



Accelerating Innovation with Unified Analytics

Ali Ghodsi
Databricks CEO
Spark Summit 2017



Apache Spark has gone mainstream!

6,000

Attendees at Spark
Summits This Year

Over

365

Meetup Group Members

K

100

%
Increase in Spark-
Related
Jobs Since 2015

Data-Driven Predictive Applications

Fraud Detection



Recommendation Engine



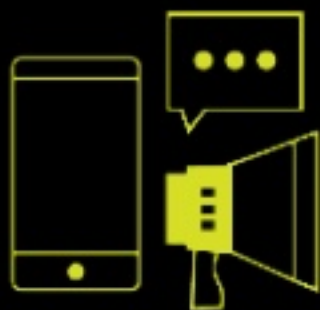
Genome Sequencing



Churn Analytics



Ad Targeting



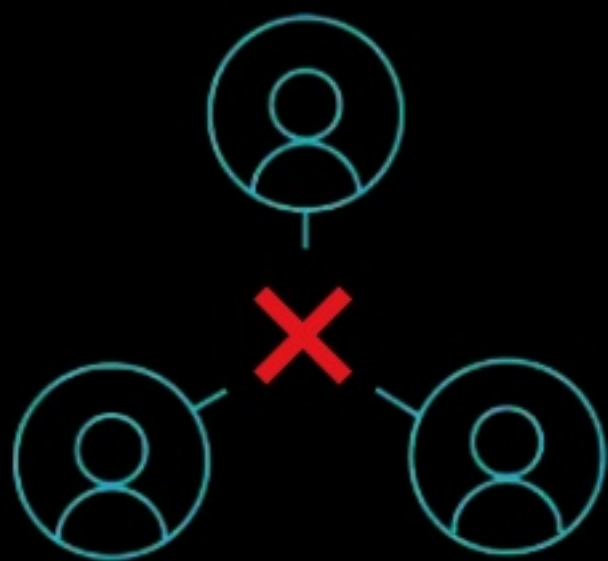
Predictive Maintenance



Medical Imaging



It's hard to collaborate across teams



Lines of Business
Data Scientists
Data Engineers

It takes too long to get results



Data Access

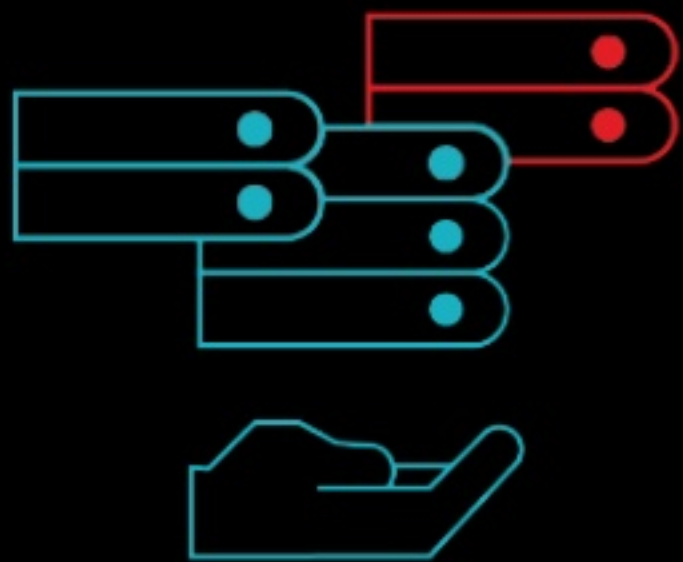
Data ETL

Data Exploration

Model Creation & Training

Moving to Production

It's hard to scale Infrastructure



Manage Complexity

Optimizing Cost

Securing Access and Control

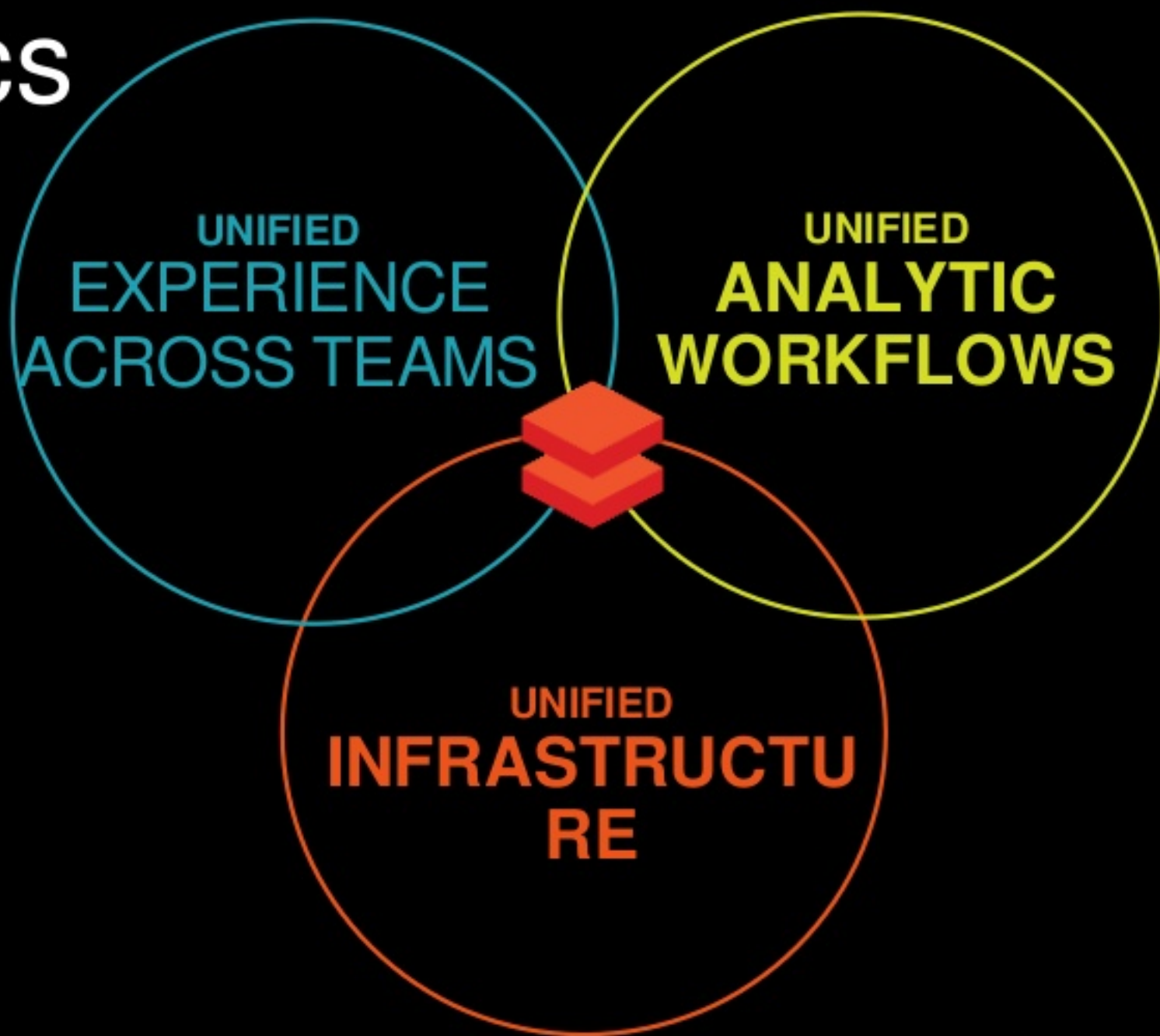


Unified Analytics Platform

Accelerate innovation by unifying data science, engineering and business.

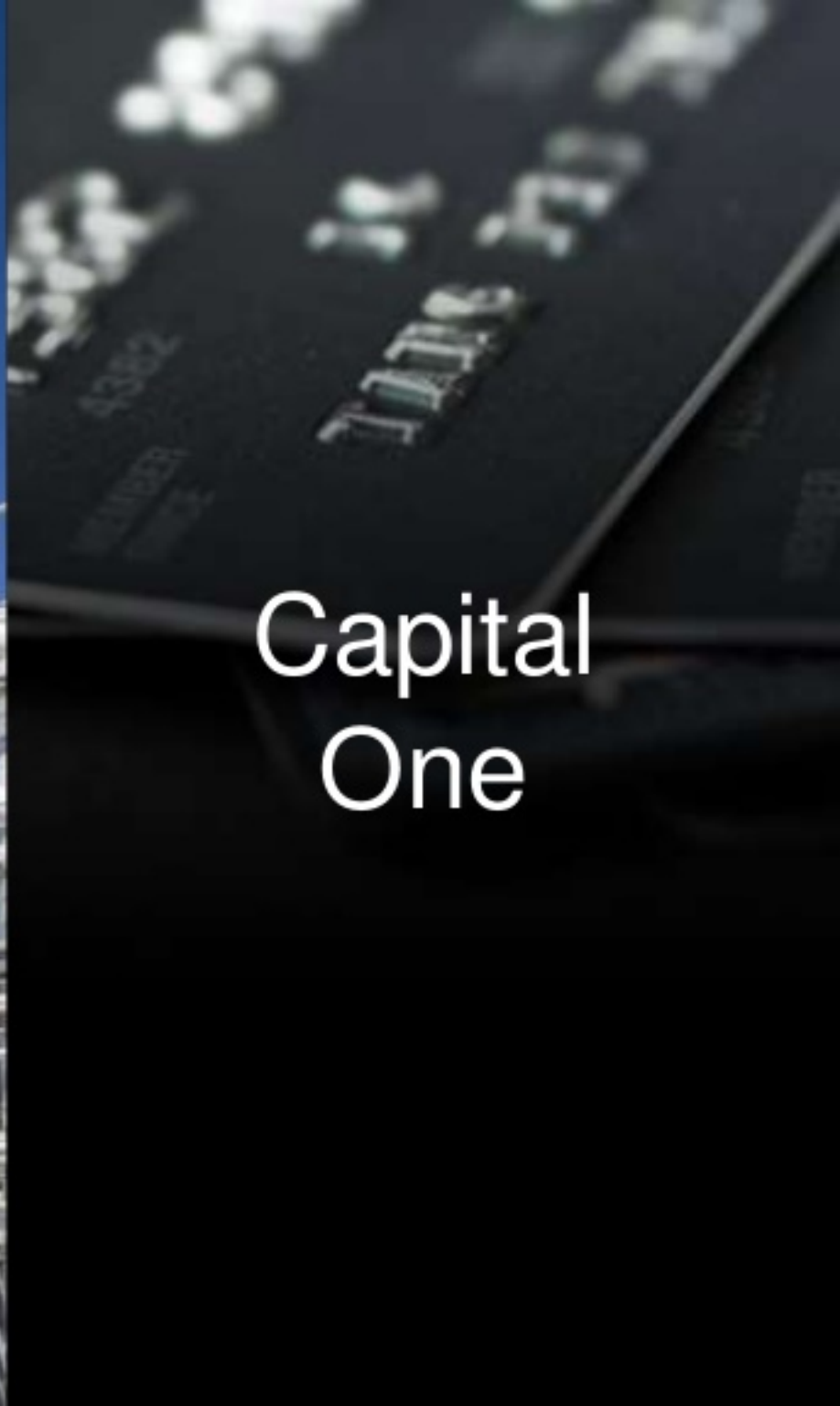
Unified Analytics Platform

Accelerate innovation by unifying data science, engineering and business.





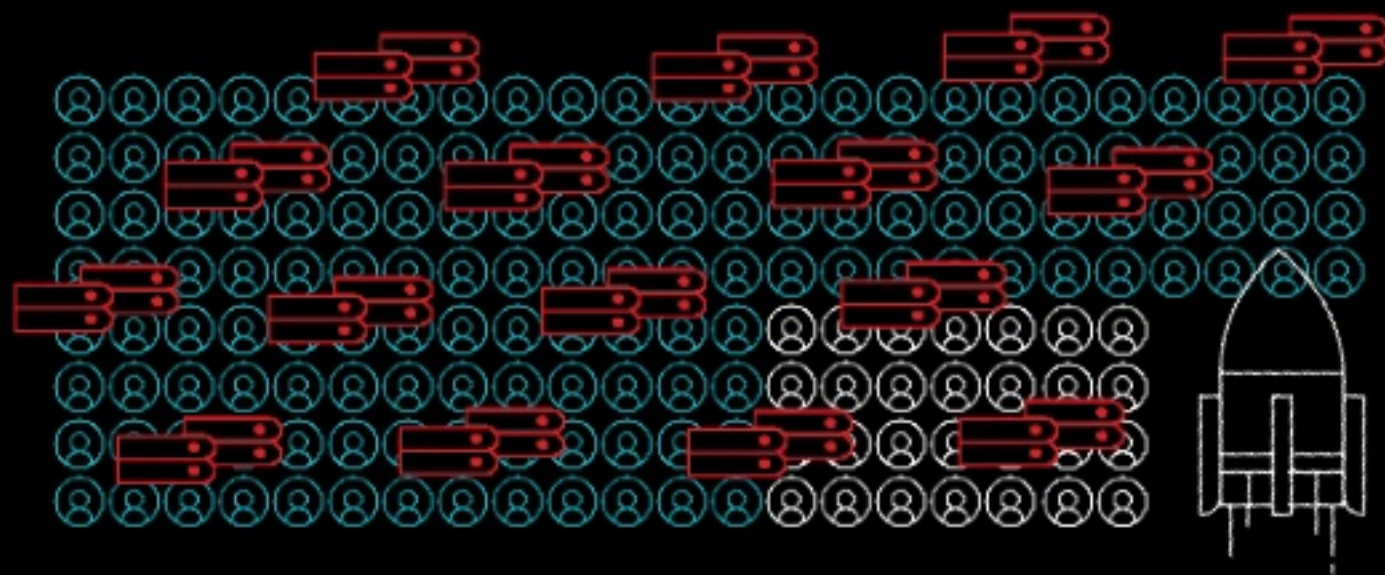
Capital
One



Scaling Across the Organization

Many organizations struggle to scale their Spark infrastructure

Most start with
a small team...



...then want to scale to 100s of data scientists,
line-of-business analysts, and data engineers.

It becomes harder to manage their own infrastructure.

Databricks Serverless

First serverless data science and big data offering

Introducing Databricks Serverless



Allows large groups of users to run workloads on a single, automatically managed pool of resources

Requires no per-user resource configuration or tuning

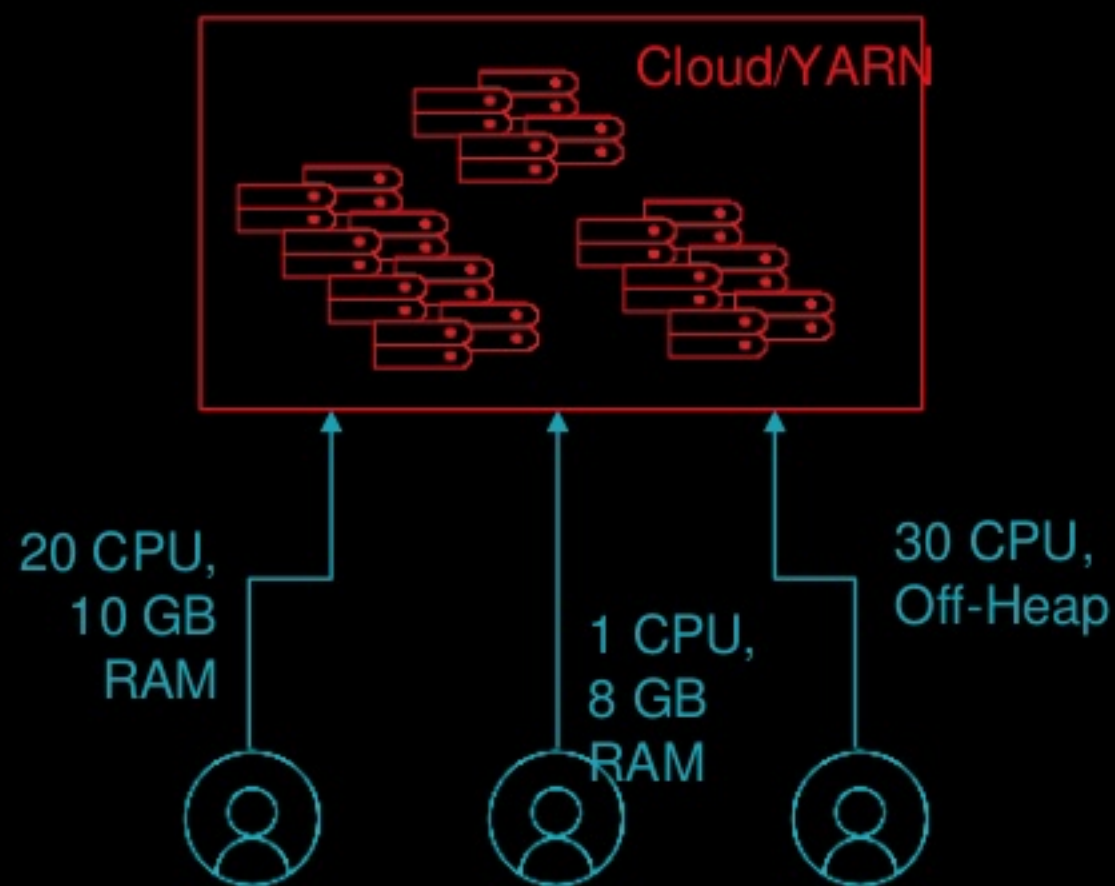
Outperforms other Spark deployments by up to 10x

Users are isolated while they share the serverless pool



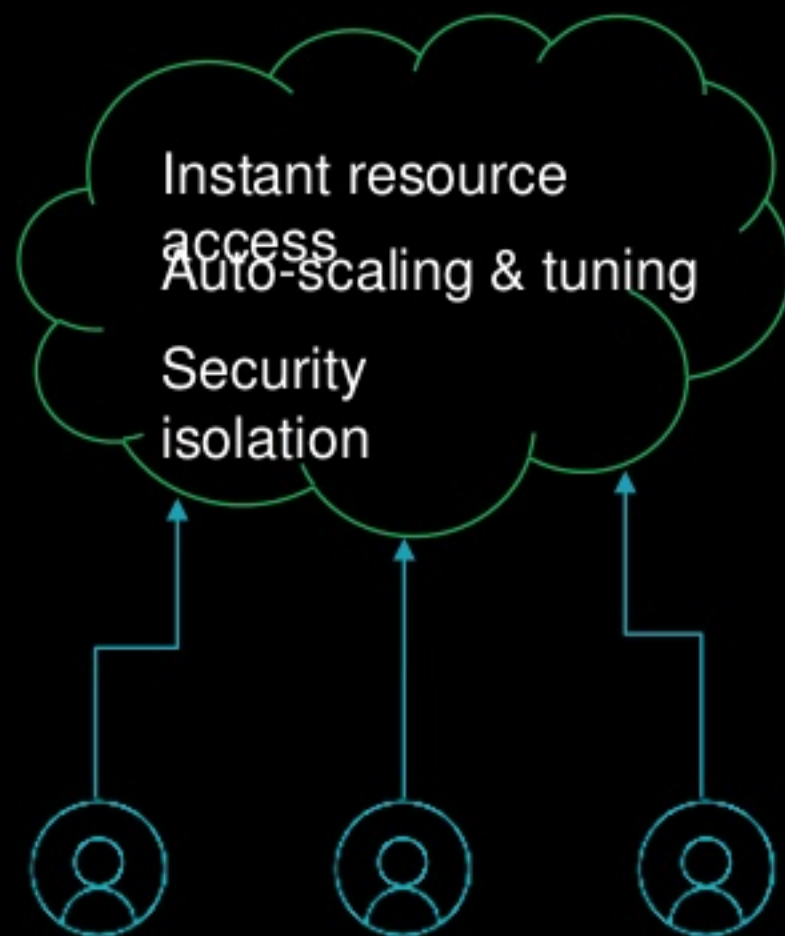
Demo

Current Spark Environments



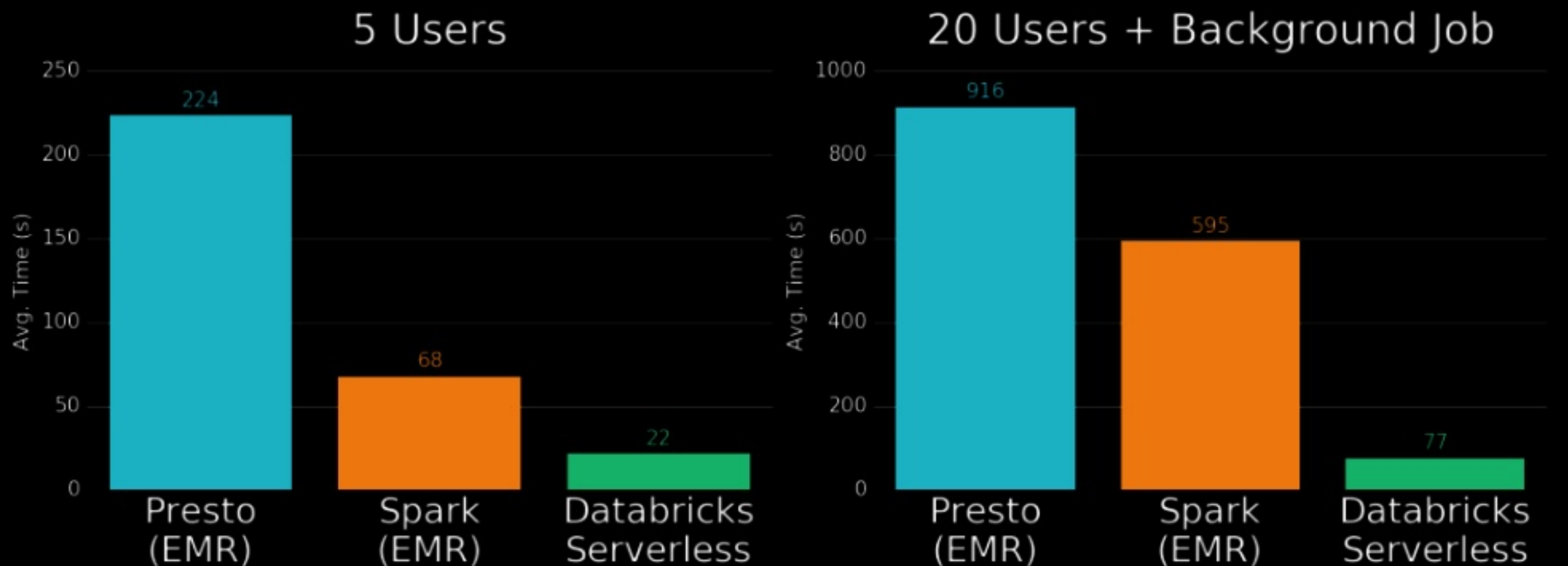
Low utilization, complex per-user config

Databricks Serverless

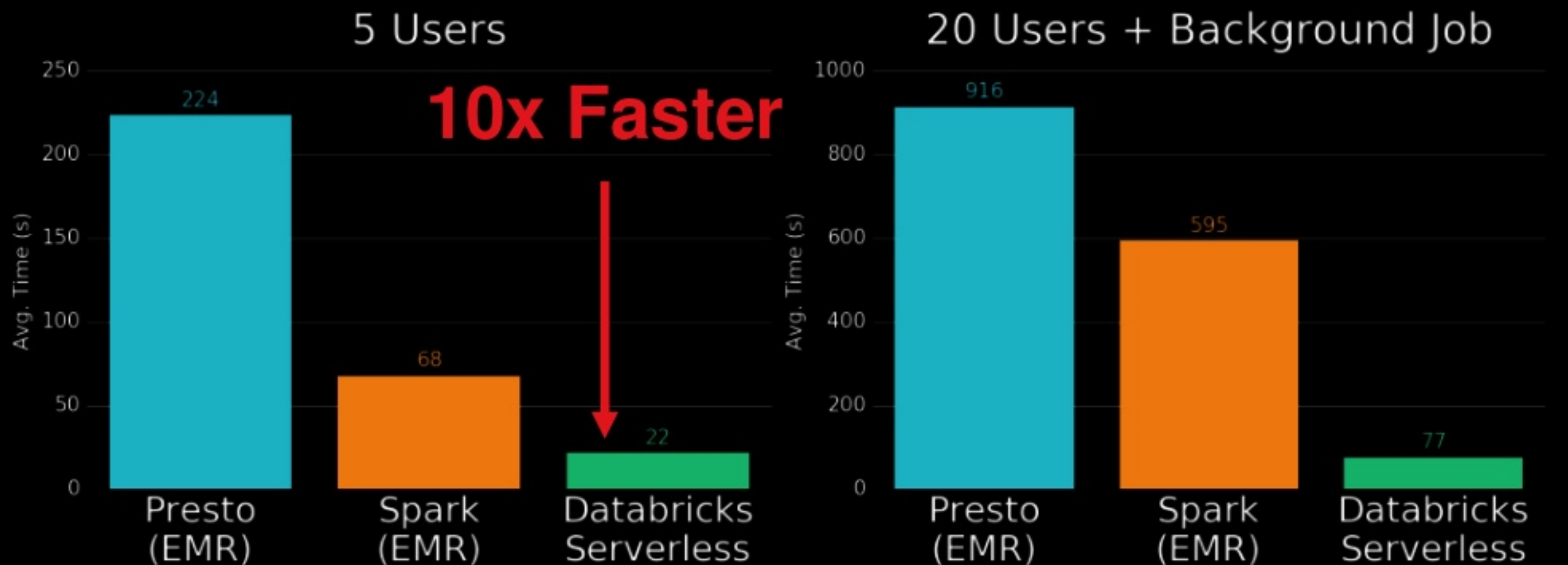


Optimal performance, minimal cost

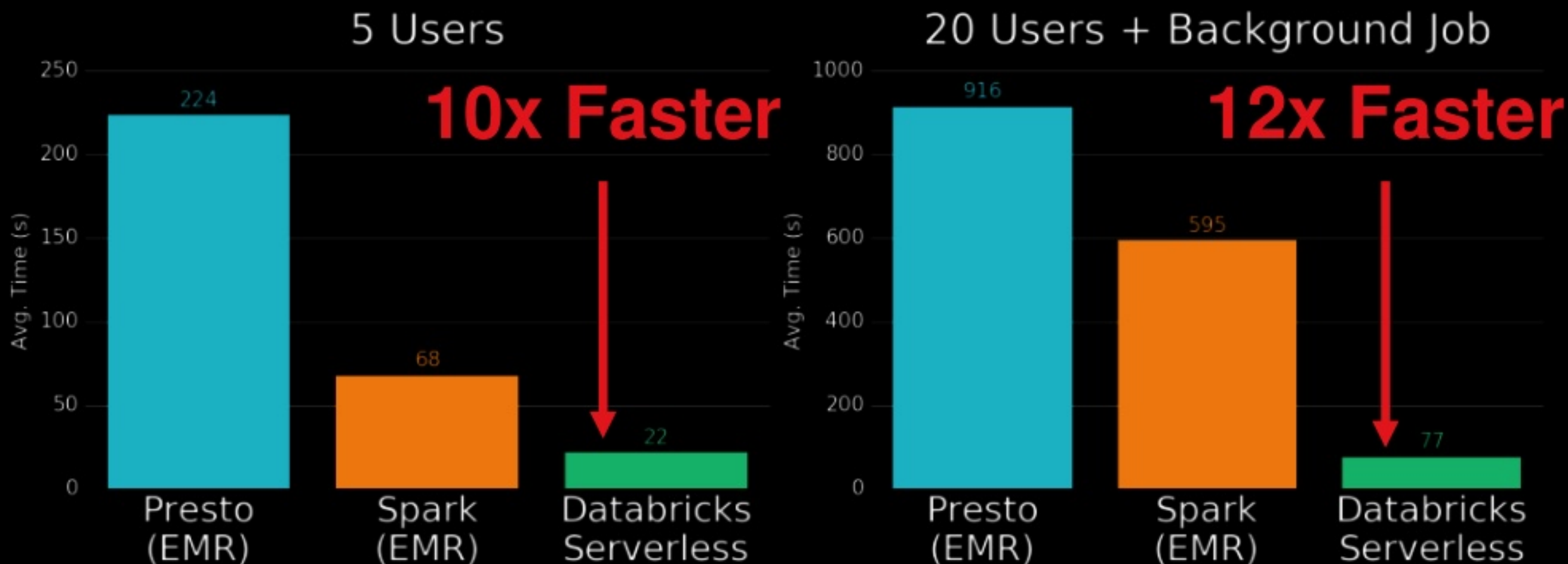
Simultaneous TPC-DS Workloads



Simultaneous TPC-DS Workloads



Simultaneous TPC-DS Workloads



Databricks Serverless

Available to all Databricks
customers this week!