nielsen

Scaling Your Skillset with Your Data

Through the Modernization of Data Science Technology

Jarrett Garcia
October 25, 2017

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

WHAT WE MEASURE

What Consumers WATCH

1.7 BILLION
VIEWING RECORDS
EACH MONTH

1.6 TRILLION
IMPRESSIONS PER YR

7 MILLION
WEB EVENTS DAILY
FROM MOBILE DEVICES

What Consumers BUY

250K
PANEL
HOUSEHOLDS

6.7 BILLION
STORE TRANSACTIONS
EACH MONTH

1.7 MILLION
STORE VISITS EACH MONTH

"We are a measurement company. Data is the core of our business. Everything we do is to help our clients and our company realize the full value of that measurement"

Mitch Barns, CEO Nielsen

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

CHALLENGE

- **Business** - sample based data collection to digital scale
- **People** - Variety of skill sets requires future proof tooling
- **Technology** - Over dependence on proprietary software ($$$)
- **Infrastructure** - old, expensive and not scalable

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

STRATEGY

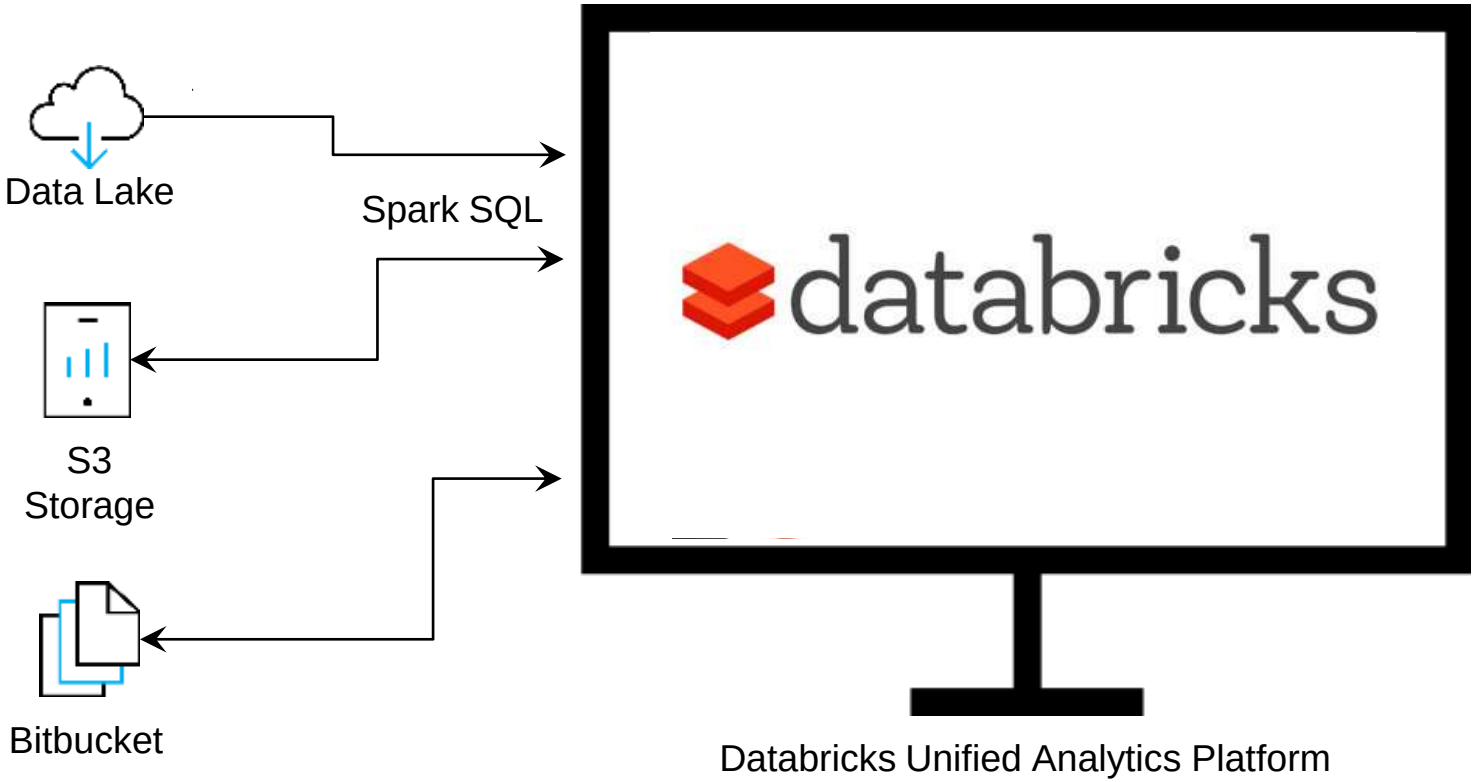
Goal: Create a modern Data Science organization that uses best practices and tools

,

Modernize technology stack	Python, SQL, Databricks, AWS
Support speed and scale of R&D	Cluster computing & data access
Break group silos and increase collaboration	SDLC collaboration software
Attract Data Science talent	Modern methods & tools

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

MODERN TECHNOLOGY STACK



Copyright © 2017 The Nielsen Company. Confidential and proprietary.

DATABRICKS SCALING DATA SCIENCE

- ✓ Data Science can utilize Python as a SAS replacement.
  Databricks' notebooks provide a Python coding platform
- ✓ Data Science can work within the Cloud ecosystem.
  Databricks provides the users access to cloud scale and elasticity
- ✓ Data Science can access data and perform analytics quickly
  Databricks democratizes data access

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

SUCCESS STORIES

- Databricks has helped Nielsen Data Science **reduce its SAS footprint by 40%** for an annual savings of just under **two million dollars**.
- **Reduced the runtime** of the Viewer Assignment model from **36 days to 4 hours.**
- The Databricks accelerated a Machine Learning model by removing manual classification process and **improve the AUC score from .6 to .8.**
- Increased Databricks users from **9 to 156 users** while partnering with Databricks to **reduce DBU usage by 45%.**

The Area Under the Curve (AUC) is a common evaluation metric for binary classification problems. If the classifier is very good, the area under the curve will be close to 1.

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

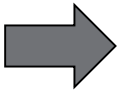
Copyright © 2017 The Nielsen Company. Confidential and proprietary.

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

Move to Cloud Computing

Requirements:

- Flexible compute options
- Separate compute and storage
- Parallelization options
- Improved deployment speed



databricks

APACHE Spark™

Copyright © 2017 The Nielsen Company. Confidential and proprietary.

CHALLENGE

- Legacy servers will not support newer versions of various languages, packages and tools
- Older and outdated processor and memory capabilities
- Limited space capabilities and costly expansions
- Long computation times for tools and algorithms
- Over dependence on proprietary software ($$$)

Copyright © 2017 The Nielsen Company. Confidential and proprietary.