

# HOW CAPITAL ONE ENABLED INNOVATION USING YBYD

---

Kajal Wood  
Jim Lebonitte



**Kajal Wood**  
Senior Director  
Capital One



**Jim Lebonitte**  
Sr. Distinguished Engineer  
Capital One

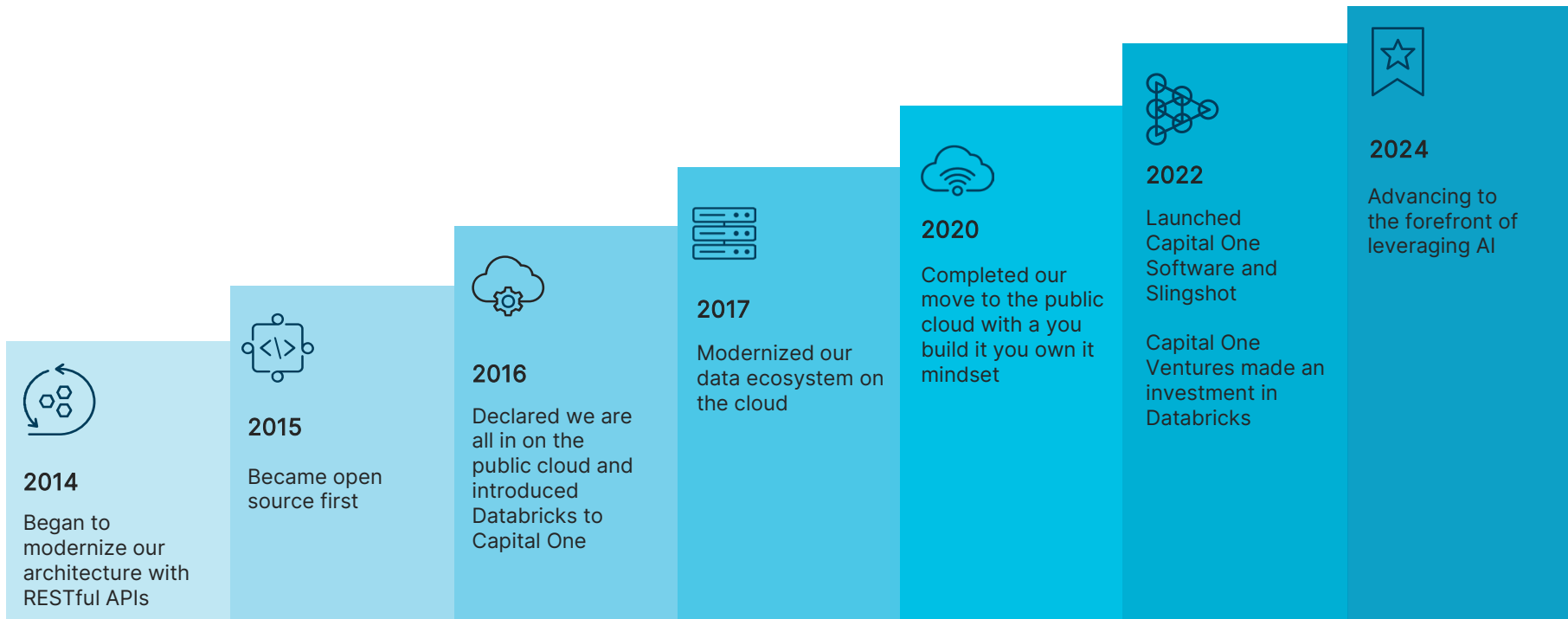
# AGENDA

- Capital One Tech Transformation
- “You Build, Your Data” Concept
- Unified Platforms
  - Self-Service Portal
  - Automated Deployment
- Deployment Models & Their Use Cases
- Key Takeaways

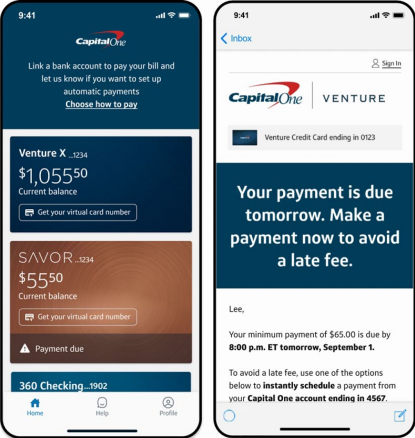
# DATA IS THE FOUNDATION OF CAPITAL ONE



# WE'RE A TECHNOLOGY-FIRST COMPANY, CHANGED FROM INSIDE OUT



# AI AND ML IS CENTRAL TO HOW DELIVER VALUE TO OUR CUSTOMERS



# ENABLING YOU BUILD IT YOU OWN IT IS KEY TO EMPOWERING OUR USERS



Governance  
& Standards



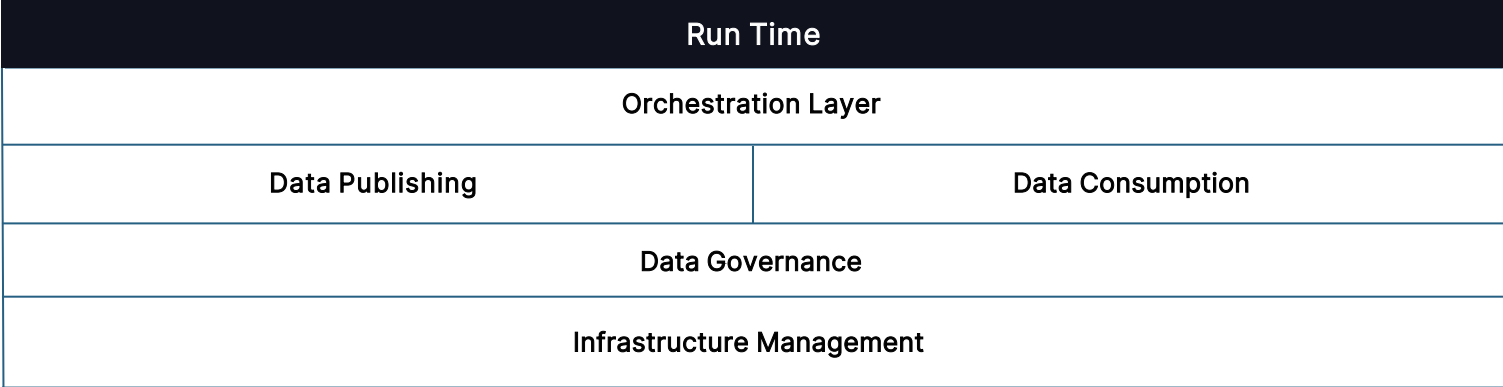
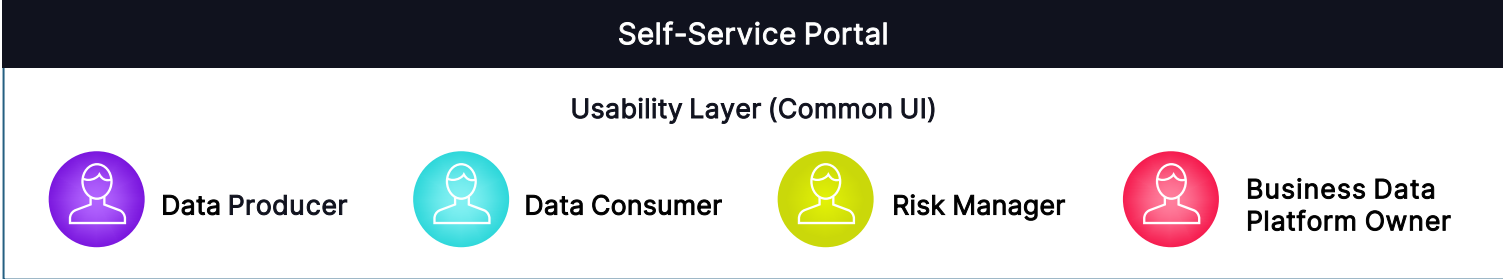
Unified, Self-Service  
Platforms



Empowered  
Users



# CREATING UNIFIED PLATFORMS ARE IMPORTANT TO ENABLE SCALE





# THESE PLATFORMS CONSIST OF 2 TYPES OF CAPABILITIES

## Self-Service Portal

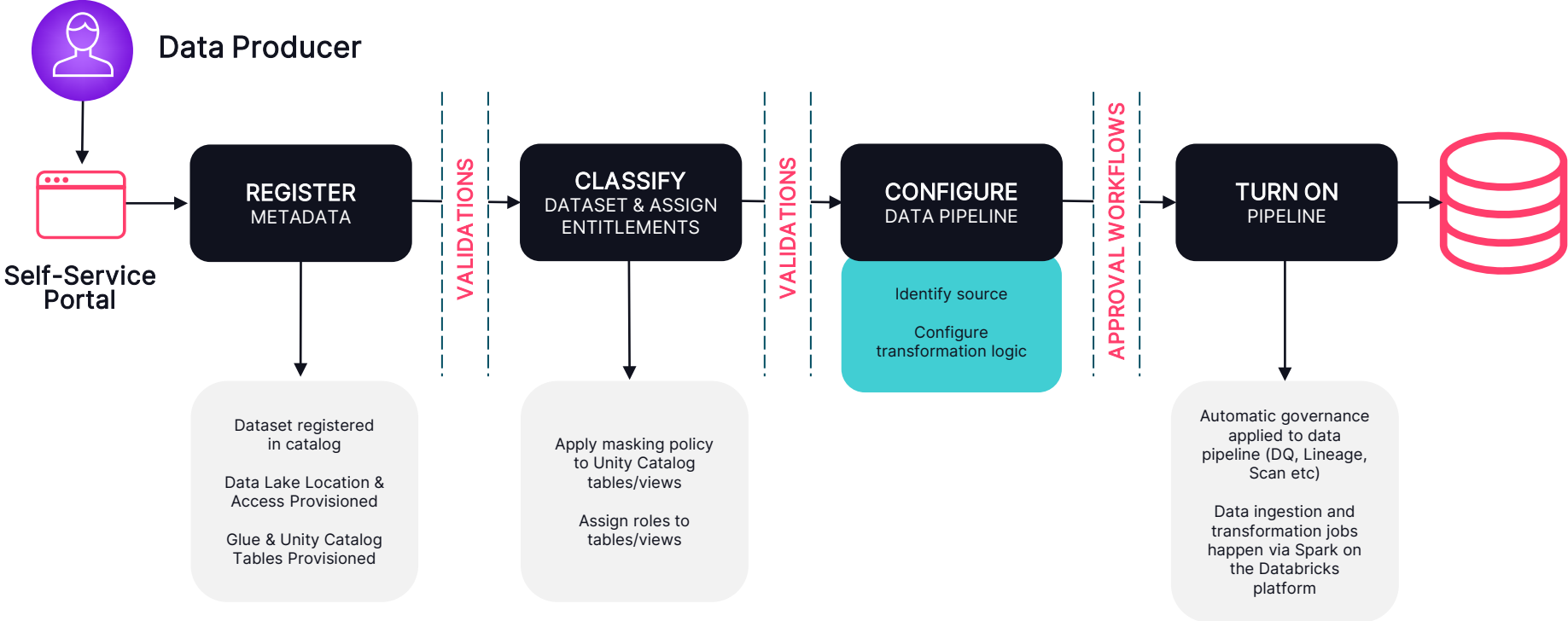
- User Interface & Experience
- Data Modeling
- Metadata Catalog
- Pipeline Design
- Data Set Deployment
- Access Approvals

## Run Time Components

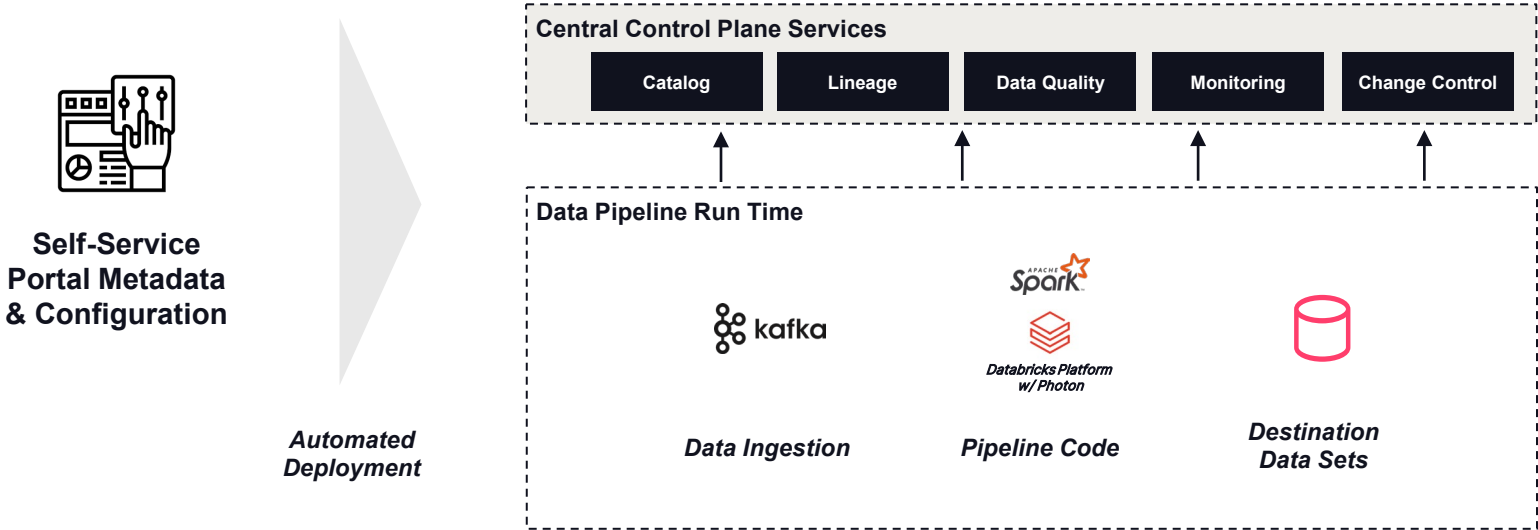
- Operational Catalogs
- Storage Locations
- Central Pipelines
- Federated Pipelines
- Tokenization
- Data Management Enforcement
- Fine Grained Access Control



# OUR SELF-SERVICE PORTAL IS THE FRONT DOOR TO DATA



# CONFIGURED ASSETS ARE AUTOMATICALLY DEPLOYED INTO AWS



# WE HAVE CREATED 2 DEPLOYMENT OPTIONS DEPENDING ON THE NEED

## Centrally Managed

### Pros

- Standardizes the approach of a tech solution for many use cases
- Easier for analysts and data scientists to leverage
- Lower overall operations / management costs

### Cons

- Larger Blast Radius
- Hard To Customize For All use cases

## Locally Managed

### Pros

- Smaller Blast Radius
- Easier to customize for your use case
- Easier for data / ml engineering teams to leverage

### Cons

- Requires tech talent to manage
- Higher operations / management costs



# THESE MODELS ENABLE DIFFERENT PERSONAS TO DEPLOY DATA APPLICATIONS

Software Engineer



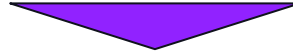
Centrally managed  
stream ingestion  
applications

Data Engineer



Locally managed  
spark data pipelines  
with custom  
transformations

Data Analyst &  
Scientist



Centrally managed  
sandbox spaces for  
data preparation and  
model build & train

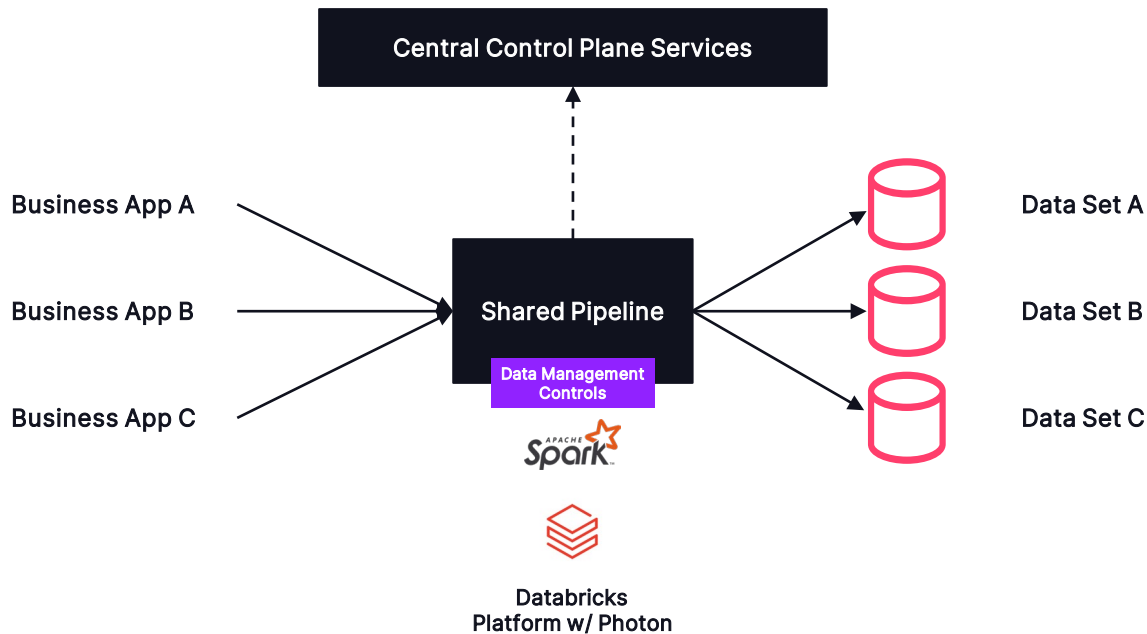
ML Engineer



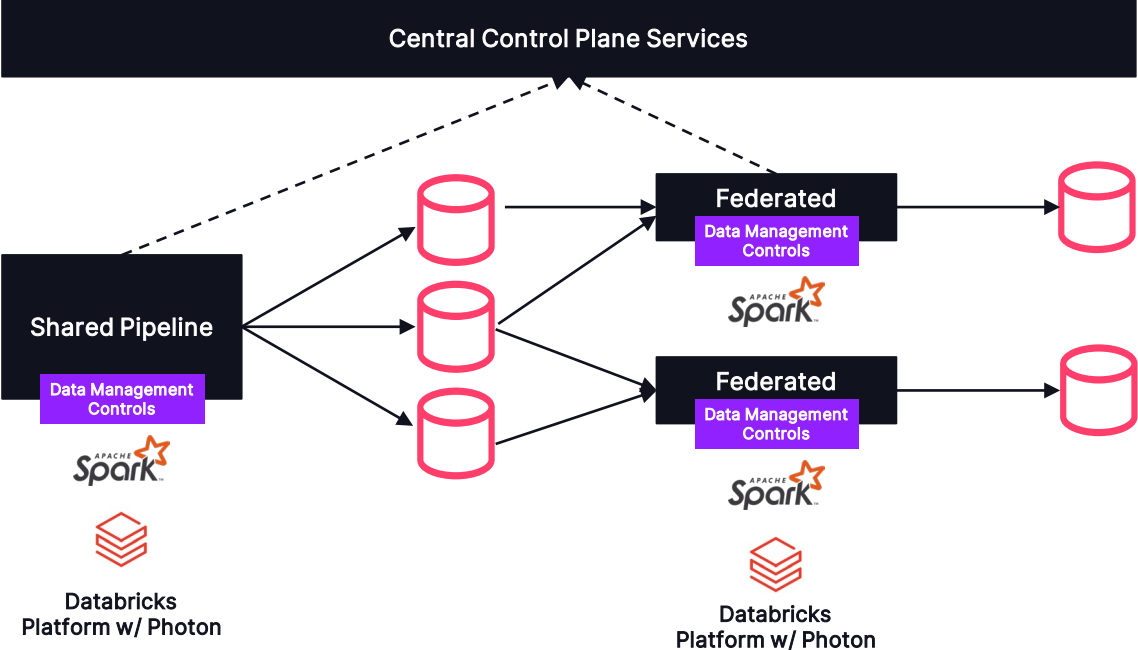
Centrally managed  
low latency data  
availability via APIs  
used for ML inference



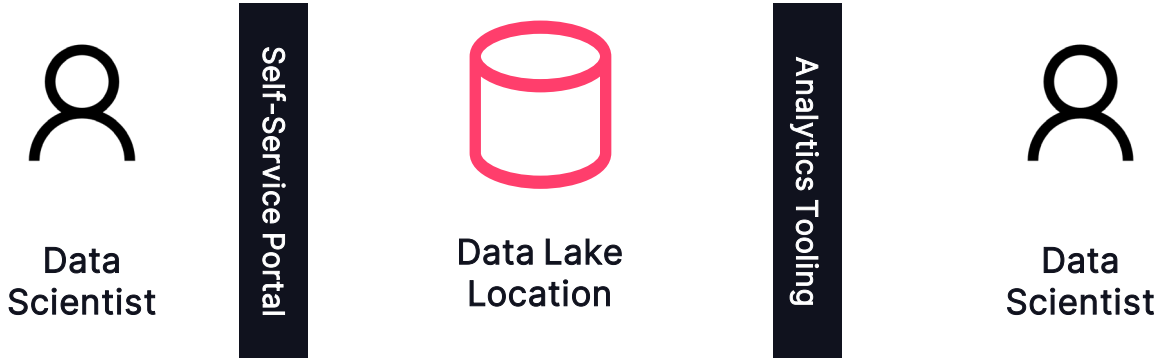
# WE HAVE ENABLED OUR BUSINESS APPS TO EASILY STREAM DATA FOR SHARING



# WE HAVE ALSO CREATED DATA PIPELINE CAPABILITIES FOR OUR ENGINEERS



# SELF SERVICE DATA SANDBOXES ENABLE RAPID MODEL BUILD AND TRAIN



Users use the self-service portal to provision a storage location

Users are able to utilize their preferred tooling to access the location and perform their day to day work





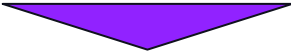
# CAPABILITIES FOR ML ENGINEERS TO DEPLOY LOW LATENCY DATA STORES



Data Lake Location



Low Latency Data Store



Once the model is ready for production our data scientists can leverage our self service portal to make them available for real-time inference



# KEY TAKEAWAYS

Enable a self-service portal and managed run time experience

Build experiences tailored to users

Automate EVERYTHING



# THANK YOU

