Democratizing AI with Databricks

Ali Ghodsi, Co-Founder & CEO

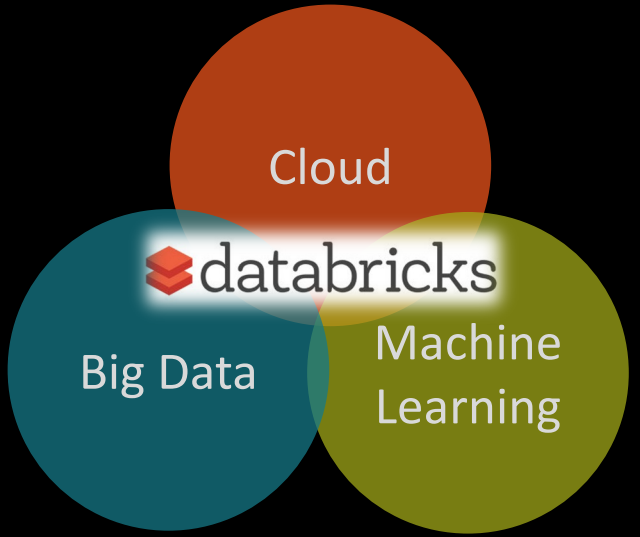
May 9, 2017

databricks

Democratizing AI with Databricks

Databricks original three bets

- Cloud computing
- Big data
- Machine learning and Artificial Intelligence



The combination of these has enabled entire new sets of use cases across many industries

500+ customers across industries

AD & MARKETING TECH | MEDIA & ENTERTAINMENT | HEALTHCARE & PHARMA | ENTERPRISE SOFTWARE

Predict where to drill for oil
based on sub-surface data

PUBLIC SECTOR | FINANCIAL SERVICES | INDUSTRIAL & IOT | RETAIL & CPG

REGENERON

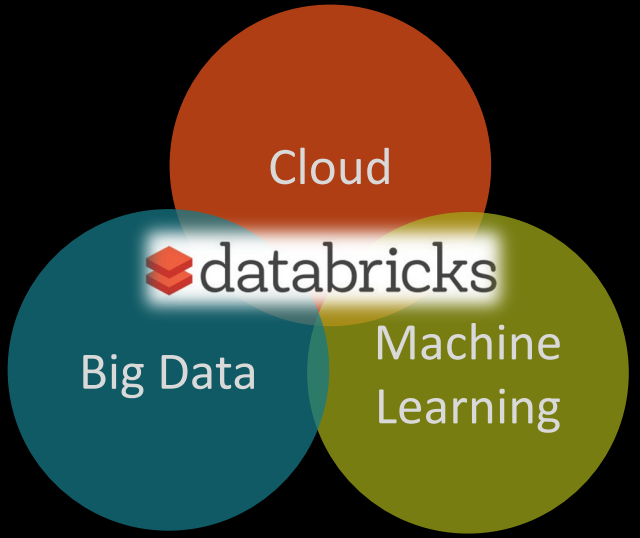
Correlate EMR of 50,000 patients
compared with their DNA

Democratizing AI with Databricks

Cloud computing
- Gartner believes this market to be $200b in 2020

Big data
- 90% of the data created in last 2 years

Machine learning and Artificial Intelligence
- Just scratched the surface of the use-cases



Likely to be a juggernaut in the analytics space,
we believe it will be Databricks

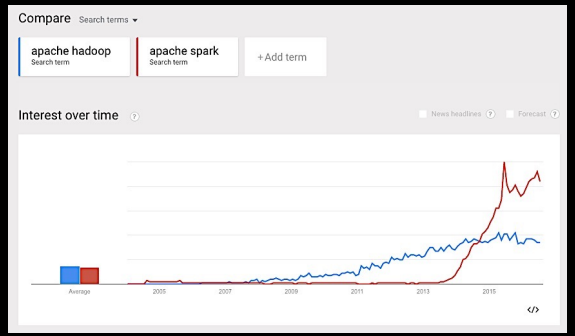
Background: Apache Spark

Virtually every company doing AI on massive data does it with Spark

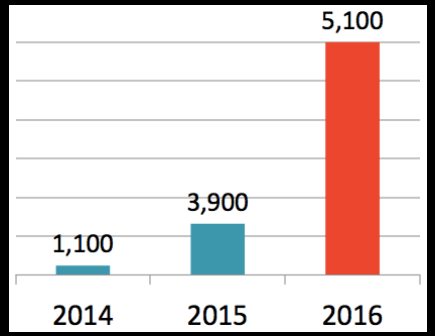
We built Spark enable the unification of

- Processing of large amounts of unstructured data (ELT)
- Make predictions on that data using machine learning (ML)
- Get insights in real-time continuously (streaming)

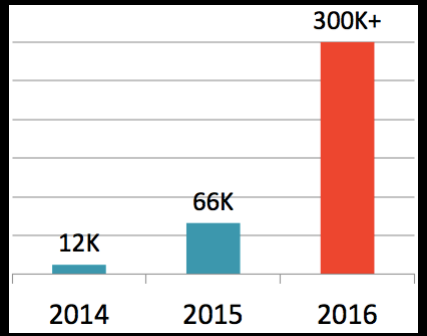
SPARK VS HADOOP



SUMMIT ATTENDEES

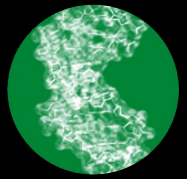


MEETUP MEMBERS

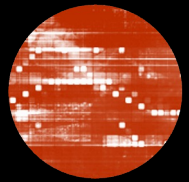


Artificial Intelligence

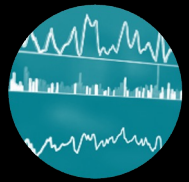
Artificial Intelligence Applications



PREDICTIONS



CLUSTERING



ANOMALIES

AI GAP

DATA WAREHOUSES

HADOOP DATA LAKES

CLOUD STORAGE

Big Data

Hardest part of AI isn't AI

"Hidden Technical Debt in Machine Learning Systems ", Google NIPS 2015

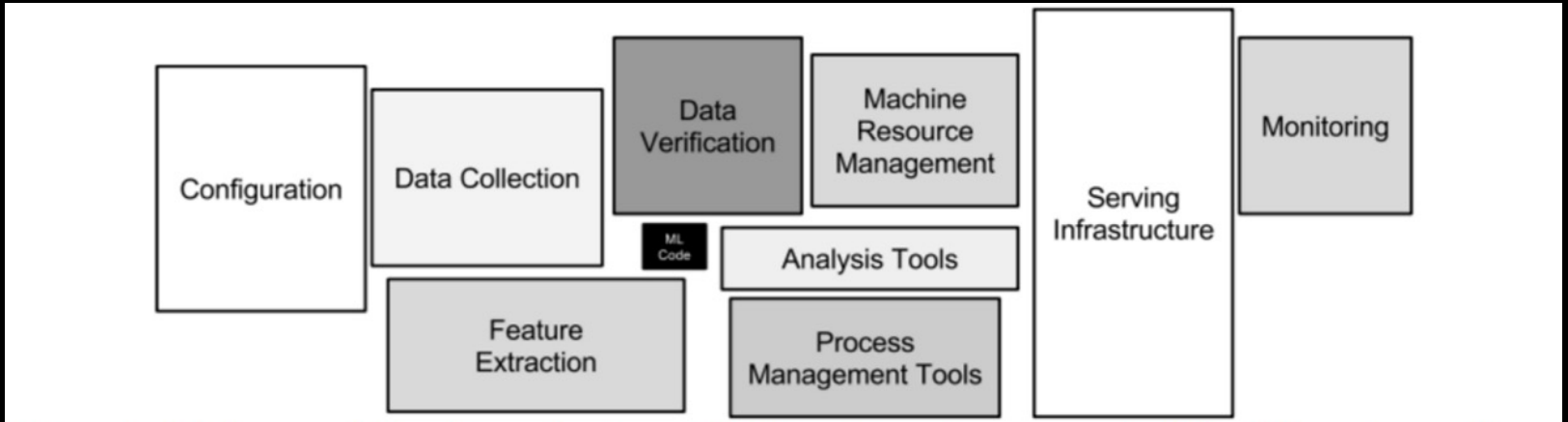


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

The hardest part of AI is Big Data

Big Data was the Missing Link for AI

BIG DATA

Customer Data

Emails/Web pages

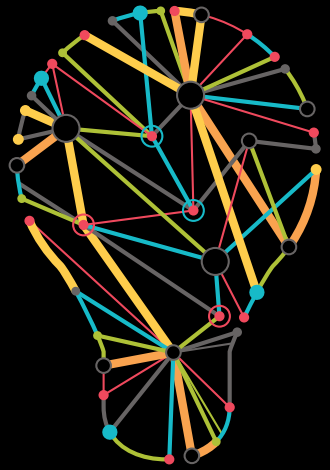
Click Streams

Sensor data (IoT)

Video/Speech

…

GREAT RESULTS

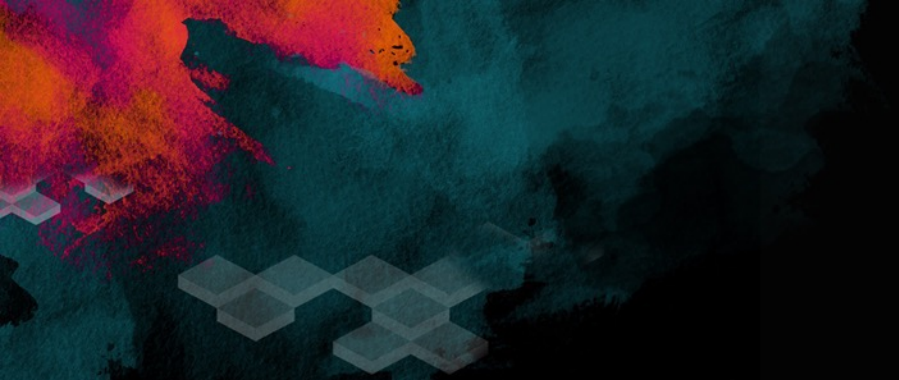


But few companies are successful with AI

Why is there a gap?

1. Difficult to manage & secure data infrastructure
2. Hard for teams to share and ask questions from the data
3. Hard to put analytics into production

Databricks Platform



databricks

Databricks Platform

1. **Difficult to create, manage, and secure Spark infrastructure**
2. **Not easy for teams to query the data, collaborate, and share the results**
3. **Hard to put analytics into production**

Databricks Serverless Spark Platform

- Auto-tuning Spark with reliability and availability
- 10-40x speedup vs Apache Spark, custom data connectors
- full governance, SOC2 & HIPAA compliance

databricks

Databricks Platform

1. Difficult to create, manage, and secure Spark infrastructure
2. **Not easy for teams to query the data, collaborate, and share the results**
3. Hard to put analytics into production

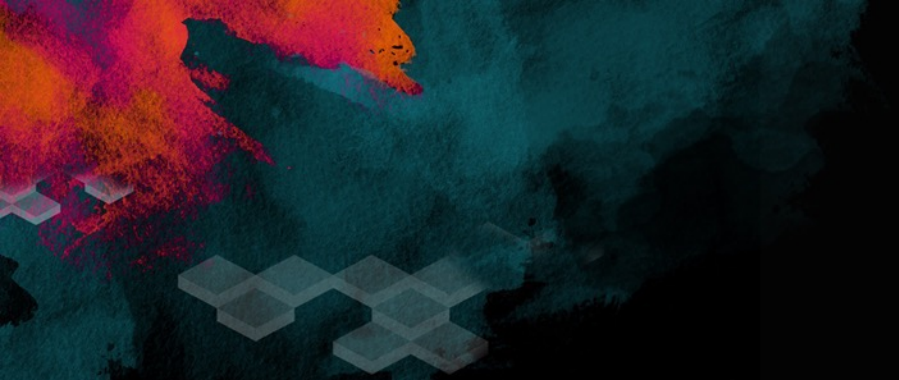
Databricks Collaborative Workspace

- Notebooks – Interactive exploration for for ETL, AI, Real-time applications
- Dashboards – interactive real-time dashboards to democratize access to Spark
- Reports – Ability to produce periodic reports with predictive insights

Databricks Serverless Spark Platform

- Auto-tuning Spark with reliability and availability
- 10-40x speedup vs Apache Spark, custom data connectors
- full governance, SOC2 & HIPAA compliance

Databricks Strategy



databricks

Making Big Data Simple with the Cloud

Databricks Data Science Workspace

- Software as a Service (SaaS)

Databricks Serverless Spark Platform

- Platform as a Service (PaaS)

Infrastructure as a Service on demand

- Simple, low-cost, scale-out, pay-as-you-go

Databricks: cloud-agnostic killer app

Databricks Data Science Workspace

- Software as a Service (SaaS)

Databricks Serverless Spark Platform

- Platform as a Service (PaaS)

On-prem market for Spark

- Huge demand for Databricks platform on-prem
  - Hadoop vendors commoditized storage through data lakes
  - Most on-prem enterprises struggling with Spark
  - Big void due to Databricks keeping enterprise value out of Spark

We plan to bring Databricks to wherever there is data

Databricks: unified analytics platform

Databricks Data Science Workspace
- Software as a Service (SaaS)

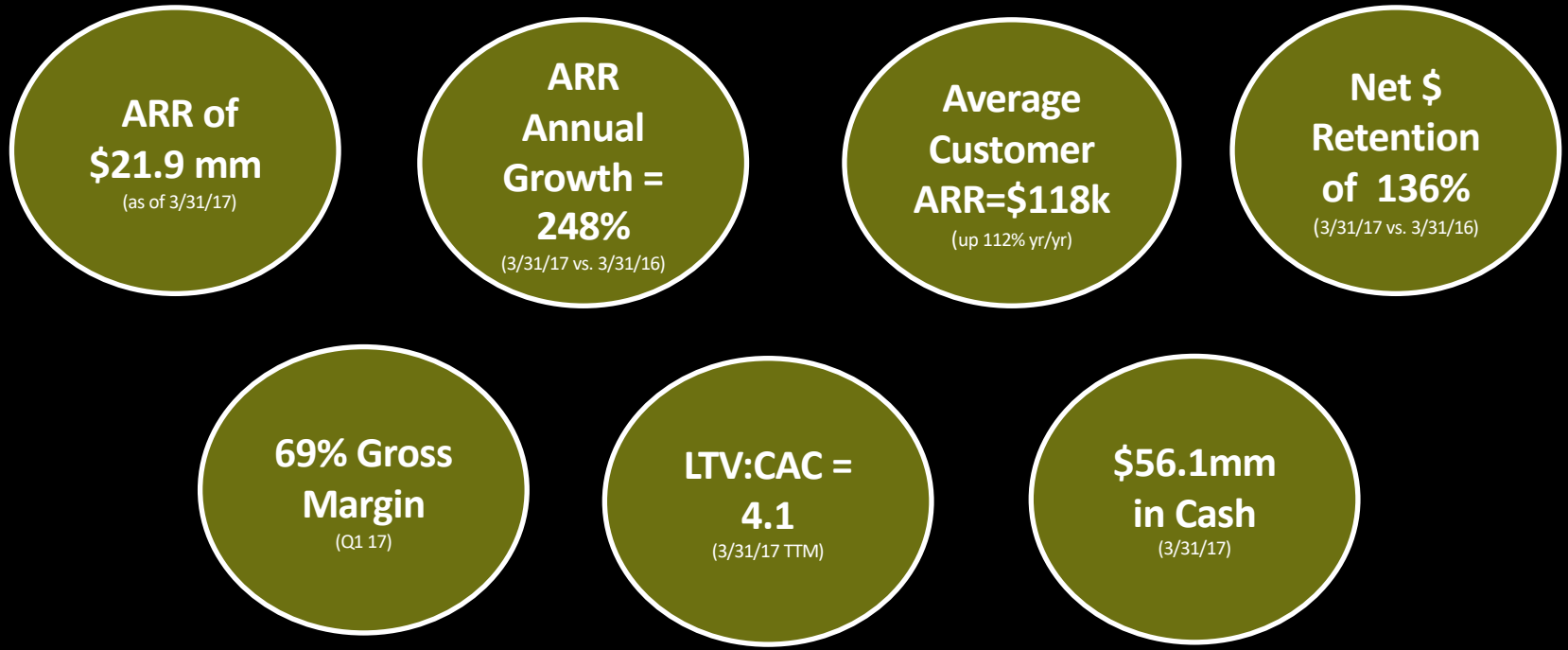
Databricks Serverless Spark Platform
- Platform as a Service (PaaS)

Unified Analytics Platform on-top of cloud and on-prem storage

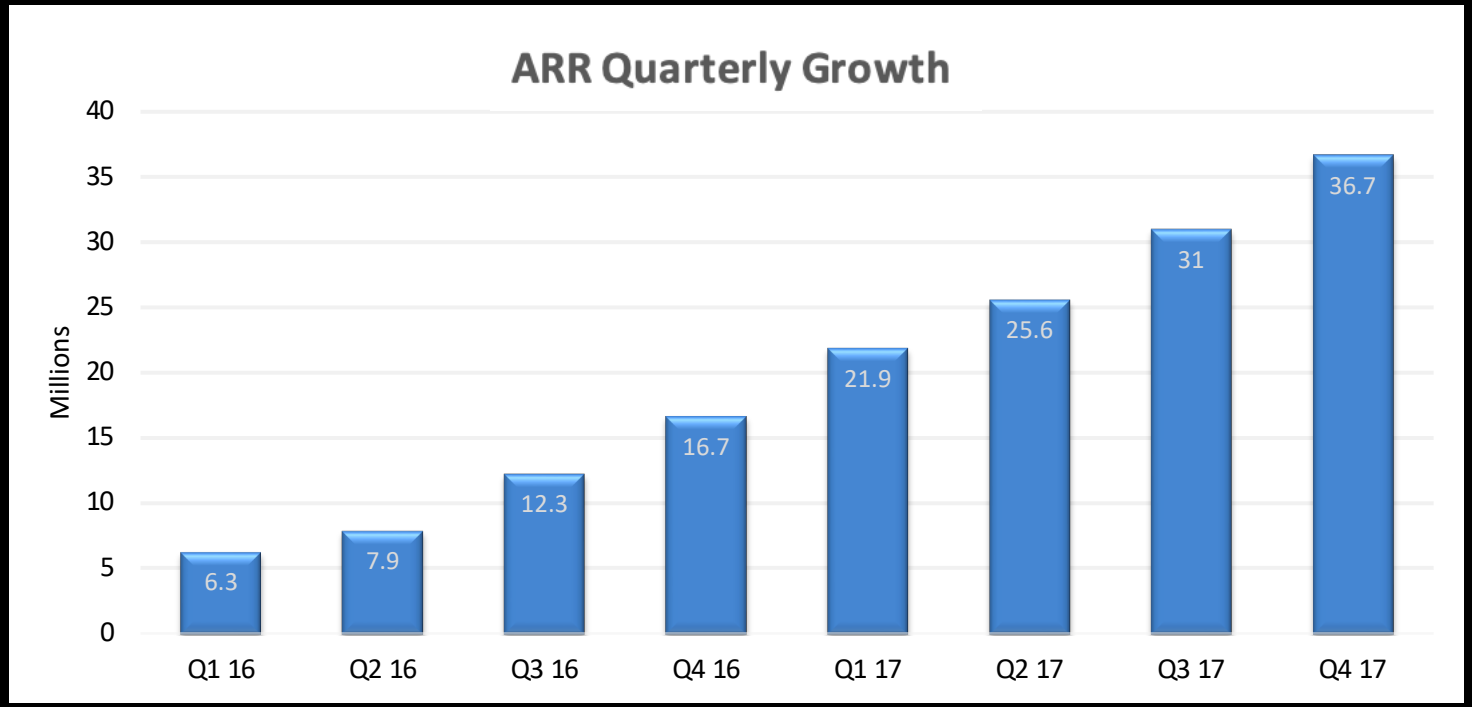
Databricks Finance

databricks

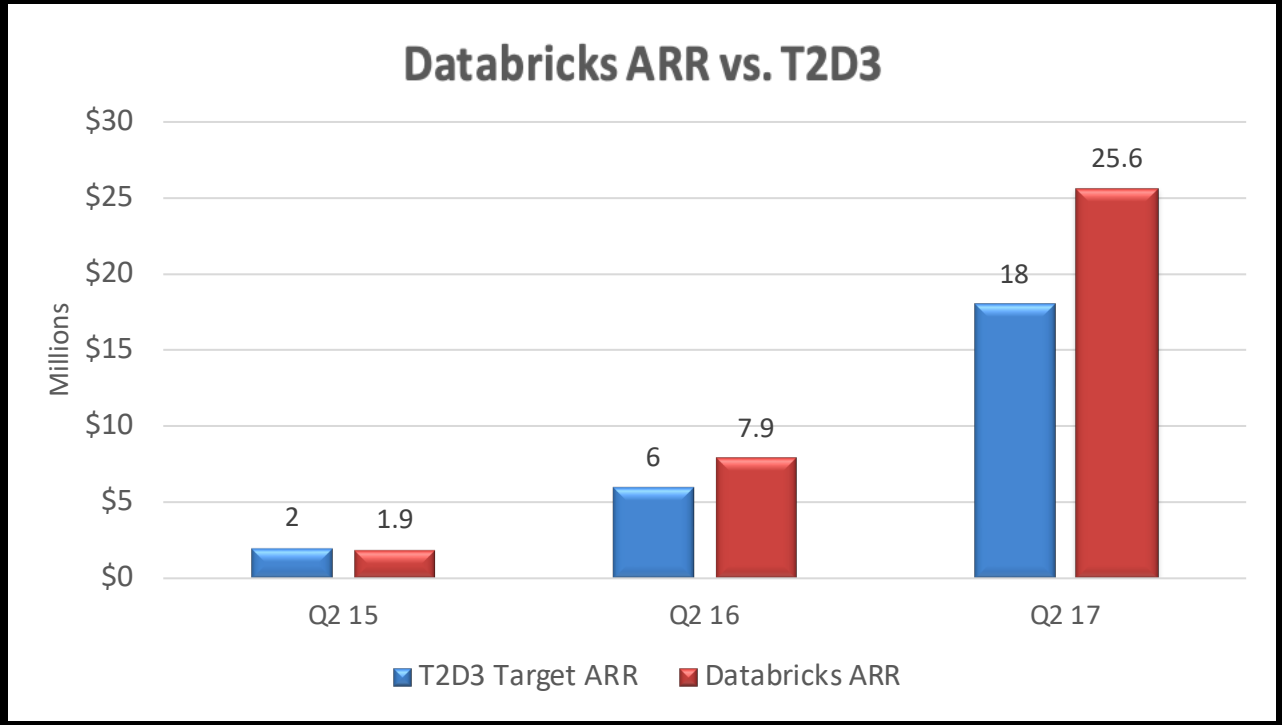
Great Metrics All Around



Incredible ARR Growth (5x growth for Q1 17 vs. Q1 16)



Databricks vs. T2D3 (triple, triple, double, double, double)



The Team

Ali Ghodsi, CEO and co-founder
- MBA, PhD, Adjunct Professor UC Berkeley, co-founder Peerialism

Ron Gabrisko, Chief Revenue Officer
- VP Sales at Cyclone pre-revenue to acquisition by Axway to IPO, Stanford M.B.A. & M.S

Rick Schultz, Chief Marketing Officer
- SVP marketing Alteryx pre-revenue to IPO, Oracle VP Product Marketing, UC Berkeley B.S.

John Winkenbach, SVP of Finance
- CFO at Jobvite, VP finance Technorati

Hatim Shafique, Chief Customer Officer
- CCO & CIO at AppDynamics pre-revenue to IPO

Patrick Wendell, VP of Engineering and co-founder
- UC Berkeley M.Sc., Princeton B.S.

Michael Hoff, SVP of BD and Partnerships
- VP of Channels at Tableau, Global VP of Sales EMC, GM Windows Azure MSFT

Try Apache Spark with Databricks

Try latest version of Apache Spark and preview of Spark 2.2

http://databricks.com/try

Thank You