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, Chairman & CEO, General Electric

"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."

John Deere

https://www.youtube.com/watch?v=uq4kQPsM4cQ

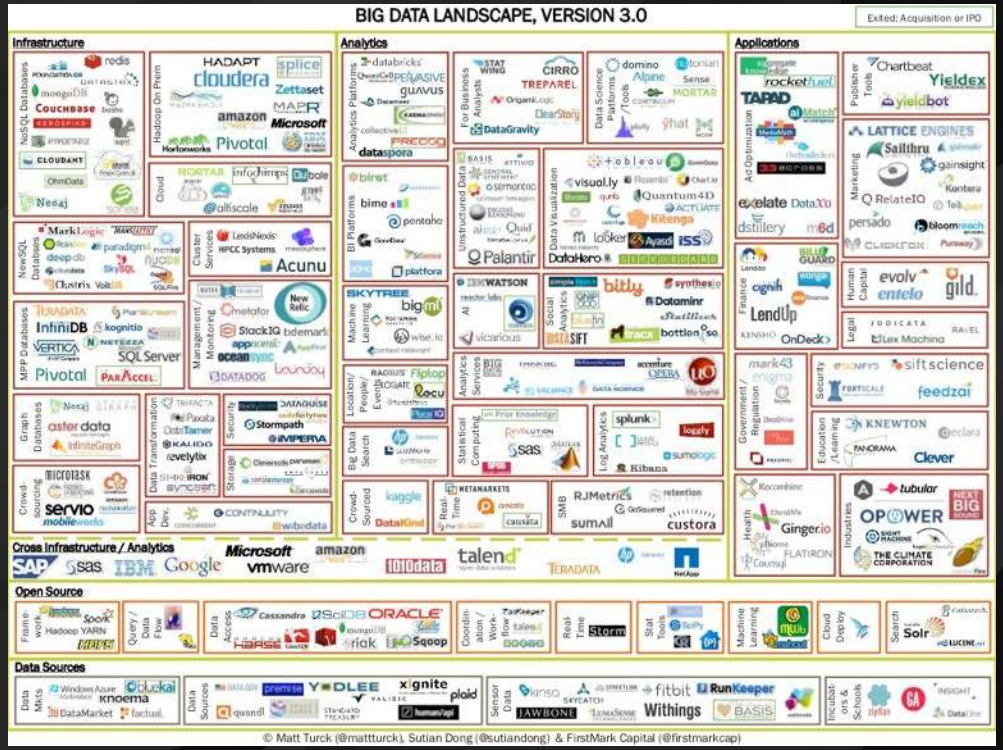
Streaming, analysis, storage and visualization of data coming from 200,000 farming machines

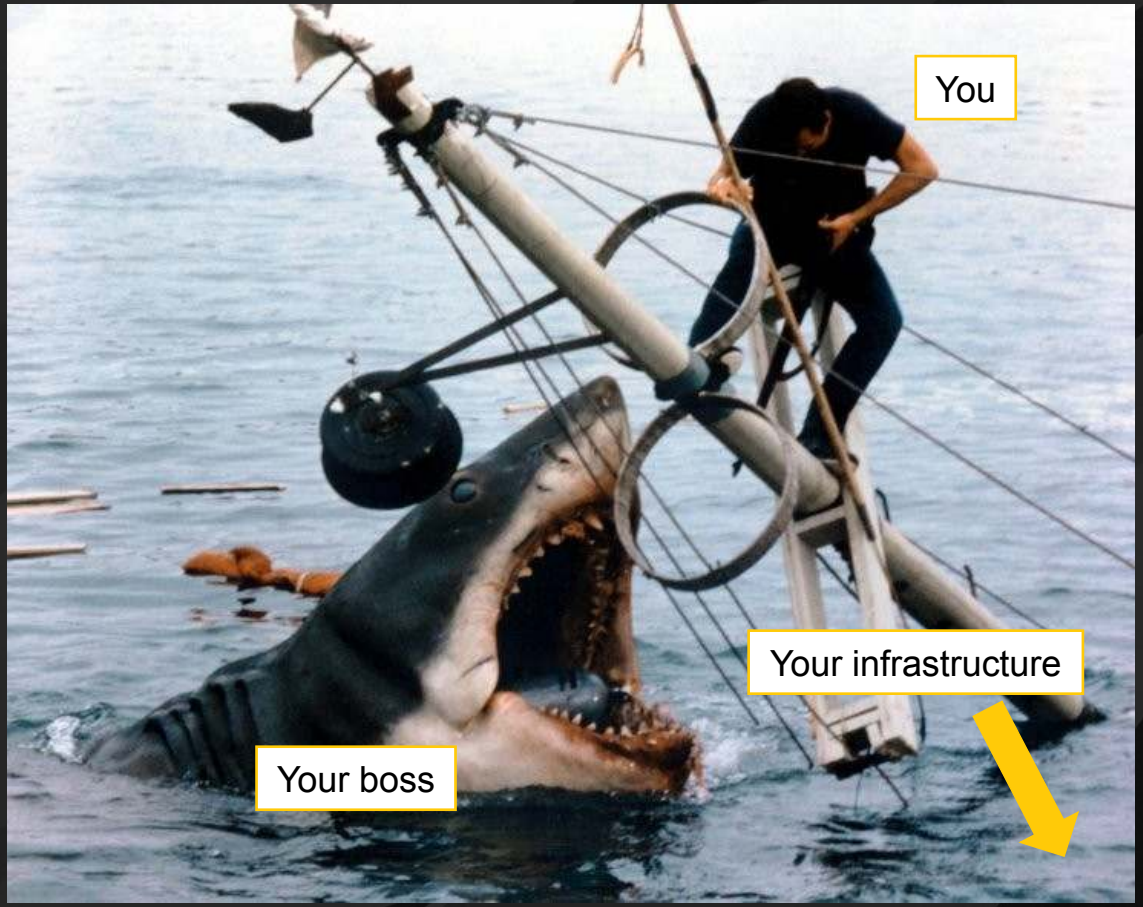
Precision agriculture and yield optimization for farmers

New business model for John Deere

Big Data: the naive view

Let's sail the seven seas of Big Data, we have a map!

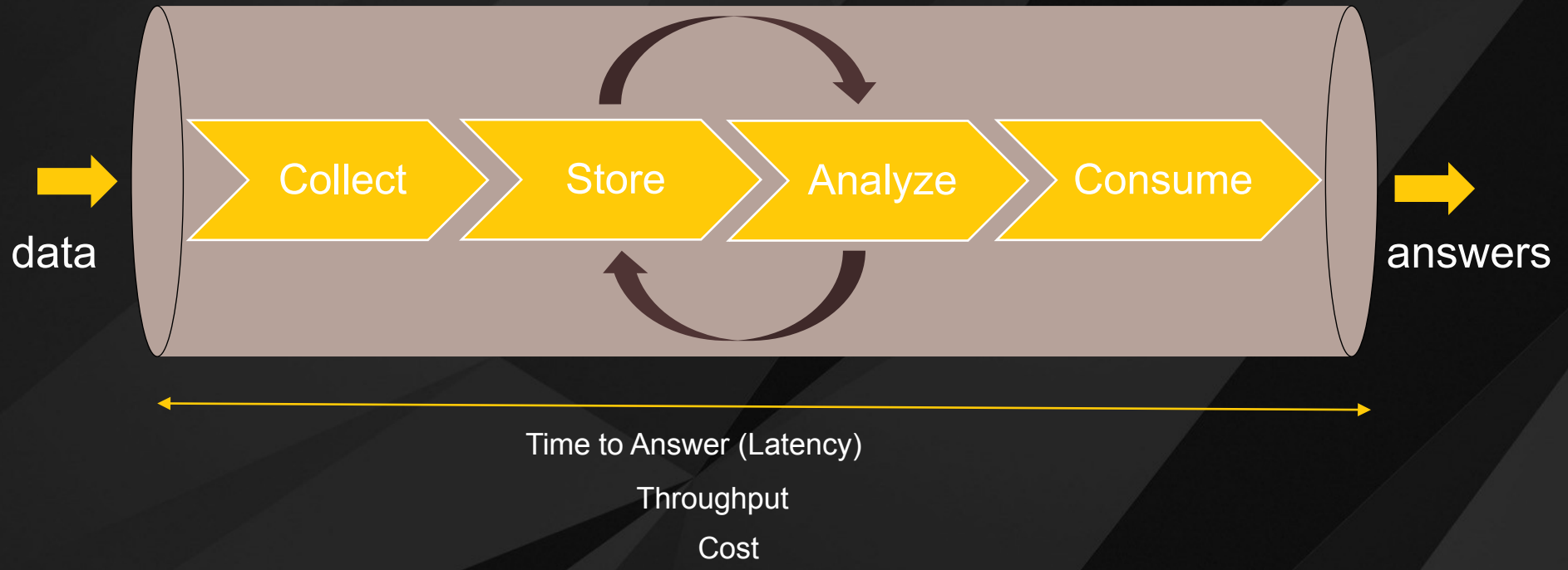


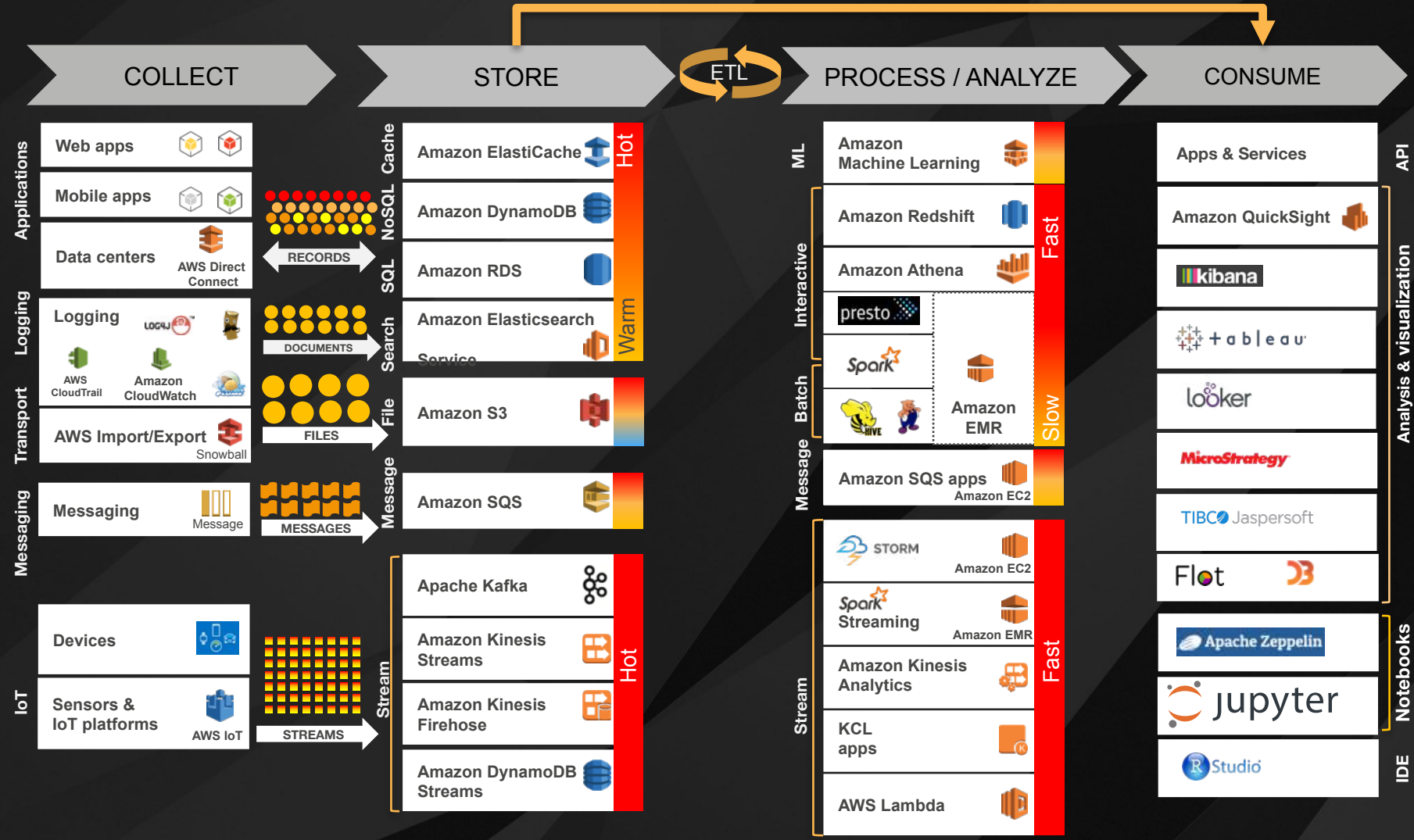

You
Your infrastructure
Your boss

Big Data: the vendor view

One solution to rule them all! … and in darkness bind you?

Big Data: the customer view




COLLECT
STORE
ETL
PROCESS / ANALYZE
CONSUME
Applications
Web apps
Mobile apps
Data centers
AWS Direct Connect
Logging
Logging
AWS CloudTrail
Amazon CloudWatch
Transport
AWS Import/Export
Snowball
Messaging
Messaging
Message
IoT
Devices
Sensors & IoT platforms
AWS IoT
RECORDS
DOCUMENTS
FILES
MESSAGES
STREAMS
Cache
Amazon ElastiCache
NoSQL
Amazon DynamoDB
SQL
Amazon RDS
Search
Amazon Elasticsearch Service
File
Amazon S3
Message
Amazon SQS
Hot
Warm
Stream
Apache Kafka
Amazon Kinesis Streams
Amazon Kinesis Firehose
Amazon DynamoDB Streams
Hot
ML
Amazon Machine Learning
Interactive
Amazon Redshift
Amazon Athena
presto
Batch
Spark
HIVE
Amazon EMR
Fast
Slow
Message
Amazon SQS apps
Amazon EC2
Stream
STORM
Amazon EC2
Spark Streaming
Amazon EMR
Amazon Kinesis Analytics
KCL apps
AWS Lambda
Fast
API
Apps & Services
Analysis & visualization
Amazon QuickSight
kibana
tableau
looker
MicroStrategy
TIBCO Jaspersoft
Flot
Notebooks
Apache Zeppelin
jupyter
IDE
RStudio

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 expanded its use of Redshift to its global exchange properties.

The Forrester Wave™
Big Data Warehouse, Q2 2017

MAP REDUCE

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

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, 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.

FINRA estimates it will save up to $20 million annually by using AWS instead of a physical data center infrastructure.

REAL-TIME ANALYTICS

understanding the present

HEARST

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 (web and mobile). 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.

Supercell

https://aws.amazon.com/solutions/case-studies/supercell /

Supercell is one of the largest mobile gaming companies.

They were the first gaming company to reach 100 million daily users.

Using AWS, they collect and process over 45 billion events and 10 TB of data every day.

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

Amazon

amazon.com

Amazon is a leader in practical machine learning solutions and uses it in hundreds of services across its various businesses.

Amazon uses Machine Learning for Customer Support, building models based on recent orders, click-stream, user devices, prime membership usage, recent cases, recent account changes, etc.

The models are used to provide efficient self-service to our customers.

Bonjour Julien. Comment pouvons-nous vous aider ?

Vos commandes

- Suivre votre colis
- Modifier ou annuler des commandes

Retours et remboursements

- Retourner ou échanger des articles
- Imprimer une étiquette de retour

Aide pour les appareils

- Obtenir aide et support pour appareils numériques
- Résoudre les problèmes avec votre appareil

Premium

Amazon Premium

- Avantages d'Amazon Premium
- Annuler votre inscription à Amazon Premium

VISA MasterCard

Options de paiement

- Ajouter ou modifier un moyen de paiement
- Modifier une carte de paiement expirée

Paramètres du compte

- Modifier votre e-mail ou mot de passe
- Mettre à jour vos identifiants de connexion

Cherchez une solution dans nos pages d'Aide

Saisissez un mot-clé puis cliquez sur Go.

Go

Amazon

Not all customer interactions can be solved in a self-service mode. Therefore, Amazon operates large customer support centers where Customer Service Representatives (CSR) handle customer requests.

The machine learning models described above are used to optimize the human interactions of these requests.

For example, they are used to route the customer call to the best CSR before the customer has even started to speak! They are also used again during the call.

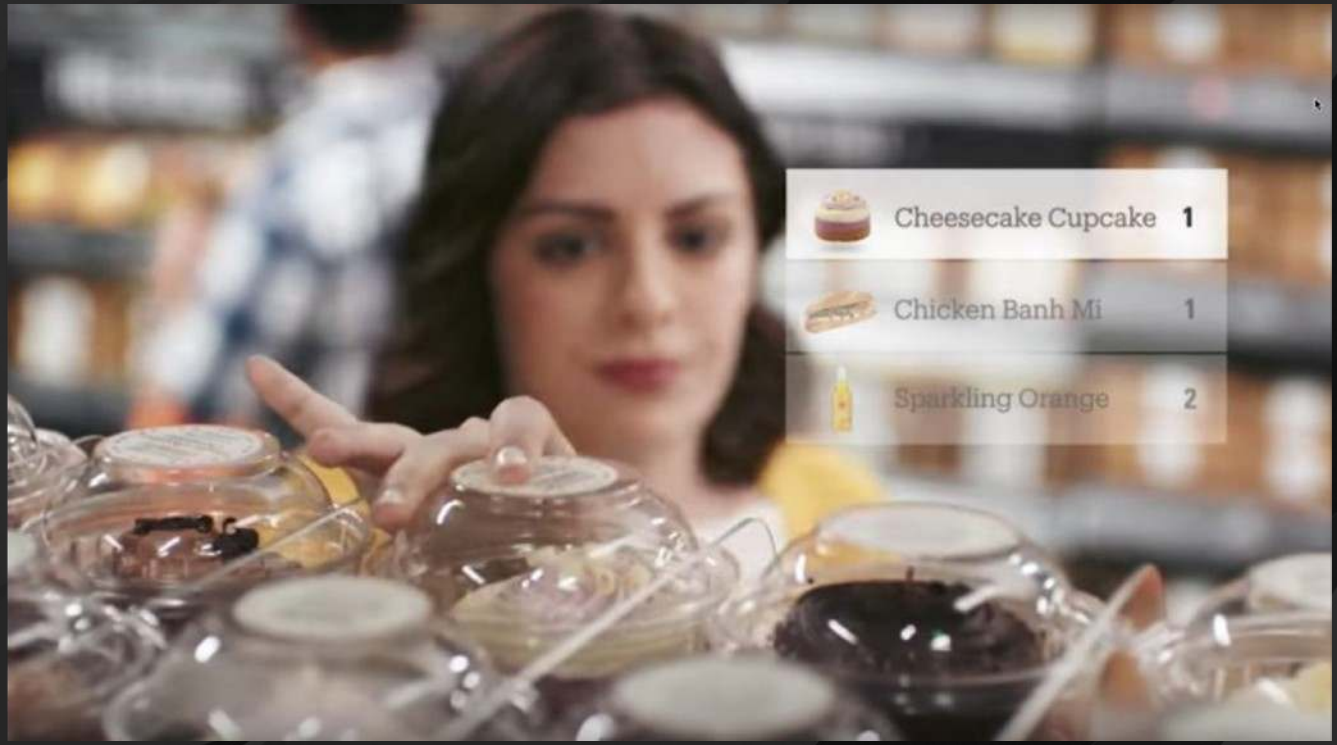
DEEP LEARNING

building human-like systems

Personal assistants

Line-free shopping

amazon go

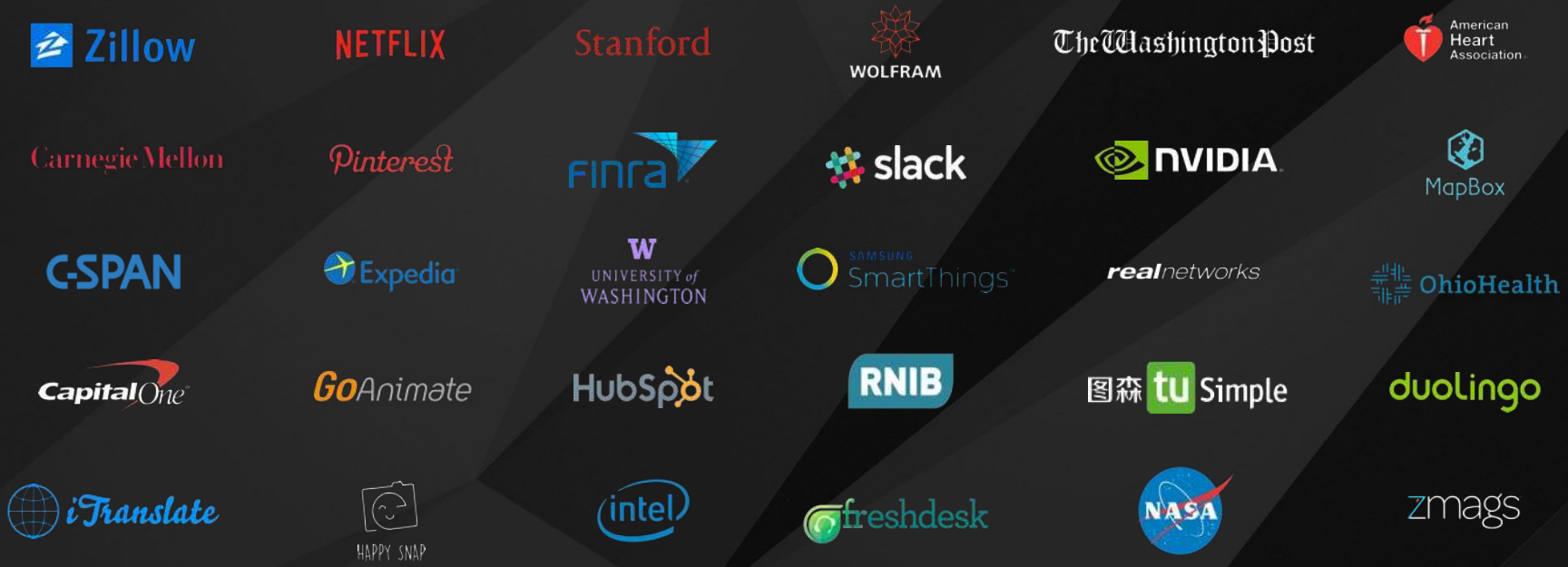


Autonomous driving

https://www.oreilly.com/ideas/self-driving-trucks-enter-the-fast-lane-using-deep-learning

This past June, a driverless truck passed a 200-mile test drive from Yuma, Arizona, to San Diego, California—a milestone for autonomous trucking in the U.S. This feat was achieved by the company TuSimple, which trained its driving system using an AI technique known as deep learning to simulate tens of millions of miles of road driving.

Selected customers running AI on AWS



Resources

Whitepaper: “Big Data Analytics Options on AWS” http://d0.awsstatic.com/whitepapers/Big_Data_Analytics_Options_on_AWS.pdf

Big Data Architectural Patterns and Best Practices on AWS

https://www.youtube.com/watch?v=K7o5OlRLtvU

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

Amazon AI: https://aws.amazon.com/ai

Thank you!

Julien Simon

Principal Technical Evangelist, AWS

@julsimon

aws