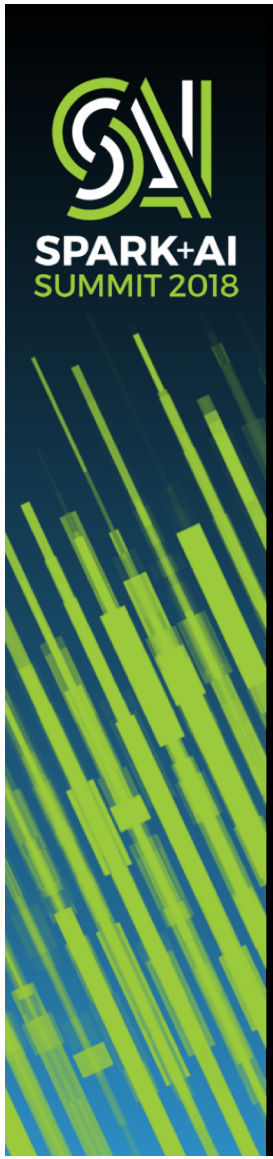


HOW APACHE SPARK CHANGED THE WAY WE HIRE PEOPLE

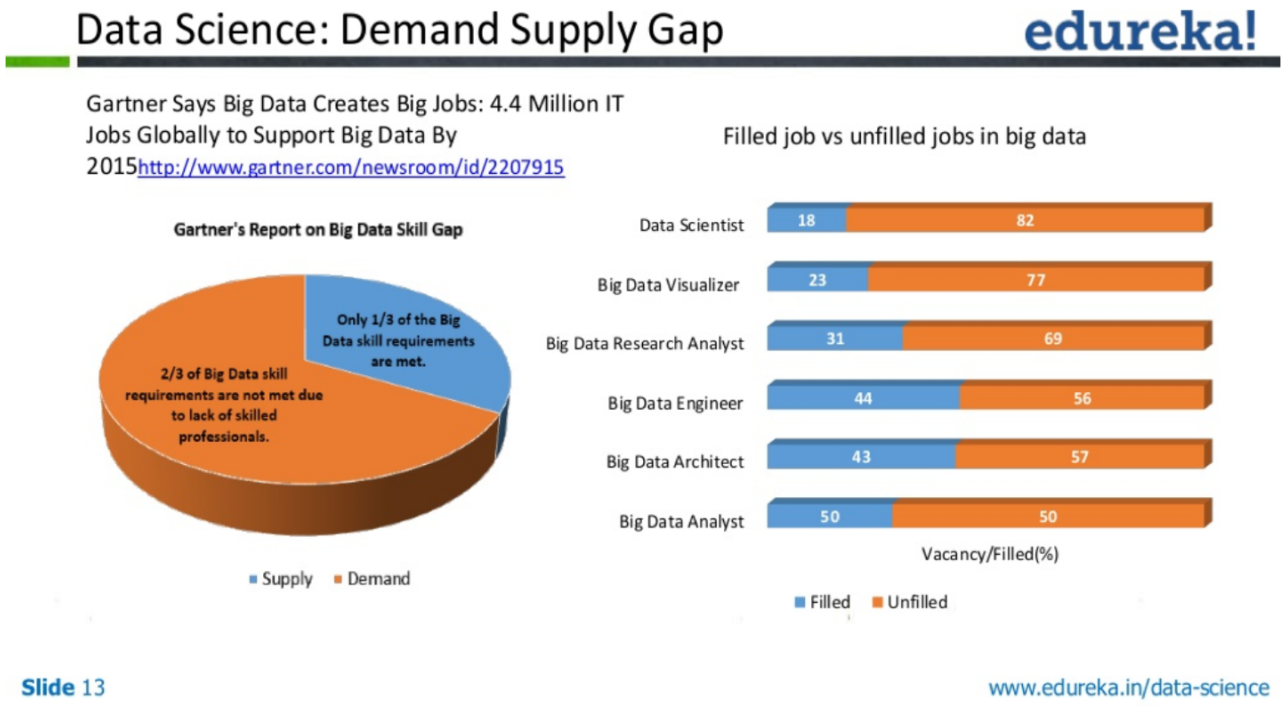
Tomasz Magdanski, iPass

#EntSAIS17

What if the war for talent ended and your company lost?

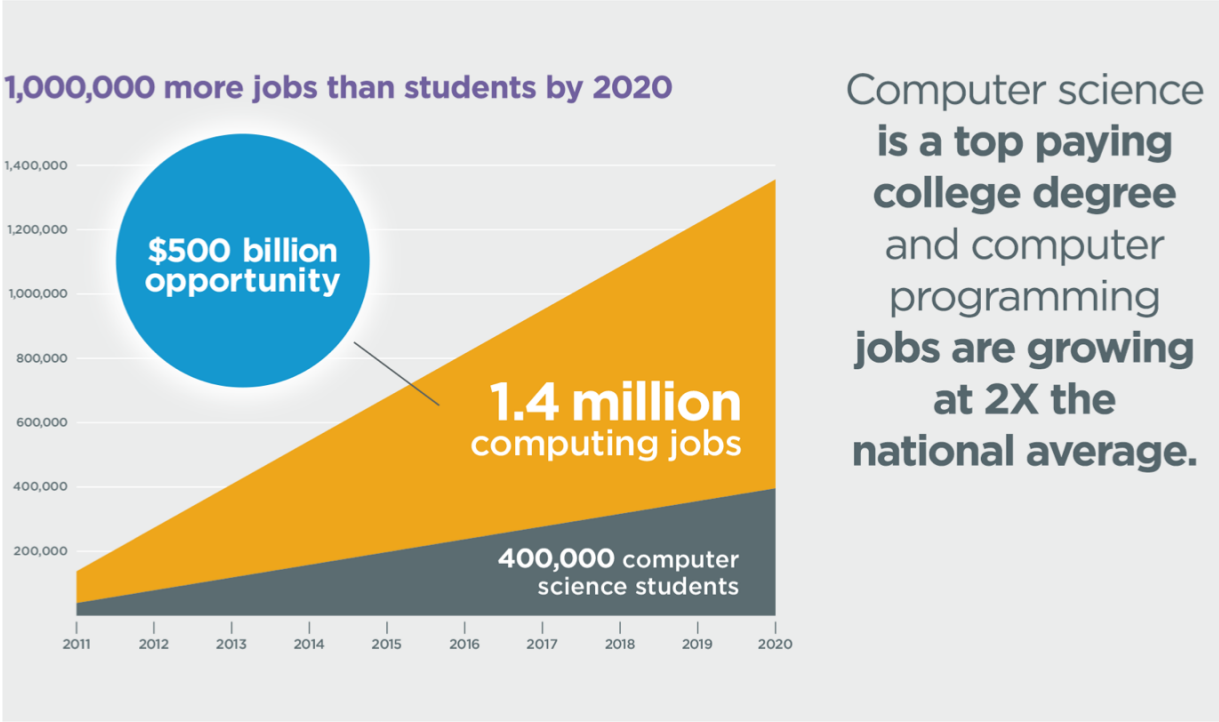
- War for talent
  - Late ’90s warning from McKinsey about talent shortage
  - Urged companies to prioritize strategies around recruiting, retaining and developing key employees
- One percent problem

Hiring is tough



Source: edureka

And its going to get worse



Source: Hour of Code

Since war for talent started we have made a full circle

- Apart from hiring skilled engineers companies look inside to fill in the gap
- Create path to grow within your organization
- But wait a minute ?
  Didn't we just say there is a big skills gap ?

What are we building ?

Goals

- Scalable platform
- Cost effective
- No data loss
- Code portability
- Easy R&D
- Extendable
- Support many languages
- Support batch, stream
- ML enabled
- Collaborative

Who we were looking for ?

- MapReduce
- Hadoop / HDFS
- Hive / Pig
- Storm
- Caching
- Avro / Parquet
- Distributed Computing
- Manage Clusters and Infrastructure
- Integrate tools
- Data Warehousing and Modeling

Who we were looking for ?

- CAP Theorem
- Data Transformation
- Data Collection
- SQL
- Cassandra / Hbase / MongoDB / mysql
- Kafka
- AWS
- Scala / Java / Python
- Understanding data structures and algorithms
- Visualization and Data Analysis
- Team player
- SPECIFIC INDUSTRY KNOWLEDGE

Spark changed what we are looking for

Spark

- Simple
- Easy to learn
- High abstraction API
- Build in connectors to major data sources
- Supports Batch, Stream
- Highly optimized and extendible
- ML library to run at scale
- Spark provides transactional writes and exact once semantic

Spark and Databricks

- A single platform that unifies data engineering and data science
- Automated cluster management / zero-management infrastructure
- Intuitive notebooks supporting multiple programming languages
- Makes collaboration easy
- Blends Data Engineering and Data Science workloads
- APIs to integrate with other tools

Spark and Databricks

- Integrated meta store
- Integrated Managed and unmanaged tables
- Workspace API
- Engineering and Customer Support including Solution Architects
- Easy dashboards
- DbUtils
- SBT tools for easy deployment

Who we are looking for now?

- Experience in programming using APIs
- Understanding data structures and algorithms
- Scala / Java / Python / R
- Visualization and Data Analysis
- Team player
- SPECIFIC INDUSTRY KNOWLEDGE

Business needs

- Wi-Fi connectivity patterns
- Our business and customers
- Existing system architecture
- Skip lengthy onboarding process
- One stack to learn

How did that change the way we hire ?

- We have internally hired:
  - QA engineer
  - App developers
  - Ex developer / product manager
  - Backend engineer
- Externally hired:
  - One very experienced senior Data Engineer
  - Few collage grads – Junior data engineers

Summary

- Thanks to Databricks we didn’t have to build a platform
- Hired mixed of internal and external candidates
- Focus on business needs
- Created 6 data products
- New seven digit revenue stream for our company
- Continue to innovate