
Data the final frontier
NCC-1701-D
The Light Work

SQREAM TECHNOLOGIES

Getting Big Data Done On a GPU-Based Database

Ori Netzer VP Product

26-Mar-14

Analytics Performance - 3 TB, 18 Billion records

Total Hardware Cost:

$15,000

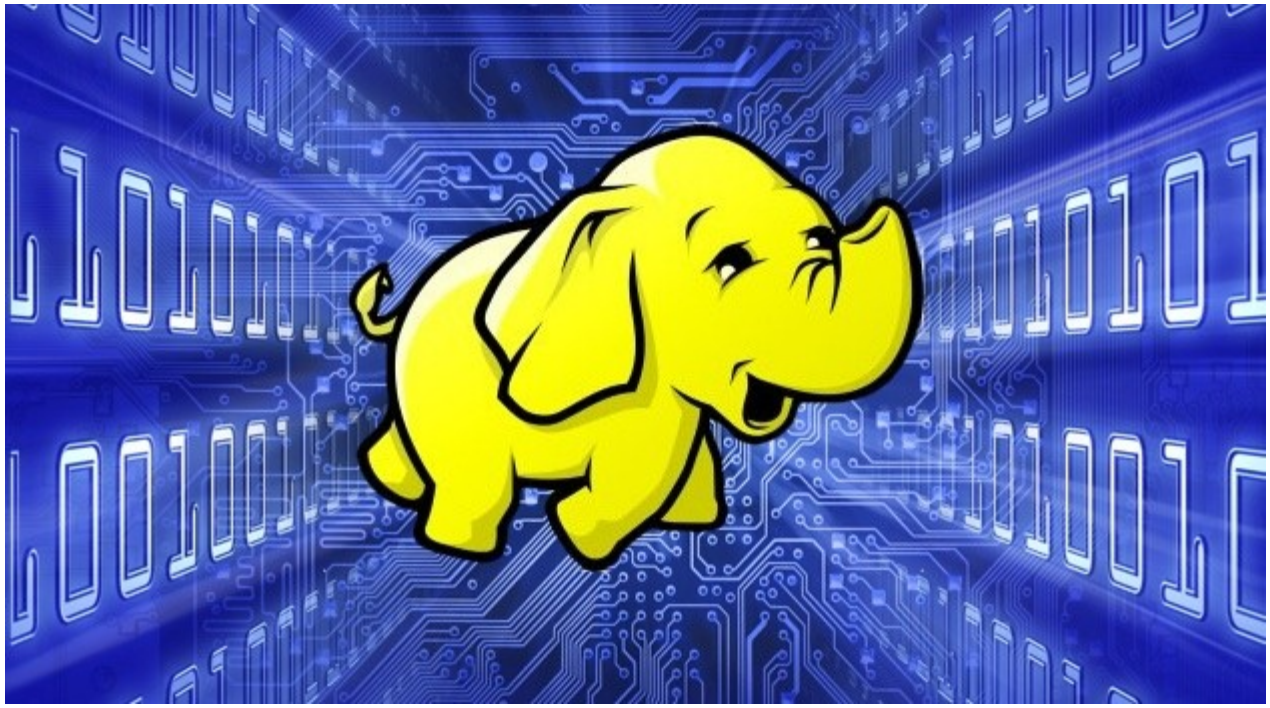
=

Same Runtime

Total Hardware Cost:

$7,000,000

Big Data 3V's - Volume



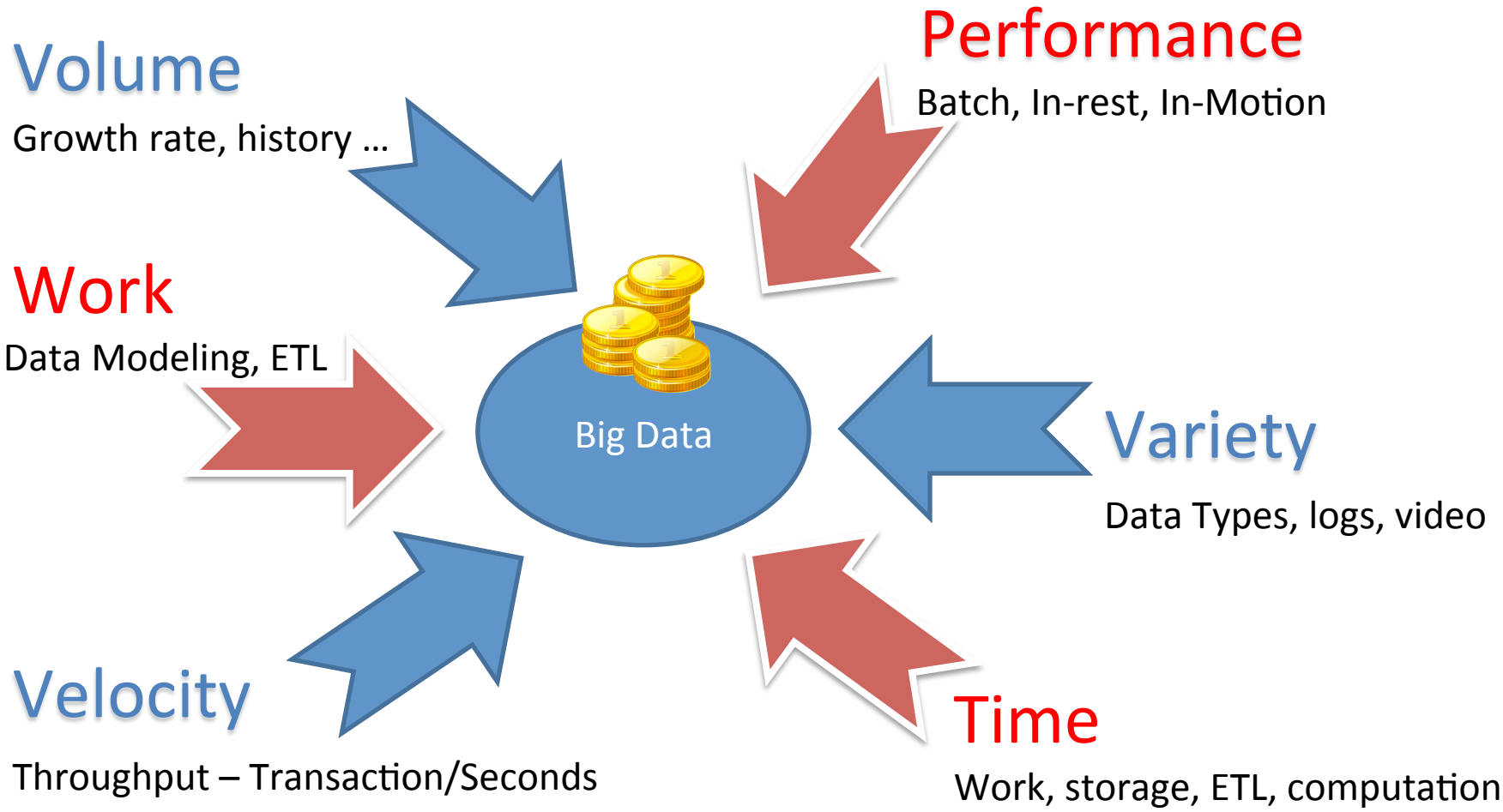
Big Data 3V's – Vareity

Big Data 3V's - velocity

Structured Data

Un-Structured Data

Big Data Challenges



Big Data – Challenges

- Hardware
  - Disk Space
  - Processing Power
  - Network
- Skill Set
  - Large clusters
  - New querying language
  - New methods

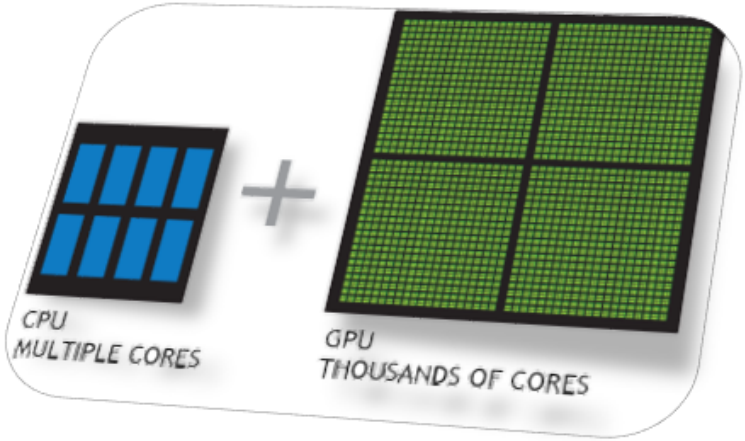
GPU advantages in a nutshell

Massive parallel computing power

Aggressive data crunching

Extreme scalability

Low power consumption

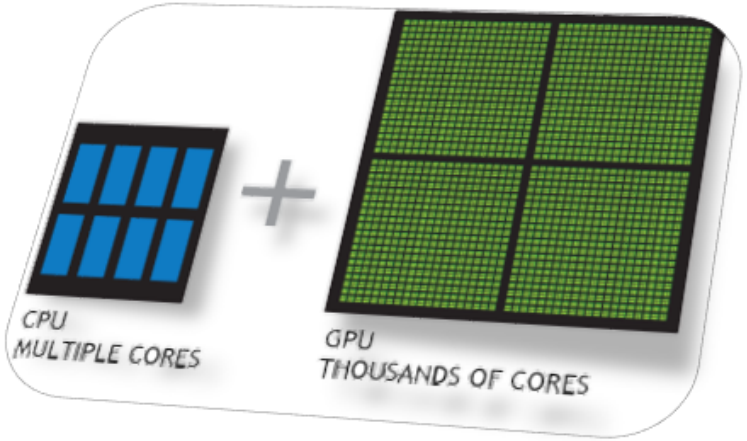


GPU Challenges in a nutshell

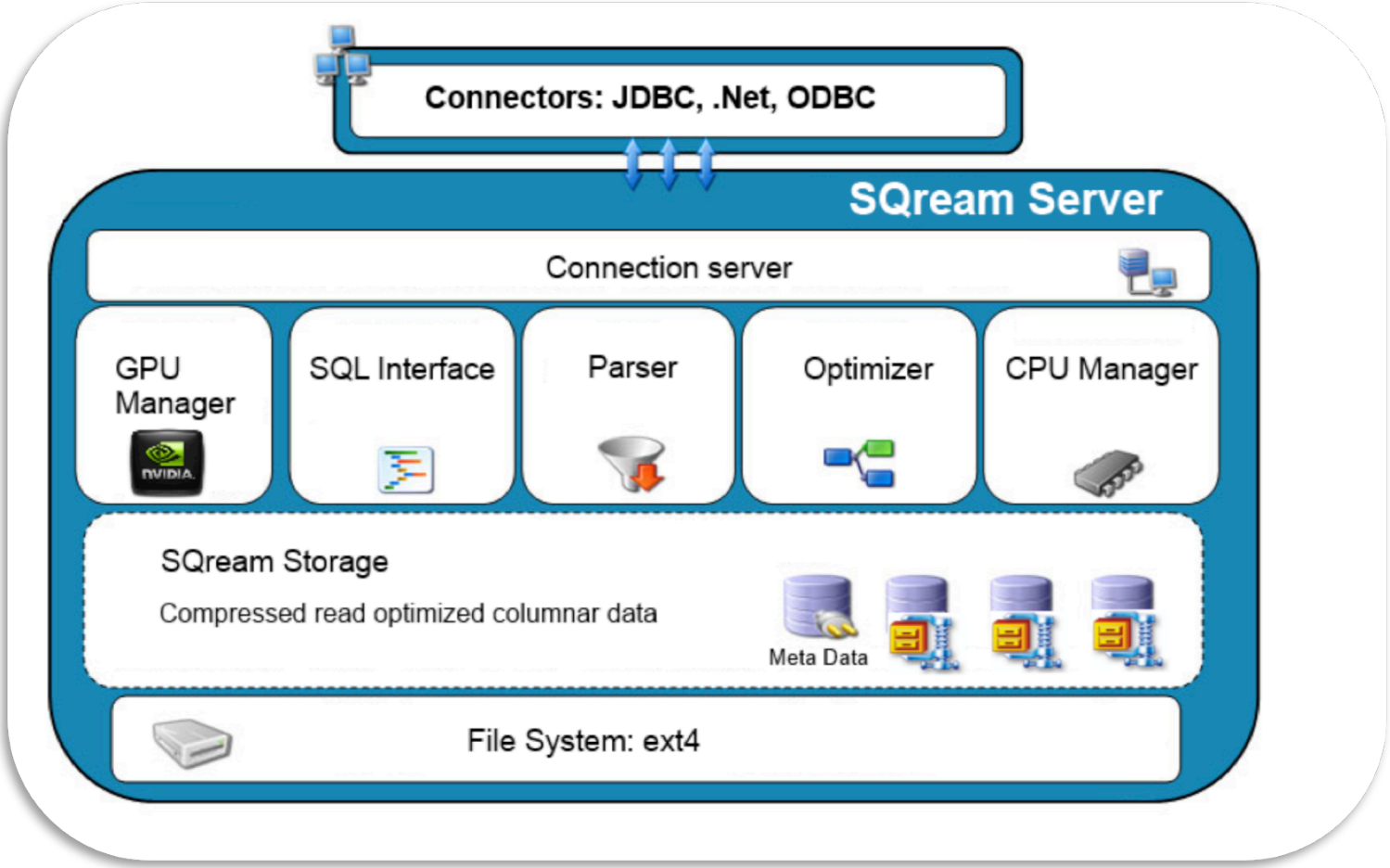
PCI-E

Limited GRAM

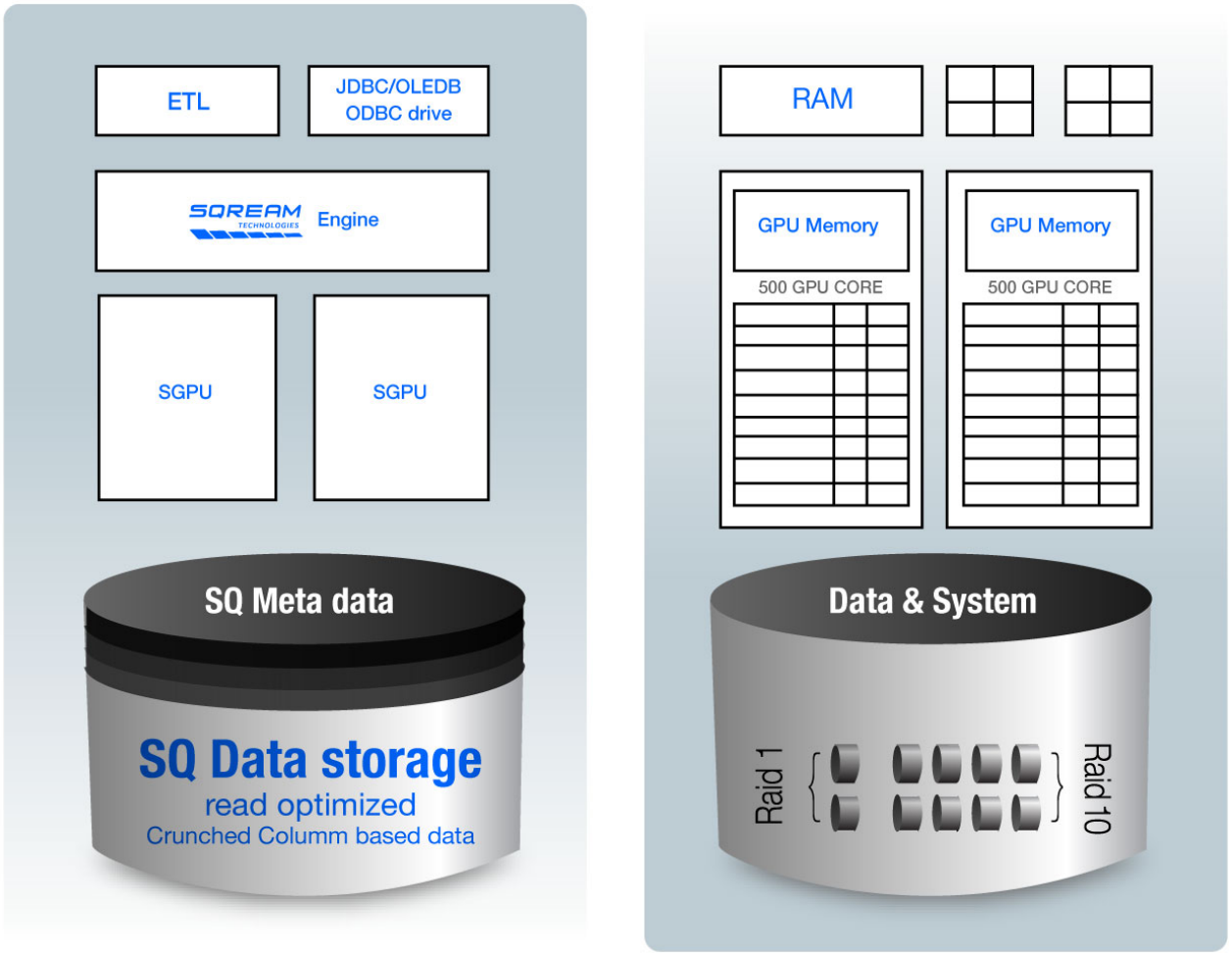
New Algorithms SIMD



SQream Top Level Architecture



The Product - under the hood

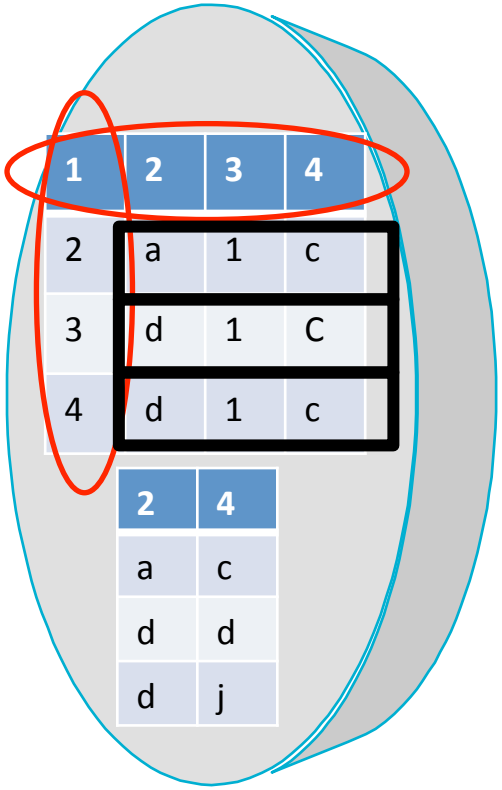


SQREAM TECHNOLOGIES

SQream Data Storage

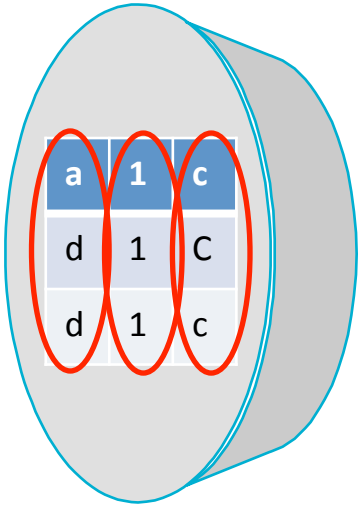
100T

Data in rows + Indices + Views



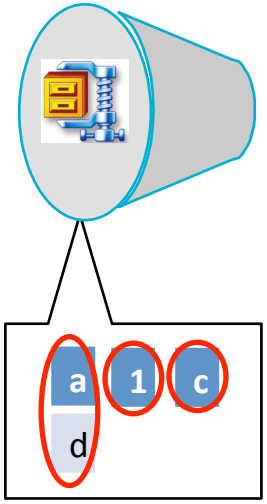
50T

Columnar Data



9T

Crunched Data +
Smart Meta Data



Storage Foot Print

Before 100T

After 9T

DBMS - Query Run Time – No Optimizations

Select T, sum(A) from Example group by T having (A) > 1000

A	B	C	D	E	F
1010	N 1	a	10	25	66
900	N 2	b	54	55	55
5000	N 3	c	754	54	58
500	N 4	f	748	554	85
800	N 5	a	47	55	88
600	N 6	r	58	555	478

CPU

2 x  = **16 cores**

Long execution time

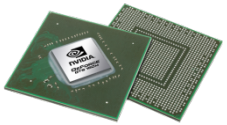
SQream - Query Run Time

Select T, sum(A) from Example group by T having (A) > 1000

Read **Pin Point** data → Process **Relevant** data → Query output

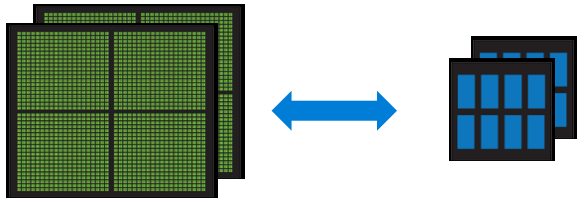
A	B	C	D	E	F
1010	N 1	a	10	25	66
900	N 2	b	54	55	55
5000	N 3	c	754	54	58
500	N 4	f	748	554	85
800	N 5	a	47	55	88
600	N 6	r	58	555	478

GPU

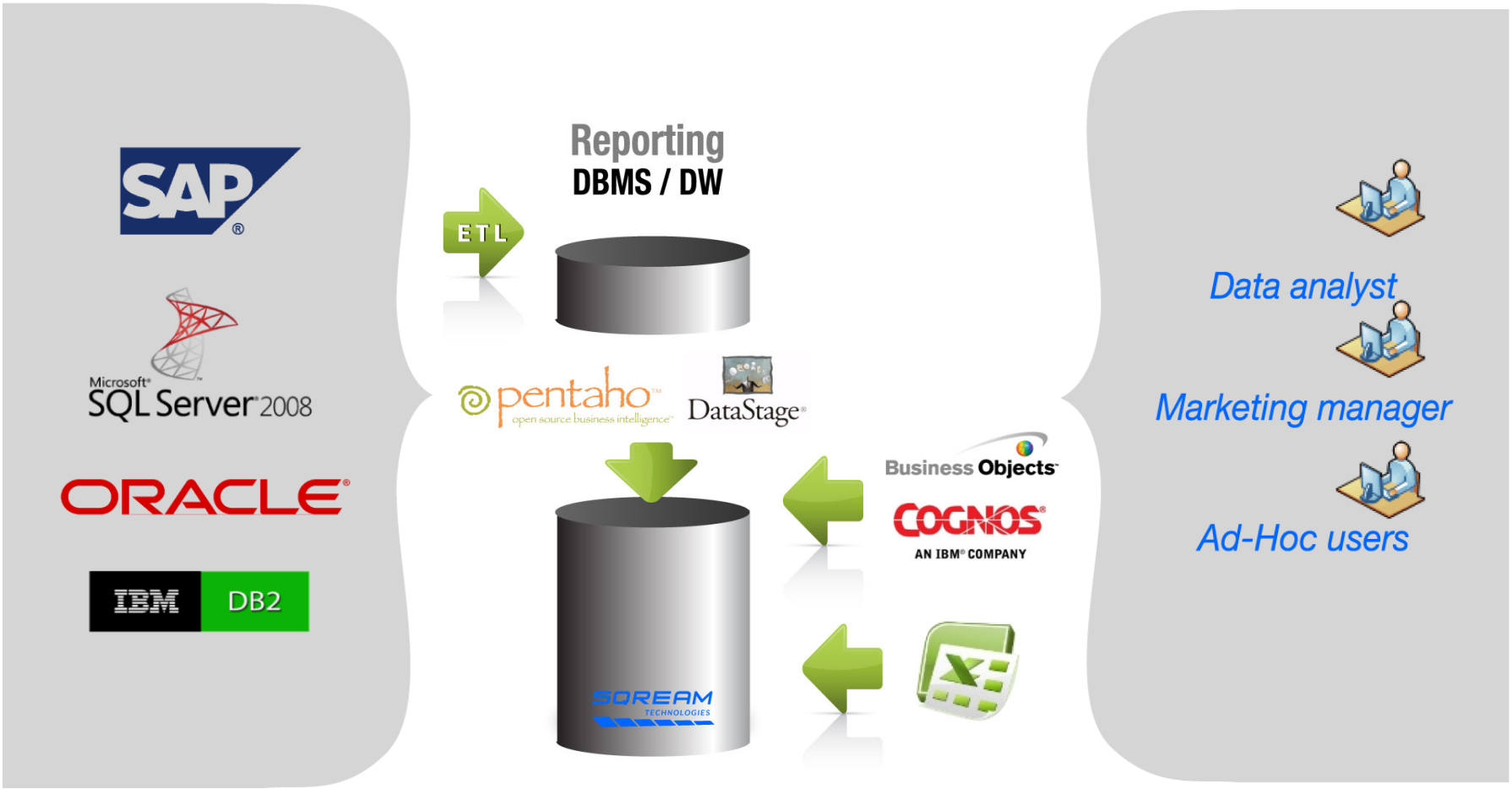
2 x 

= **~5600 cores**

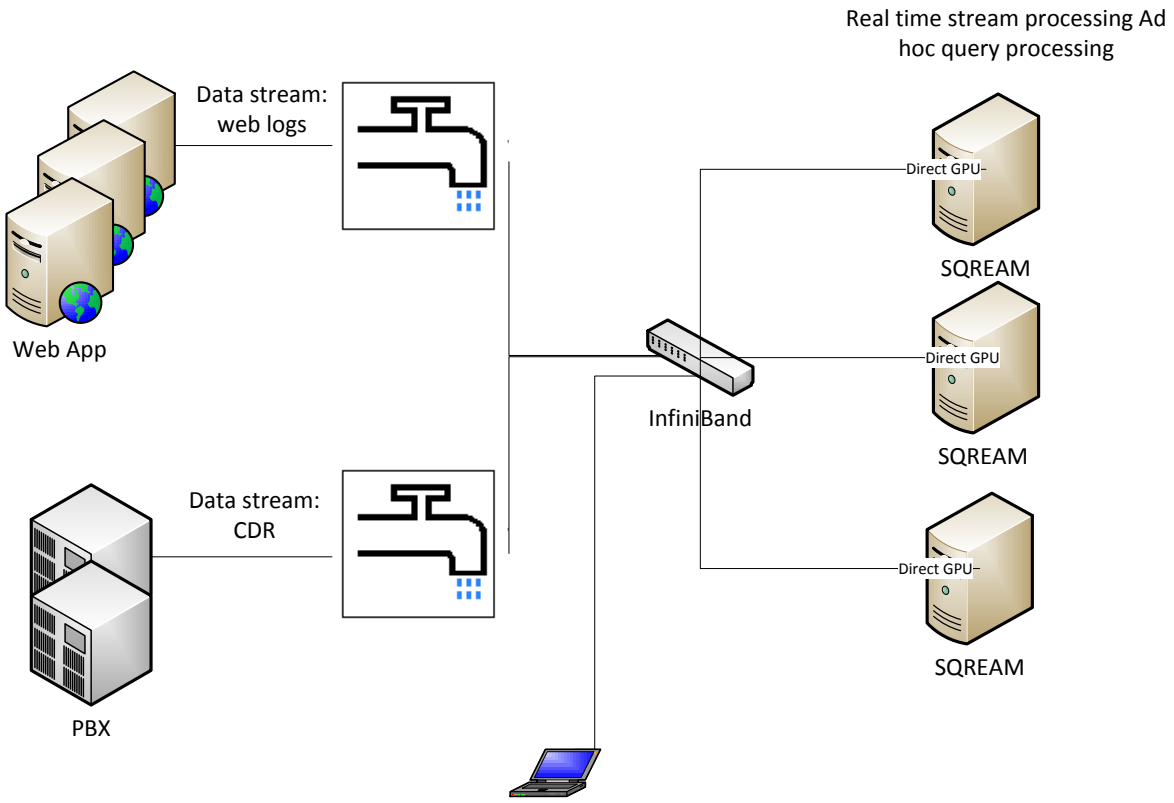
Shortest execution time



SQream as Data Warehouse

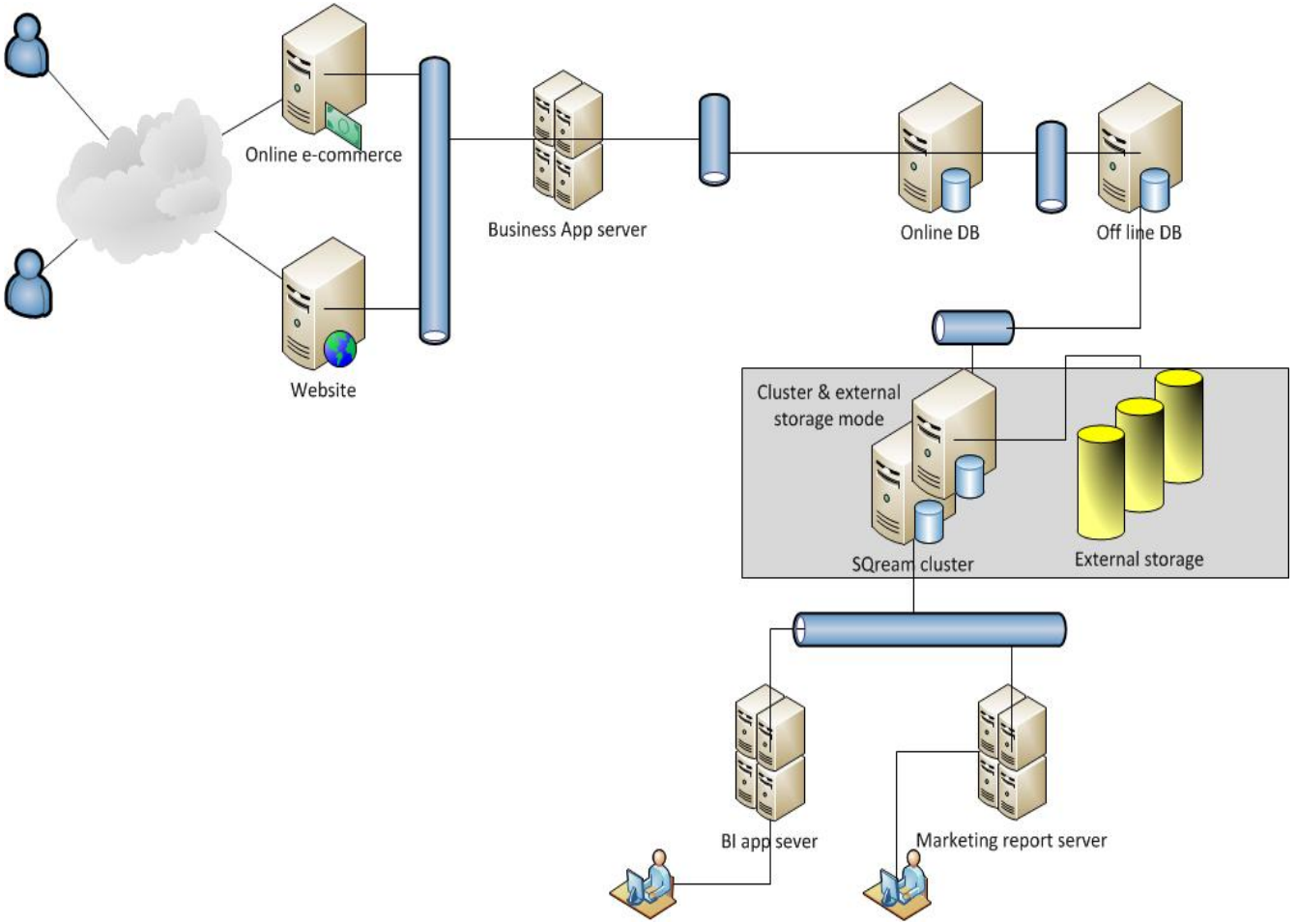


CEP Cluster

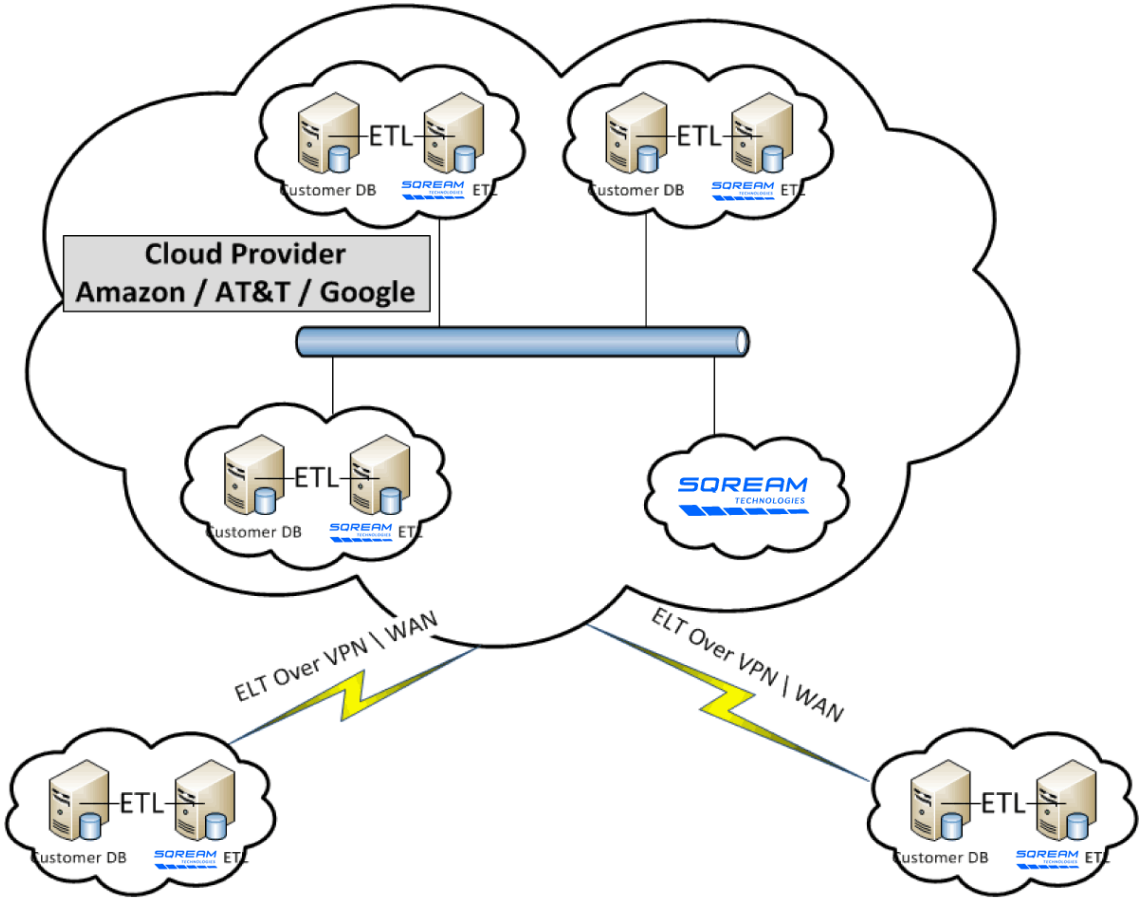


- Apply new query on Data in Motion
- Apply Real time Ad Hoc query on all data in rest

High Availability Cluster

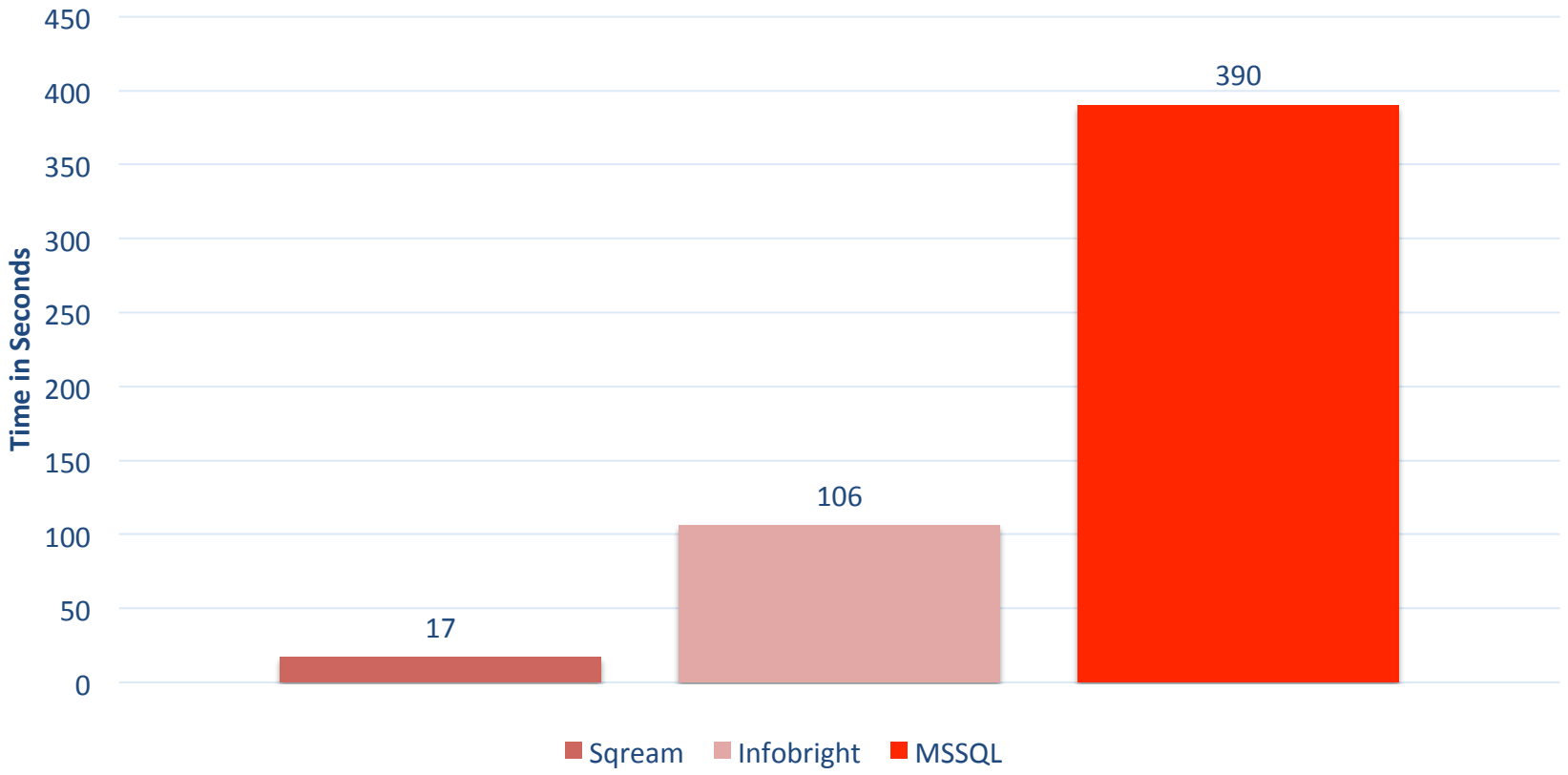


Analytics as a Services (AaaS)



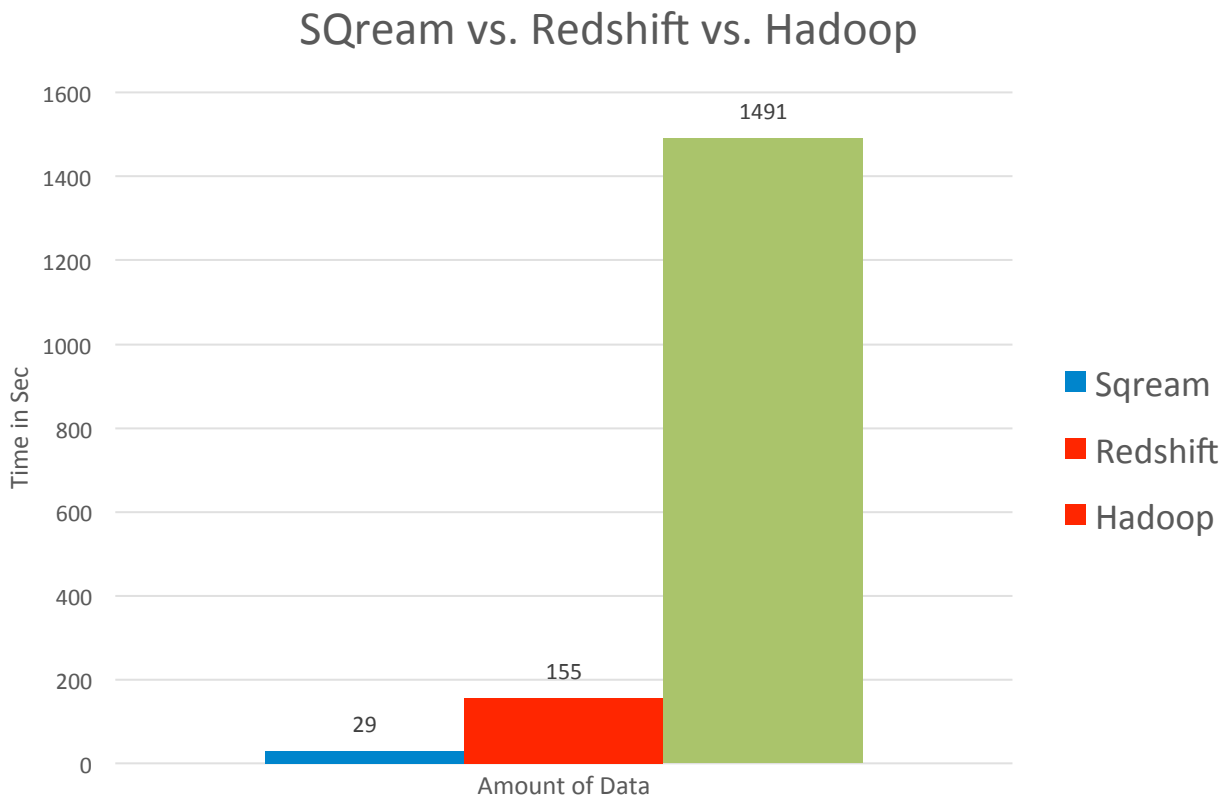
USE CASES & BENCHMARKS

Lab Results – 100GB TPC H



SQream vs. Redshift vs. Hadoop*

Query Run time – based on Hapyrus benchmark

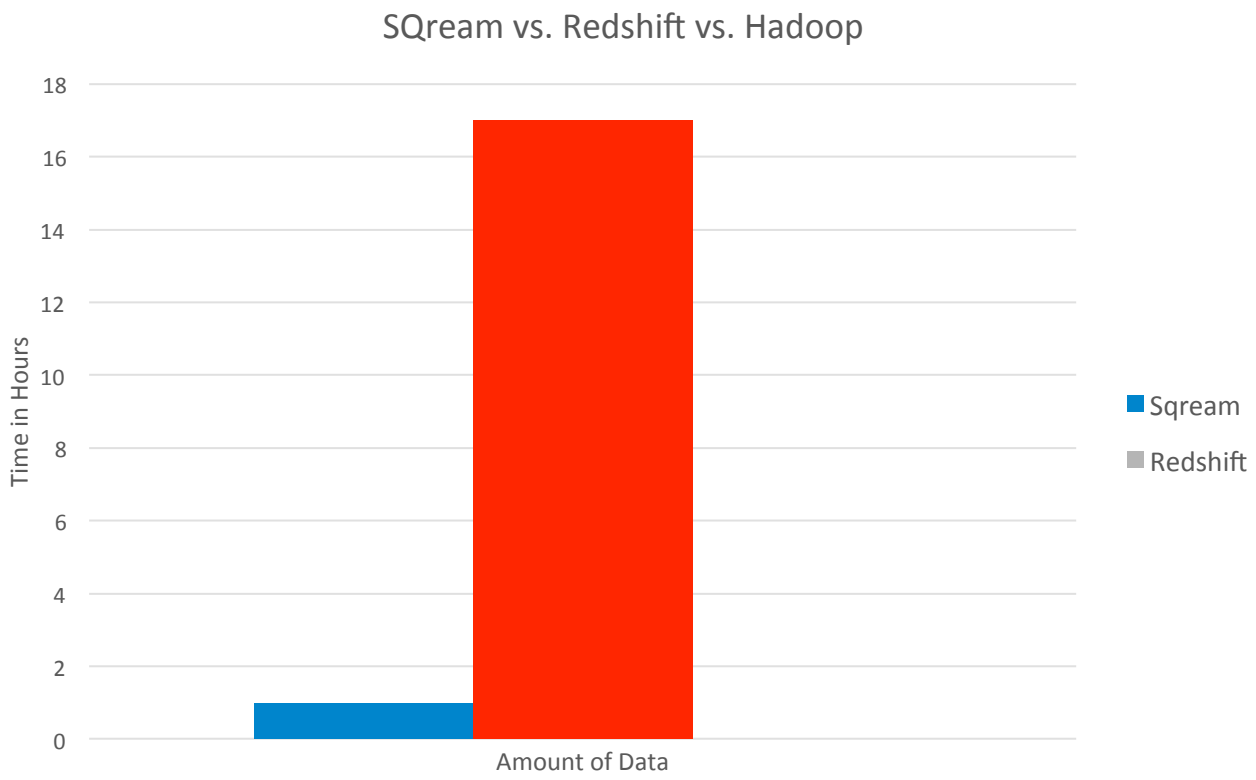


http://www.hapyrus.com/blog/posts/behind-amazon-redshift-is-10x-faster-and-cheaper-than-hadoop-hive-slides

SQREAM TECHNOLOGIES

SQream vs. Redshift vs. Hadoop*

Data Load Time – based on Hapyrus benchmark



http://www.hapyrus.com/blog/posts/behind-amazon-redshift-is-10x-faster-and-cheaper-than-hadoop-hive-slides

SQREAM TECHNOLOGIES

Cyber Analysis – Pattern Matching

Use Case

- load and analyze massive network traffic

The Test

- Load 6000 IP CDR per second
- Run queries under 10 seconds

Cyber Use Case – Existing Status

Hardware

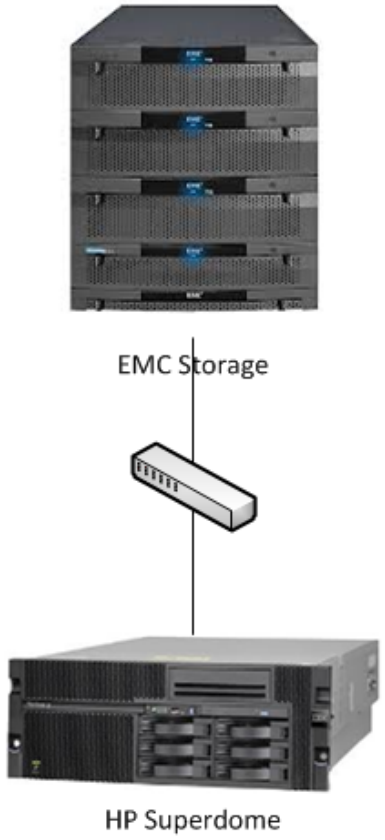
- HP Superdome 32 Cores
- EMC Storage

Total Hardware cost: ~$150,000

Software

- Oracle 11g

Existing Configuration



Cyber Use Case – SQream Solution

Hardware

- Dell T7600

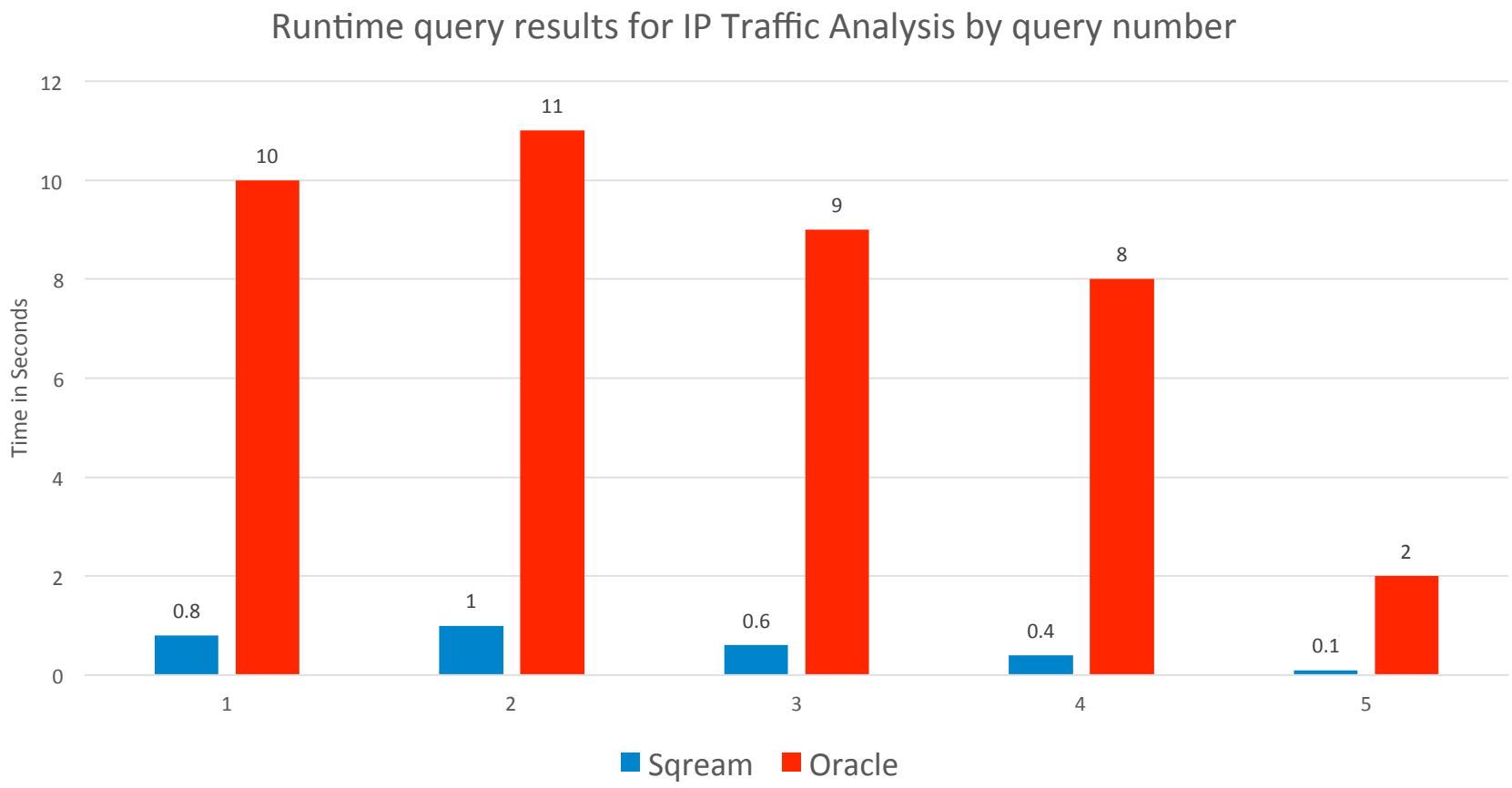
Total Hardware cost: ~$7,000

Dell T7600

Software

- SQream DB

Results for 8.5B Records ~17TB



SQREAM TECHNOLOGIES

SQream POC on Databases with Orange Silicon Valley

Soumik Sinharoy

Sr. Product Manager at Orange

SQream's DB Saves ~$6,000,000 on Big Data insights

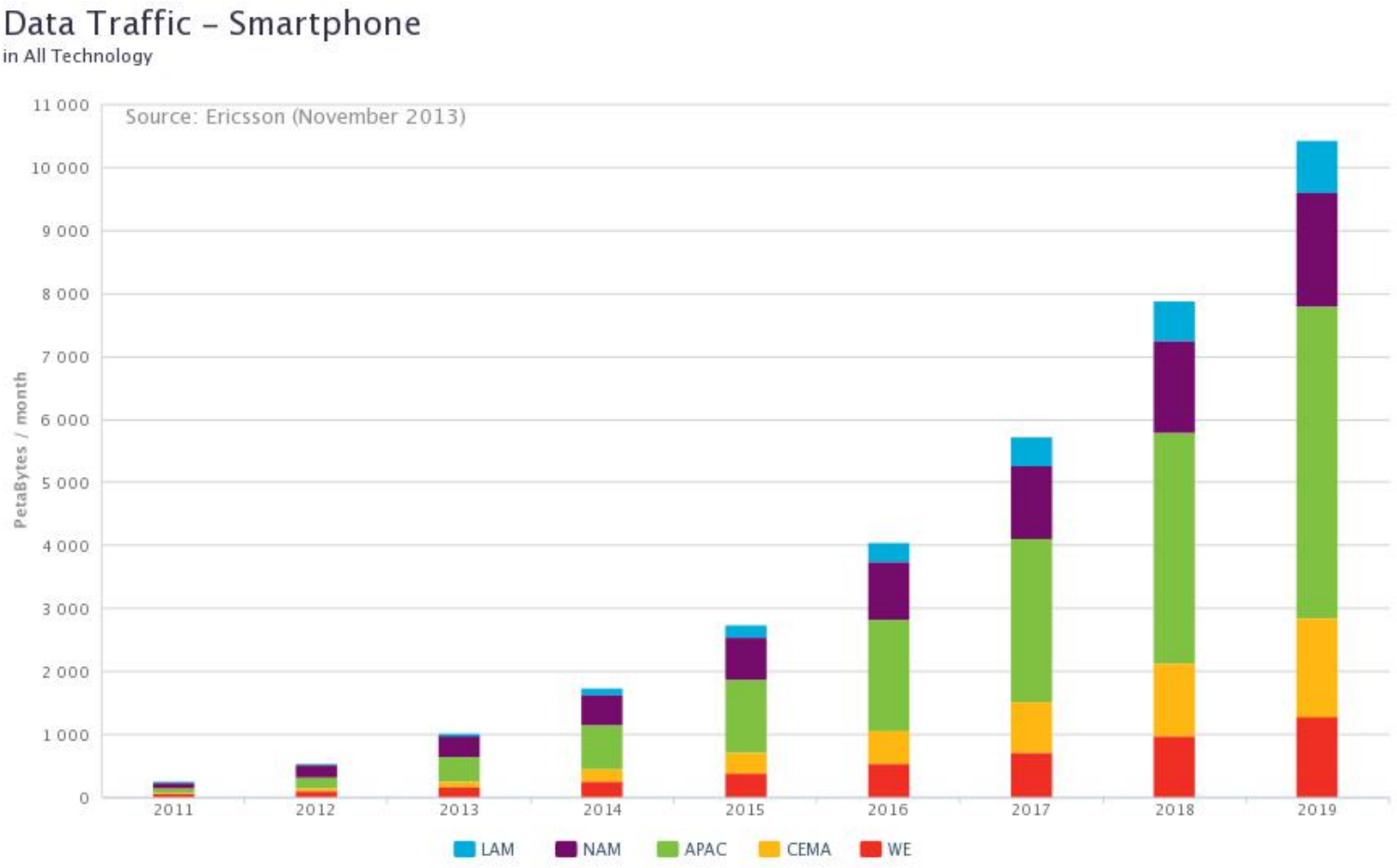
$6,300,000 Price list

SQREAM TECHNOLOGIES

$250,000 Price list

X40 cost performance

Data Is Growing World Wide IP Traffic Growth



SQREAM TECHNOLOGIES

Orange

Orange today

Orange is the story of a brand that has built a unique relationship with its customers; a relationship built on optimism, universal appeal and a focus on people rather than technology, and that allows Orange to engage customers around the world by understanding their aspirations. It's also a brand that stands out, with a logo and identity born from our values (friendly, honest, straightforward, dynamic and refreshing) that are recognised and have meaning all over the world.

@orange
the leading twitter account for CAC 40 quoted groups with 30,000 followers and 600,000 followers for all Orange accounts

more than 7 million fans across all local Orange **Facebook** pages

4G
in 8 countries

2,400 employees actively volunteering in 30 countries for the **Foundation**

7 million
Orange Money
customers in 13 countries

230 million
customers

43.5 billion euros in turnover

1.4 million used mobiles collected

172 million mobile customers

a **presence**
in more than 30 countries

over 200,000
fibre customers

170,000
employees

a portfolio of 7,493
patents

812 million Euros invested in research and
innovation

Orange is also...

60th global brand

6th brand in the telecoms sector

29 countries where brand awareness exceeds 80%

Millward Brown 2013

Orange employees at the heart of the business

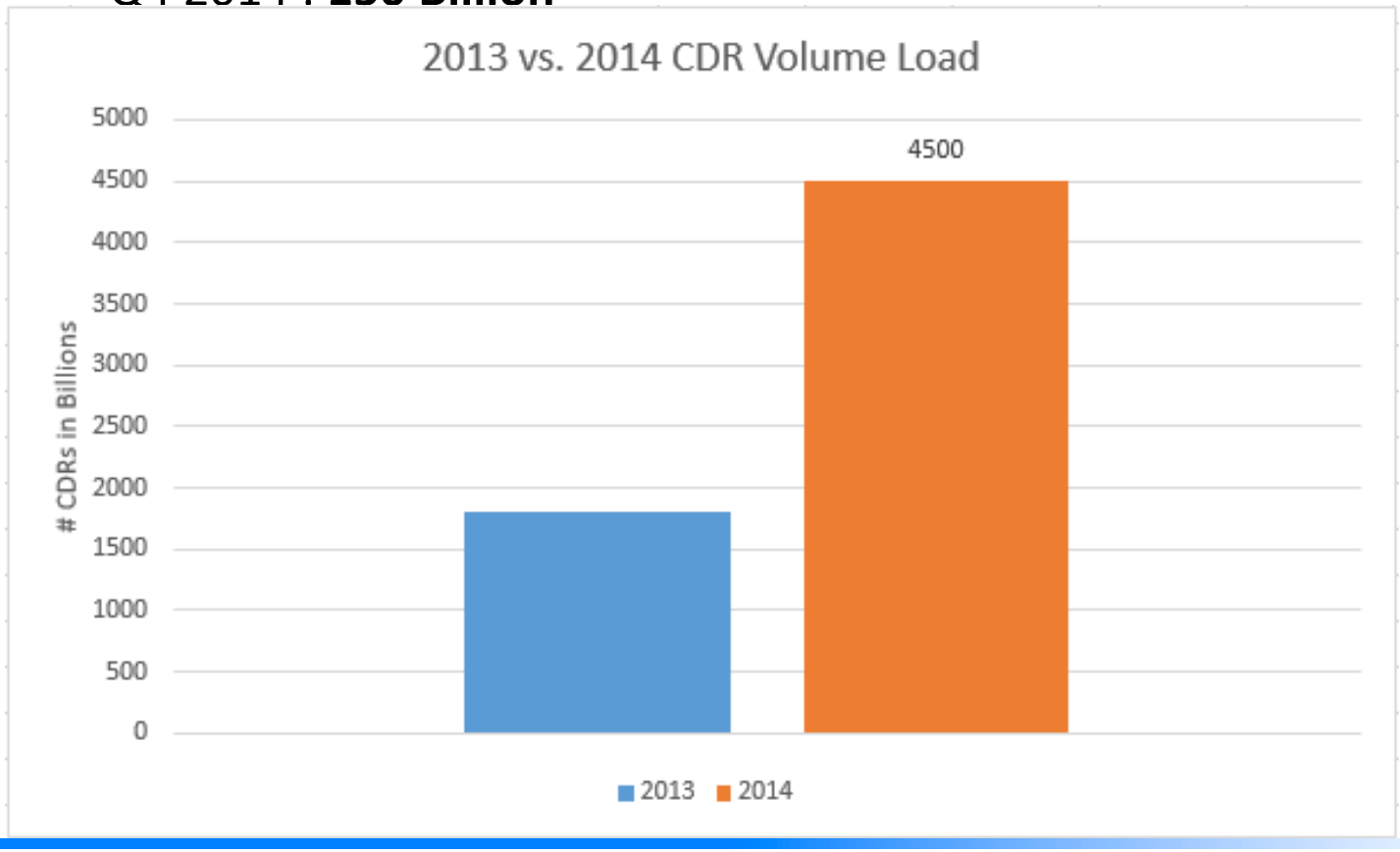
Orange is a global community of 170,000 people from all walks of life and speaking dozens of languages, reflecting the extraordinary number of markets where we operate, and representing the Group's greatest strength. It is thanks to the individual contribution of everyone that Orange can work towards a collective goal: to be the preferred operator in each country where the Group is present

more than 5 000 employees working on innovation

SQREAM TECHNOLOGIES

Our Data Is Growing Orange France IP Traffic Growth

- Monthly CDR Volume
  - Today : 60 **Billion**
  - Q4 2014 : **150 Billion**



SQREAM TECHNOLOGIES

What We Plan To Do With Big Data

Airborne Datacenters
- SuperComputing

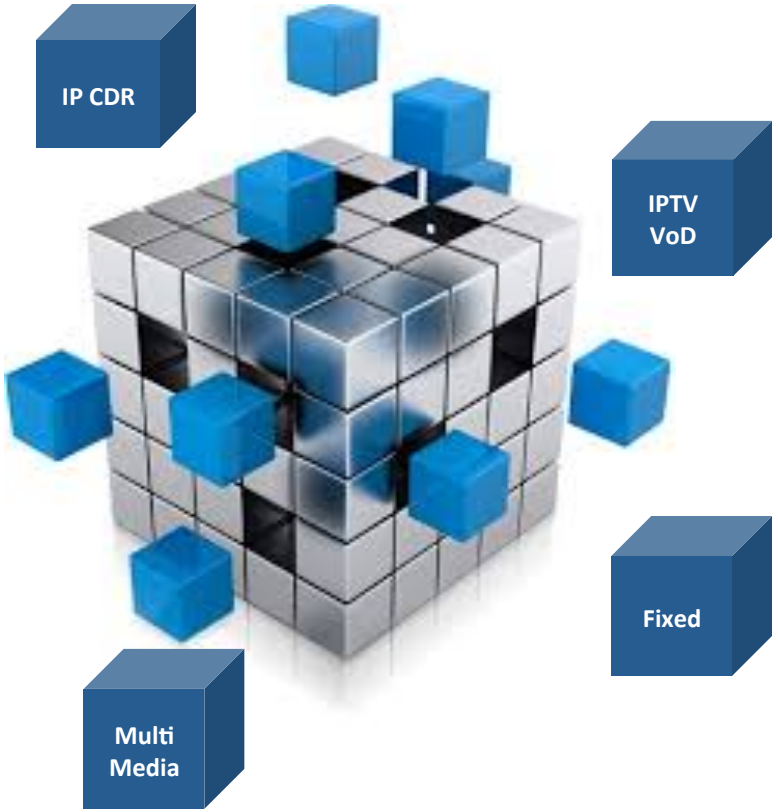
Realtime processing of sensor data

Massive database consolidation : 400 X savings in CapEX

Current Challenges

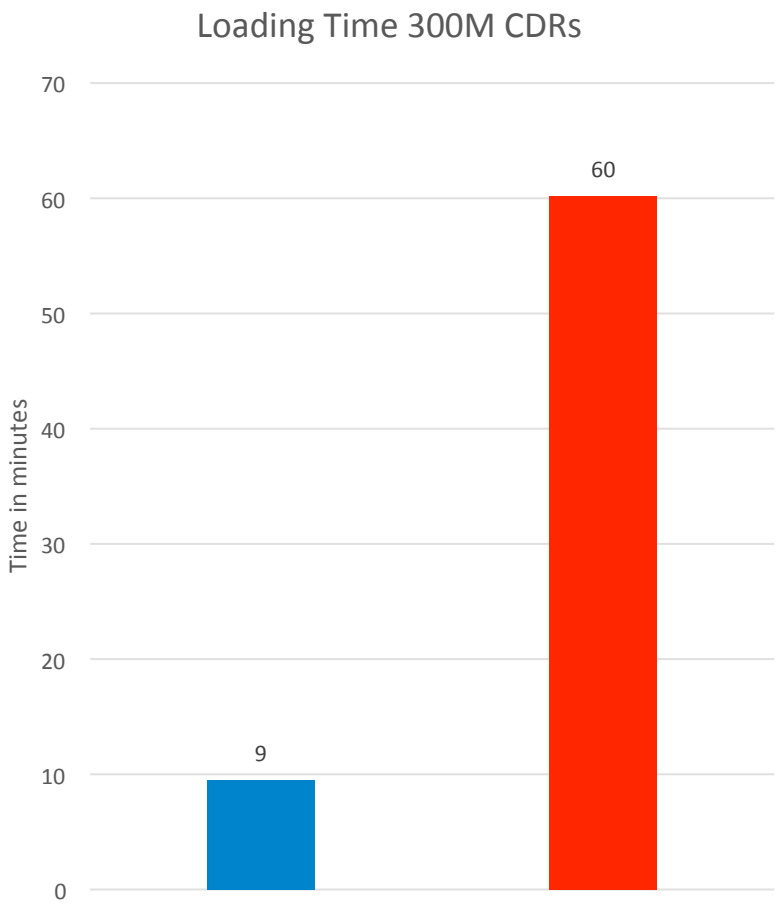
Problem Statement

- Massive Data Store
- Multi Dimensional Queries
- 4-9 Billion CDR
- Reduce execution time
- Reduce TCO



Revenue Assurance – POC vs. Leading RDBMS

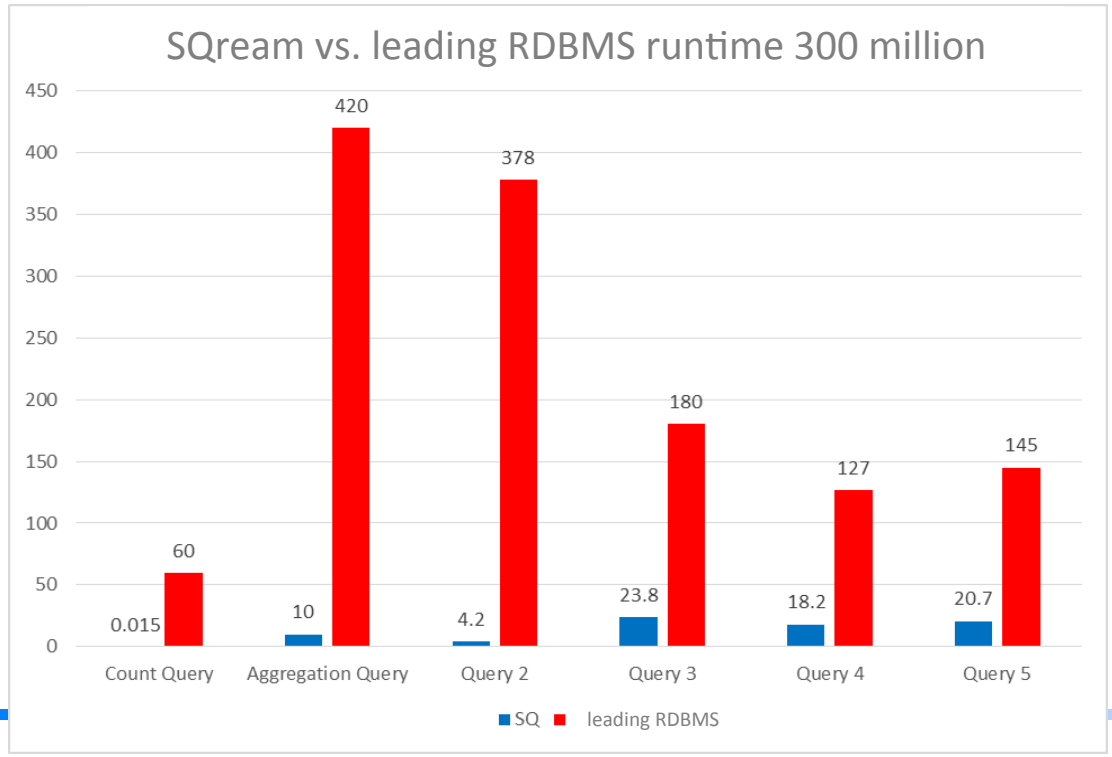
- Scan & Join 35 Billion CDR and 30M Subscribers
- Challenges:
  - Require massive investment in Hardware DW system
  - Require have manual process for optimizing the software
  - Doesn't scale



SQREAM TECHNOLOGIES

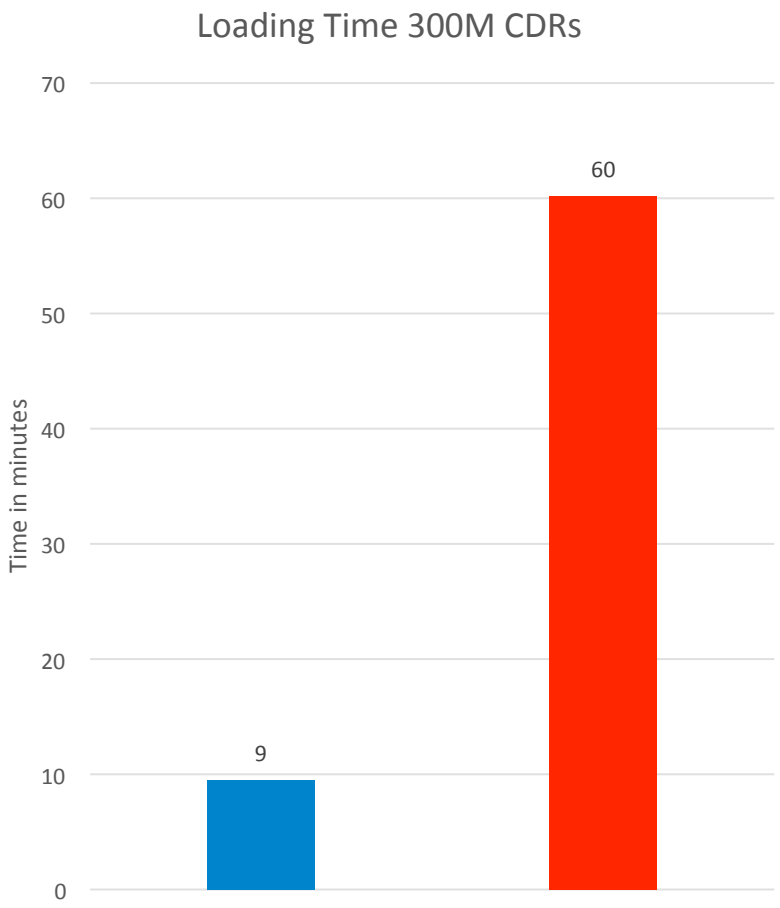
Query Run Time 300 Million CDRs

- Q2 – Map load on the network by hour, total call, minutes
- Q3 – Map by hour and call type, break down for the load on the network by call type. Sms, data, voice
- Q4 – Map load on the network in specific time
- Q5 - Map load on the network in specific time and specific service: voice, sms, data



Multi Media Reporting – IP Traffic Analysis on anonymized IP CDR from Orange France

- Data coming from MMT are monthly received on MMDM server
- Data are processed for a dynamic reporting application
- Challenges:
  - Require massive investment in Hardware DW system
  - Require have manual process for optimizing the software



SQREAM TECHNOLOGIES

~4,300,000,000 Call records
~30,000,000 Subscribers
4 Months
1 Operator - Orange

14 Servers
168 CPU cores
1,344GB RAM
67,200GB SSD
403,200GB HDD

875 Seconds

SQREAM TECHNOLOGIES

~4,300,000,000 Call records
~30,000,000 Subscribers
4 Months
1 Operator - Orange

1x K40 GPU

1 Server
8 CPU cores
128GB RAM
3,200GB SSD
10,000GB HDD

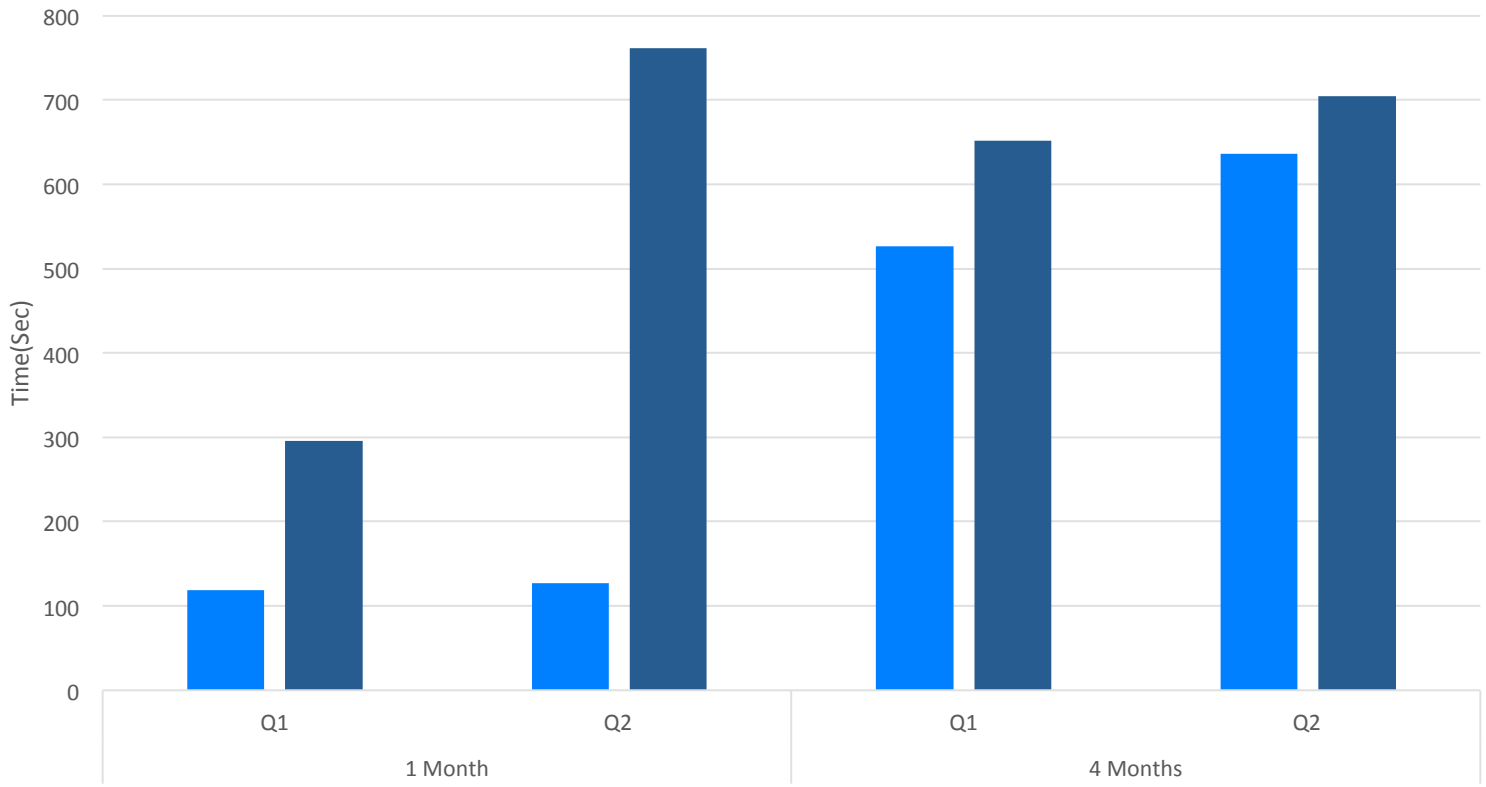
$250,000 Price list

636 Seconds

- 18 month of data was also tested Execution Time = **2484** Seconds

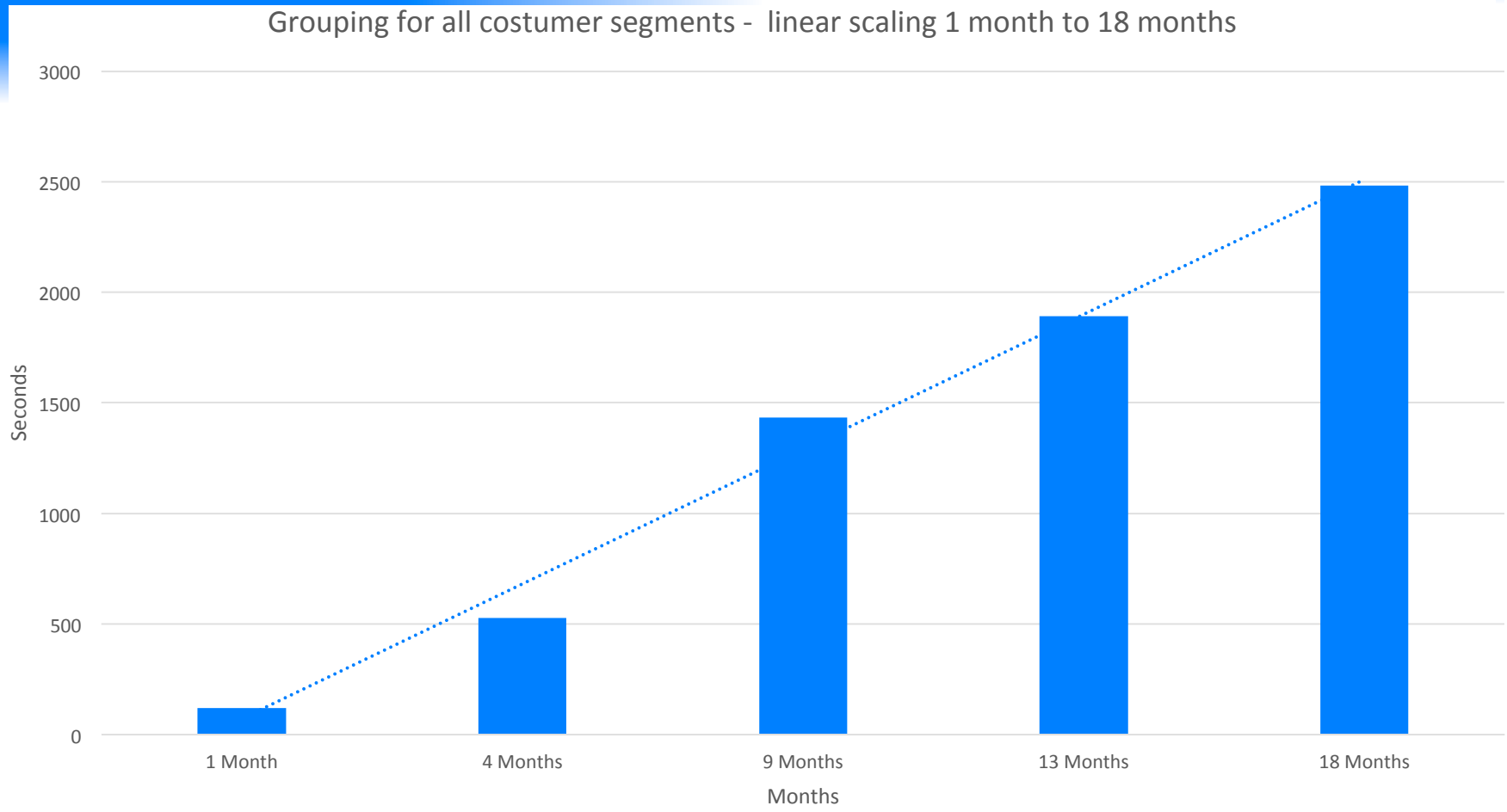
Less is More… SQream vs. Leading EDW

- Q1 – Aggregation query for IP utilization by users
- Q2 – Group By query for 1 dimension



SQREAM TECHNOLOGIES

SQream – Yes, We Scale Linearly!



QUESTIONS ???

SQream Technologies Ltd - Confidential

Why SQream? Technology

- Data Crunching:
- Faster compression time **x20**
- Faster decompression time **x50-70**
- Higher compression ratio **x5-15**
- Compute:
- Faster MPP in a node **x20**
- Higher scalability **x1 node x3000 cores**
- Lower hardware cost **$7,000,000 > $15K**

Why SQream? Hassle free Solution

TCO

- Lower power consumption **1 node vs. 10 nodes**
- Lower storage and license cost **x10 – x50**
- Lower footprint **2U vs. 20U**

Deployment advantages

- **No indexes**, no cubicles, no projections.
- Built-in **full scan ad-hoc queries** support.
- No code changes. Simple **SQL**.

SQREAM TECHNOLOGIES

SQream the fastest analytics ever

Thank you