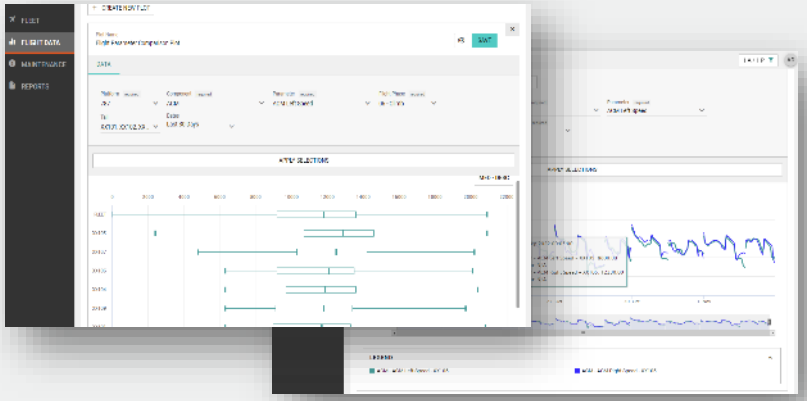
- Monitor fleet health and issues requiring attention
- Drill into issues by looking into specific parameters
- View Mx recommendations
- Set thresholds on specific parameters via custom alerts

Collins Aerospace



Fleet Overview

- Fleet health – Fault Codes, PHM Alerts
- Issues requiring attention
- Drill down to aircraft issue details

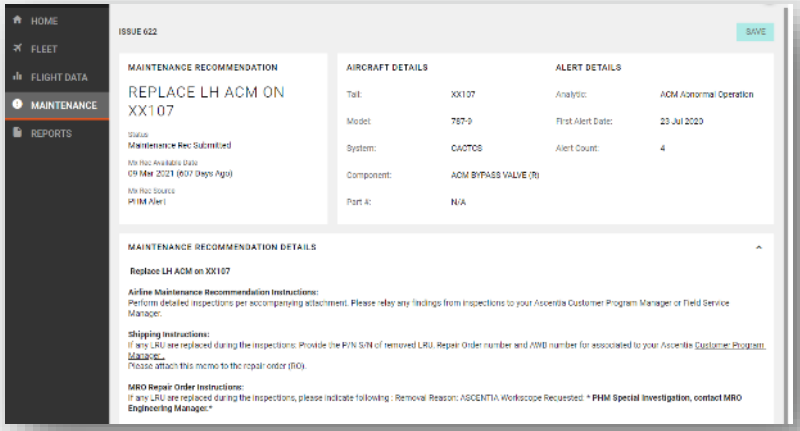


Data Visualization & Analysis

- Aircraft full flight & snapshot data plotting
- Maintenance & fault timelines
- Custom alerts

Fleet Health Score & Analytics

- Develop your own analytics
- Filter by flight phase and apply statistical aggregations
- Monitor your fleet



Maintenance Recommendations

- Aircraft issue documentation
- Prognostic notifications
- Issue life cycle tracking
- Direct communication with Ascentia team

© 2023 Collins Aerospace. | Collins Aerospace Proprietary. | This document does not include any export controlled technical data.

How We Learned To Develop and Deploy Predictive Analytics

The Early Days

Trying to fly with what we had.

Circa 2017- 2018

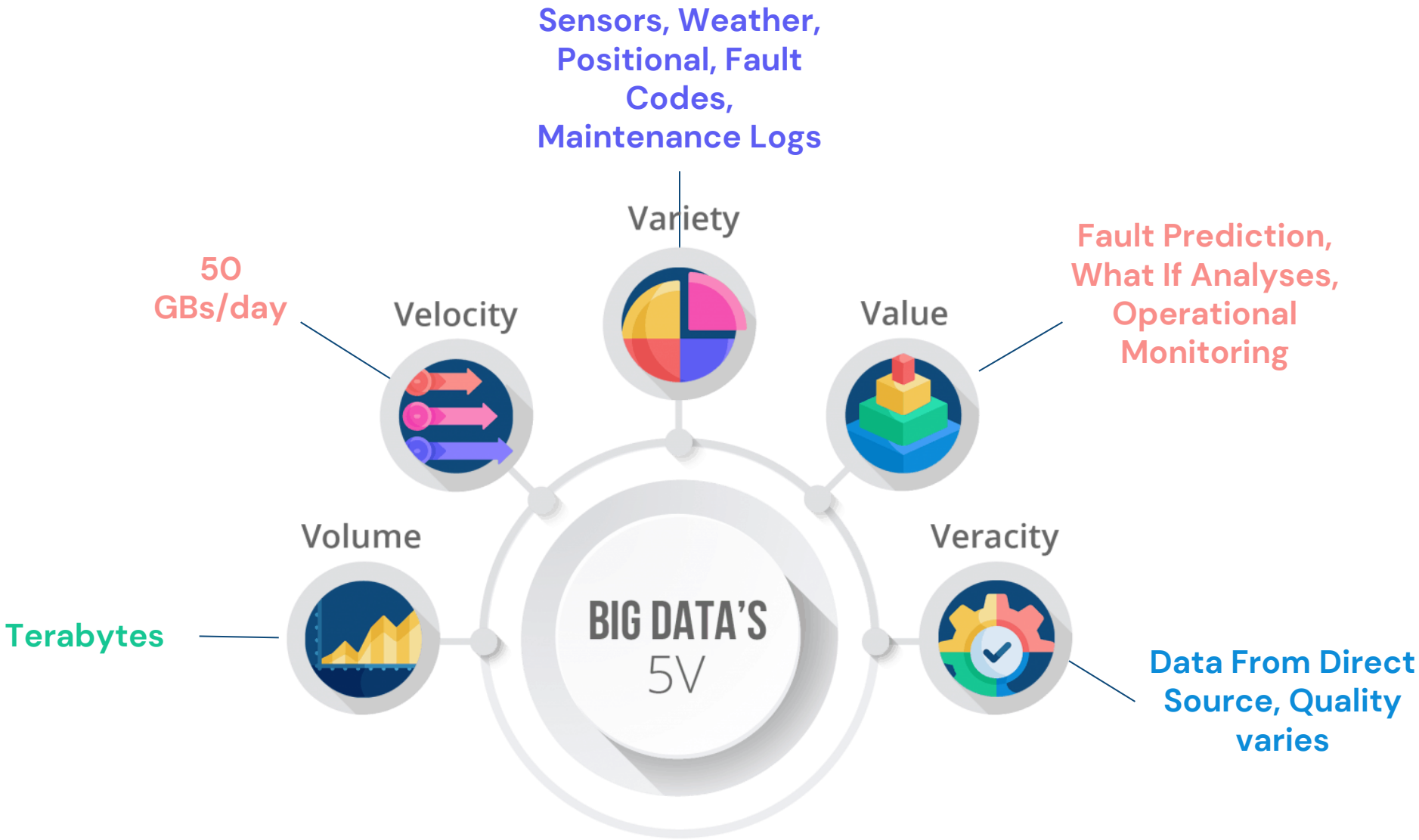
The Early Days

Trying to fly with what we had.

- Ingestion
- Data Preparation
- Analysis
- Presentation

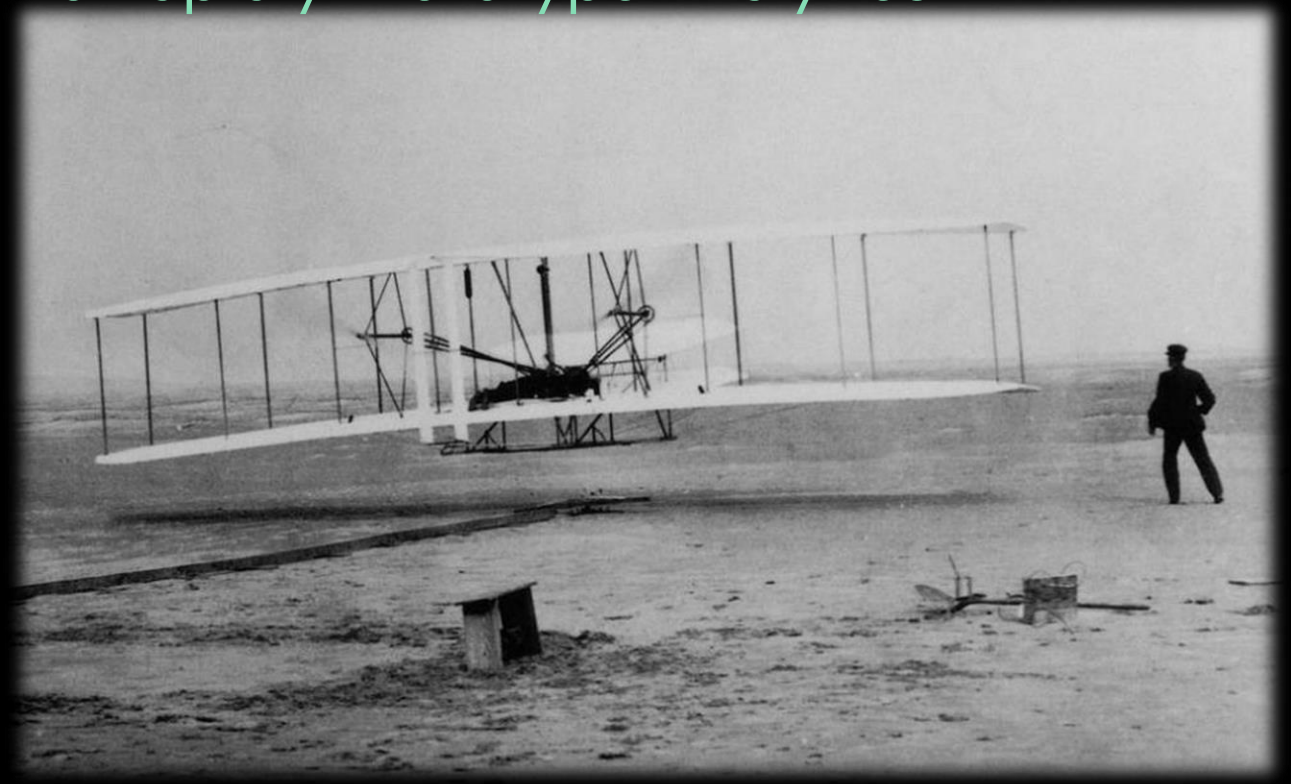
- Compute power limited to individual corporate issued laptops.
- 30 Mins to Process 1 Month of Data From Two Aircraft
- We needed Scalability for Big Data and Rapid Prototyping

Big Data Is In The Air



Our Wright Brothers Moment

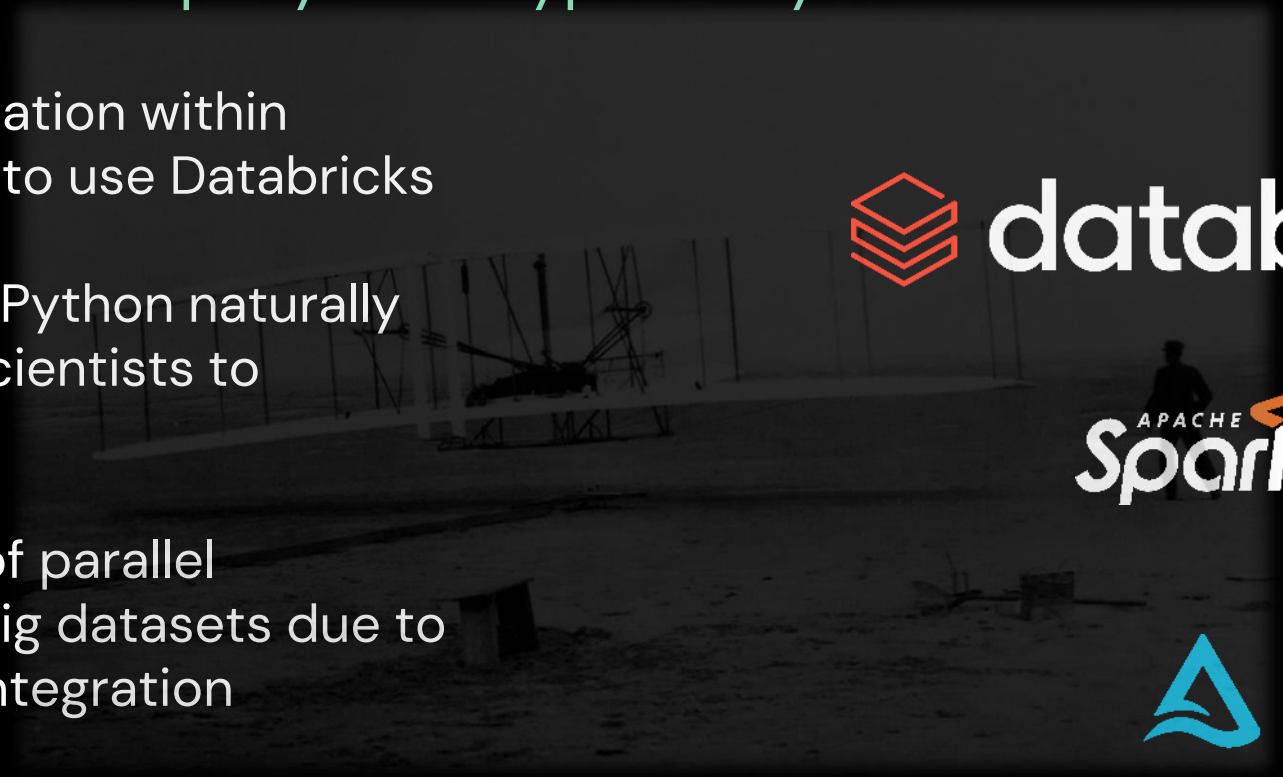
When We Began to Rapidly Prototype Analytics



Dec 2019

Our Wright Brothers Moment

When We Began to Rapidly Prototype Analytics

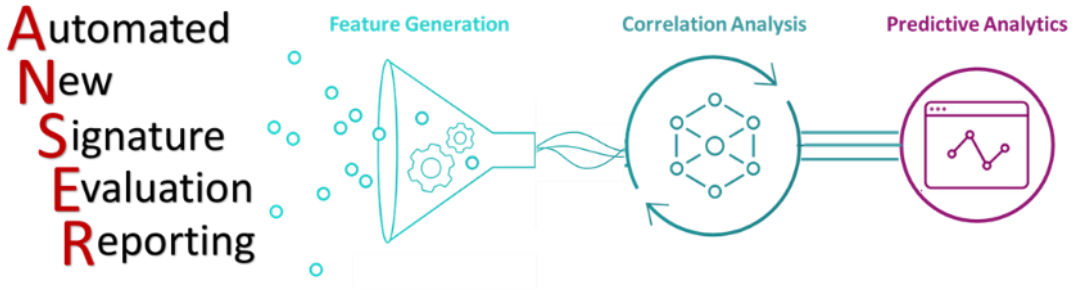


- The first organization within Collins and RTX to use Databricks
- Notebooks and Python naturally enabled Data Scientists to analyze data
- Easy adoption of parallel computing on big datasets due to Apache Spark integration
- Aircraft data stored in as delta tables in the Data Lake introduced additional efficiency

databricks

APACHE Spark™

We saved cost while going faster



- A home-grown scouting tool for finding potential analytics for Ascentia
- 168,000 Features Produced Per Flight

BEFORE

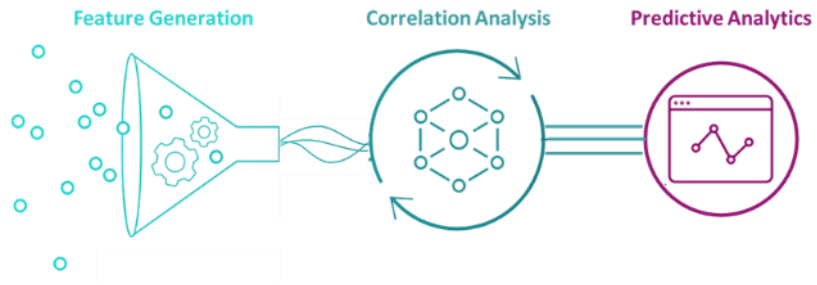
Cycles (tail + quarter): 3,741

Total Cost~ $13,929.30

Total Time~ 105.53 hr.

We saved cost while going faster

Automated
New
Signature
Evaluation
Reporting



- Leveraged Notebook widgets with Azure Data Factory for dynamic meta-data driven orchestration
- Implemented more efficient Spark processing techniques
- Optimization of delta tables.

BEFORE

Cycles (tail + quarter): 3,741

Total Cost~ $13,929.30

Total Time~ 105.53 hr.

Optimize

```
display(spark.sql(f"""

OPTIMIZE anser.dash_{customer}_{body}_signature_features_new_schema

"""))
```

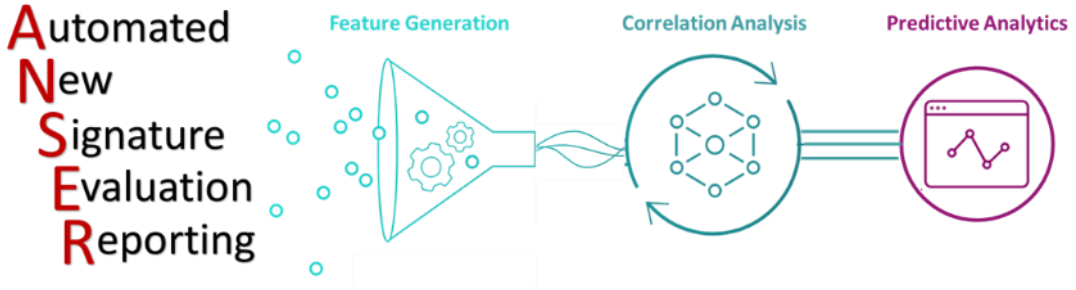
THEN Partition

```
(signature_features
  .write
  .mode("overwrite")
  .partitionBy('key', 'tailNumber')
  .option("overwriteSchema", "true")
  .saveAsTable(f"anser.dash_{customer}_{body}_signature_features_new_schema"))
```

Insight®

A Databricks Consulting Partner Company

We saved cost while going faster



- Leveraged Notebook widgets with Azure Data Factory for dynamic meta-data driven orchestration
- Implemented more efficient Spark processing techniques
- Optimization of delta tables.

BEFORE

Cycles (tail + quarter): 3,741
Total Cost~ $13,929.30
Total Time~ 105.53 hr.

AFTER

Cycles (tail only): 262
Total Cost~ $2,880.90
Total Time~ 21.83 hr.

Insight®

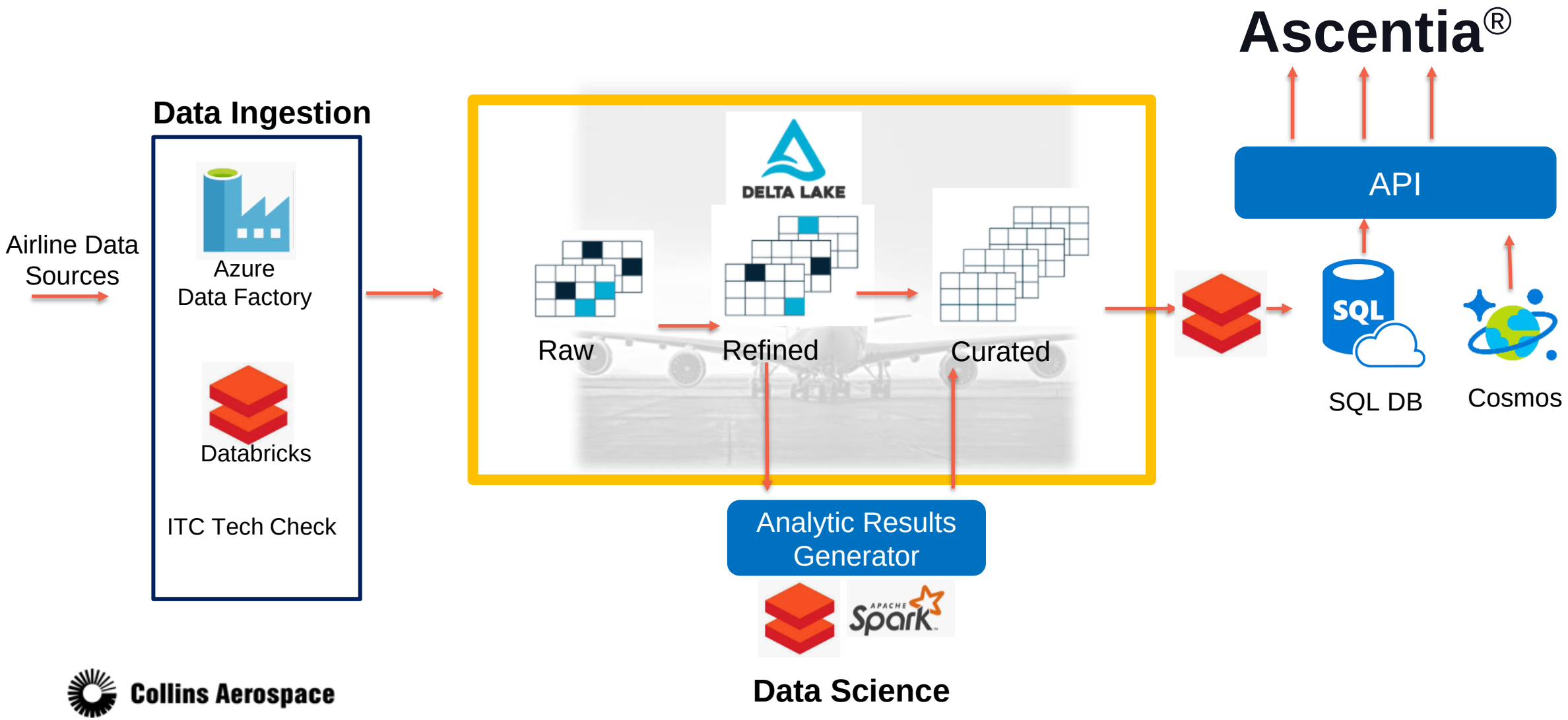
A Databricks Consulting Partner Company

The Jumbo Jet Era

When We Began to Rapidly Deploy Analytics

2022 - Current

The Jumbo Jet Era

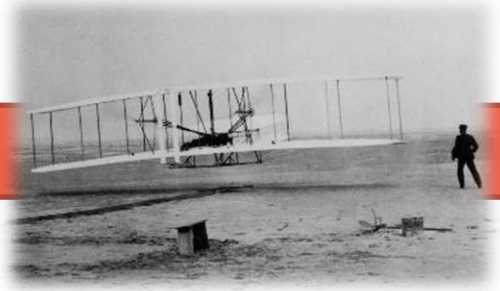


Collins Aerospace

Where We Departed vs Where We Are

- 2 Aircraft
- 7GB's of Data
- 0 Analytics

- 80 Aircraft
- 7.5 TB's of Data
- 12 Analytics



- 768 Aircraft
- 300 TB's of Data
- 55 Analytics

Ascentia®

DATA DRIVEN. PREDICTIVE. PROACTIVE.

~30%

decrease in potential delays

~20%

decrease in unplanned maintenance

Where We Are Headed

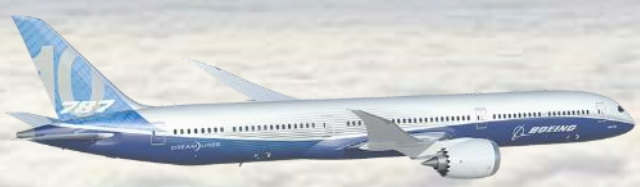
Continue adoption of platform solutions to reach new heights

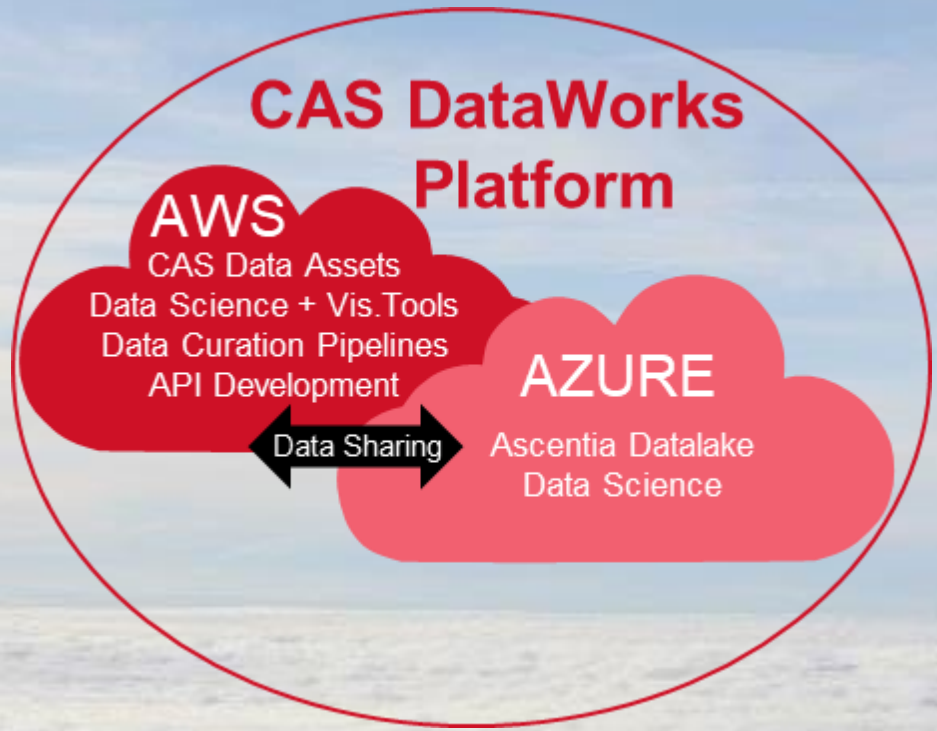
Where We Are Headed

Continue adoption of platform solutions to reach new heights

Multi-cloud Interoperability

- Data Sharing Across the Organization
- Increase opportunity for data science and analytics
- Promote Growth of Data Citizens





Where We Are Headed

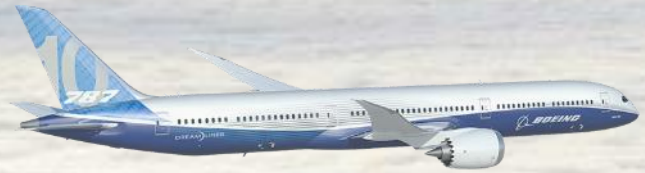
Continue adoption of platform solutions to reach new heights

Unity Catalog

CAS Data Governance

How does Governance bring value to the business and Collins?

Establish compliant, trusted, and quality data throughout the data ecosystem



Data Management

- Provide foundational data management support in the discovery across CAS with profiling, cataloging, enrichment & lineage of data sources

Data Privacy

- Support teams around privacy regulatory impacts and changing landscape of legislation

Data Rights

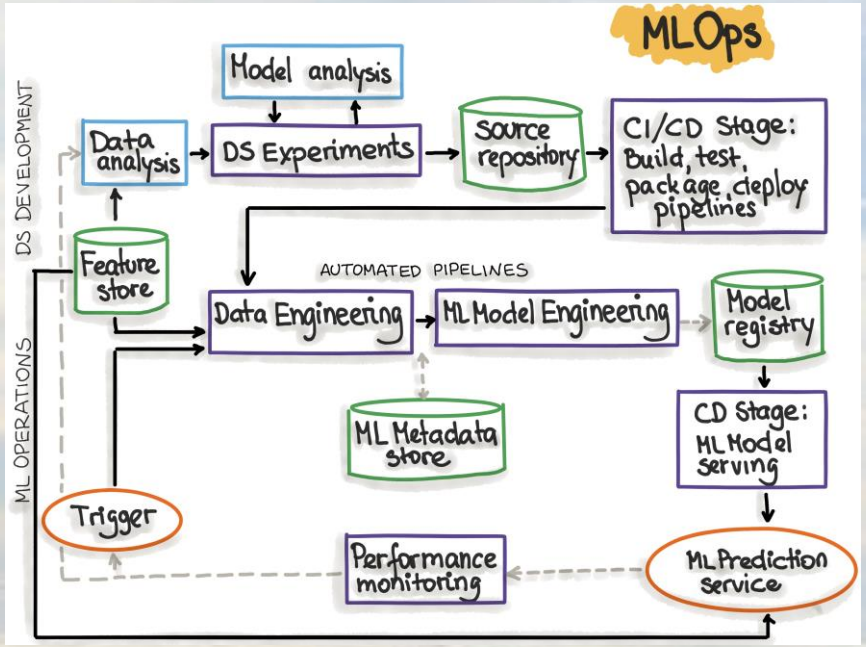
- Support business teams in discussions around access to data sets including how data can be shared/reused

Data Quality

- Collaborate with the business to ensure data is fit for purpose and establish the processes to measure and improve it

Where We Are Headed

Continue adoption of platform solutions to reach new heights



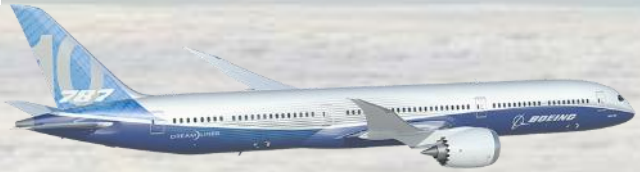
https://ml-ops.org/content/mlops-principles

Implement MLOps

Level 0 Everything is Manual!

Level 1 Automate the Model Development and Deployment

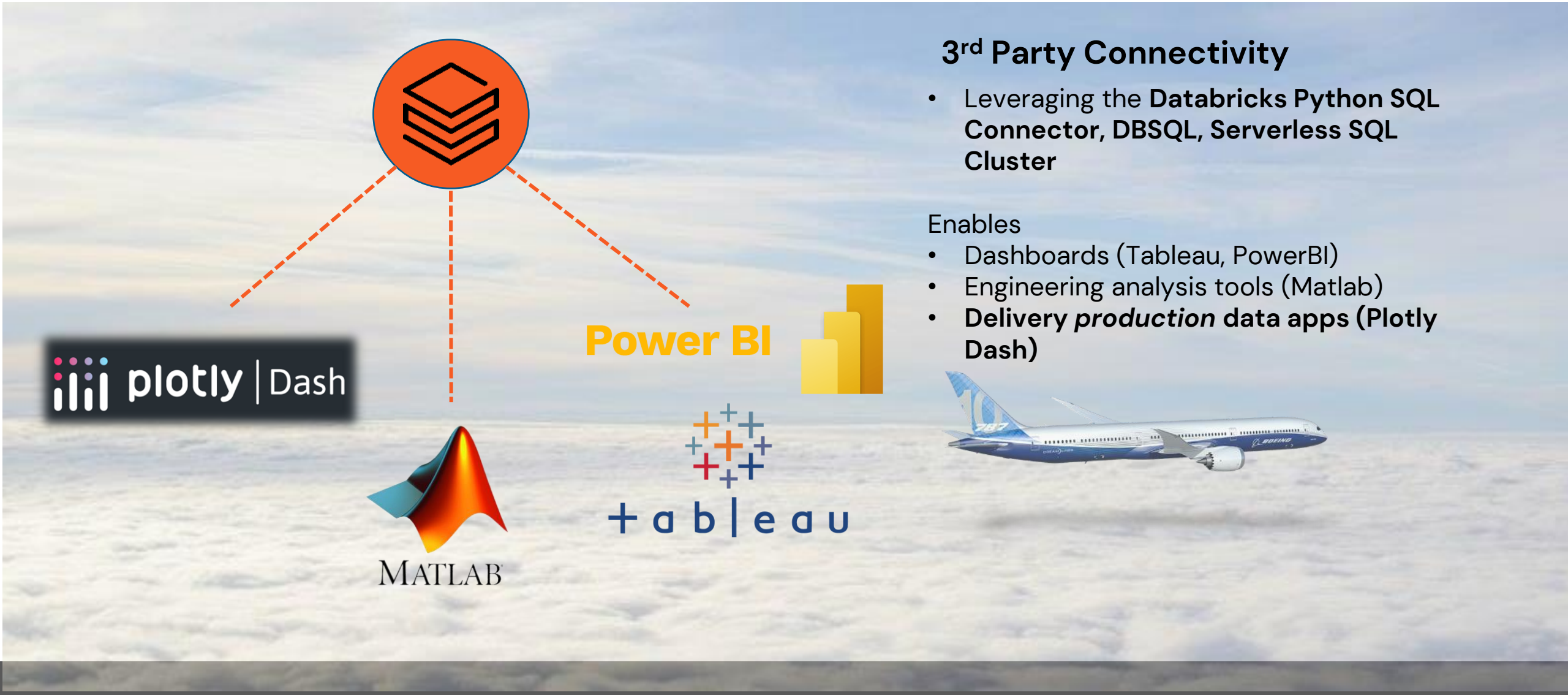
Level 2 Automate the Monitoring and Updating with Reinforcement Learning



Where We Are Headed

Continue adoption of platform solutions to reach new heights

Partner Connect



3rd Party Connectivity

- Leveraging the **Databricks Python SQL Connector, DBSQL, Serverless SQL Cluster**

Enables

- Dashboards (Tableau, PowerBI)
- Engineering analysis tools (Matlab)
- **Delivery *production* data apps (Plotly Dash)**

Where We Are Headed

Transforming Data into Value

Thank you