Accelerating your Machine Learning Projects with

Rodney Joyce

Practice Director @ Data-Driven AI

LinkedIn: bit.ly/rodneyjoyce

www.data-driven.ai

data-driven ai

Thanks to our Sponsors:

Agenda

- AWS versus Azure
- The Modern Data Platform
- Comparing ML Projects to Software Engineering
- *Before* Architecture
- *After* Architecture with Databricks
- Where to host your models
- Summary & Questions

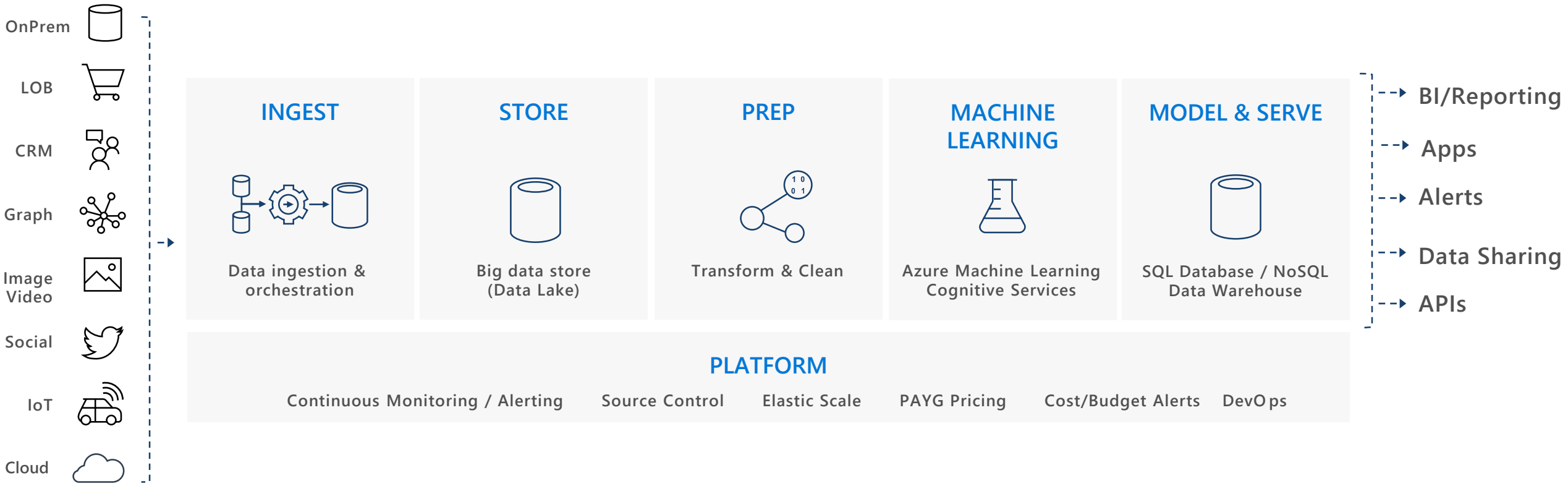
ML Concepts are the same in AWS

- Databricks
- Azure Machine Learning
- Data Factory
- Blob Storage
- KeyVault
- Azure DevOps

=

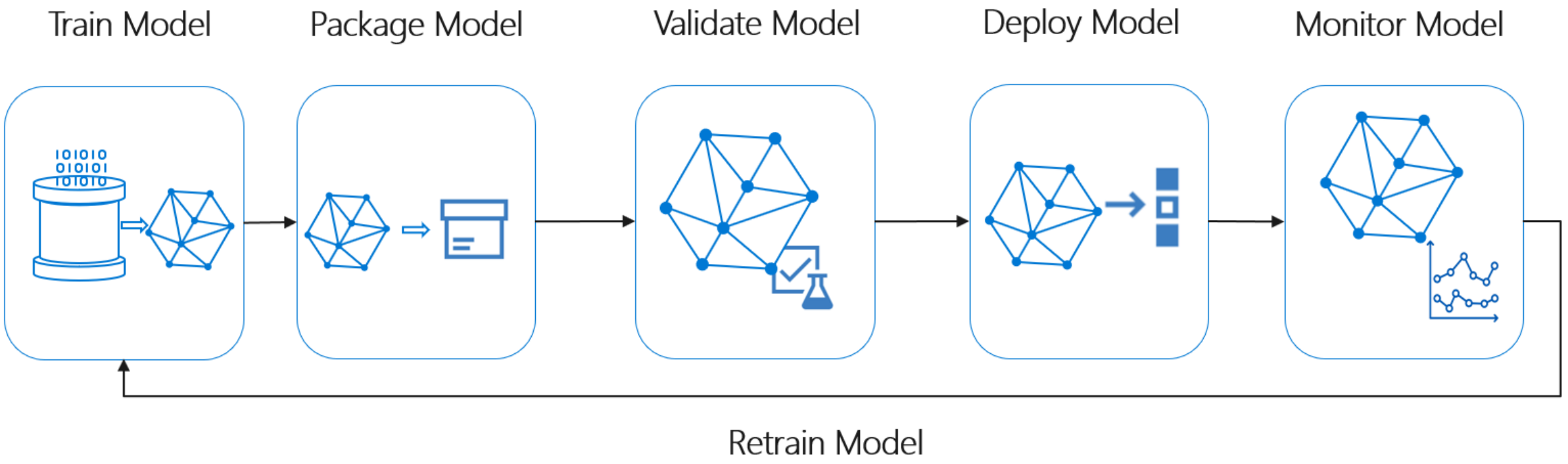
- Databricks
- AWS Sage Maker
- Glue
- S3
- Secrets Manager
- CloudFormation

The Modern Data Platform



Machine Learning enables us to do Predictive Analytics

Difference - Workflow to create Model artifact



- Data Scientist
- Exploratory
- Data Engineer
- Pickle/Onnx
- Data Scientist
- Compare Metrics
- Data Engineer
- Batch or Realtime
- Data Engineer via MLOps monitoring

An ML model is just a function like any other service. It takes some input and gives back a result…

Differences - Roles

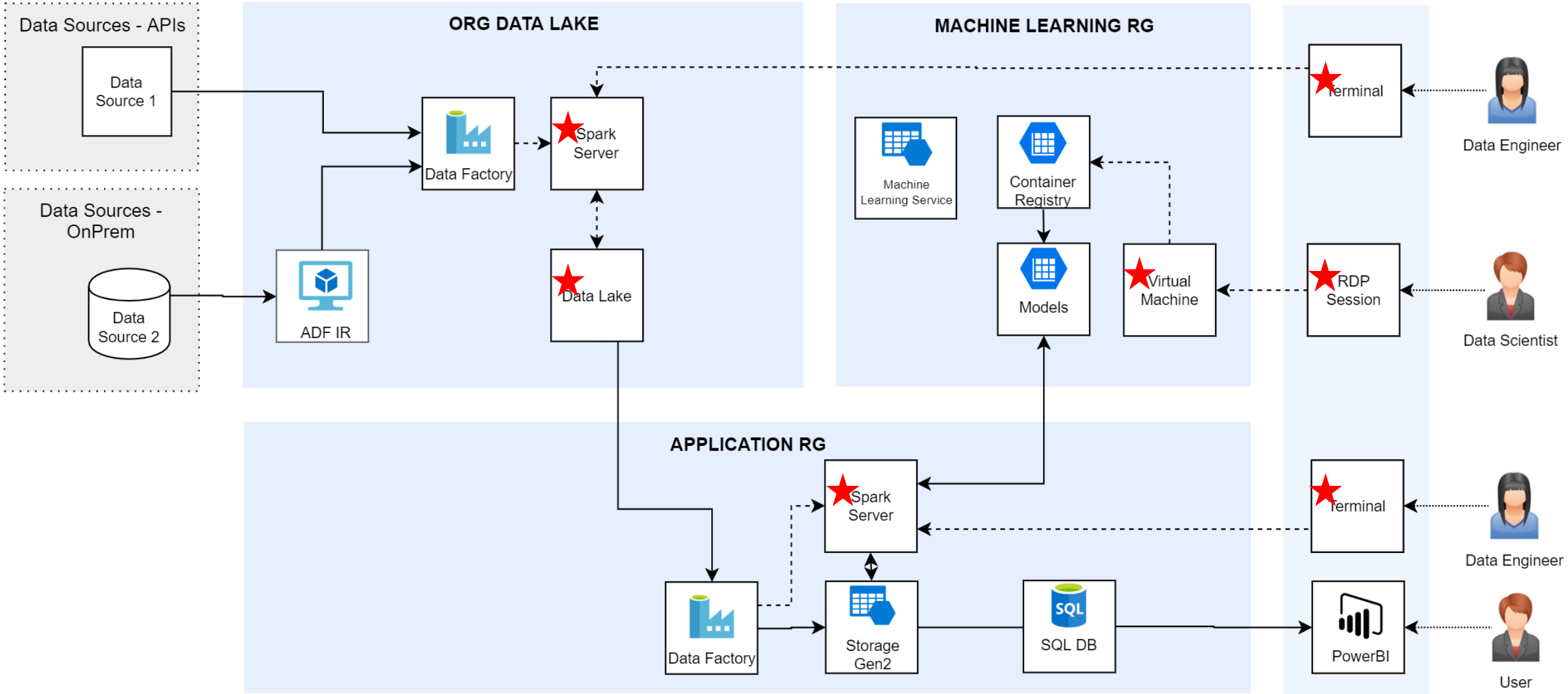
Data Scientist and Data Engineers is a bigger leap than a Front and Backend or FullStack devs

Differences in ML projects versus App Dev

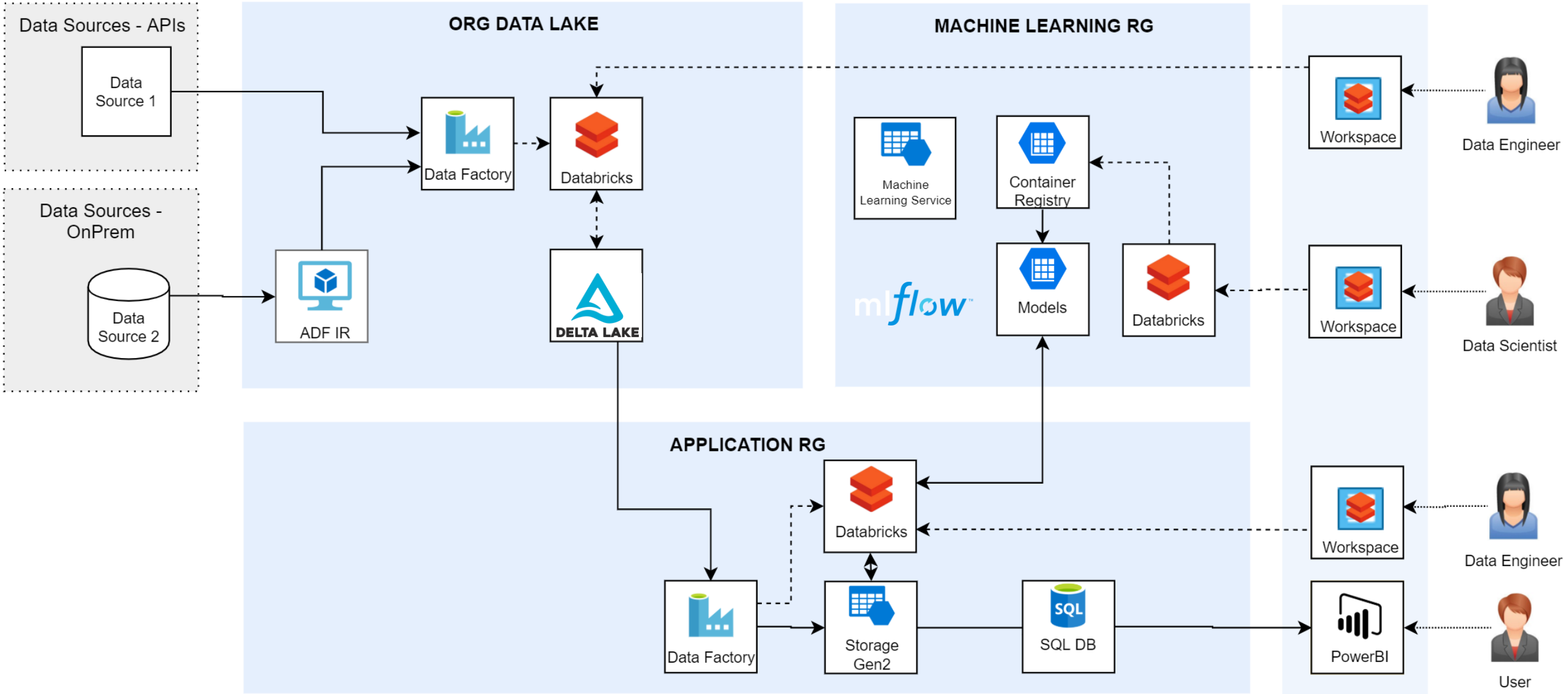
- Exploration versus engineering code requires different coding style and approach
- Lack of Operationalization maturity in Machine Learning projects
- Data Science dev workflow is different to App workflow with different artifacts
- The capturing of these artifacts needs to mature in MLOps

The Databricks ecosystem solves some of these problems by providing a unified analytics platform

BEFORE – without Databricks



AFTER – with Databricks



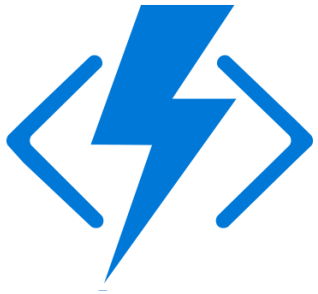
ML Ops – Where to host your model

- What format?
- How does another team member debug locally?
- How to test your models automatically
- How to deploy your models
- How to monitor your models
- How to retrain

PaaS Deployment Options on Azure

Azure Container Services

Azure Kubernetes



Azure Functions (serverless)

All containerized (Docker) and have pros and cons, so how to choose?

- Budget
- POC/DEV versus PROD
- Size of input payload
- Frequency/Scale of service usage (batch, realtime, sync)
- Number of services/ML Models (Generally 1 container per model service)

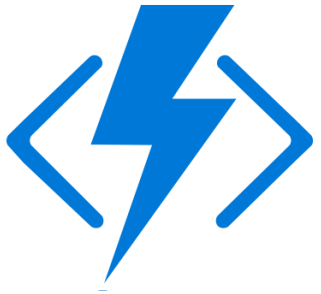
PaaS Deployment Options on Azure

Azure Container Services

- Simple, quick
- Cheap
- Cannot auto-scale
- Good for DEV Environment
- Good for debugging ML services
- Good for small number of services

Azure Kubernetes

- Huge scale (many microservices)
- Relatively expensive
- Scales automatically and self-heals
- Good for PROD Environment
- Requires K8 experience to manage
- Additional complexity & skillsets



Azure Functions (Preview)

- Serverless API
- Very cheap
- Good for PROD Environment
- Good for intermittent usage

Summary & Benefits of Databricks

- Databricks helps facilitate the handover and reduce friction between DE, DS and the business
- It gives Data Scientists an exploration tool with source control, tracking and much more
- It gives Data Engineers huge scale transformations integrated with cloud data services
- It reduces friction by allowing the DE and DS to share code in a single, familiar environment
- Use ML Flow for Model tracking and training – integrates with SageMaker and AMLS
- Use Delta Lake to bring some structure to your data lake

Get started with the Databricks Community Edition for free

We’re hiring… free cash

data-driven ai

ML Engineering & MLOps

Data Engineering

Advanced Training

$2000 bonus for any successfully placed referral…

Questions?

Download the Slides:
https://data-driven.ai/events/sydney-databricks-meetup-feb-mlops-and-databricks/

Rodney Joyce

Practice Director @ Data-Driven AI

LinkedIn: bit.ly/rodneyjoyce

www.data-driven.ai

data-driven ai