



The AWS Big Data combo

Amazon Redshift

Amazon QuickSight

Amazon Machine Learning

Amazon DSSTNE

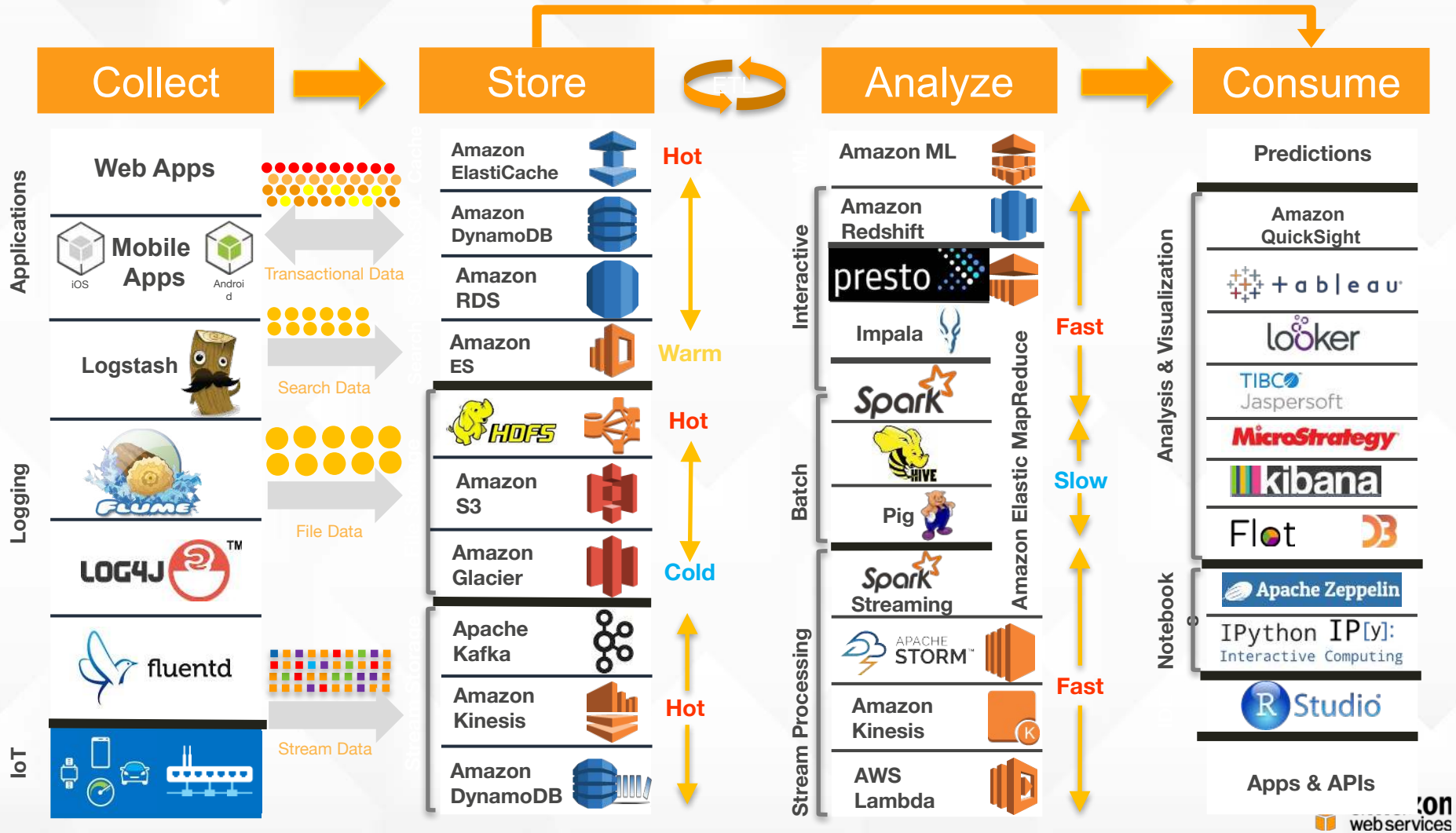


Julien Simon, Principal Technical Evangelist

julsimon@amazon.fr - [@julsimon](https://twitter.com/julsimon)

Agenda

- Data warehouse: Amazon Redshift
- Business Intelligence: Amazon QuickSight
- Prediction models: Amazon Machine Learning
- Recommendation models: Amazon DSSTNE ('Destiny')



Amazon Redshift

a relational, petabyte scale, fully managed data warehousing service

- Postgres SQL is all you need to know
- ODBC and JDBC drivers available
- Parallel processing on multiple nodes
- No system administration
- Free tier: 750 hours / month for 2 months
- Available on-demand from \$0.25 / hour / node
- As low as \$1,000 / Terabyte / year



“ Come for the cost, stay for the performance ”

Amazon Redshift architecture

Parallel processing

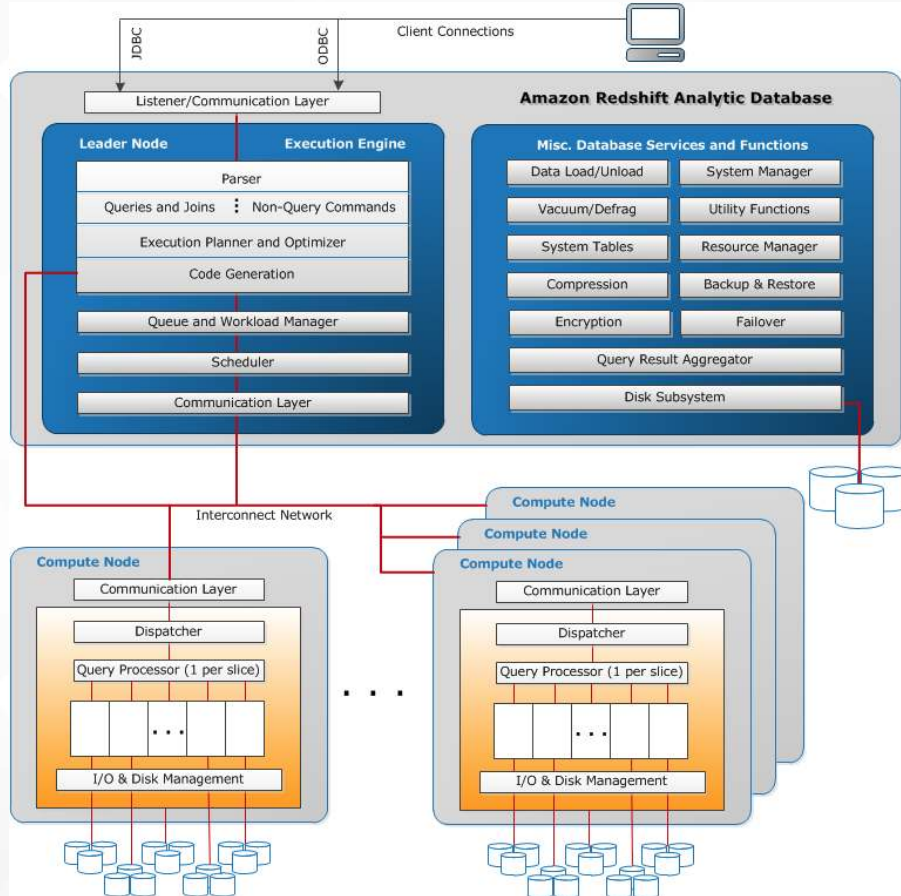
Columnar data storage

Data compression

Query optimization

Compiled code

Workload management



Case study: Financial Times

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



- BI analysis of reader traffic, in order to decide which stories to cover
- Conventional data warehouse running on Microsoft technologies
- Scalability issues, impossible to perform real-time analytics → Amazon Redshift PoC
- Amazon Redshift performed so quickly that some analysts thought it was malfunctioning 😊

John O'Donovan, CTO: *"Amazon Redshift is the single source of truth for our user data."*

"Some of the queries we're running are 98 percent faster, and most things are running 90 percent faster (...) and the ability to try Redshift out before having to invest a significant amount of capital was a huge bonus."

"Being able to explore near-real-time data improves our decision making massively. We can make decisions based on what's happening now rather than what happened three or four days ago."

- TCO divided by 4



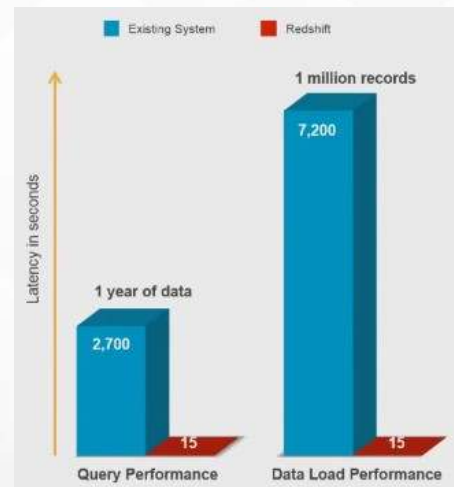
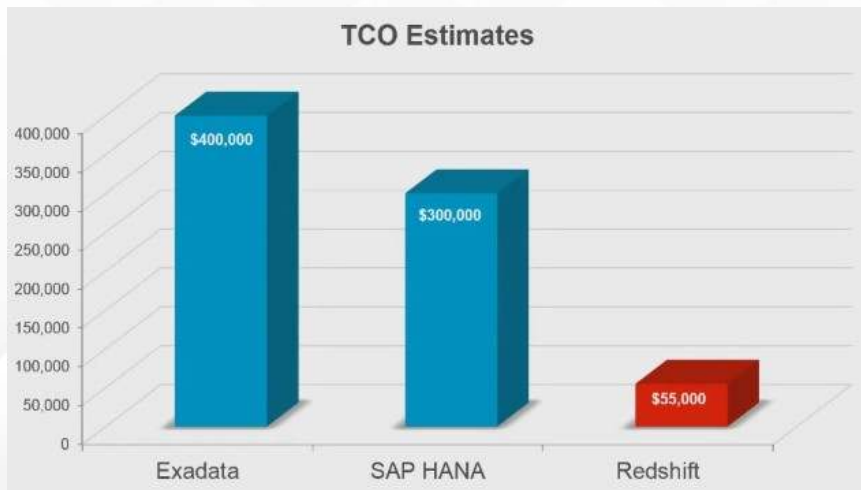
Case study: Boingo

<https://www.youtube.com/watch?v=58URZbp1voY>



- Largest operator of airport wireless hotspots in the world: 1M+ hotspots, 100+ countries
- About 15 TB of data, growing at 2-3 TB per year
- Platform running on SAP (ETL) & Oracle 11g: low performance, heavy admin, high cost
- Evaluated Oracle Exadata, SAP HANA and Amazon Redshift
- Selected Amazon Redshift and migrated in 2 months

6-7x less expensive than alternatives



Queries
180x faster

Data load
480x faster



Amazon Redshift performance

No indexes, no partitioning, no wizardry.

Distribution key

- How data is spread across nodes
- EVEN (default), ALL, KEY

Sort key

- How data is sorted inside of disk blocks
- Compound and interleaved keys are possible

Both are crucial to query performance!



Universal Pictures

DEMO #1

*Demo gods, I'm your humble servant,
please be good to me*

16-node Amazon Redshift cluster (dc1.large)
1 billion lines of CSV data (45GB) in a single table
Run SQL queries



Amazon QuickSight

- Easy exploration of AWS data
 - Fast insights with SPICE
 - Intuitive visualizations with AutoGraph
 - Native mobile experience
 - Sharing and collaboration through StoryBoards
-
- Fully managed: no hardware or software to license
 - \$9 per user per month



Amazon EMR



Amazon RDS



Amazon Kinesis



Amazon S3



Amazon DynamoDB



File Upload



Amazon Redshift

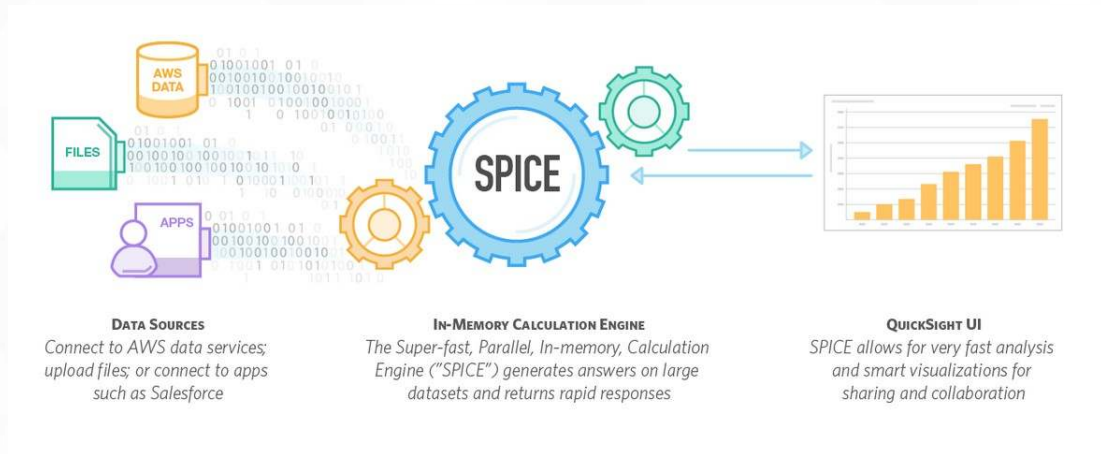


Third Party

SPICE

Super-fast, Parallel, In-memory optimized Calculation Engine

- 2x to 4x compression columnar data
- Compiled queries
- Rich calculations with SQL-like syntax
- Very fast response time



DEMO #2

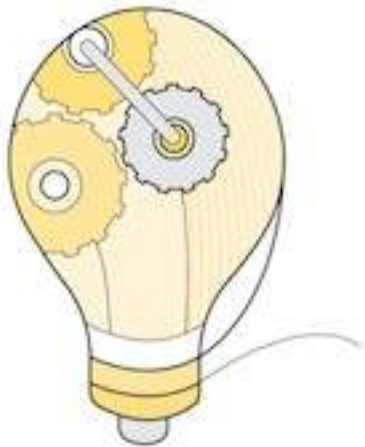
*Demo gods, I know QuickSight is still in preview,
but I need it to work, ok?*

Explore our Redshift data
with Amazon QuickSight



Amazon Machine Learning

*A managed service for building ML models
and generating predictions*



Integration with Amazon S3, Redshift and RDS
Data transformation, visualization and exploration
Model evaluation and interpretation tools
API for batch and real-time predictions

\$0.42 / hour for analysis and model building (eu-west-1)
\$0.10 per 1000 batch predictions
\$0.0001 per real-time prediction

Amazon ML prediction algorithms

- Binary attributes → binary classification
- Categorical attributes → multi-class classification
- Numeric attributes → linear regression
- Code samples on <https://github.com/aws-labs/machine-learning-samples>

Case study: BuildFax

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

BuildFax:
On-Demand Property Condition.

More than
8,000
Cities &
Counties

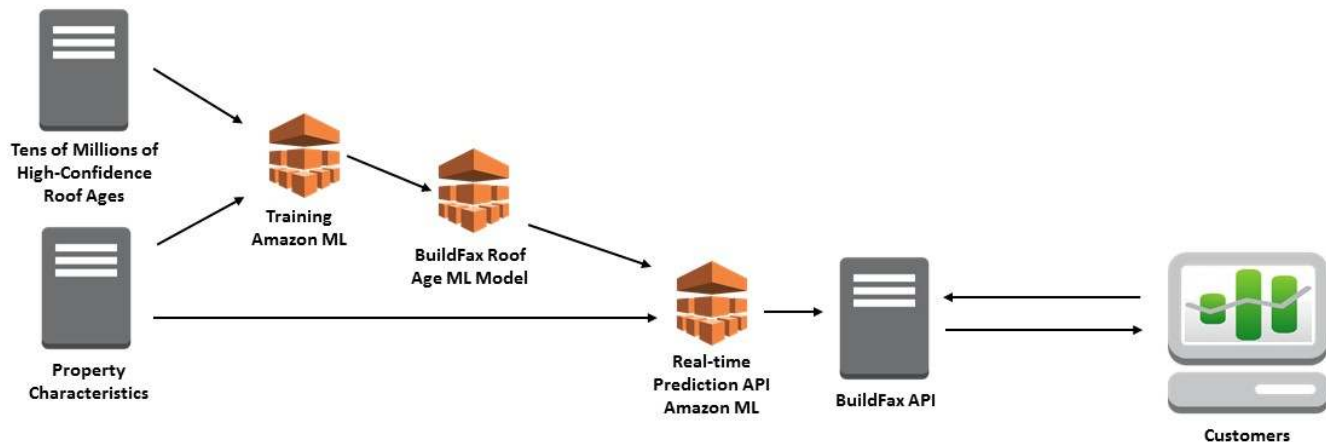
90+
MILLION
Properties

185+
MILLION
Building
Permits

10+
BILLION
Data Points

“Amazon Machine Learning democratizes the process of building predictive models. It's easy and fast to use, and has machine-learning best practices encapsulated in the product, which lets us deliver results significantly faster than in the past”

Joe Emison, Founder &
Chief Technology Officer



Case study: 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



DEMO #3

*Demo gods, I know I'm pushing it,
but please don't let me down now*

Load data from Amazon Redshift

Train and evaluate a regression model with Amazon ML

Create a real-time prediction API

Perform real-time predictions from a Java app



Amazon DSSTNE (aka 'Destiny')

- Deep Scalable Sparse Tensor Network Engine
- Open source software library for training and deploying 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 with GPU instances

DEMO #4

Demo gods, just 10 more minutes...

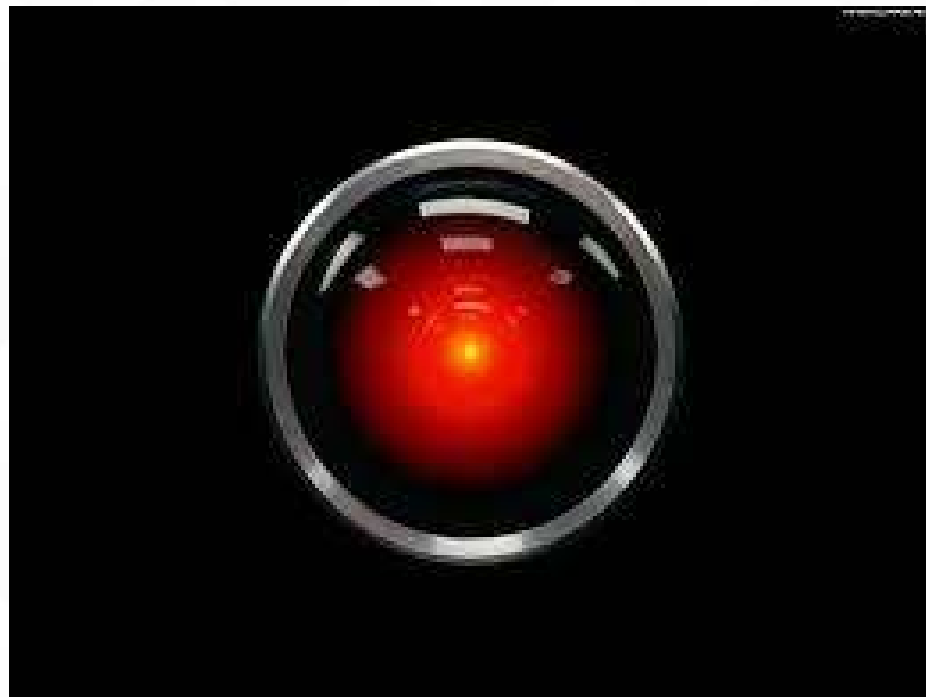
Create a GPU instance

Build the DSSTNE library

Load the MovieLens data set

Train a recommendation model

Compute movie recommendations



AWS User Groups



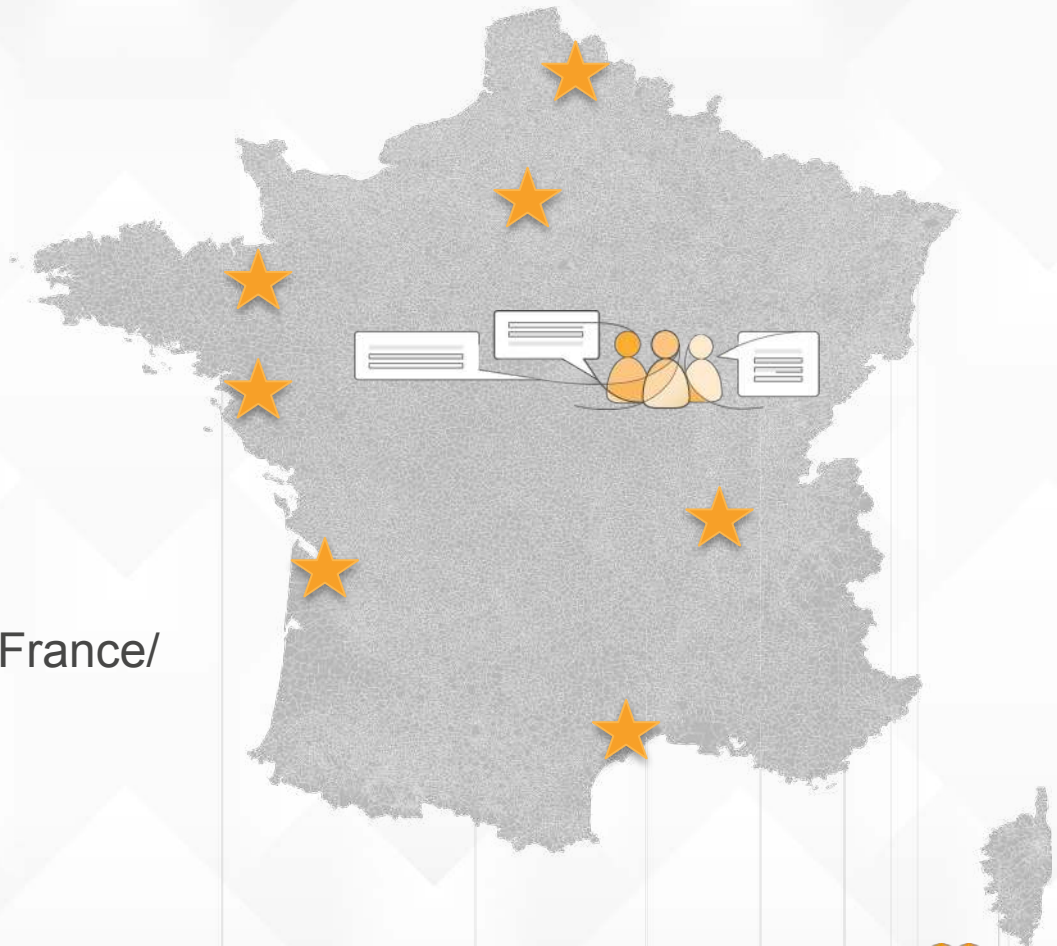
Lille
Paris
Rennes
Nantes
Bordeaux
Lyon
Montpellier



facebook.com/groups/AWSFrance/



[@aws_actus](https://twitter.com/aws_actus)



Save the date: AWS Paris Summit, 31/05/2016

<https://aws.amazon.com/fr/summits/paris/>





Thank you !

Julien Simon, Principal Technical Evangelist

julsimon@amazon.fr - [@julsimon](https://twitter.com/julsimon)