

# DATA+AI SUMMIT

BY  databricks

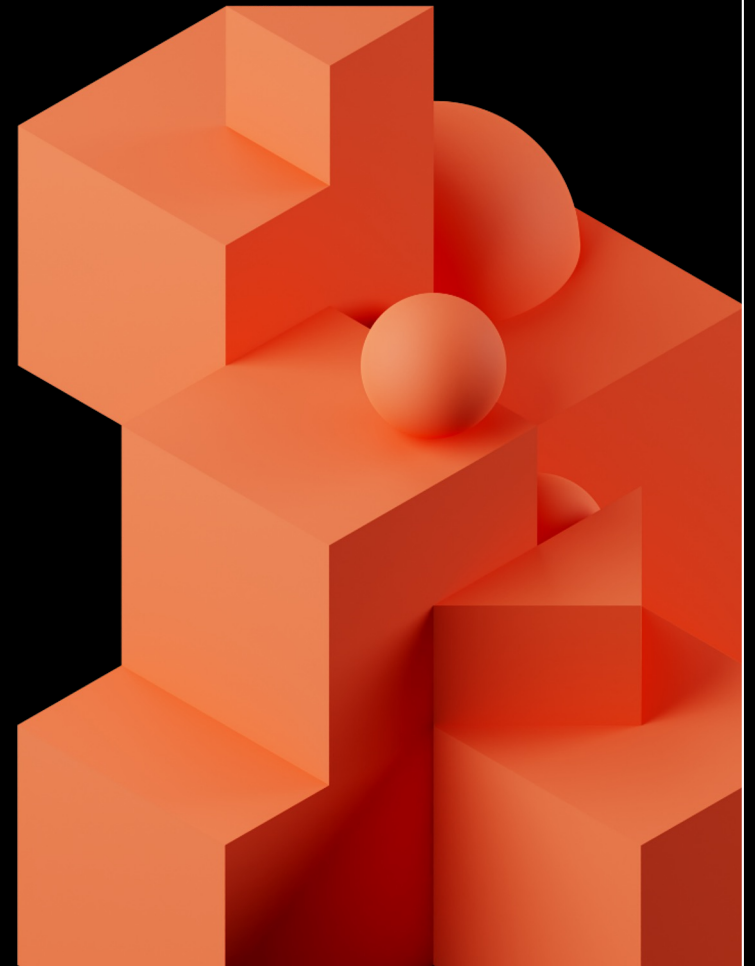
## JetBlue's real-time AI & ML digital twin journey using Databricks

Derrick Olson

Rob Bajra

---

Databricks  
2023

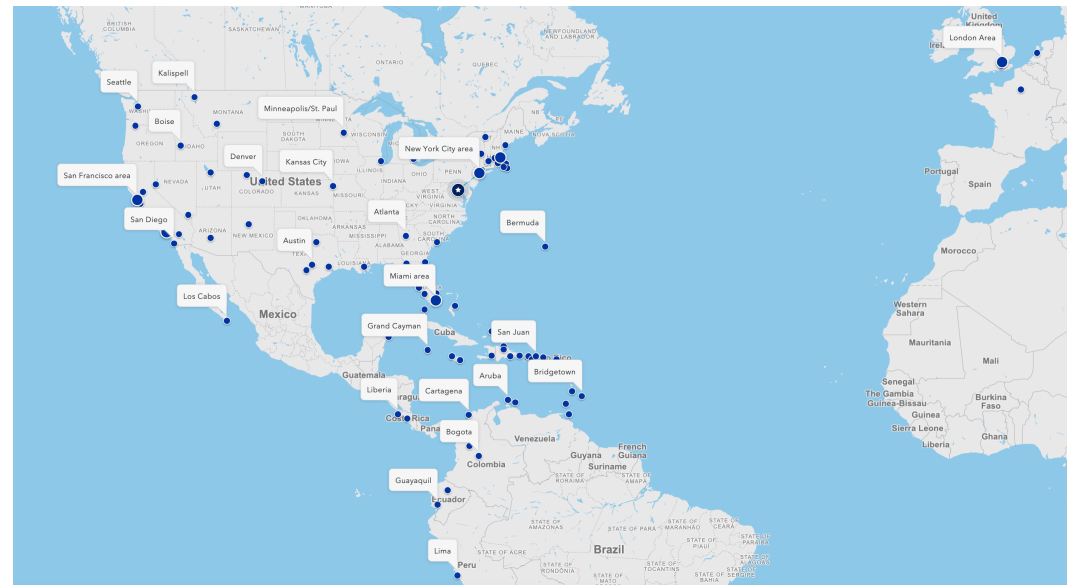


# Jet who?

## Bringing humanity back to air travel

- 100,000 Customers rely on us on more than 1,000 daily flights on 288 aircraft
- Operate in the most congested airspace in the world
- Disruptive low fares enabled by high aircraft utilization rate
- Limited margin of error for operations

JetBlue network



# Operational Challenges

## Decision making inherently reactive

### Situational awareness

- Limited situational awareness around delay drivers
- Difficulty looking into past conditions that led to current outcomes
- Many personas and roles

### Complexity of data

- Data is too large for humans to comprehend quickly
- Identifying known or potential risks
- High number of relationships between entities in network

### Cascading effects

- Any disruption can potentially have cascading effects which can cause further delays and cancellations
- It can take several hours to even days to recover



# JetBlue's journey

## Operational Digital twin to enable proactive decision making

- Lakehouse + ML platform + operational decision-making platform
- Near real time representation of JetBlue's network
- Flight phase models trained on ML Platform to predict operations
- Cascade logic and simulation

**2021**

- Tech stack Snowflake, dbt, Azure Data Factory
- Limited use of real time data
- Foundational machine learning models

**2022**

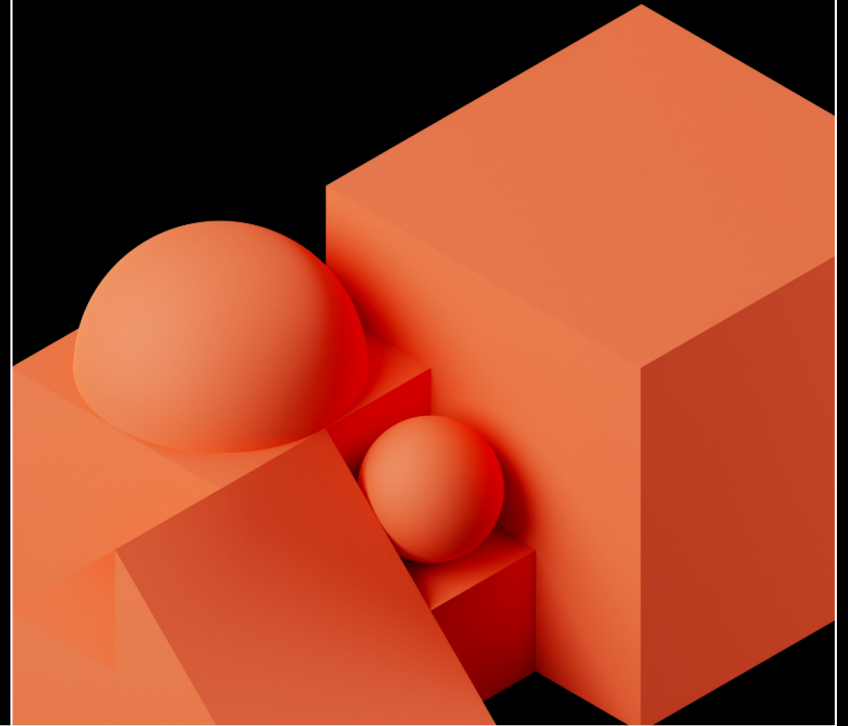
- ✓ Digital Twin kickoff
- ✓ BlueML integration
- ✓ Initial models and pipelines
- ✓ Use cases tested with users using BI front end

**2023**

- ✓ Real time data pipelines and inference at scale
- ✓ Data Lakehouse
- ✓ UI/UX for Situational awareness app
- ✓ Unity Catalog
- ✓ Simulator
- ✓ Chatbot

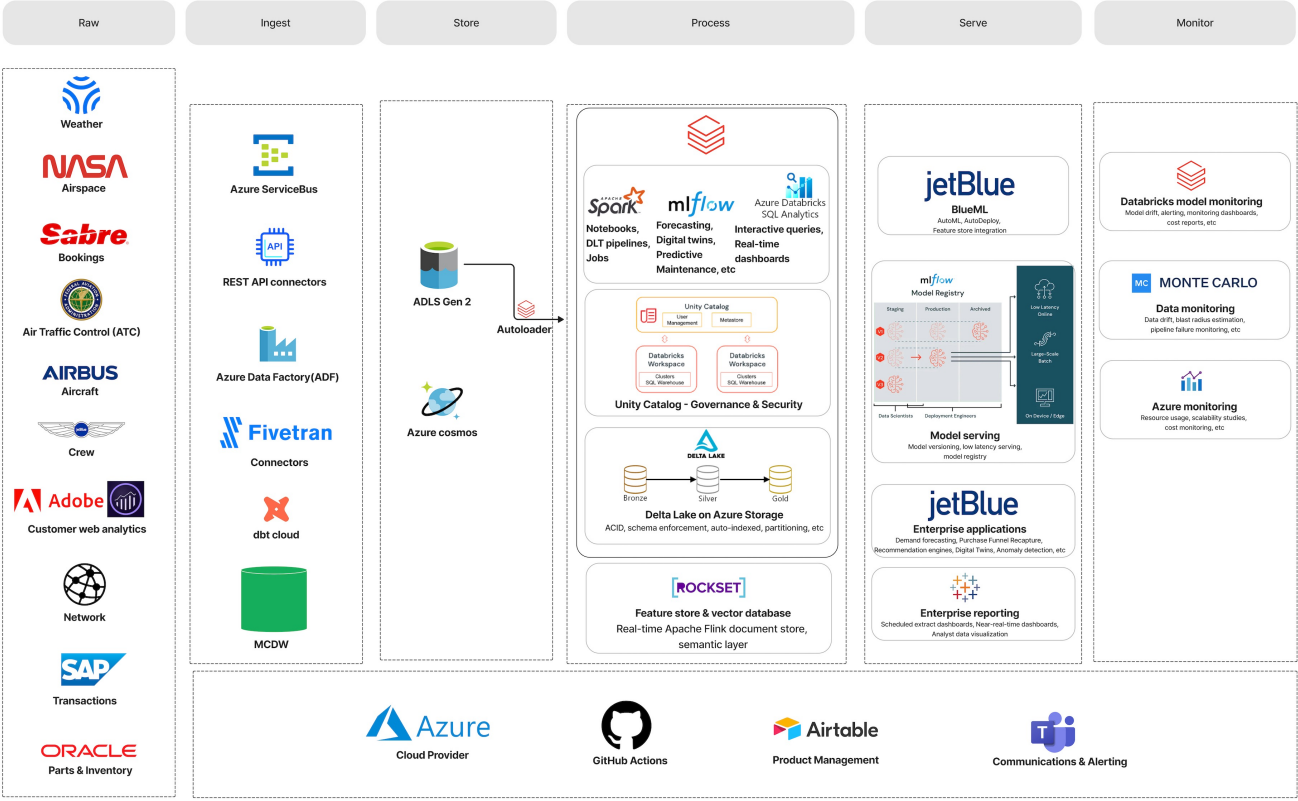


# Building blocks



# JetBlue Data Landscape

jetBlue Data, Analytics and Machine Learning Architecture



Databricks is a huge part of both our Lakehouse and ML platform

Other key technologies

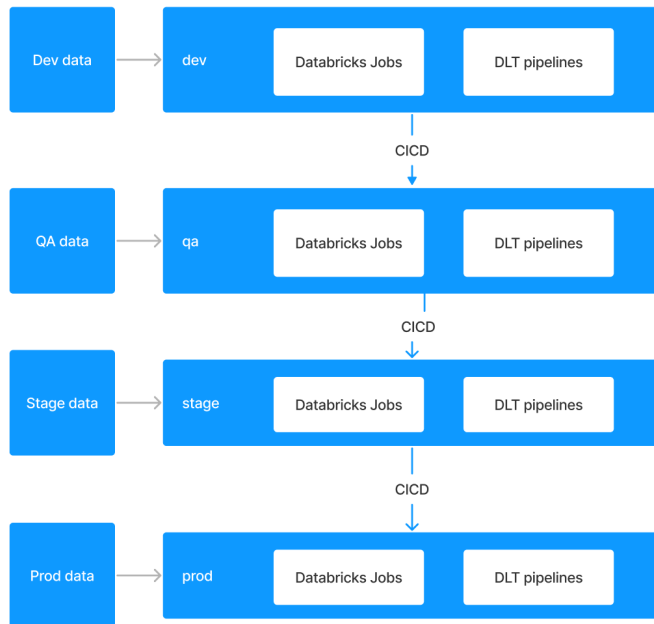
- **Rockset**
- Document store, online store, vector store
- **Tableau**
- Business Intelligence dashboards
- **Kubernetes**
- Application front end, model endpoints
- **Monte Carlo**
- Monitoring
- **GitHub Actions**
- CI/CD and version control



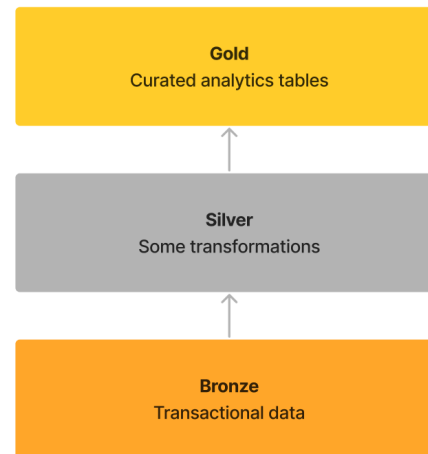
# Pipeline as code and medallion framework

Databricks APIs integrate with our CICD framework

## Separation of data by environment



## Medallion framework



## CICD

- Databricks APIs integrate with Github Actions workflows
- Checks are performed prior to promotion
  - Configuration checks
  - Workflow checks
  - Repo checks
  - Scripts/notebooks exist
  - Compute exists
  - Data elements exist
- Other checks (Linters, Unit tests, Integration tests)



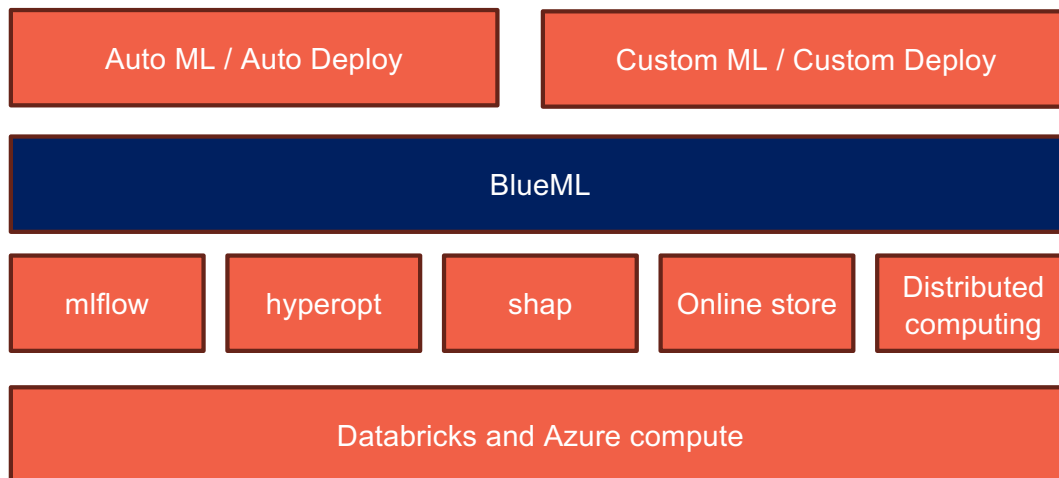
# More than 15 products are powered by ML

Team	# of ML powered products	Use cases
Commercial Data Science	6+	Pricing, forecasting, segmentation, A/B testing
Operations Data Science	2+	Optimization, real time prediction, chatbots
AI and ML engineering	2+	Chatbots, MLOps
Cyber Security	5	Threat detection, News Scraping, forecasting, security reviews
Revenue Management	2	Pricing, forecasting



# BlueML introduction

BlueML is JetBlue's in-house ML development and deployment platform



- BlueML is a machine learning platform made up of a common library and infrastructure configurations maintained by the data science teams at JetBlue
- It has enabled us to rapidly develop and iterate ML models
- BlueML also makes the deployment process easier



# Why BlueML?

BlueML has made the path to production faster and smoother

## Key features

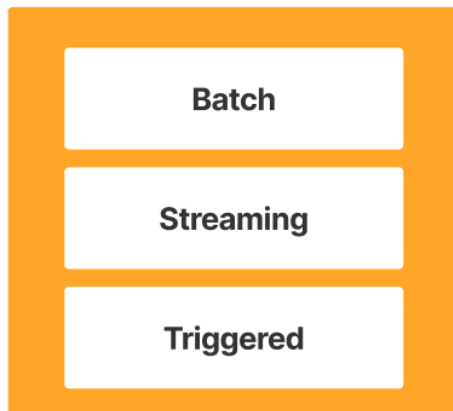
- ✓ 6 contributors
  - ✓ 3 teams
  - ✓ Streaming inference
  - ✓ Batch inference
  - ✓ Event based inference
  - ✓ HyperOpt
  - ✓ SHAP explainability
  - ✓ Pandas dataframes
  - ✓ Spark dataframes
- BlueML has made ML development faster
    - Allows Data Scientists to focus on core modeling and feature engineering iterations
    - Abstracts away boilerplate and complex code into simpler API
    - Model tuning, explainability modules
  - Standardization of ML pipelines
    - All pipelines can be deployed to existing Databricks and Azure infrastructure
    - The same model pipeline can be used to score batch, streaming data or be hosted as an endpoint
    - Modular code and integrates with JetBlue's infra and needs
  - Contributions from anyone at JetBlue!



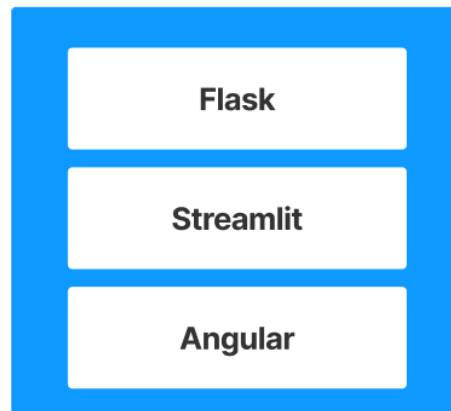
# Machine learning in production

JetBlue uses Databricks and Kubernetes for ML in production

## Databricks



## Azure Kubernetes



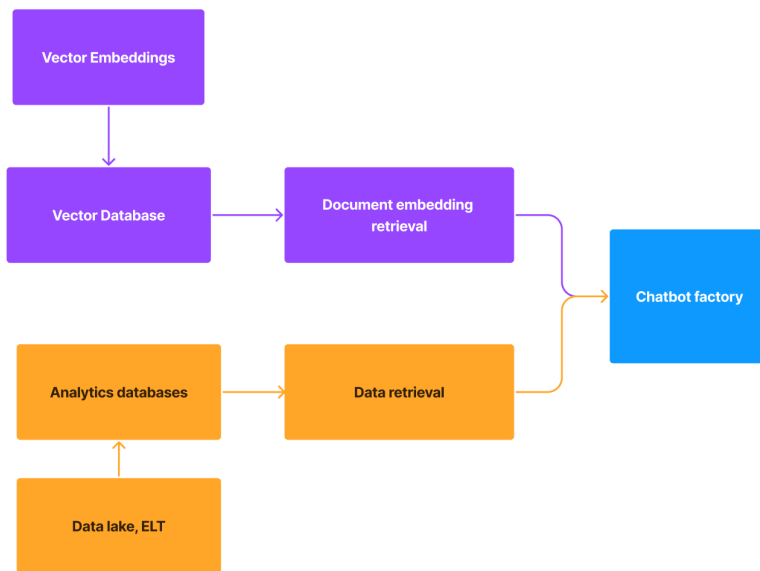
- JetBlue uses either Databricks or Azure Kubernetes to orchestrate models in production, depending on the use case
- Streamlit, Flask and Angular for front end



# Language models and chatbots at JetBlue

JetBlue is using Open AI and Dolly for internal use cases

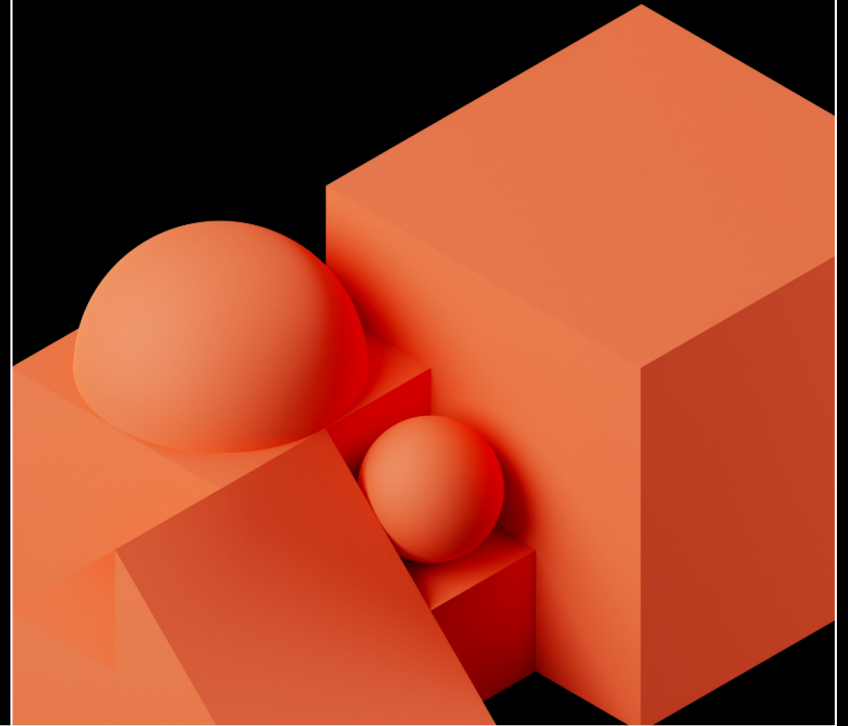
## High level use cases



- Language models have changed how users interact with data
- JetBlue has developed chatbots for 2 use cases - document retrieval and data retrieval
- We are using with Open AI and Dolly LLMs and experimenting with MosaicML
- Chatbot factory is a common set of tools, developed internally, that can be used to create multiple chatbots rapidly
- Working towards improving evaluation, context and LLMOps framework

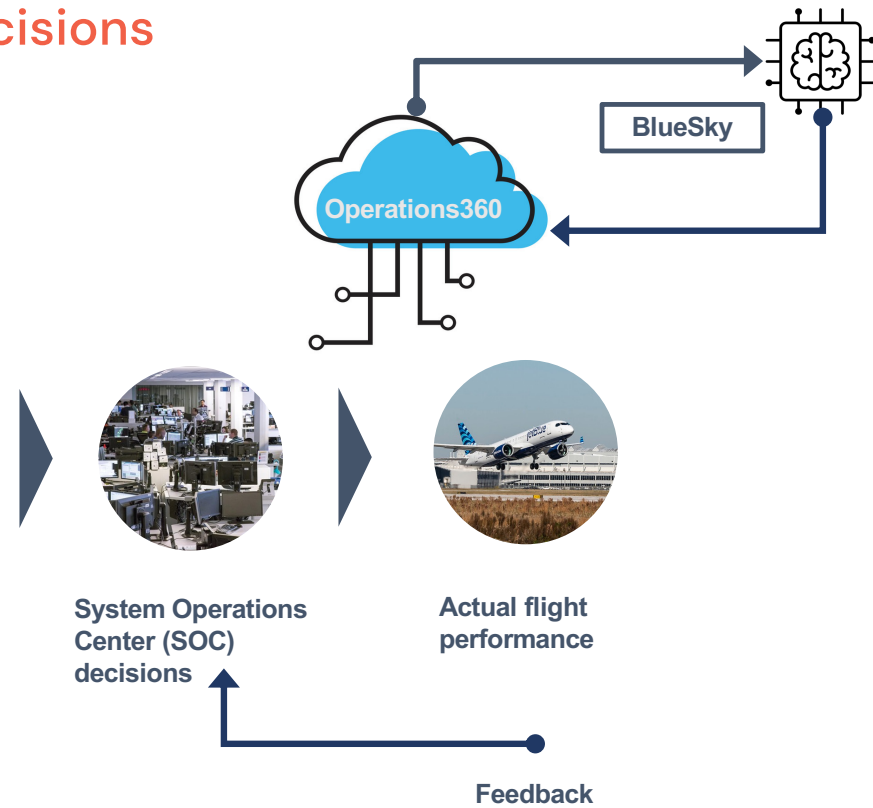
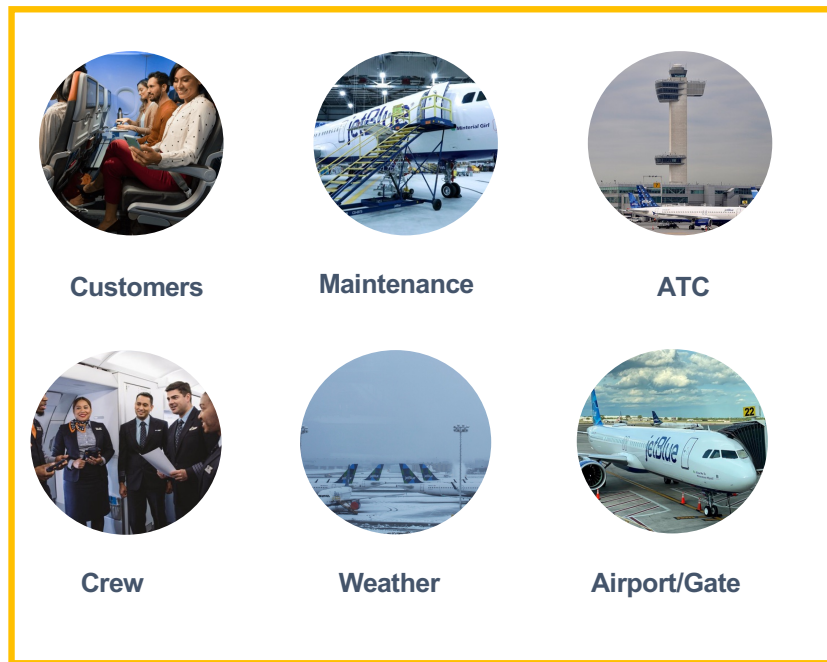


# BlueSky Digital Twin



# Datasets

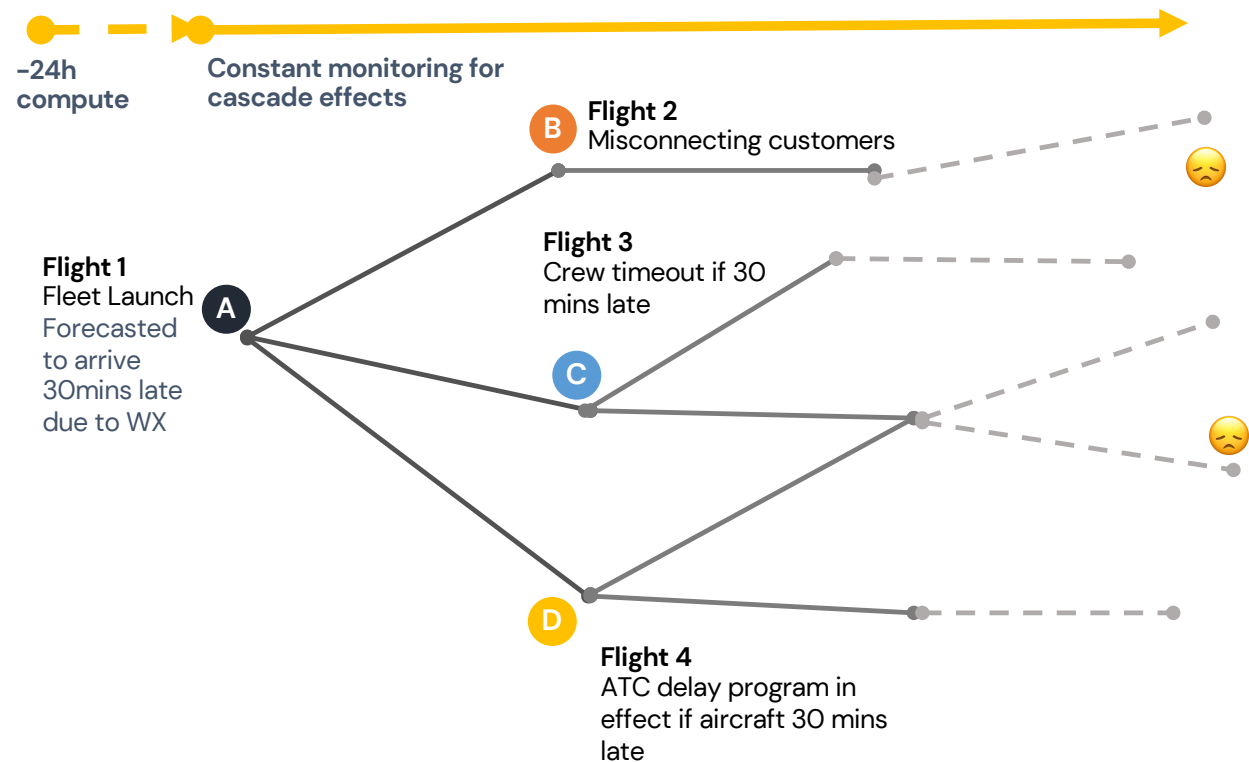
Real time data needed to make decisions



# Flight Network

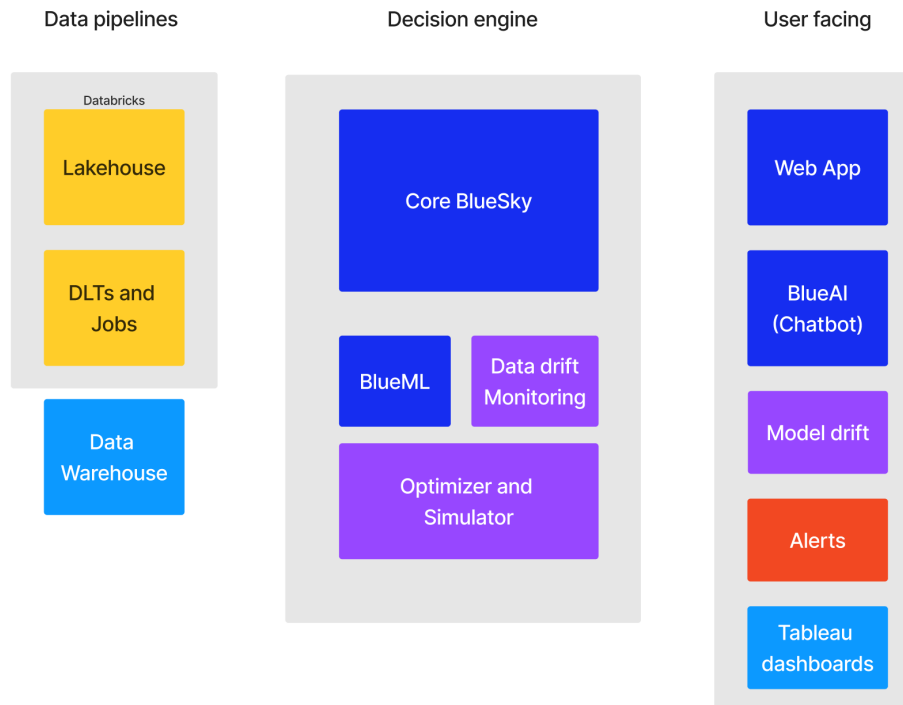
## Constant monitoring for cascading effects

- Flights in the system are connected via dependencies such as customers, crew and aircraft
- Forecasting the status of a flight requires taking into consideration its dependencies



# BlueSky in Databricks

From data ingestion to operational insight

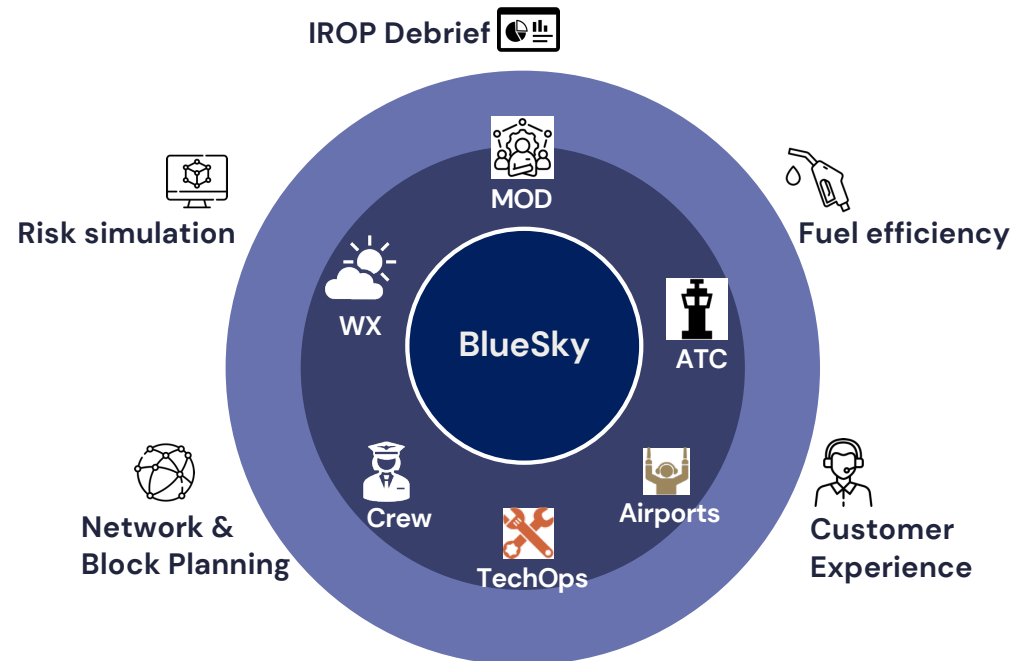




# Digital Twin Layers

BlueSky is the core part of JetBlue's operational digital twin

- BlueSky Ecosystem
- Immediate Use Case Applications
  - Solve day-of disruptions
- Push back decision window
  - Optimize planning



# BlueSky app

Intuitive platform for operational decision makers

**Current Conditions**

Select City

UTC 14:25 05-01-2023 Local 14:25 05-01-2023 Data 14:25 05-01-2023 Refreshed 14:25 05-01-2023

United States Map

- Chicago Windy 82° / 56°
- New York Foggy 82° / 56°
- Las Vegas Partly Sunny 74° / 57°
- Boston Sunny 82° / 56°
- Seattle Freezing Rain 82° / 61°
- Fort Lauderdale Fair 82° / 56°
- San Francisco Raining 80° / 58°
- Los Angeles Snowing 82° / 56°
- Washington Partly Sunny 79° / 53°
- Orlando Partly Sunny 82° / 56°
- New Orleans Sunny 74° / 53°
- San Juan Windy 82° / 56°

**Today's Metrics**

Day's Plan	Departure / Arrival	Taxi & Turn	Completion	Utilization
Scheduled Flights 985	FLDO 85.1%		Completed 389	Crew Utilization 70.9%
Load Factor 80.5%	DO 32.9%		Completion Factor 99.7%	Aircraft Utilization 40.9%
MEL's 30	AO 29.6%			Average Fuel Cost \$30.5
	A14 42.2%			

**Alerts**

On-Time Performance National Airspace Air Traffic Crew Airports TechOps Customers Weather

UTC 14:25 05-01-2023 Local 14:25 05-01-2023 Data 14:25 05-01-2023 Refreshed 14:25 05-01-2023

Advanced Filter

Estimated/Actual OUT (30 min interval)

Date	Tail	Flight	Alert	Decision Window	Flight Impacted	Crew Impacted	Passenger Impacted
02-20-2023	3424	29930	--	--	--	--	--
02-20-2023	2004	29944	--	--	--	--	--
02-20-2023	3992	29492	--	--	--	--	--
02-20-2023	3992	29492	--	--	--	--	--
02-20-2023	3992	29492	--	--	--	--	--
02-20-2023	3992	29492	--	--	--	--	--
02-20-2023	3992	29492	--	--	--	--	--



# BlueAI chatbot

BlueAI is an experimental chatbot for retrieving insights from BlueSky

The screenshot displays the BlueAI chatbot interface. At the top, there is a blue navigation bar with the 'jetBlue' logo and links for 'Home' and 'Documentation'. Below this, the title 'BlueAI' and subtitle 'BlueSky chatbot' are shown. The chat history consists of two messages:

- User:** What are the details of flight 1555?  
**Chatbot:** The details for flight 1555 on June 21, 2023 are as follows:
  - Flight Number: 1555
  - Local Departure Date: 2023-06-21
  - Departure Airport: John F. Kennedy International Airport (JFK)
  - Scheduled Departure Time: 07:00 AM EDT
  - Departure Terminal: 5
  - Arrival Airport: Los Angeles International Airport (LAX)
  - Scheduled Arrival Time: 10:22 AM PDT
  - Arrival Terminal: 5
  - Predicted Turn Variance: -8.0
  - Predicted Air Time: NonePlease note that the predicted air time is not available for this flight.  
Tool: database\_action
- User:** Why is flight 1555 delayed?  
**Chatbot:** Flight 1555 is delayed due to airspeed, ground control distance, and arrival total gates. Is there anything else I can help you with?  
Tool: delay\_reasons\_tool

At the bottom of the interface, there is a text input field with the placeholder 'Your message' and a 'Clear history' button to its right.



# Next steps

More data in the Lakehouse, more AI use cases

## Lakehouse

- Streaming Booking and Customer data ingestion
- Feature store integration
- More extensive data drift

## BlueML

- Distributed deep learning
- LLM support
- Reliability monitoring

## BlueSky

- Chatbot iterations
- Graph optimization
- Model champion/challenger
- Model drift



# We're hiring!

Solve fun complex problems with JetBlue!



We're hiring across Product Management, AI & ML engineering, Data Science, Data Engineering and more!

