Amazon Redshift,
a data warehouse for the masses

Julien Simon

Principal Technical Evangelist, AWS

@aws_actus @julsimon

Meetup Les Nouvelles Organisations
11/02/2016, Paris

Amazon Redshift

a relational, petabyte scale, fully managed data warehousing service

- 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 "

Case study: Photobox

Maxime Mezin, Data & Photo Science Director:

"*L'entrepôt de données ne comportait que les données du site e-commerce liées aux ventes. Alors que nous avions la volonté d'intégrer des données du service clients et des données d'analyse (…) Nous avions atteint la limite du stockage de la base Oracle, et cela ne marchait pas très bien en termes de performances.*"

"*Avec Redshift, la rapidité d'exécution des traitements a été multipliée par 10. Sans parler de la vitesse de chargement des données.*"

"*On paie en fonction de la quantité de données que l'on va stocker. Chez Google, cela était plus compliqué.*"

- 2 Redshift clusters : 1 for historical data, 1 for real-time processing
- Total Cost of Ownership divided by 7 (90K€→13K€)

Case study: 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.*"
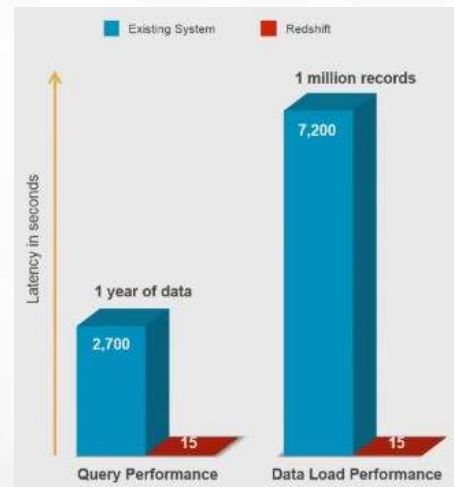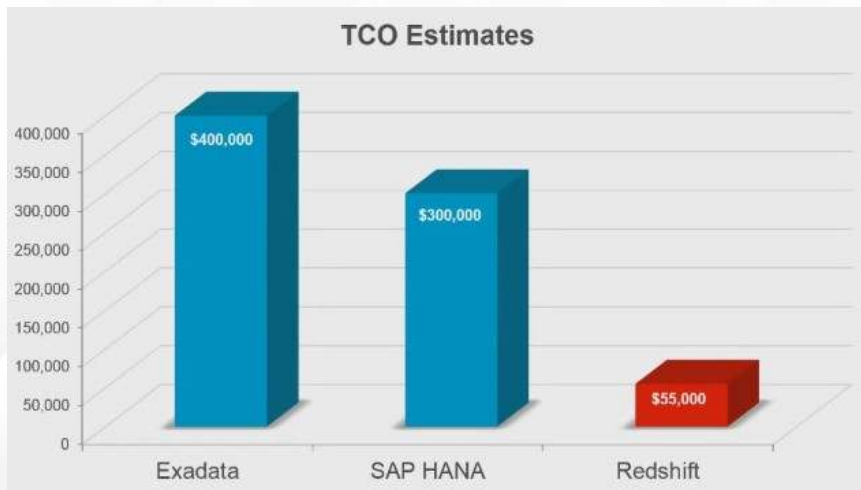
- Total Cost of Ownership divided by 4

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

DEMO

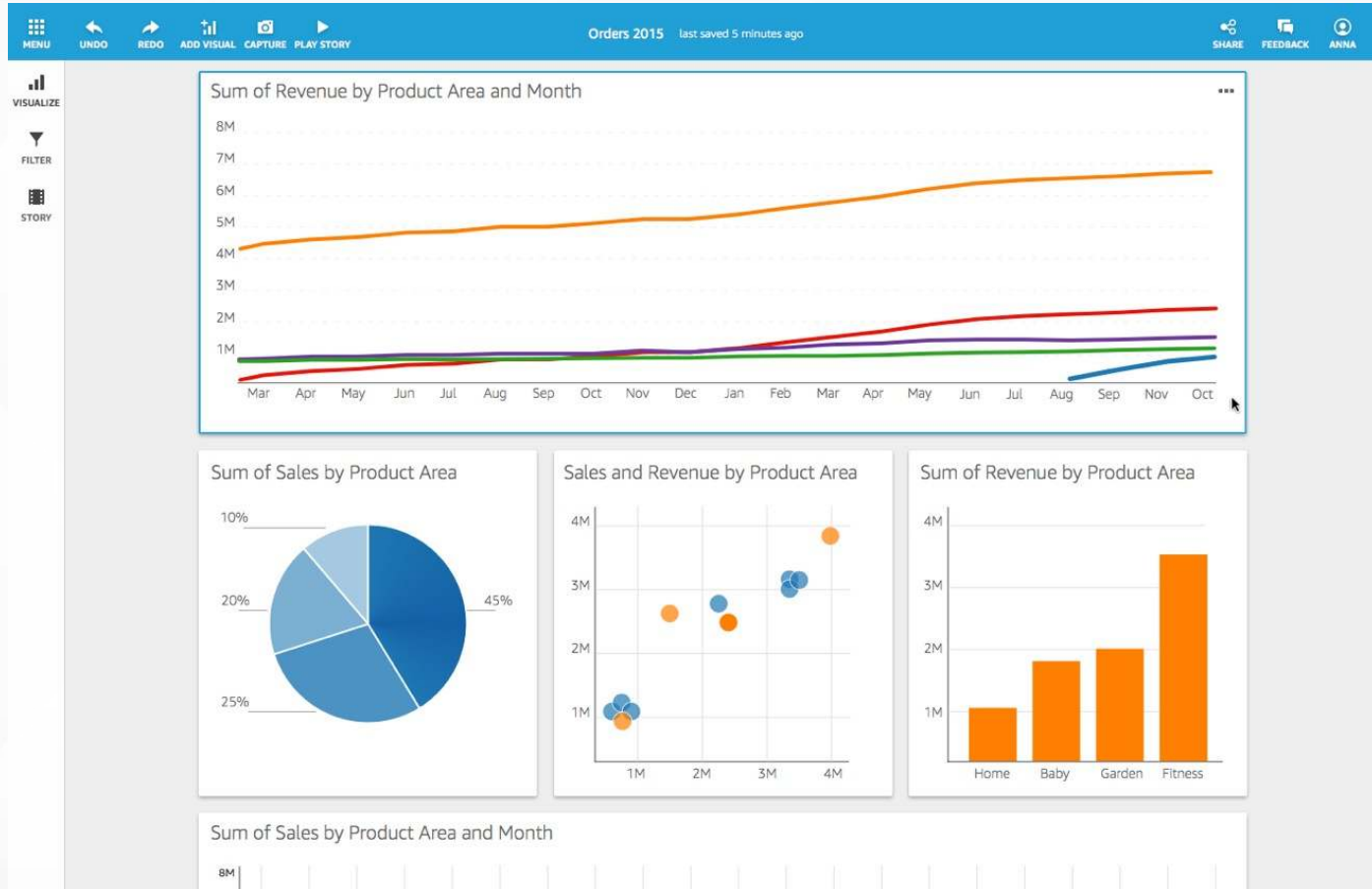*Demo gods, I'm your humble servant, please be good to me*

4-node cluster
1 billion lines of CSV data (45GB)
→ 1 billion rows in a single table

Amazon QuickSight

Thank you! Let's keep in touch

Many events and meetups in 2016 all across France
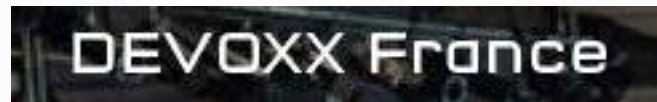
@aws_actus @julsimon

March 7-8

March 23-24

April 6-7 (Lyon)

April 20-22

April 25

AWS Summit
May 31st

AWS User Groups in Paris,
Lyon, Nantes, Lille & Rennes
(meetup.com)