

# A Data Journey with AWS

Julien Simon

Principal Technical Evangelist, AWS

@julsimon



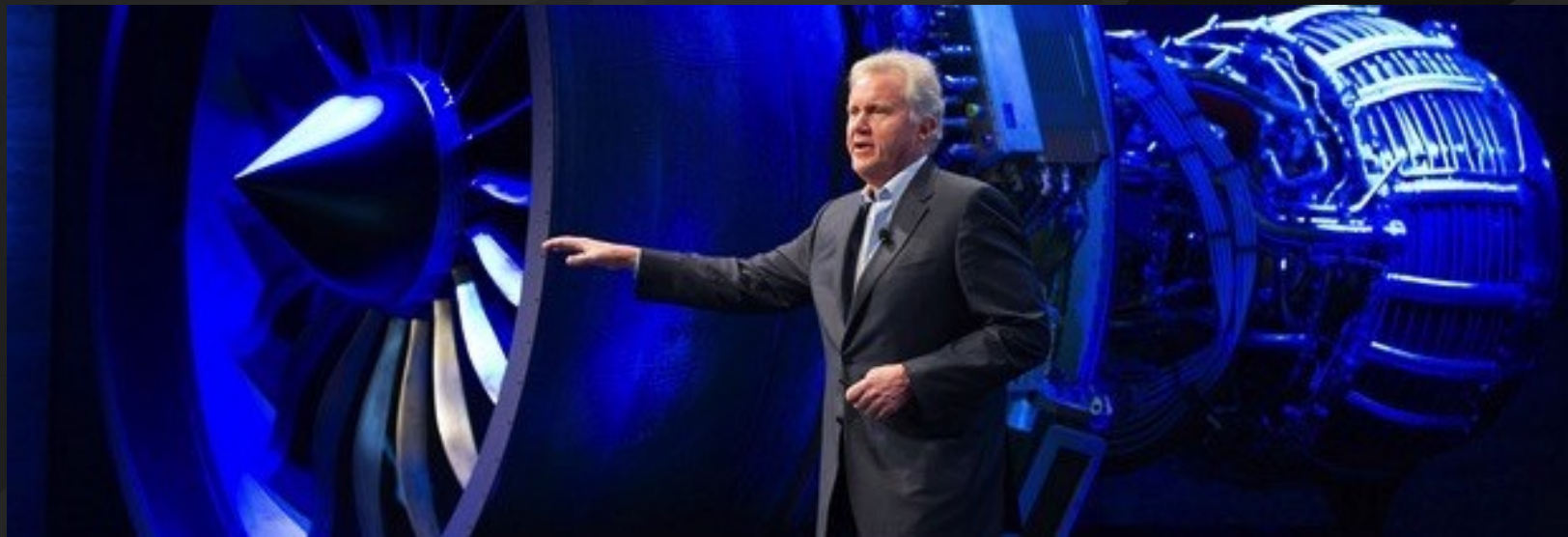
What does  
“Digital”  
really mean?



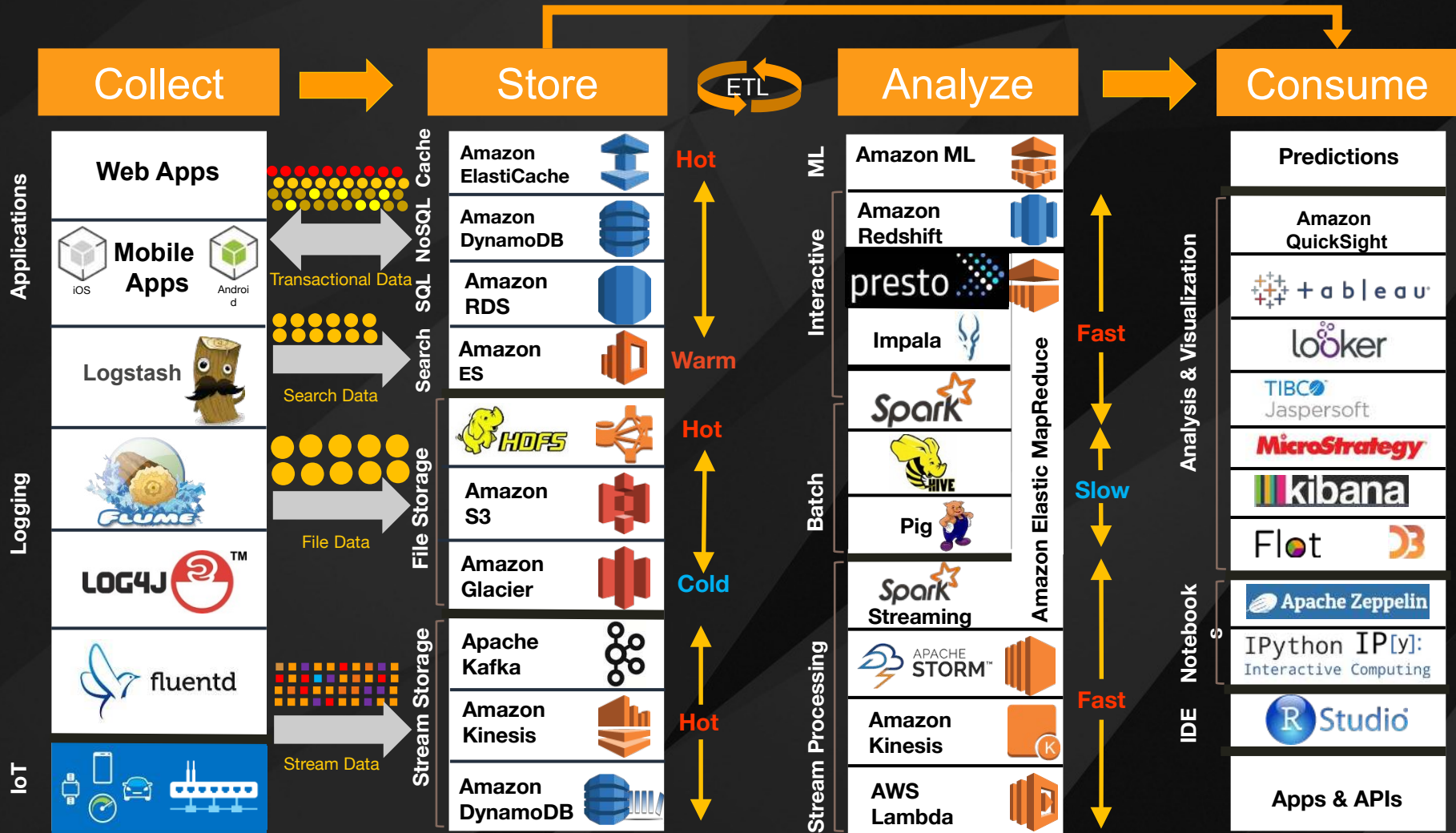
Turning  
**Data**  
into  
**Business Value**

...through  
Software

Jeff Immelt, GE Chairman & CEO



*“If you went to bed last night as an industrial company, you’re going to wake up this morning as a software and analytics company.”*



# DATA WAREHOUSING

understanding the past



# Nasdaq

<https://aws.amazon.com/solutions/case-studies/nasdaq-finqloud/>

In 2014, Nasdaq replaced the existing data warehouses for its US equities and options exchanges with Amazon Redshift.

On a nightly basis, Nasdaq loads approximately 5 billion rows of data into Redshift within a 4-6 hour window.

Amazon Redshift now powers a number of data analytics applications at Nasdaq, including its **billing system** for US customers. In 2015, Nasdaq is expanding its use of Redshift to its global exchange properties.



# MAP REDUCE

processing huge amounts of data  
for fun and profit (hopefully)

# FINRA

<https://aws.amazon.com/solutions/case-studies/finra/>

FINRA, the primary regulatory agency for broker-dealers in the US, uses AWS extensively in their IT operations and has migrated key portions of its technology stack to AWS including Market Surveillance and Member Regulation.

For market surveillance, each night FINRA loads approximately **35 billion rows of data into Amazon S3 and Amazon EMR (up to 10,000 nodes)** to monitor trading activity on exchanges and market centers in the US.

# REAL-TIME ANALYTICS

understanding the present

# Hearst

<https://aws.amazon.com/solutions/case-studies/hearst-data-anayltics/>

Hearst Corporation is one of the largest diversified communications company in the world.

The company began migrating **10 of its 29 global data centers** to AWS to reduce its IT infrastructure footprint.

Hearst Corporation monitors trending content on **250+ sites worldwide**. To facilitate this, Hearst built a clickstream analytics platform on AWS that transmits and processes over **30 TB of data a day** using AWS resources.

# LARGE-SCALE SIMULATIONS

looking at possible futures



# Kellogg Company

<https://aws.amazon.com/solutions/case-studies/kellogg-company/>

Kellogg needed a solution that could accommodate terabytes of data, scale according to infrastructure needs, and stay within its budget.

Amazon Web Services (AWS) offered **a fully SAP-certified HANA environment** on a public cloud platform. Kellogg decided to start immediately with test and development environments for its US operations.

These Amazon EC2 instances process **16 TB of sales data weekly** from promotions in the US, modeling **dozens of data simulations a day**.

# AON Benfield

<https://aws.amazon.com/solutions/case-studies/aon/>

Aon Benfield is the world's leading reinsurance intermediary and full-service capital advisor.

Aon Benfield Analytics offers industry-leading catastrophe management, actuarial, rating agency advisory and risk and capital strategy expertise.

By using AWS GPU instances, Aon Benfield is able to perform actuarial calculations with **greater computing power**, in **shorter time frames**, and for **less cost** than on-premise deployments and CPU cores: *"Using AWS helps us reduce a 10-day process to 10 minutes"*



# MACHINE LEARNING

using the past to predict the future

# ICAO



The International Civil Aviation Organization (ICAO) is a specialized agency of the United Nations. ICAO collects accident and incident data every day, including information about the location, aircraft, operator, flight phase, as well as narratives around these data points.

Amazon Machine Learning is used to determine the **classification** and **risk categorization** of each event based on **patterns** contained in the flight phase information and the narratives.

Amazon ML ingests the raw data and the narratives every day to provide ICAO a daily list of accidents categorized by risk. This allows ICAO to provide updated accidents statistics to its customers on a daily basis.

# Fraud.net

<http://aws.amazon.com/fr/solutions/case-studies/fraud-dot-net/>

*"We considered five other platforms, but Amazon Machine Learning was the best solution.*

*Amazon keeps the effort and resources required to build a model to a **minimum**.*

*Using Amazon Machine Learning, we've quickly created and trained **a number of specific, targeted models**, rather than building a single algorithm to try and capture all the different forms of fraud."*

Whitney Anderson,  
CEO



# DEEP LEARNING

Finding complex patterns

# Tech 2015: Deep Learning And Machine Intelligence Will Eat The World



**Anthony Wing Kosner,** CONTRIBUTOR

*Quantum of Content and innovations in user experience*

[FOLLOW ON FORBES \(924\)](#)



Opinions expressed by Forbes Contributors are their own.

FULL BIO ▼

Despite what [Stephen Hawking](#) or Elon Musk say, [hostile Artificial Intelligence](#) is not going to destroy the world anytime soon. What is certain to happen, however, is the continued ascent of the practical applications of AI, namely deep learning and machine intelligence. The word is spreading in all corners of the tech industry that the biggest part of big data, the unstructured part, possesses learnable patterns that we now have the computing power and algorithmic leverage to discern—and in short order.

# Amazon DSSTNE (aka 'Destiny')

- Deep Scalable Sparse Tensor Network Engine
- Open source software library for deep neural networks using GPUs : <https://github.com/amznlabs/amazon-dsstne>
- Used by Amazon.com for product recommendation
- Multi-GPU scale for training and prediction
- Large Layers: larger networks than are possible with a single GPU
- Sparse Data: optimized for fast performance on sparse datasets
- Can run locally, in a Docker container or on AWS (multi-instance possible)

# Amazon Destiny vs Google TensorFlow

## First DSSTNE Benchmarks TLDR: Up to Almost 15x Faster than TensorFlow

<https://medium.com/@scottlegrand/first-dsstne-benchmarks-tldr-almost-15x-faster-than-tensorflow-393dbeb80c0f>

*“DSSTNE on a single virtualized K520 GPU (released in 2012) is faster than TensorFlow on a bare metal Tesla M40 (released in 2015)”*

*“TensorFlow does not provide the automagic model parallelism provided by DSSTNE”*



# Demo: Amazon Destiny

<http://grouplens.org/datasets/movielens/>

27,000 movies

138,000 users

20 million movie recommendations

(Matrix is 99.5% sparse)

→ Start a **g2.8xlarge** instance

4 K520 GPUs: **6144** CUDA cores

→ Train a neural network

Input & output layers: **27,000** neurons

3 hidden layers, **128** neurons each

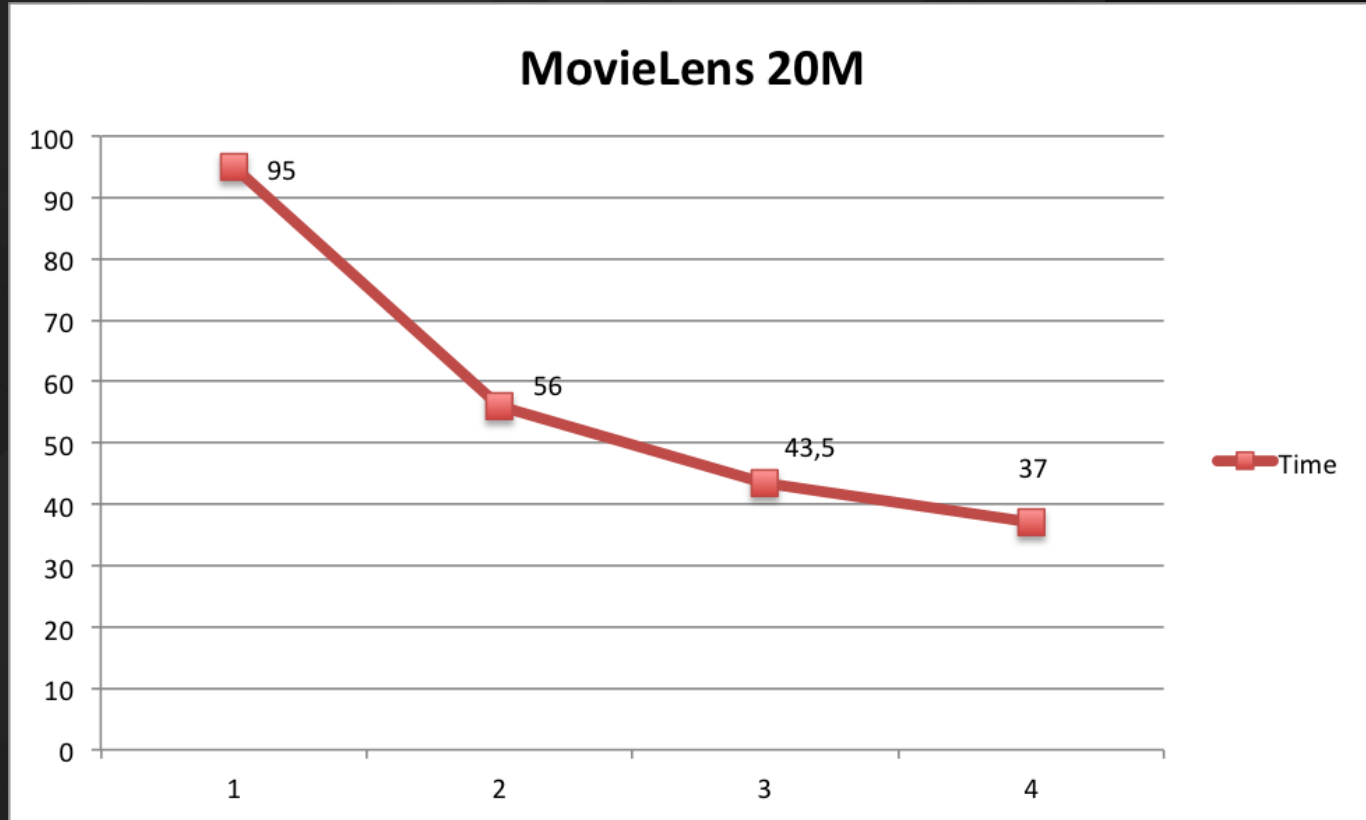
→ Recommend 10 movies per user

	Alice	Bob	Charlie	David	Ernest
Star Wars		1		1	1
Lord of the Rings	1		1		1
Inception		1			
Bambi	1			1	
Pretty Woman	1				1



	Alice	Bob	Charlie	David	Ernest
Star Wars	0.23	1	0.12	1	1
Lord of the Rings	1	0.34	1	0.89	1
Inception	0.8	1	0.43	0.76	0.45
Bambi	1	0.42	0.5	1	0.34
Pretty Woman	1	0.09	0.67	0.04	1

# Training Amazon Destiny on multiple GPUs





The technology  
platform of choice

# Resources

AWS Big Data blog: <https://blogs.aws.amazon.com/bigdata/>

Big Data Architectural Patterns and Best Practices on AWS

<https://www.youtube.com/watch?v=K7o5OIRLtvU>

Real-World Smart Applications With Amazon Machine Learning

<https://www.youtube.com/watch?v=sHJx1KJf8p0>

Deep Learning: Going Beyond Machine Learning

<https://www.youtube.com/watch?v=Ra6m70d3t0o>

# Thank you!

Julien Simon

Principal Technical Evangelist, AWS

@julsimon

