AWS Summit 2013 Milan

31 Ottobre 2013

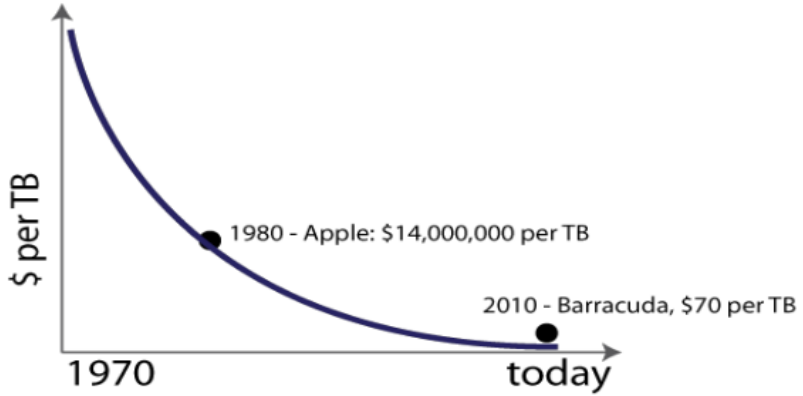
DATA ANALYSIS ON AWS

Hakan Gurel

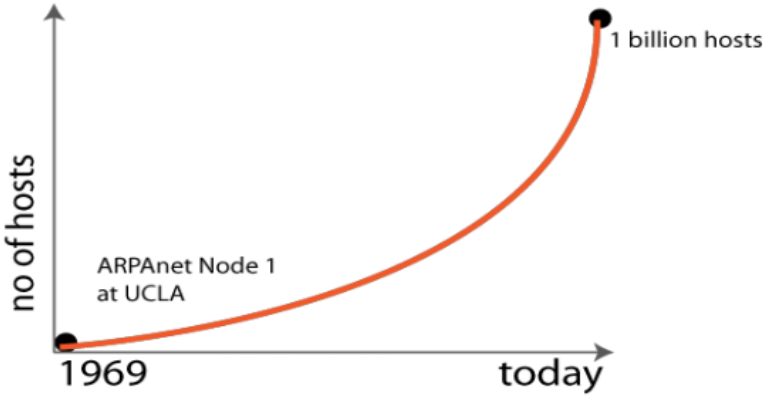
Solutions Architecture

THE COST OF GENERATING DATA IS FALLING

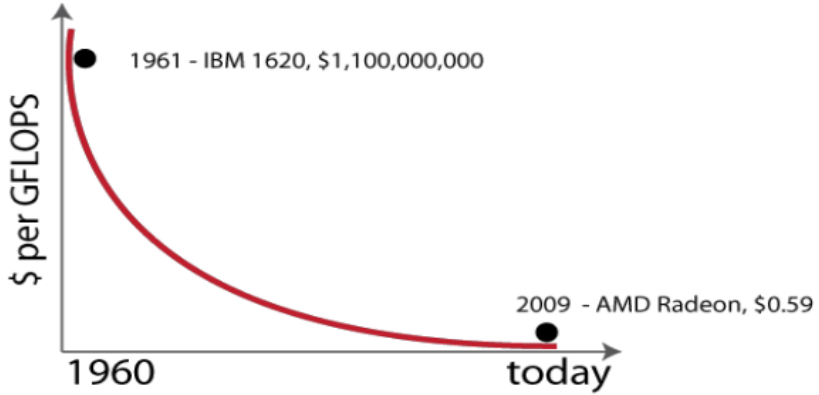
storage cost



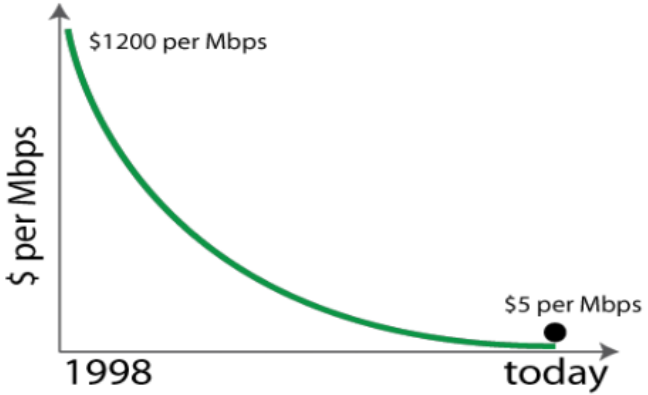
network access



CPU cost



bandwidth cost



THE MORE DATA YOU COLLECT
THE MORE VALUE YOU CAN
DERIVE FROM IT

amazon.fr Premium

Bonjour Conde Carlos. Découvrez nos conseils personnalisés. (Vous n'êtes pas Conde ?)

Soldes : -30%, -50%, -70%...

Chez Conde | Promotions | Chèques-cadeaux | Listes et idées cadeaux

Votre compte | Aide

Toutes nos boutiques

Rechercher Toutes nos boutiques GO Panier Liste d'envies

- MP3 et Cloud Player
- Amazon Cloud Drive
- App-Shop pour Android
- Livres
- Kindle
- Musique, DVD et Blu-ray
- High-Tech
- Informatique et bureau
- Jeux vidéo et Consoles
- Jouets, Enfants et Bébés
- Cuisine et Maison
- Beauté, Hygiène et Santé
- Vêtements et Chaussures
- Montres et Bijoux
- Sports et Loisirs
- Bricolage

Découvrez la toute nouvelle famille Kindle Fire

Kindle Fire 159€

Kindle Fire HD 199€

Toute nouvelle liseuse Kindle 79€ seulement

du 17 au 23 septembre

Soldes et promotions -30%, -50%, -70%...

Cliquez ici

amazon cloud player

Votre musique sur tous vos lecteurs : PC, Mac, téléphones et tablettes Android, iPhone et iPod Touch. Plus d'informations

Ventes Flash

Stocks limités Cliquez ici

A découvrir

Vous avez regardé | Vous apprécierez peut-être également

Mattel - V6827 - Poupées mannequins...
EUR 23,81

Barbie - W4469 - Poupée et...
EUR 23,39

Robes pour Barbie - Collection conte...
Simply Exquisite
EUR 14,95

Barbie - V6913 - Poupée Mannequin...
EUR 23,90

Vous en rêvez

Linchpin: Are You Indispensable?

"This is what the future of work (and the world) looks like. Actually, it's already happening...

Lire la suite

EUR 8,55

Voir plus d'articles dans votre liste d'envies

Boutique Gaming

Cliquez ici

amazon.com Prime

Hello, Jon Jenkins. We have recommendations for you. (Not Jon?)

Jon's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Up to 50% off select Father's Day gifts
Presented by DEWALT

Your Digital Items | Your Account | Help

Shop All Departments | Search | All Departments | GO | Cart | Wish List

Your Account > Your Orders

Orders Listed By Date | Open Orders | Digital Orders | Order History Reports

Search Your Orders: Title, Department, Recipient... Search Orders

Show: Orders placed in 1999 Go

- -Select different orders to view-
- Orders placed in the last 30 days
- Orders placed in the past 6 months
- Orders placed in 2011
- Orders placed in 2010
- Orders placed in 2009
- Orders placed in 2008
- Orders placed in 2007
- Orders placed in 2006
- Orders placed in 2005
- Orders placed in 2004
- Orders placed in 2003
- Orders placed in 2002
- Orders placed in 2001
- Orders placed in 2000
- Orders placed in 1999

4 orders placed in the last **30 days**

1 of 1

Order Placed:
June 7, 2011
View Order Details | View Invoice

Order Number: 103-7182269-7289827
Recipient: Jon Jenkins
Shipping Speed: One-Day Shipping
Order Total: $57.13

Shipment 1 of 1
Shipped
Delivery Estimate: June 8, 2011

adidas Men's Samba Classic Soccer Shoe,Black/White,14 M
Sold by: Amazon.com, LLC
Available actions

Order Placed:
June 5, 2011
View Order Details | View Invoice

Order Number: 103-0008122-0201048
Recipient: Jon Jenkins
Shipping Speed: Standard
Order Total: $159.90

Shipment 1 of 1
Shipped
Delivery Estimate: June 13, 2011 - June 16, 2011

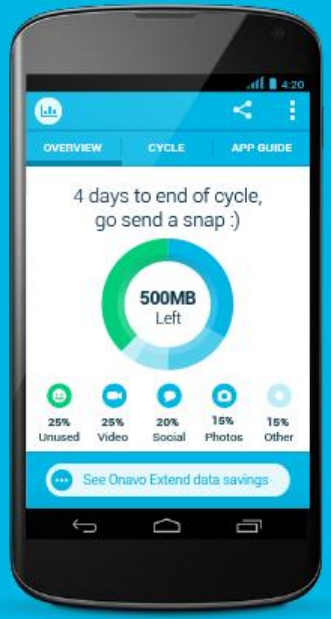
Track package

Livex Lighting Mission 1033-91
Sold by: Blue Marble Lighting (seller profile)
Available actions

Contact seller
File a claim
Return items
Leave seller feedback

onavo

HOME | APPS | NEWS | HELP | INSIGHTS | ABOUT



ONAVO COUNT 2.0

Your life is going mobile, so keep Count

With sleek and simple reporting, understand your data usage and how you compare with everyone else.

Available now on Google Play

Watch the video

Get it on Google play

WSJ

"iPhone App Doubles Your Data Plan"

Ben Rooney, Wall Street Journal

Onavo AppRank™ Top iPhone Apps

Top AppRank | Biggest Gains | Biggest Dips

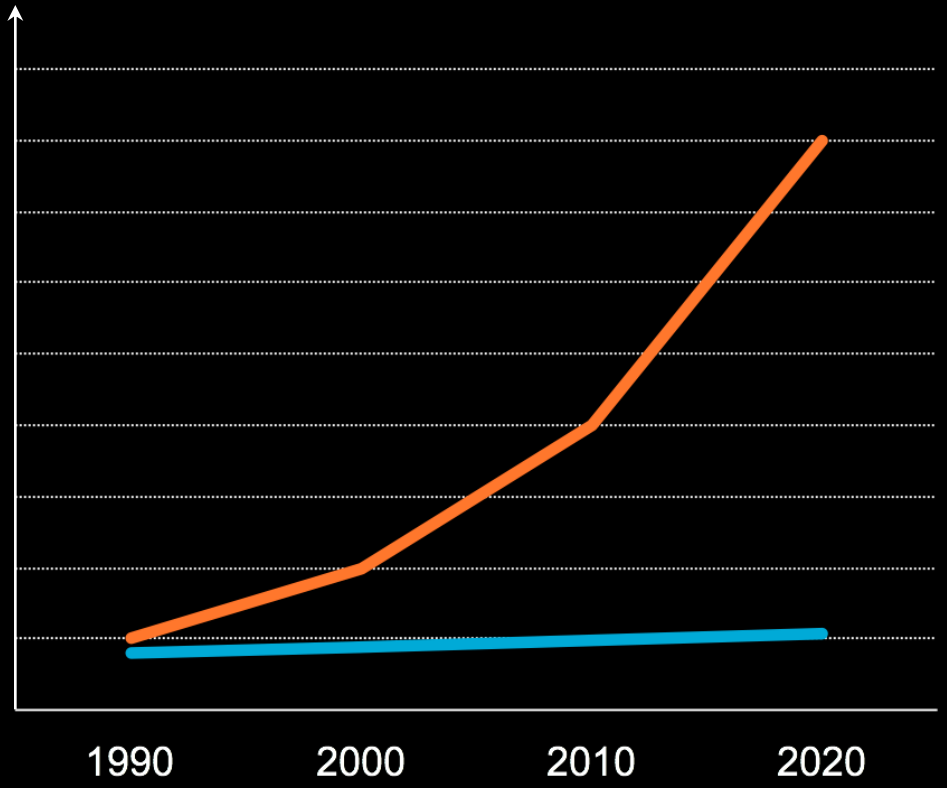
August 2013

ALL CATEGORIES
Books
Business
Catalogs
Education
Entertainment
Finance
Food & Drink
Games
Health & Fitness

App Name	Onavo AppRank™	Market Share
Facebook Facebook, Inc.	1 0	73.4% + 0.94%
YouTube Google, Inc.	2 0	47.6% - 0.69%
Instagram Burbn, Inc.	3 0	38.3% + 2.48%
Google Maps Google, Inc.	4 0	35.0% - 0.64%
Pandora Radio Pandora Media, Inc.	5 0	31.4% + 0.6%
Twitter Twitter, Inc.	6 0	27.0% + 0.84%


Lower cost,
higher throughput
GENERATE
STORE → ANALYZE → SHARE
Highly
constrained

DATA VOLUME



Generated data

Available for analysis

Gartner: User Survey Analysis: Key Trends Shaping the Future of Data Center Infrastructure Through 2011
IDC: Worldwide Business Analytics Software 2012–2016 Forecast and 2011 Vendor Shares

GENERATE → STORE → ANALYZE → SHARE

+ ELASTIC AND HIGHLY SCALABLE
+ NO UPFRONT CAPITAL EXPENSE
+ PAY FOR ONLY WHAT YOU USE
+ AVAILABLE ON-DEMAND

= REMOVE CONSTRAINTS

AWS Import / Export
AWS Direct Connect

GENERATE ➔ STORE ➔ ANALYZE ➔ SHARE

- Generated and stored in AWS
- Inbound data transfer is free
- Multipart upload to S3
- Physical media
- AWS Direct Connect
- Regional replication of AMIs and snapshots

Amazon S3,
Amazon Glacier,
Amazon DynamoDB,
Amazon RDS,
Amazon Redshift,
AWS Storage Gateway,
Data on Amazon EC2

GENERATE ➔ **STORE** ➔ ANALYZE ➔ SHARE

AMAZON S3

SIMPLE STORAGE SERVICE


Spotify
Listen to millions of songs for free.
Play, discover and share with your friends.
Download Spotify
By clicking download, you agree to Spotify's terms & conditions and privacy policy.
CASE STUDY

AMAZON DYNAMODB

HIGH-PERFORMANCE, FULLY MANAGED NoSQL DATABASE SERVICE

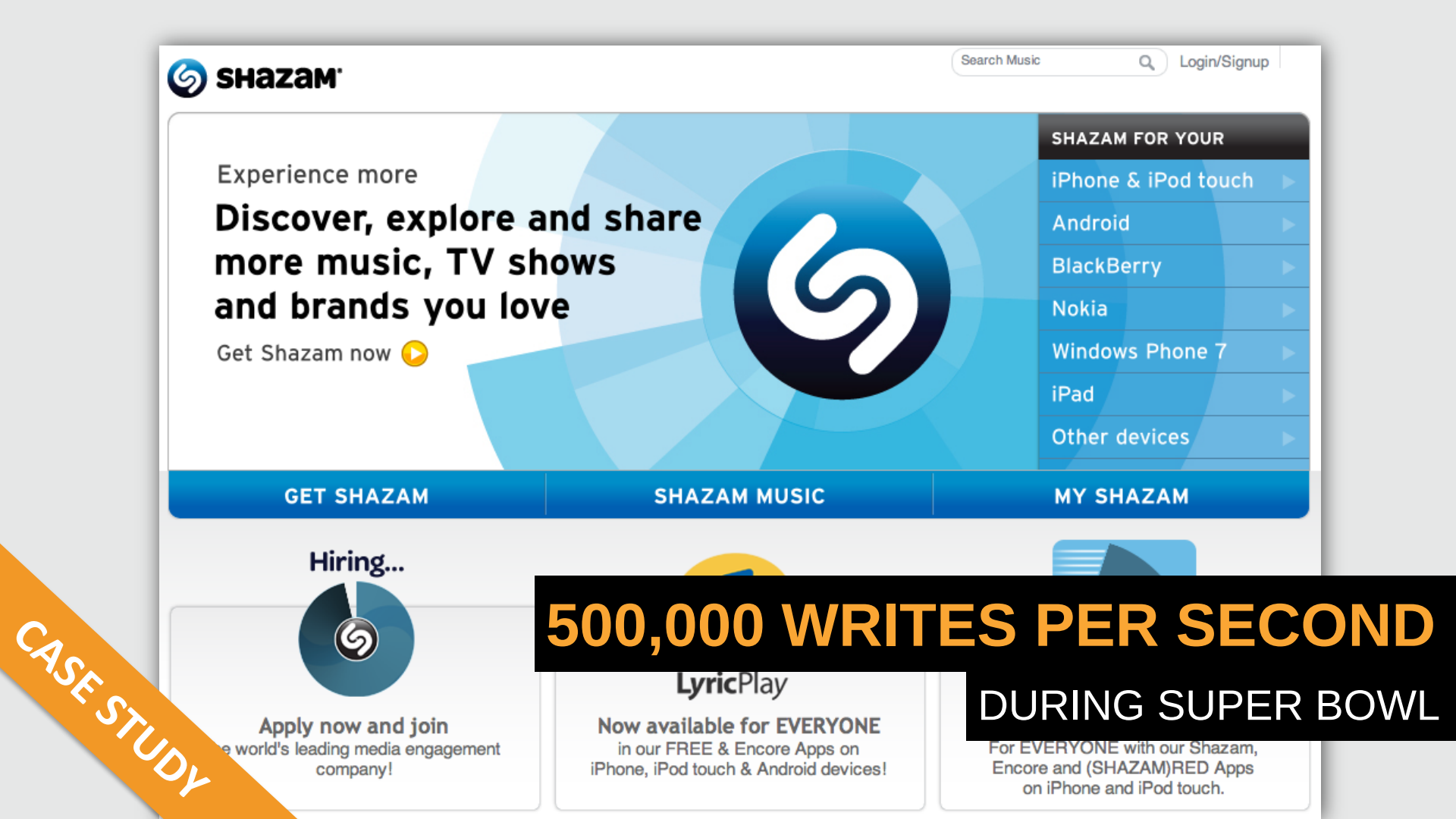
DURABLE & AVAILABLE

CONSISTENT, DISK-ONLY WRITES (SSD)

LOW LATENCY

AVERAGE READS < 5MS, WRITES < 10MS

NO ADMINISTRATION


SHAZAM
Search Music
Login/Signup
Experience more
Discover, explore and share more music, TV shows and brands you love
Get Shazam now
SHAZAM FOR YOUR
iPhone & iPod touch
Android
BlackBerry
Nokia
Windows Phone 7
iPad
Other devices
GET SHAZAM
SHAZAM MUSIC
MY SHAZAM
Hiring...
Apply now and join
e world's leading media engagement company!
LyricPlay
Now available for EVERYONE
in our FREE & Encore Apps on iPhone, iPod touch & Android devices!
For EVERYONE with our Shazam, Encore and (SHAZAM)RED Apps on iPhone and iPod touch.
CASE STUDY
500,000 WRITES PER SECOND
DURING SUPER BOWL

AMAZON REDSHIFT

FULLY MANAGED, PETA-BYTE SCALE DATAWAREHOUSE ON AWS

Enterprises average between 3 and 4 DBAs per data warehouse.

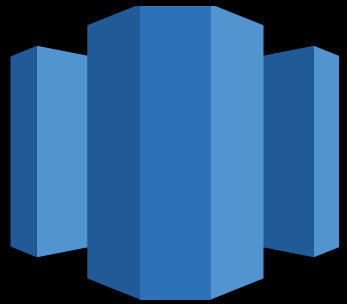
Gartner

Source: Gartner. Critical factors in calculating the data warehouse TCO, July 2009

DESIGN OBJECTIVES:

A petabyte-scale data warehouse service that was…

AMAZON REDSHIFT



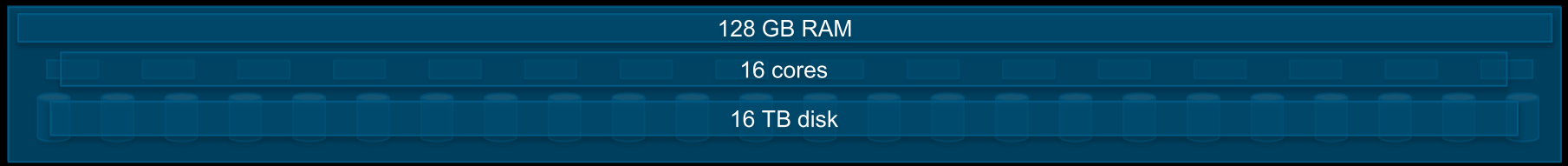
A Lot Faster

A Lot Cheaper

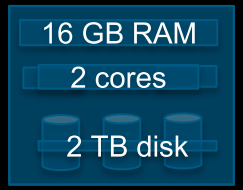
A Whole Lot Simpler

AMAZON REDSHIFT
RUNS ON OPTIMIZED HARDWARE

HS1.8XL: 128 GB RAM, 16 Cores, 16 TB compressed user storage, 2 GB/sec scan rate



HS1.XL: 16 GB RAM, 2 Cores, 2 TB compressed customer storage



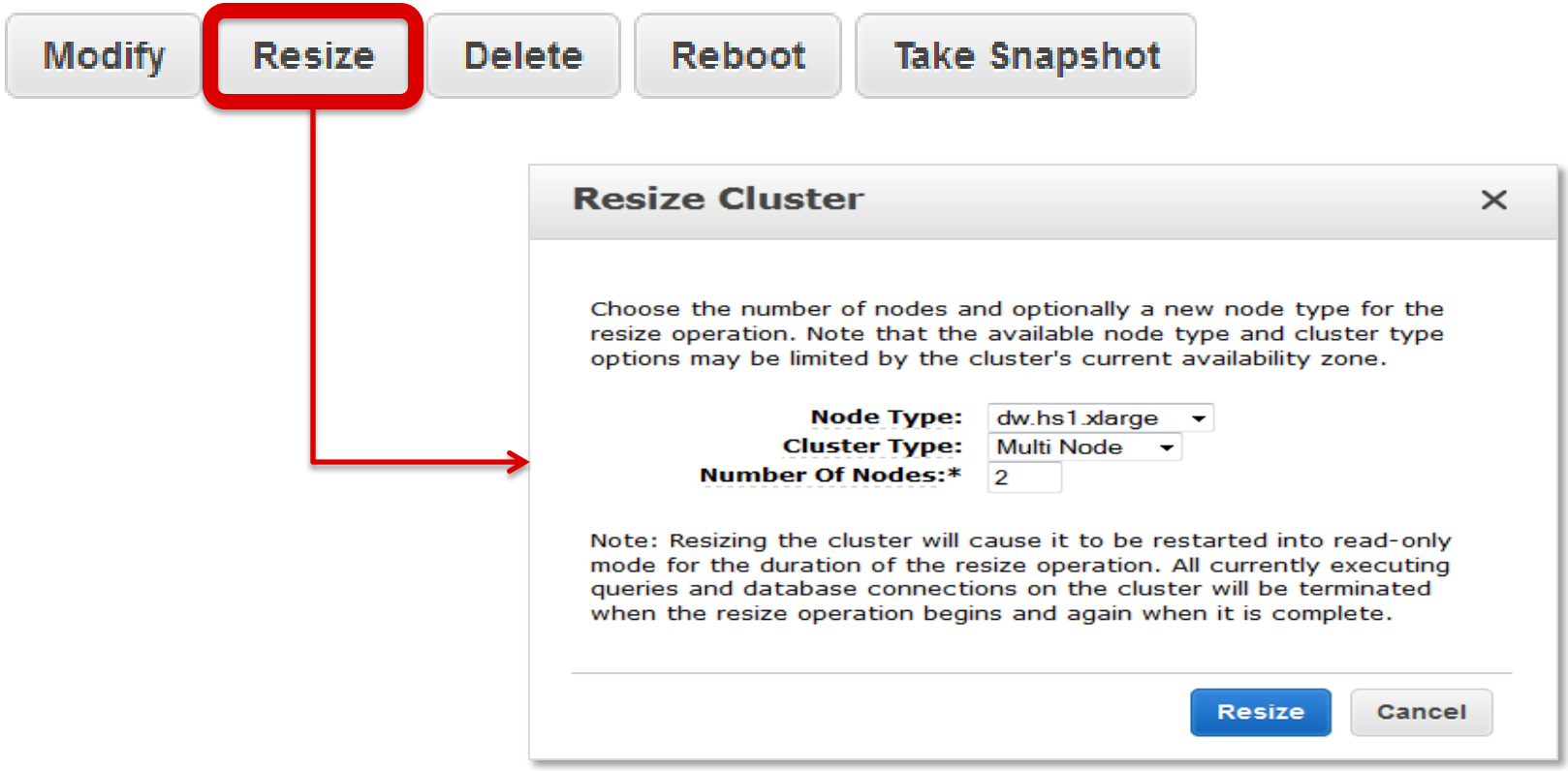
2 billion row dataset. 6 representative queries.

Amazon Redshift: 2 instance cluster

Compared to 32 nodes. 128 CPUs. 4.2 TB RAM. 1.6 PB storage. 2 billion row data set.

12x to 150x faster

30 MINUTES
DOWN TO
12 SECONDS


Modify
Resize
Delete
Reboot
Take Snapshot
Resize Cluster
Choose the number of nodes and optionally a new node type for the resize operation. Note that the available node type and cluster type options may be limited by the cluster's current availability zone.
Node Type:
dw.hs1.xlarge
Cluster Type:
Multi Node
Number Of Nodes:*
2
Note: Resizing the cluster will cause it to be restarted into read-only mode for the duration of the resize operation. All currently executing queries and database connections on the cluster will be terminated when the resize operation begins and again when it is complete.
Resize
Cancel

AMAZON REDSHIFT LETS YOU START SMALL AND GROW BIG

Extra Large Node (HS1.XL)

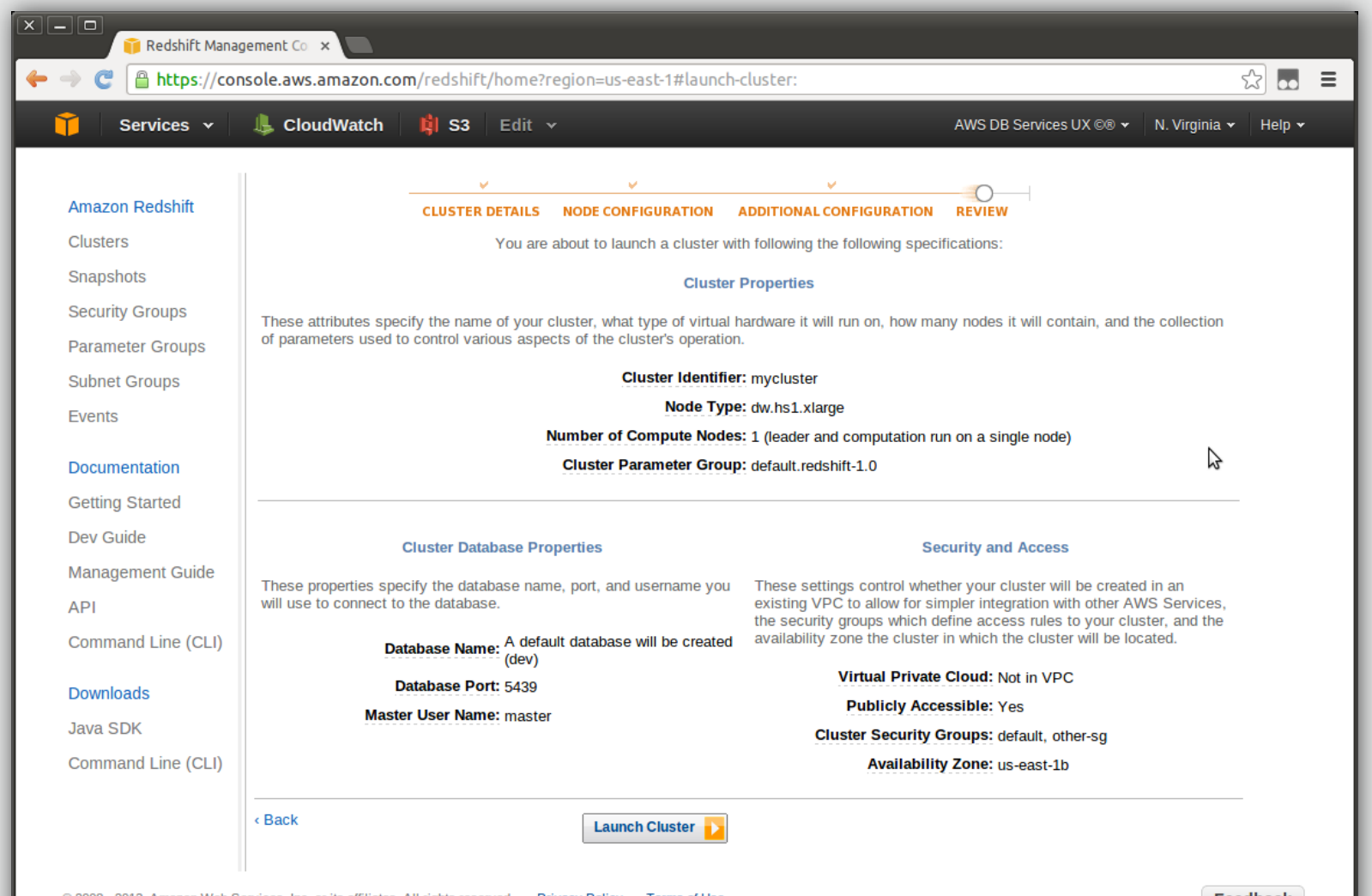
Single Node (2 TB)

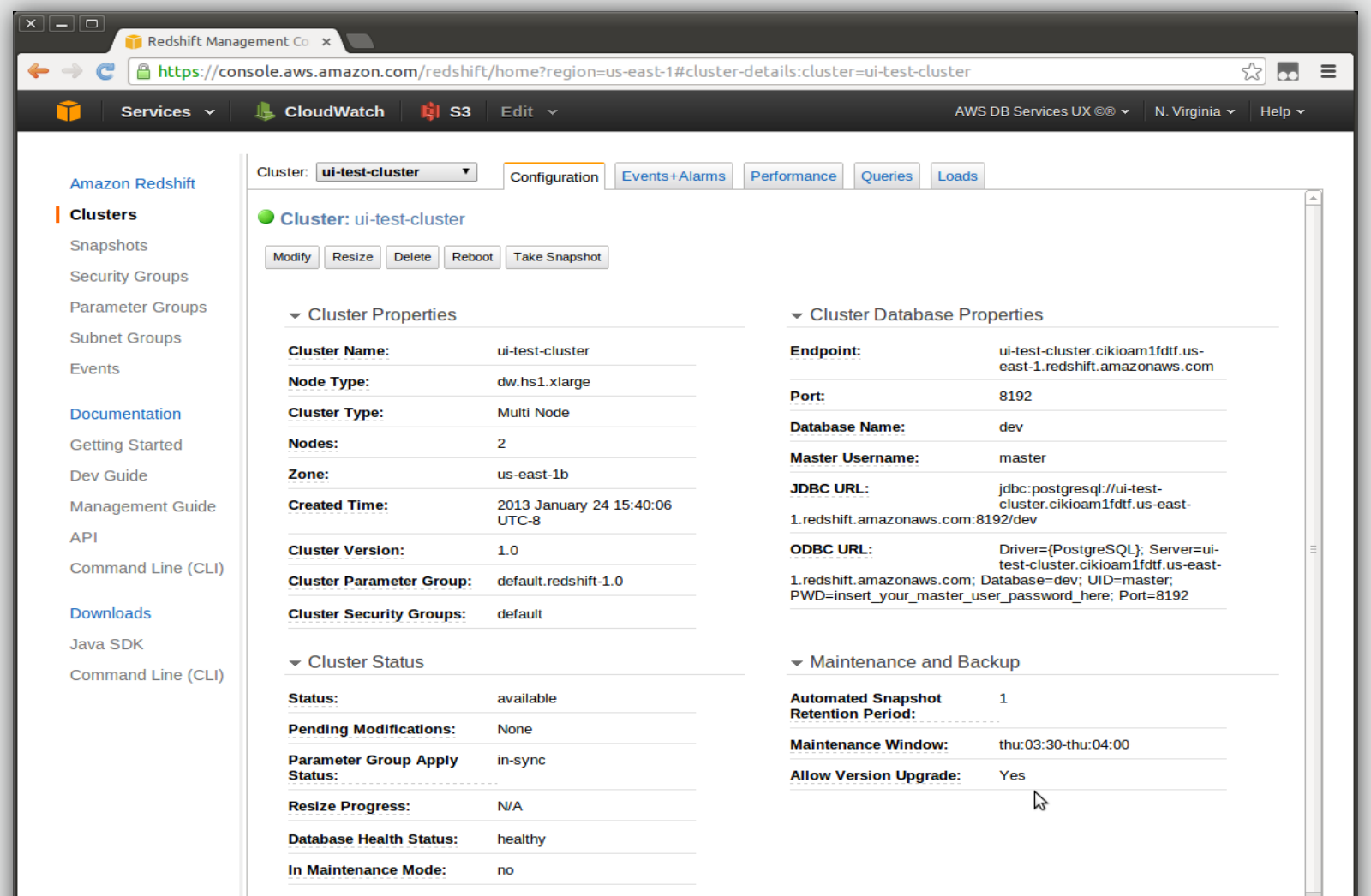
Cluster 2-32 Nodes (4 TB – 64 TB)

Eight Extra Large Node (HS1.8XL)

Cluster 2-100 Nodes (32 TB – 1.6 PB)

CREATE A DATAWAREHOUSE IN MINUTES


Redshift Management Co
https://console.aws.amazon.com/redshift/home?region=us-east-1#launch-cluster:
Services
CloudWatch
S3
Edit
AWS DB Services UX ©®
N. Virginia
Help
Amazon Redshift
Clusters
Snapshots
Security Groups
Parameter Groups
Subnet Groups
Events
Documentation
Getting Started
Dev Guide
Management Guide
API
Command Line (CLI)
Downloads
Java SDK
Command Line (CLI)
CLUSTER DETAILS
NODE CONFIGURATION
ADDITIONAL CONFIGURATION
REVIEW
You are about to launch a cluster with following the following specifications:
Cluster Properties
These attributes specify the name of your cluster, what type of virtual hardware it will run on, how many nodes it will contain, and the collection of parameters used to control various aspects of the cluster's operation.
Cluster Identifier: mycluster
Node Type: dw.hs1.xlarge
Number of Compute Nodes: 1 (leader and computation run on a single node)
Cluster Parameter Group: default.redshift-1.0
Cluster Database Properties
These properties specify the database name, port, and username you will use to connect to the database.
Database Name: A default database will be created (dev)
Database Port: 5439
Master User Name: master
Security and Access
These settings control whether your cluster will be created in an existing VPC to allow for simpler integration with other AWS Services, the security groups which define access rules to your cluster, and the availability zone the cluster in which the cluster will be located.
Virtual Private Cloud: Not in VPC
Publicly Accessible: Yes
Cluster Security Groups: default, other-sg
Availability Zone: us-east-1b
‹ Back
Launch Cluster


Redshift Management Co
https://console.aws.amazon.com/redshift/home?region=us-east-1#cluster-details:cluster=ui-test-cluster
Services
CloudWatch
S3
Edit
AWS DB Services UX ©®
N. Virginia
Help
Amazon Redshift
Clusters
Snapshots
Security Groups
Parameter Groups
Subnet Groups
Events
Documentation
Getting Started
Dev Guide
Management Guide
API
Command Line (CLI)
Downloads
Java SDK
Command Line (CLI)
Cluster: ui-test-cluster
Configuration
Events+Alarms
Performance
Queries
Loads
Cluster: ui-test-cluster
Modify
Resize
Delete
Reboot
Take Snapshot
Cluster Properties
Cluster Name: ui-test-cluster
Node Type: dw.hs1.xlarge
Cluster Type: Multi Node
Nodes: 2
Zone: us-east-1b
Created Time: 2013 January 24 15:40:06 UTC-8
Cluster Version: 1.0
Cluster Parameter Group: default.redshift-1.0
Cluster Security Groups: default
Cluster Status
Status: available
Pending Modifications: None
Parameter Group Apply Status: in-sync
Resize Progress: N/A
Database Health Status: healthy
In Maintenance Mode: no
Cluster Database Properties
Endpoint: ui-test-cluster.cikioam1fdtf.us-east-1.redshift.amazonaws.com
Port: 8192
Database Name: dev
Master Username: master
JDBC URL: jdbc:postgresql://ui-test-cluster.cikioam1fdtf.us-east-1.redshift.amazonaws.com:8192/dev
ODBC URL: Driver={PostgreSQL}; Server=ui-test-cluster.cikioam1fdtf.us-east-1.redshift.amazonaws.com; Database=dev; UID=master; PWD=insert_your_master_user_password_here; Port=8192
Maintenance and Backup
Automated Snapshot Retention Period: 1
Maintenance Window: thu:03:30-thu:04:00
Allow Version Upgrade: Yes

Redshift Management Co

https://aws-db-ux.integ.amazon.com/redshift/home?region=us-east-1#performance:cluster=test-load-files;metrics=

Services | Edit | AWS DB Services UX ©® | N. Virginia | Help

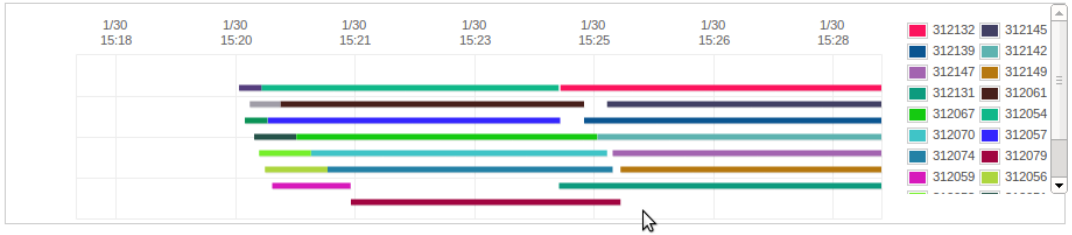
- Amazon Redshift
- **Clusters**
- Snapshots
- Security Groups
- Parameter Groups
- Subnet Groups
- Events
- Documentation
- Getting Started
- Dev Guide
- Management Guide
- API
- Command Line (CLI)
- Downloads
- Java SDK
- Command Line (CLI)

Cluster: **test-load-files** | Configuration | Events+Alarms | Performance | Queries | Loads

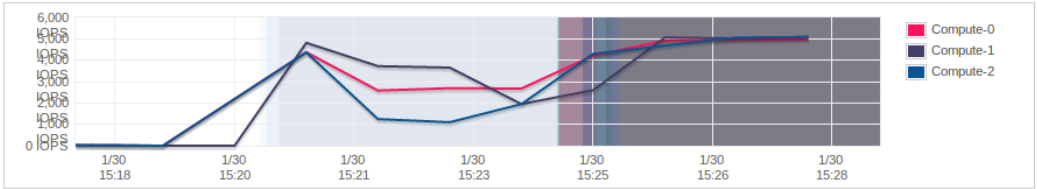
Time Range: 01/30 15:17—01/30 15:29 UTC-8 | Period: 1 Minute | Statistic: Average | Custom Metrics Selection | Nodes | Refresh

Queries

Hover over the queries graph or click on a query ID in the legend to inspect queries. Click and drag on any graph to zoom in.



Read IOPS



Query ID: 312054
Type: Query
User: master
Run Time: 4m 8.71s
Start Time: Wed Jan 30 15:20:21 GMT-800 2013
End Time: Wed Jan 30 15:24:30 GMT-800 2013
SQL:

```
select ps_partkey,
sum(ps_supplycost *
ps_availqty) as value from
partsupp, supplier, nation
where ps_suppkey = s_suppkey
and s_nationkey = n_nationkey
and n_name = 'GERMANY' group by
ps_partkey having
sum(ps_supplycost *
ps_availqty) > ( select
sum(ps_supplycost *
ps_availqty) * 0.0001000000
from partsupp, supplier, nation
where ps_suppkey = s_suppkey
and s_nationkey = n_nationkey
and n_name = 'GERMANY' ) order
by value desc LIMIT 1;
```

Query ID: 312057
Type: Query
User: master
Run Time: 4m 4.7s
Start Time: Wed Jan 30 15:20:26 GMT-800 2013

Redshift Management Co

https://aws-db-ux.integ.amazon.com/redshift/home?region=us-east-1#

Services | Edit | AWS DB Services UX ©® | N. Virginia | Help

- Amazon Redshift
- **Clusters**
- Snapshots
- Security Groups
- Parameter Groups
- Subnet Groups
- Events

- Documentation
- Getting Started
- Dev Guide
- Management Guide
- API
- Command Line (CLI)

- Downloads
- Java SDK
- Command Line (CLI)

Cluster: test-load-files ▾ | Configuration | Events+Alarms | Performance | Queries | Loads | Query ×

Cluster:	test-load-files
User:	master
Run Time:	1h 26m 4.74s
Start Time:	Fri Jan 25 11:08:28 GMT-800 2013
End Time:	Fri Jan 25 12:34:33 GMT-800 2013

SQL

```
-- using default substitutions select c_count, count(*) as custdist from ( select c_custkey, count(o_orderkey) from
customer left outer join orders on c_custkey = o_custkey and o_comment not like '%special%requests%' group by c_custkey )
as c_orders (c_custkey, c_count) group by c_count order by custdist desc, c_count desc LIMIT 1
```

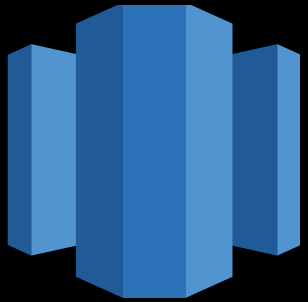
Explain Plan

```
XN Limit  (cost=132071117583790.86..132071117583790.86 rows=1 width=8)
  ->  XN Merge  (cost=132071117583790.86..132071117583791.36 rows=200 width=8)
        ->  XN Network  (cost=132071117583790.86..132071117583791.36 rows=200 width=8)
              ->  XN Sort  (cost=132071117583790.86..132071117583791.36 rows=200 width=8)
                    ->  XN HashAggregate  (cost=131071117583782.72..131071117583783.22 rows=200 width=8)
                          ->  XN Subquery Scan c_orders  (cost=131071114895782.72..131071116815782.72 rows=153600000 width=8)
                                ->  XN HashAggregate  (cost=131071114895782.72..131071115279782.72 rows=153600000 width=16)
                                      -
>  XN Hash Right Join DS_DIST_BOTH  (cost=1920000.00..131071107215988.88 rows=1535958769 width=16)
                                            ->  XN Seq Scan on orders  (cost=0.00..19200000.00 rows=1535958769 width=16)
                                            ->  XN Hash  (cost=1536000.00..1536000.00 rows=153600000 width=8)
                                                  ->  XN Seq Scan on customer  (cost=0.00..1536000.00 rows=153600000 width=8)
```

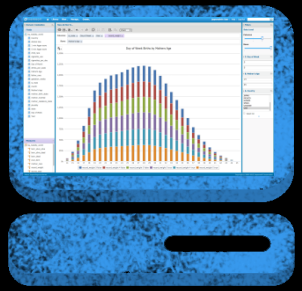
Cluster Performance During Query Execution

Cluster performance is shown from 3 minutes prior to query execution through 3 minutes after query completion.

CPU Utilization



JDBC/ODBC



ACTUATE
The BIRT Company™

birst

JASPERSOFT
the intelligence inside

MicroStrategy

PERVASIVE

pentaho

tableau
SOFTWARE

Firefox

AWS Marketplace: Amazon Redshift

https://aws.amazon.com/marketplace/b/6297422011/ref=gtw_navlft_node_6297422011#

awsmarketplace

Amazon Web Services Home

Sign in or Create a new account

Your Account | Help

Shop All Categories | Search AWS Marketplace | GO | Your Software

Categories

All Categories

Business Software

Amazon Redshift

Filters

Operating System

Windows releases:
All

Linux/UNIX distributions:
All

Amazon Redshift

Amazon Redshift is a fast and powerful, fully managed, petabyte-scale data warehouse service in the cloud. Amazon Redshift offers you fast query performance when analyzing virtually any size data set using the same SQL-based tools and business intelligence applications you use today. With a few clicks in the AWS Management Console, you can launch an Amazon Redshift cluster, starting with a few hundred gigabytes of data and scaling to a petabyte or more.

Read more

Resources

Amazon Redshift
Fast and Powerful, Petabyte-Scale Data Warehouse Service

AWS Partners
AWS Partners can help you with data warehousing, BI and more

Featured Products

JASPERSOFT the intelligence inside

Jaspersoft Reporting and Analytics fo...
Jaspersoft
64-bit Amazon Machine Image
$0.00 to $5.54/hr for software

MicroStrategy

Free MicroStrategy Suite
MicroStrategy
64-bit Amazon Machine Image
$0.00/hr for software

ATTUNITY CloudBeam

Attunity CloudBeam
Attunity
Software as a Service
From $4.95 per month and $0.05 per GB

birst

Birst
Birst
Software as a Service
From $995 per month

Firefox

Jaspersoft Reporting and Analytics for A...

https://aws.amazon.com/marketplace/pp/B00B527JQ0

awsmarketplace

Amazon Web Services Home

Sign in or Create a new account

Your Account | Help

Shop All Categories | Search AWS Marketplace | GO | Your Software

Jaspersoft Reporting and Analytics for AWS

Sold by: Jaspersoft | See product video

PROMOTION - June 15- July 15, 2013 - AWS customers who run Jaspersoft Reporting and Analytics for AWS using a Standard XL instance type through AWS Marketplace will receive $175 of AWS Promotional Credit if they run it for a total of 200 hours between June 15, 2013 and July 31, 2013. For qualifying customers, the Promotional Credit will be sent to the email address registered to their AWS Marketplace account. Restrictions apply; see https://aws.amazon.com/marketplace/help/201193990 . Jaspersoft BI Professional for AWS is a commercial open source reporting and analytics server built for AWS that ... Read more

Customer Rating	★★★★★ (2 Customer Reviews)
Latest Version	5.1.0
Base Operating System	Linux/Unix, Amazon Linux 2013.03
Delivery Method	64-bit Amazon Machine Image (AMI) (Learn more)
Support	See details below

Continue

You will have an opportunity to review your order before launching or being charged.

Pricing Details

For region US East (Virginia)

Hourly Fees

Base Operating System Linux/Unix, Amazon Linux 2013.03

Delivery Method 64-bit Amazon Machine Image (AMI) (Learn more)

Support See details below

AWS Services Required Amazon EC2, Amazon EBS

Highlights

- PROMOTION - June 15- July 15, 2013 - AWS customers who run Jaspersoft Reporting and Analytics for AWS using a Standard XL instance type through AWS Marketplace will receive $175 of AWS Promotional Credit if they run it for a total of 200 hours between June 15, 2013 and July 31, 2013. For qualifying customers, the Promotional Credit will be sent to the email address registered to their AWS Marketplace account. Restrictions apply; see https://aws.amazon.com/marketplace/help/201193990 .
- 10 Minutes to Your AWS Data: purpose-built for AWS, our reporting and analytics server allows you to quickly and easily connect to your Amazon RDS and Redshift data. In under 10 minutes you can be reporting on and analyzing your data.
- BI for Your Business or App: built to modern web standards with a HTML5 UI and web service APIs, our flexible BI suite can be used to analyze your business or deliver stunning interactive reports and dashboards inside your app.

Product Description

PROMOTION - June 15- July 15, 2013 - AWS customers who run Jaspersoft Reporting and Analytics for AWS using a Standard XL instance type through AWS Marketplace will receive $175 of AWS Promotional Credit if they run it for a total of 200 hours between June 15, 2013 and July 31, 2013. For qualifying customers, the Promotional Credit will be sent to the email

Pricing Details

For region **US East (Virginia)**

Hourly Fees

Total hourly fees will vary by instance type and EC2 region.

EC2 Instance Type	Software	EC2 *	Total
Standard Medium (m1.medium)	$0.40/hr	$0.12/hr	$0.52/hr
Standard Large (m1.large)	$0.80/hr	$0.24/hr	$1.04/hr
Standard XL (m1.xlarge)	$0.00/hr	$0.48/hr	$0.48/hr
High-Memory XL (m2.xlarge)	$1.39/hr	$0.41/hr	$1.80/hr
High-Memory 2XL (m2.2xlarge)	$2.77/hr	$0.82/hr	$3.59/hr
High-Memory 4XL (m2.4xlarge)	$5.54/hr	$1.64/hr	$7.18/hr
High-CPU XL (c1.xlarge)	$2.03/hr	$0.58/hr	$2.61/hr

EBS Storage Fees **

$0.10 / GB / Month for Standard EBS Storage

* Assumes On-Demand EC2 pricing; prices for Reserved and Spot Instances will be lower. See pricing details.

** Data transfer fees not included.

Learn about instance types

Recent Product Reviews

★★★★★ 06/03/2013

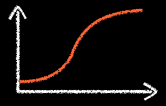
Great Dashboards, super fast to deploy, very cost effective

This is truly a disruptive product offering. The pricing is extremely cost effective and I had it setup with dashboar...
Read more

	Price Per Hour for HS1.XL Single Node	Effective Hourly Price Per TB	Effective Annual Price per TB
On-Demand	$ 0.850	$ 0.425	$ 3,723
1 Year Reservation	$ 0.500	$ 0.250	$ 2,190
3 Year Reservation	$ 0.228	$ 0.114	$ 999

DATA WAREHOUSING DONE THE AWS WAY

Easy to provision and scale up massively



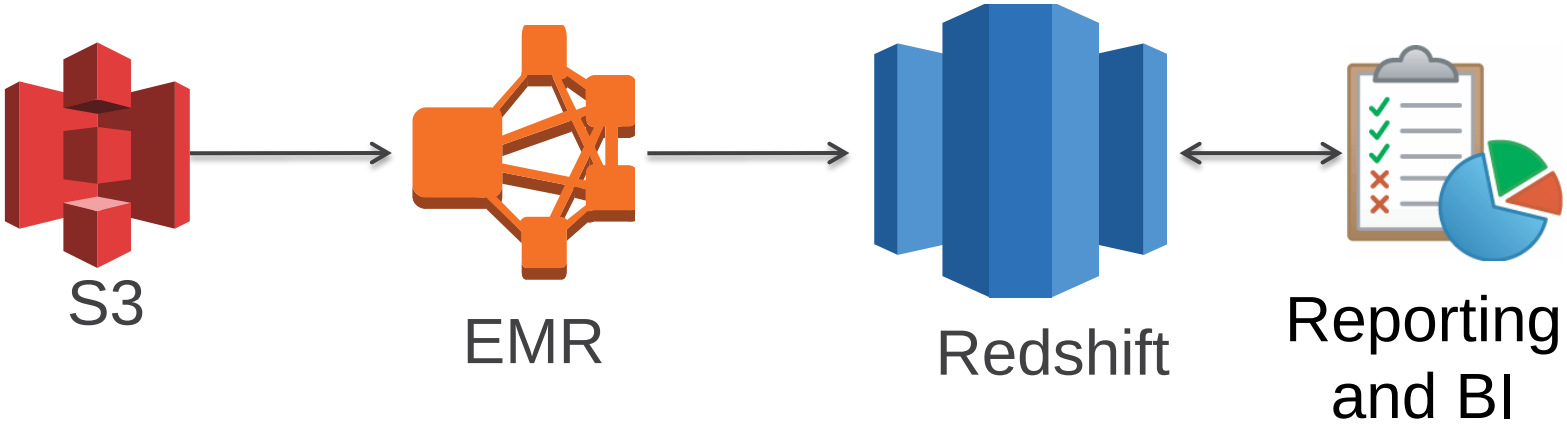
No upfront costs, pay as you go

Really fast performance at a really low price

Open and flexible with support for popular tools

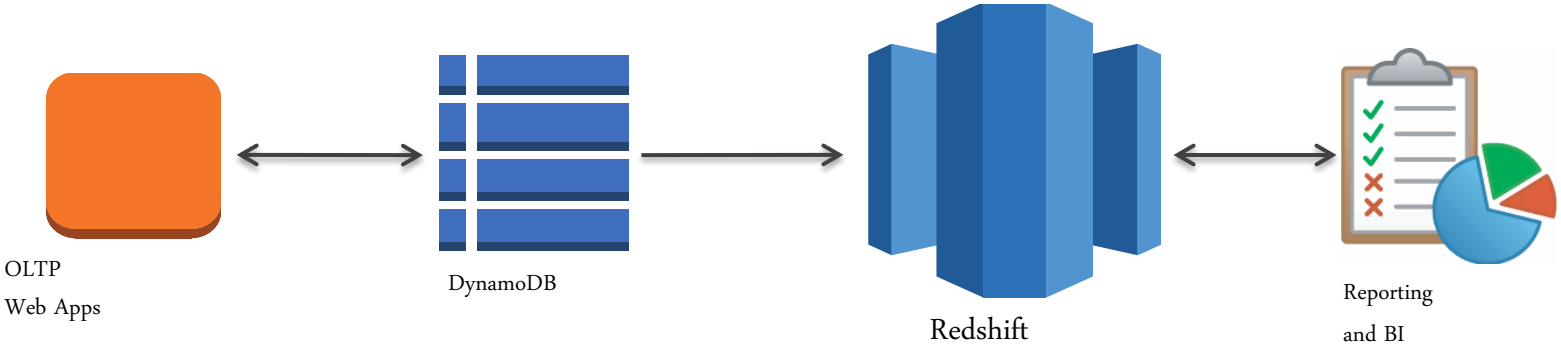
USAGE SCENARIOS

Cloud ETL for Big Data



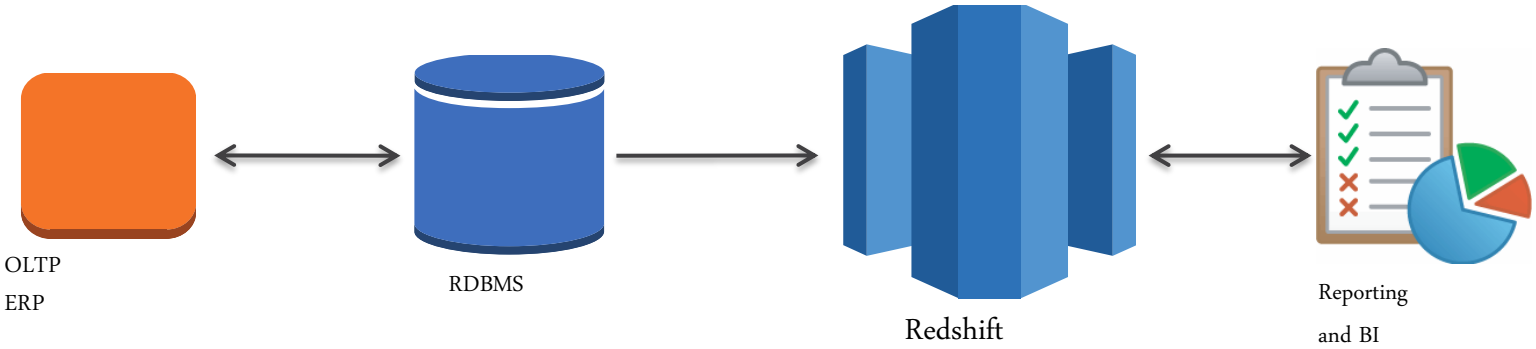
- Maintain online SQL access to historical logs
- Transformation and enrichment with EMR
- Longer history ensures better insight

Live archive for (structured) Big Data



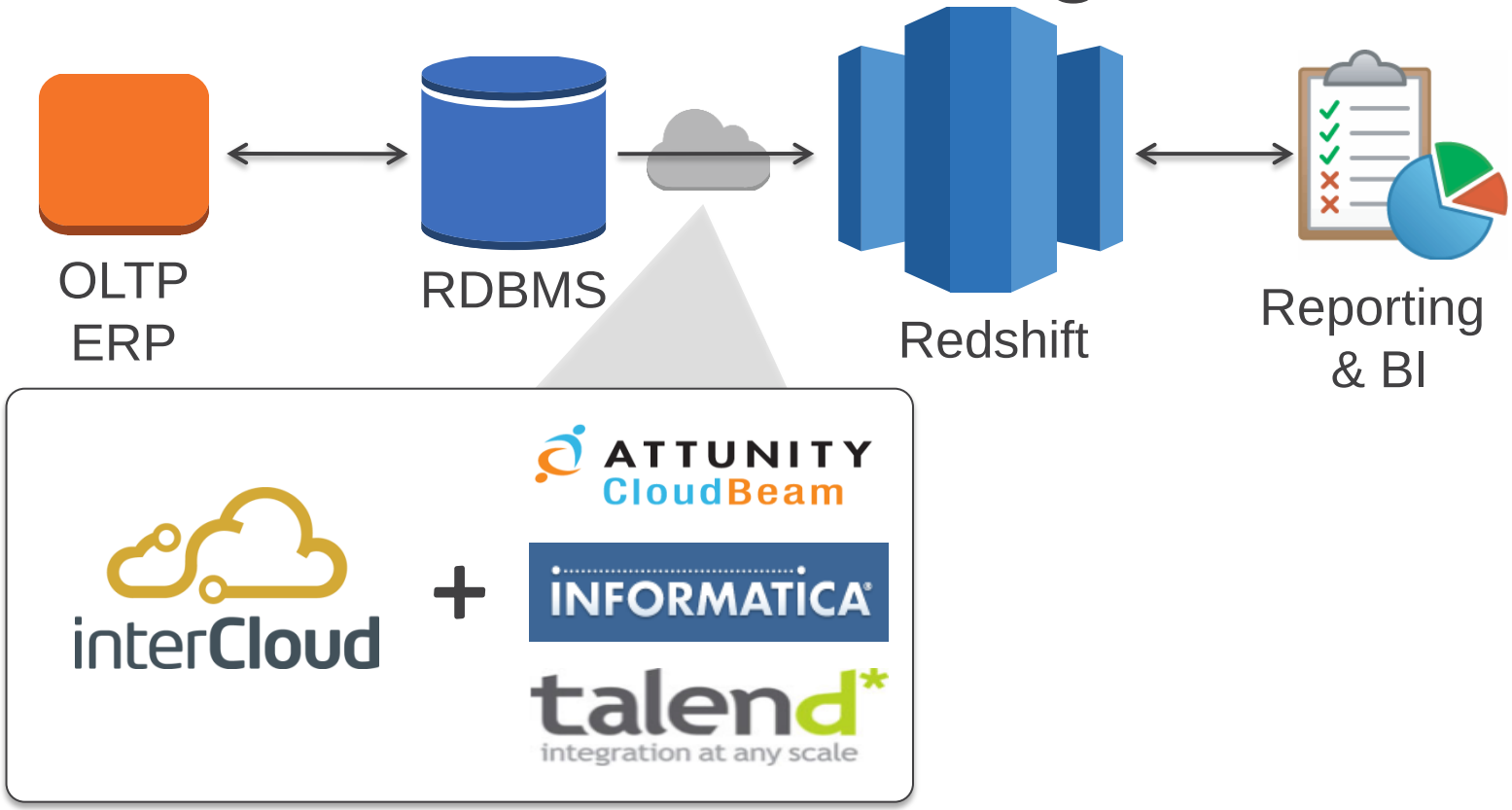
- Direct integration with copy command
- High velocity data
- Data ages into Redshift
- Low cost, high scale option for new apps

Reporting Warehouse



- Accelerated operational reporting
- Support for short-time use cases
- Data compression, index redundancy

On-Premises Integration



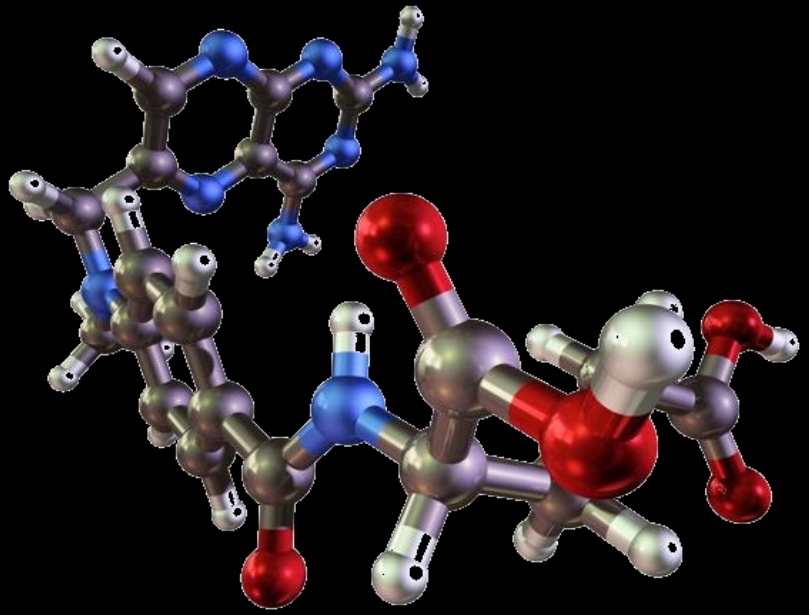
GENERATE ➔ STORE ➔ **ANALYZE** ➔ SHARE

Amazon EC2
Amazon Elastic MapReduce

AMAZON EC2

ELASTIC COMPUTE CLOUD

SCHRÖDINGER®



CLUSTER GPU

QUADRUPLE EXTRA LARGE

2x Intel Xeon X5570, quad-core **Nehalem** architecture

2x NVIDIA Tesla Fermi **M2050 GPUs**

22 GB of memory – 1.7 TB of storage

ON A SINGLE INSTANCE

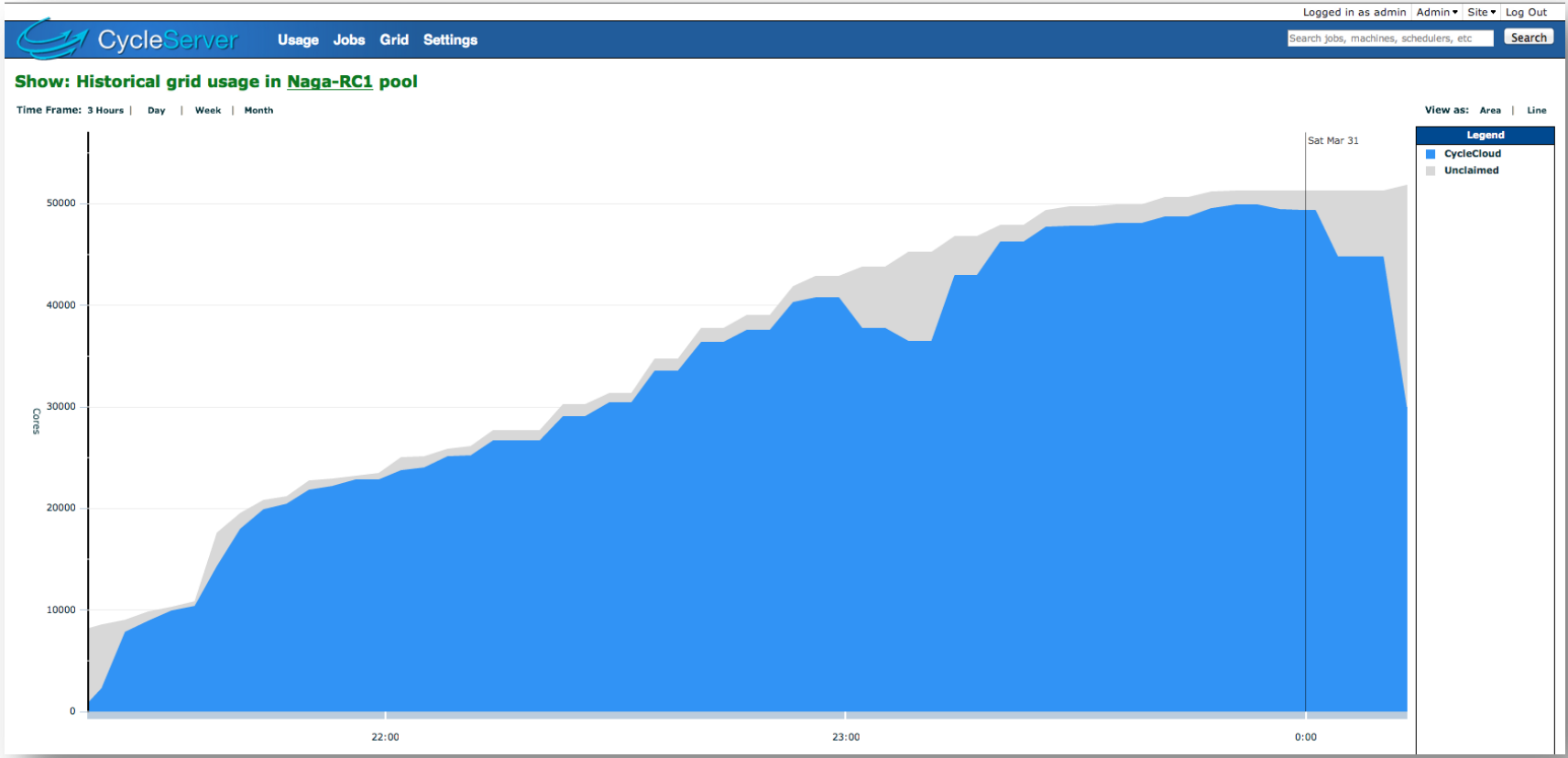
COMPUTE TIME: **4h**

COST: 4h x $2.1 = **$8.4**

ON MULTIPLE INSTANCES

COMPUTE TIME: **1h**

COST: 1h x 4 x $2.1 = **$8.4**


Logged in as admin
Admin
Site
Log Out
CycleServer
Usage
Jobs
Grid
Settings
Search jobs, machines, schedulers, etc
Search
Show: Historical grid usage in Naga-RC1 pool
Time Frame: 3 Hours | Day | Week | Month
View as: Area | Line
Legend
CycleCloud
Unclaimed
Sat Mar 31
Cores
50000
40000
30000
20000
10000
0
22:00
23:00
0:00

For 3 hours

$**4828.85**/hr

instead of

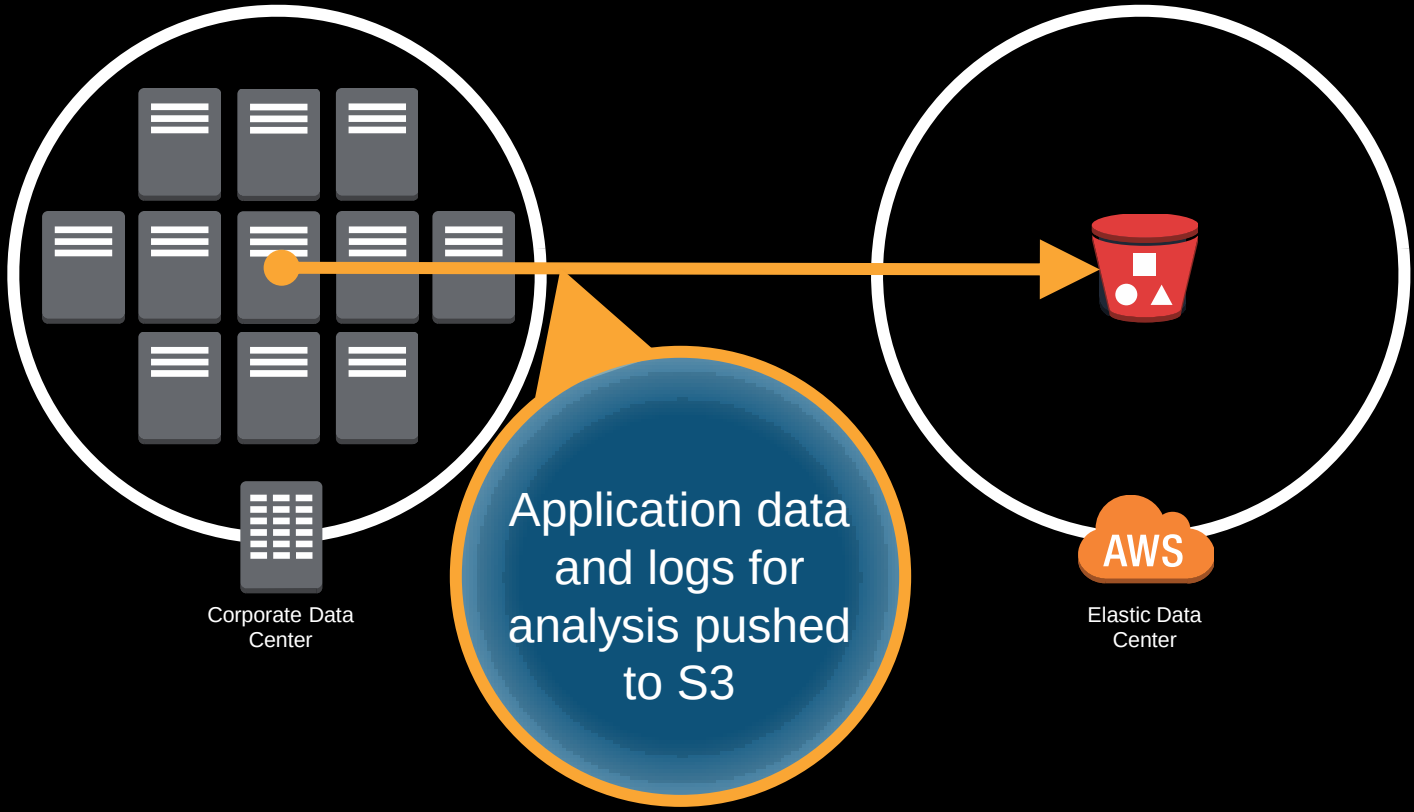
$**20+ MILLIONS**

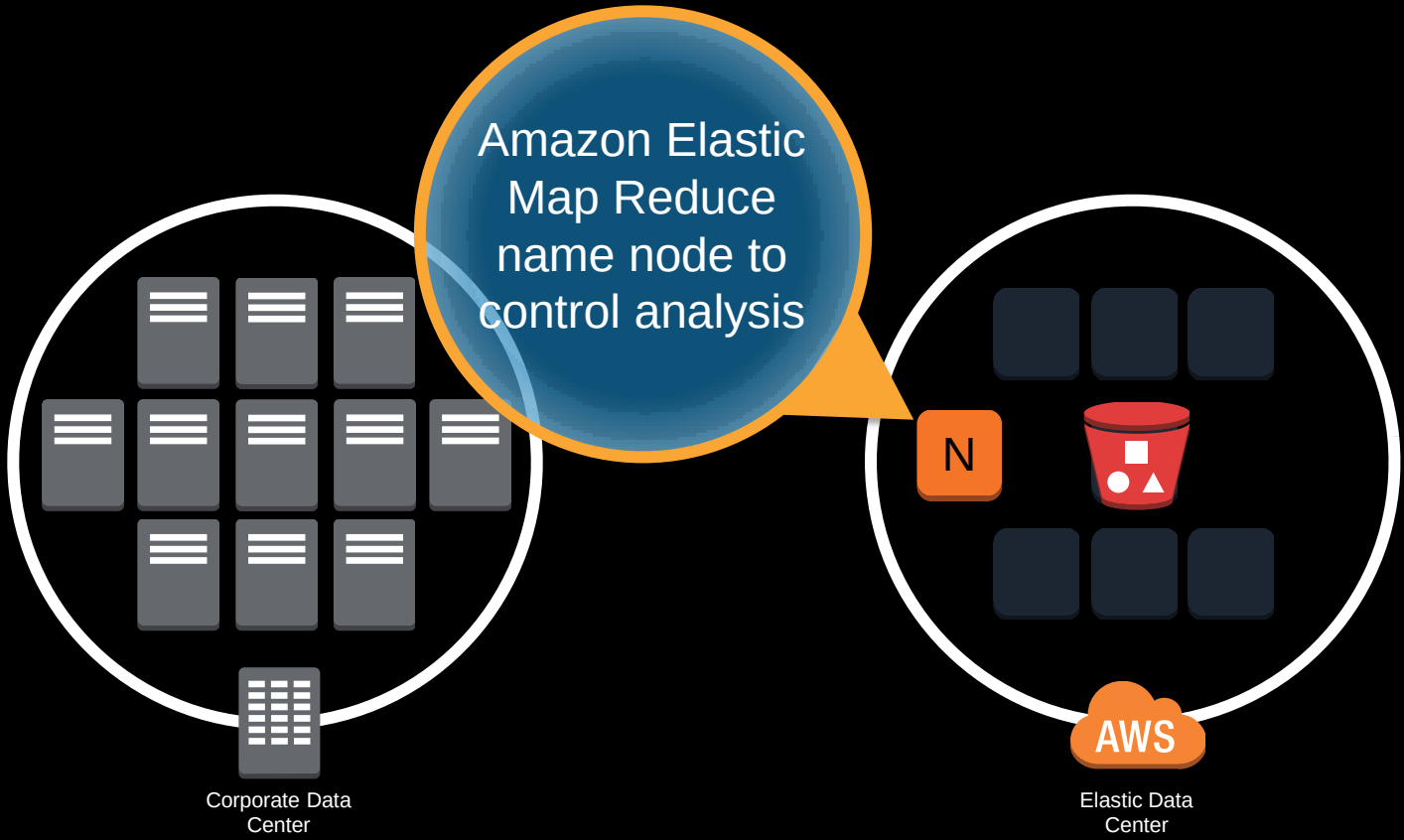
in infrastructure

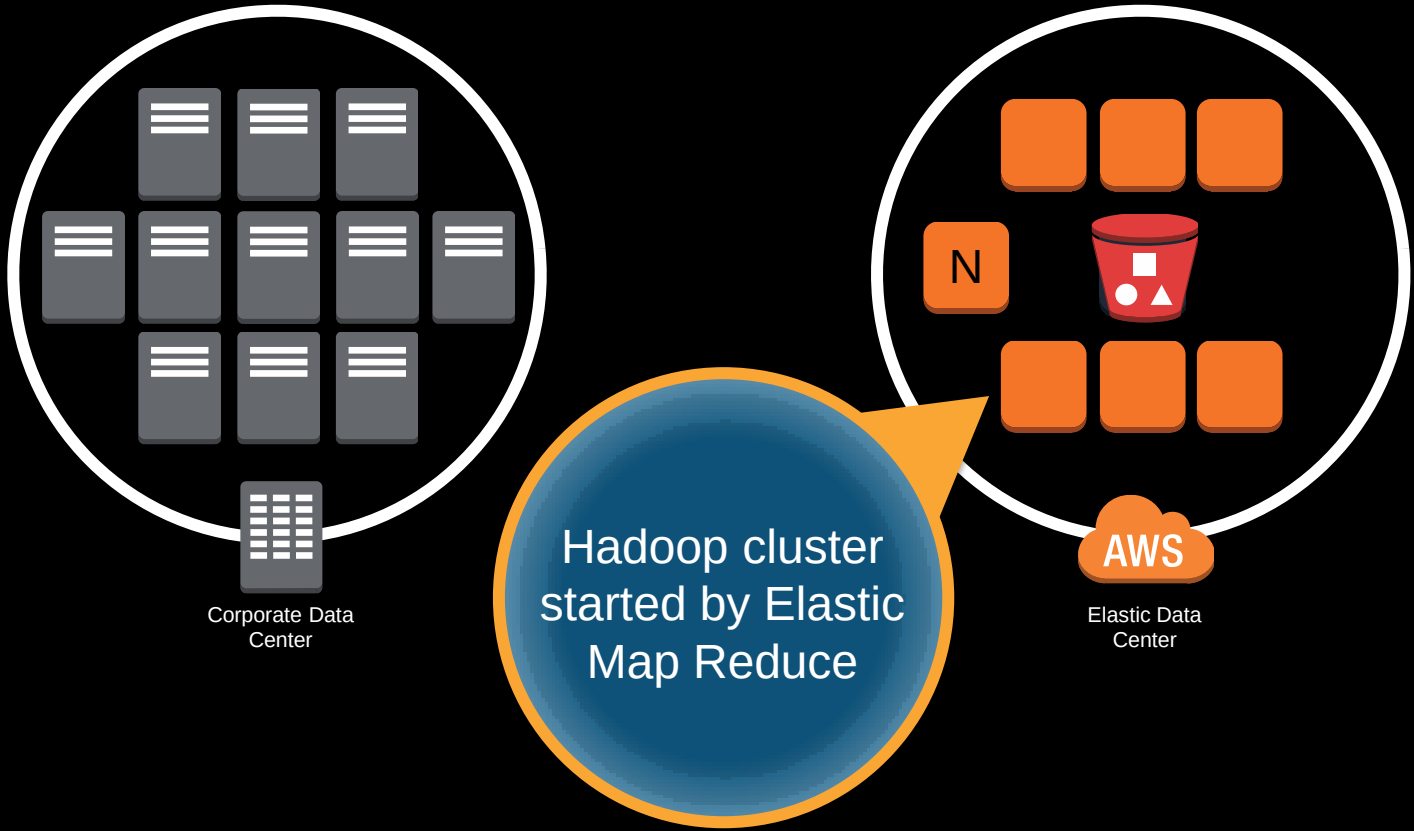
AMAZON ELASTIC MAPREDUCE

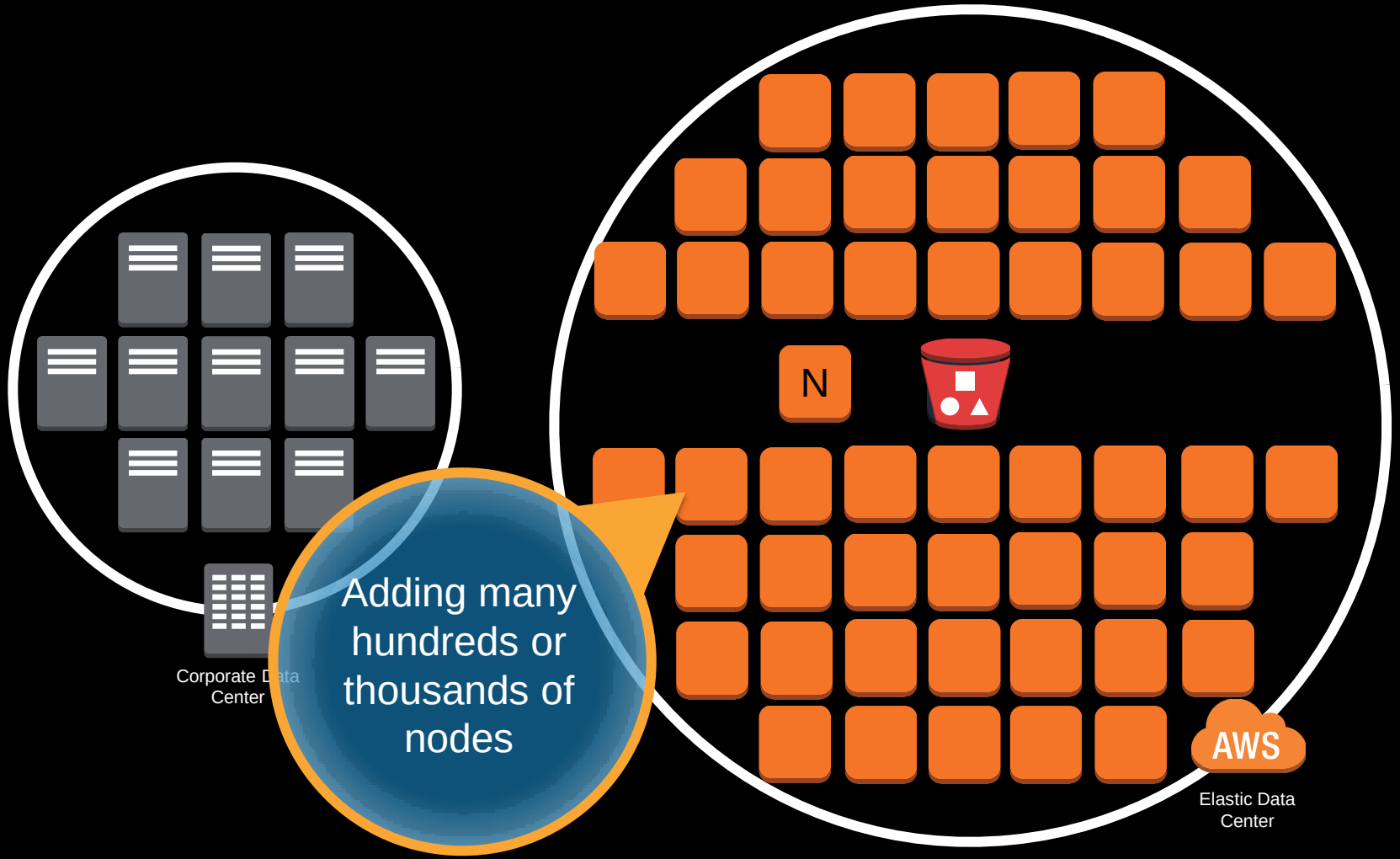
HADOOP AS A SERVICE

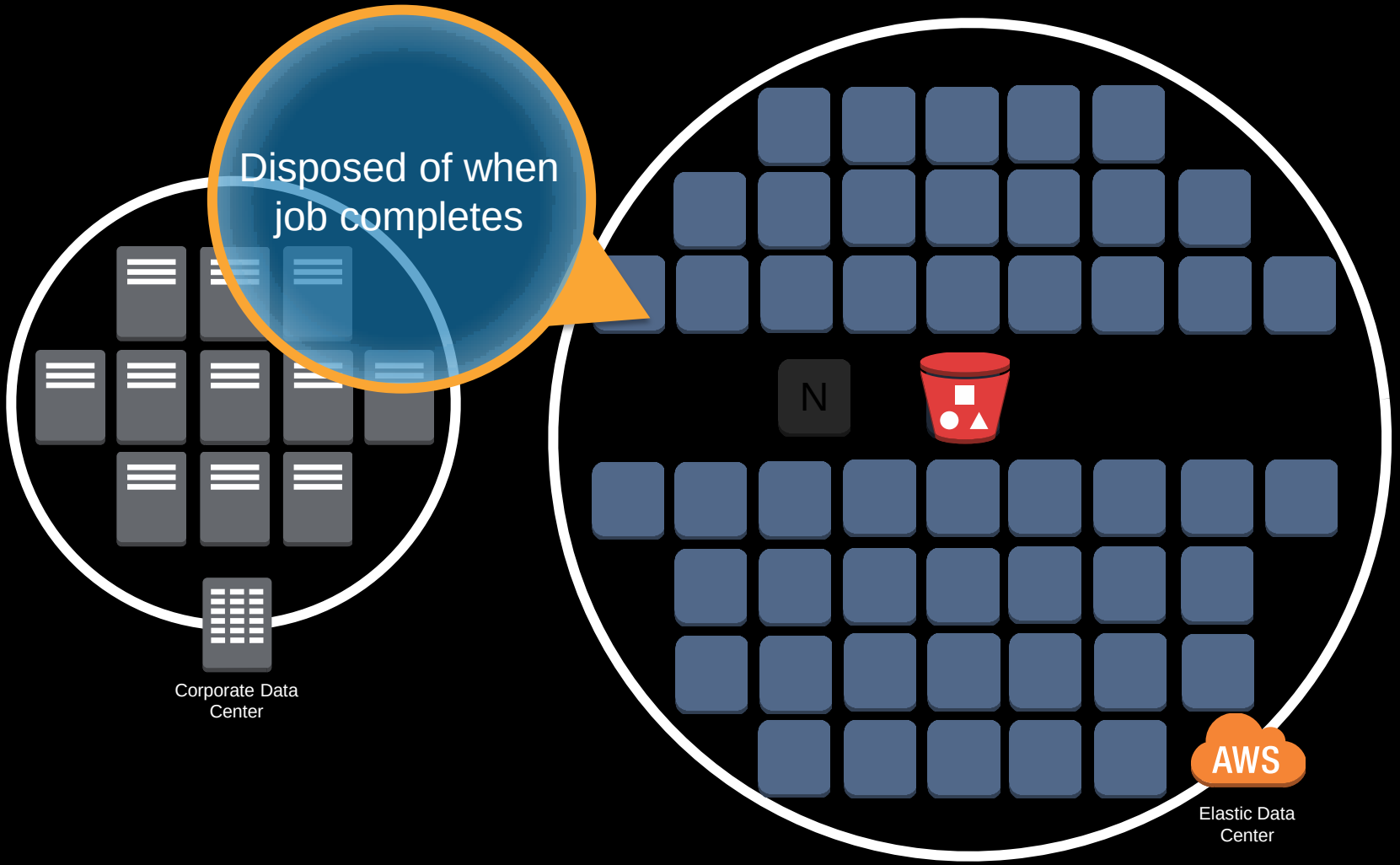
- A FRAMEWORK
- SPLITS DATA INTO PIECES
- LETS PROCESSING OCCUR
- GATHERS THE RESULTS

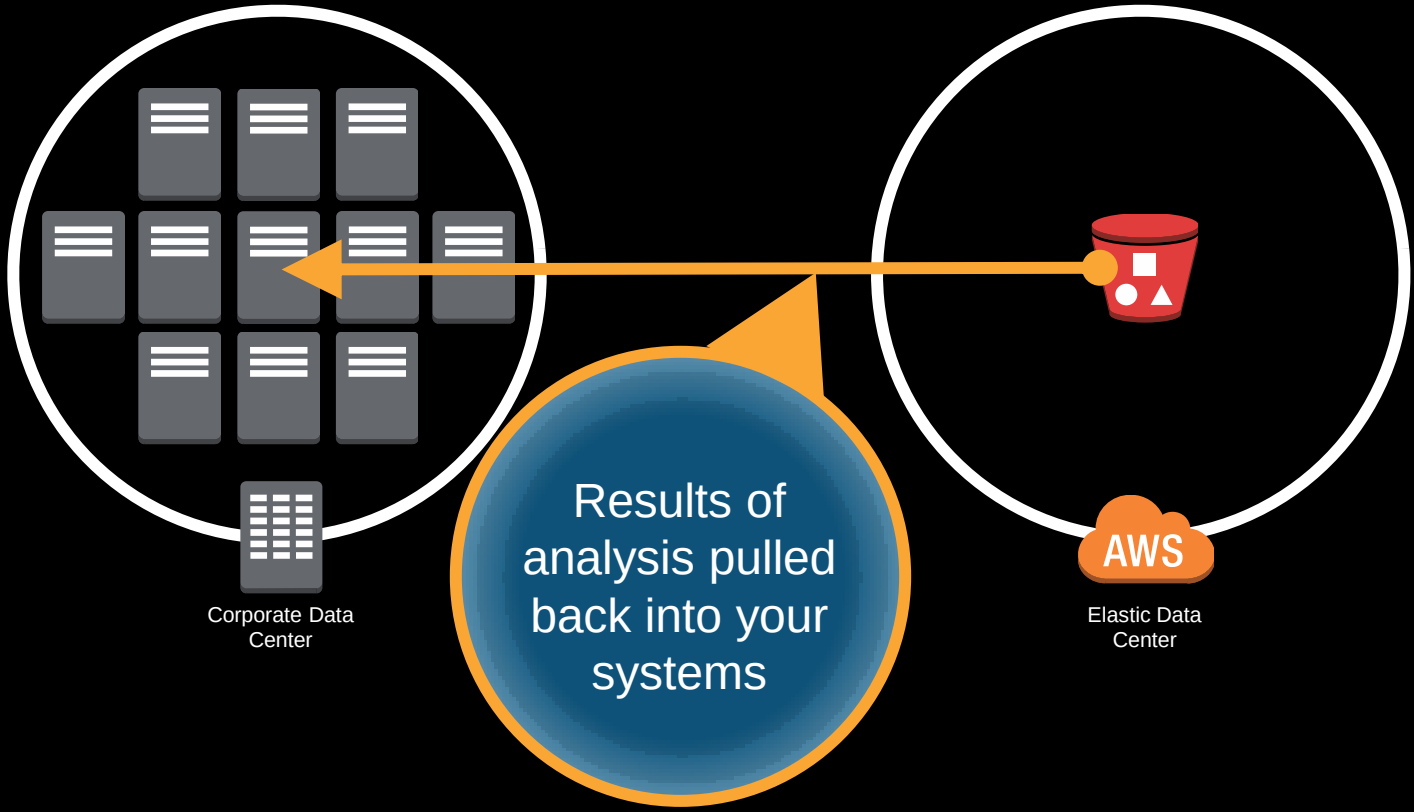

Application data and logs for analysis pushed to S3
AWS
Corporate Data Center
Elastic Data Center


Amazon Elastic Map Reduce name node to control analysis
N
AWS
Corporate Data Center
Elastic Data Center


N
AWS
Hadoop cluster started by Elastic Map Reduce
Corporate Data Center
Elastic Data Center


N
Adding many hundreds or thousands of nodes
Corporate Data Center
AWS
Elastic Data Center


Disposed of when job completes
N
AWS
Corporate Data Center
Elastic Data Center


Corporate Data Center
Results of analysis pulled back into your systems
AWS
Elastic Data Center

LOGIN

waze OUTSMARTING TRAFFIC, TOGETHER.

LIVE MAP SUPPORT BLOG ABOUT

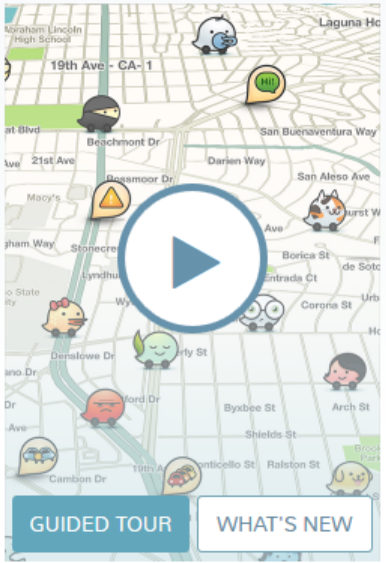
GET THE BEST ROUTE, EVERY DAY, WITH REAL TIME HELP FROM OTHER DRIVERS.

Waze is the world's fastest-growing community-based traffic and navigation app. Join other drivers in your area who share real-time traffic and road info, saving everyone time and gas money on their daily commute.

WAZE. OUTSMARTING TRAFFIC, TOGETHER.

Download on the App Store

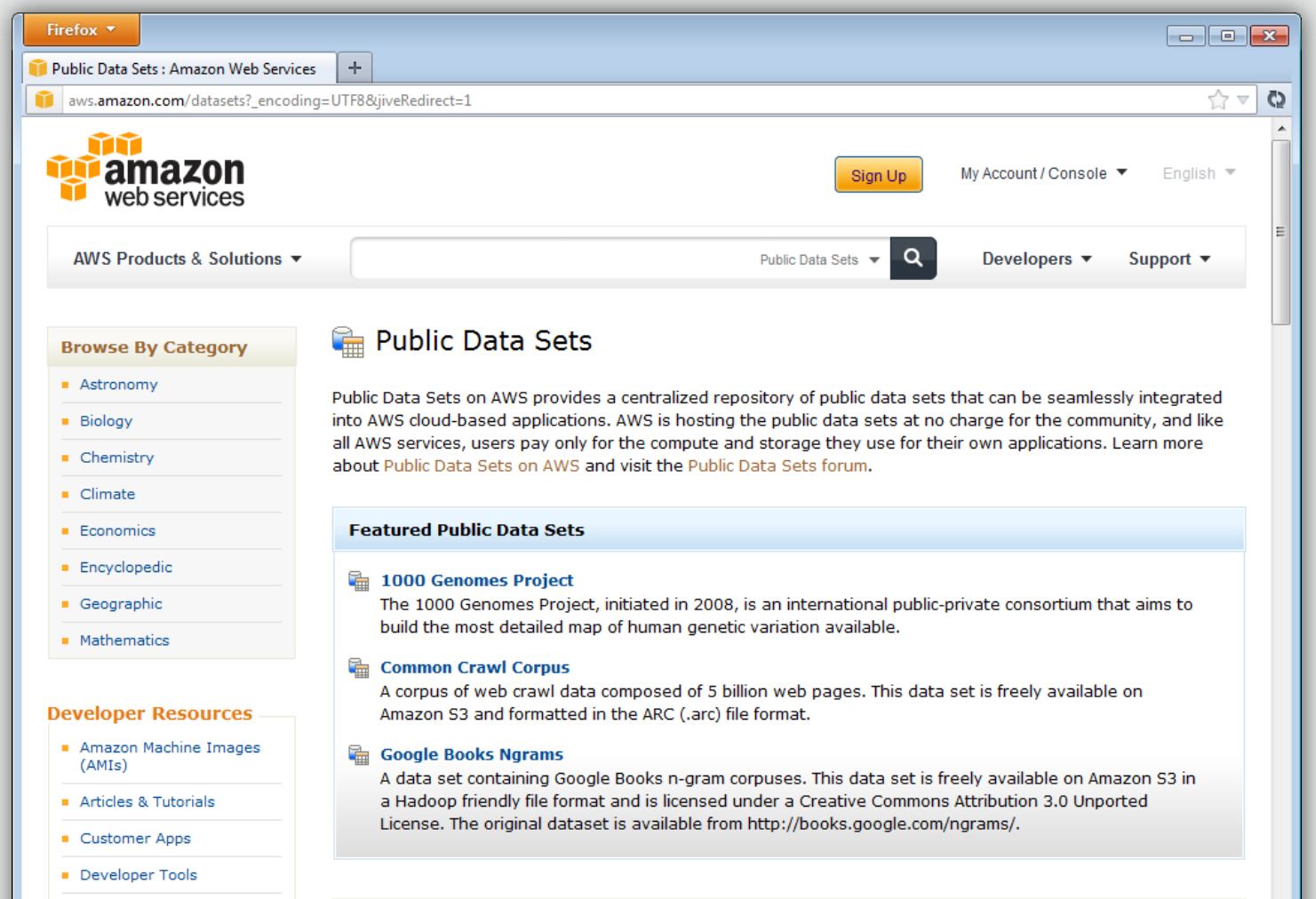
GET IT ON Google play



Nothing can beat real people working together

Amazon S3,
Amazon DynamoDB,
Amazon RDS,
Amazon Redshift,
Data on Amazon EC2

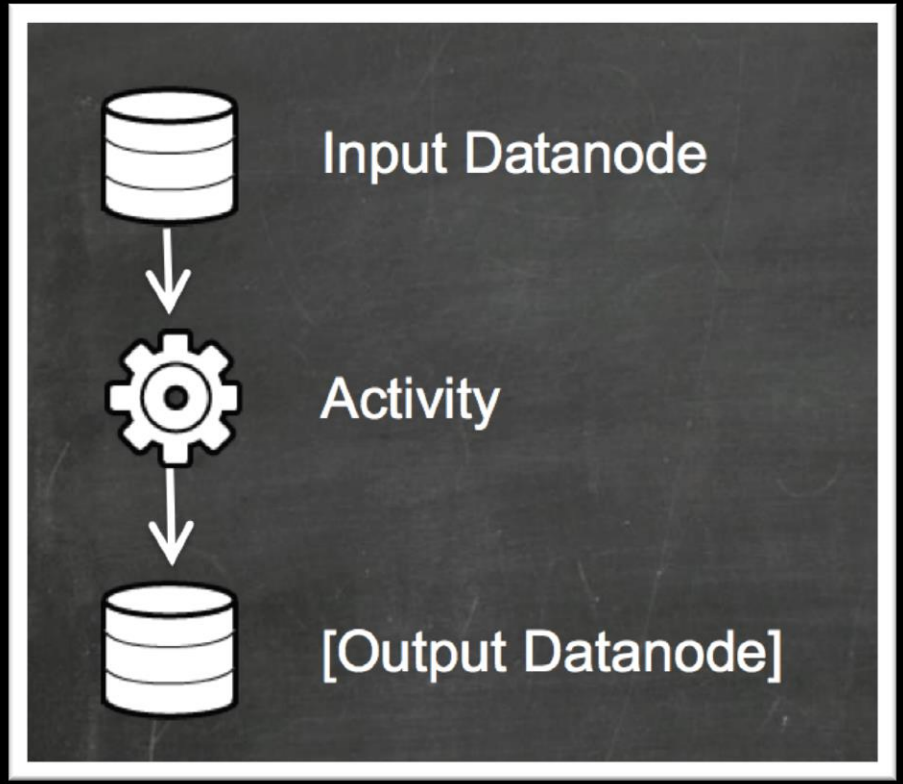
GENERATE ➔ STORE ➔ ANALYZE ➔ **SHARE**

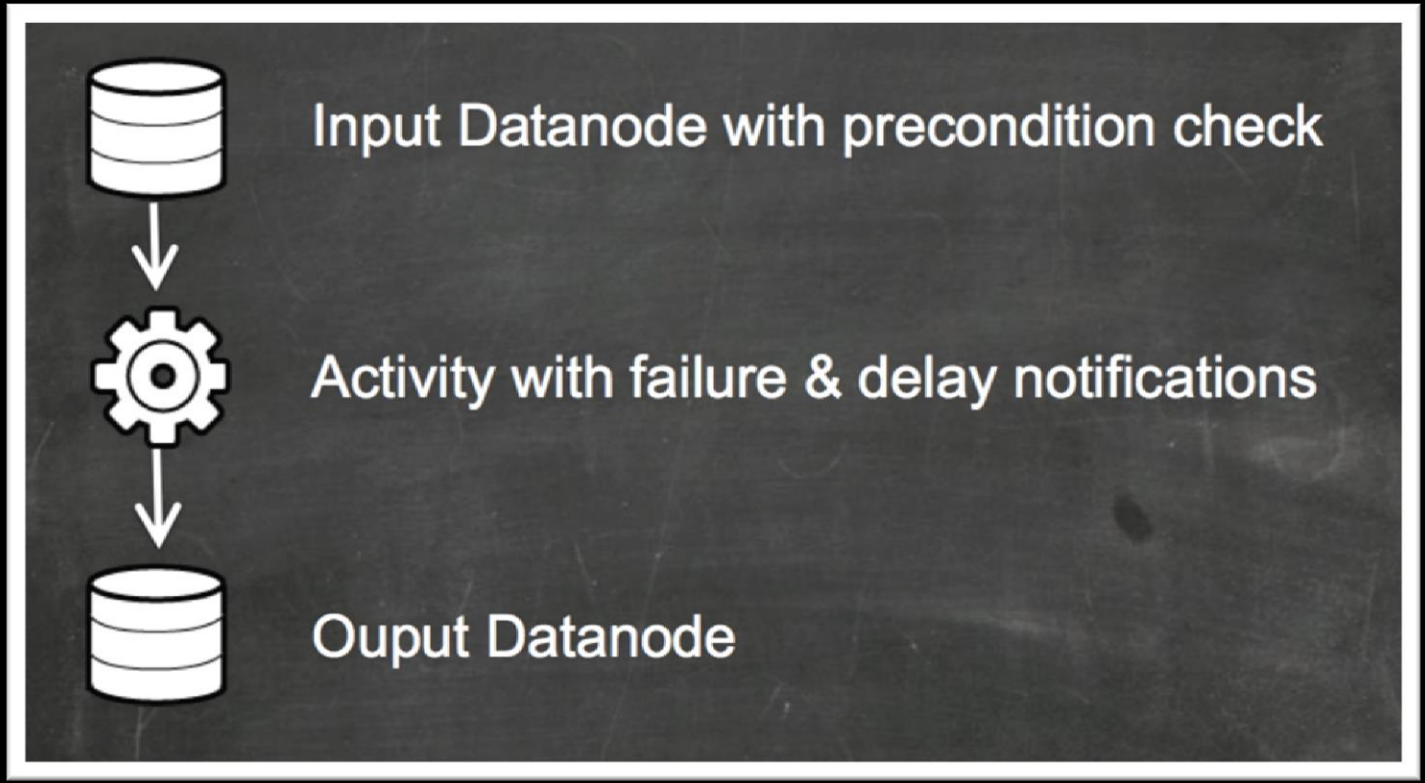

Firefox
Public Data Sets : Amazon Web Services
aws.amazon.com/datasets?_encoding=UTF8&jiveRedirect=1
amazon web services
Sign Up
My Account / Console
English
AWS Products & Solutions
Public Data Sets
Developers
Support
Browse By Category
Astronomy
Biology
Chemistry
Climate
Economics
Encyclopedic
Geographic
Mathematics
Developer Resources
Amazon Machine Images (AMIs)
Articles & Tutorials
Customer Apps
Developer Tools
Public Data Sets
Public Data Sets on AWS provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications. AWS is hosting the public data sets at no charge for the community, and like all AWS services, users pay only for the compute and storage they use for their own applications. Learn more about Public Data Sets on AWS and visit the Public Data Sets forum.
Featured Public Data Sets
1000 Genomes Project
The 1000 Genomes Project, initiated in 2008, is an international public-private consortium that aims to build the most detailed map of human genetic variation available.
Common Crawl Corpus
A corpus of web crawl data composed of 5 billion web pages. This data set is freely available on Amazon S3 and formatted in the ARC (.arc) file format.
Google Books Ngrams
A data set containing Google Books n-gram corpuses. This data set is freely available on Amazon S3 in a Hadoop friendly file format and is licensed under a Creative Commons Attribution 3.0 Unported License. The original dataset is available from http://books.google.com/ngrams/.

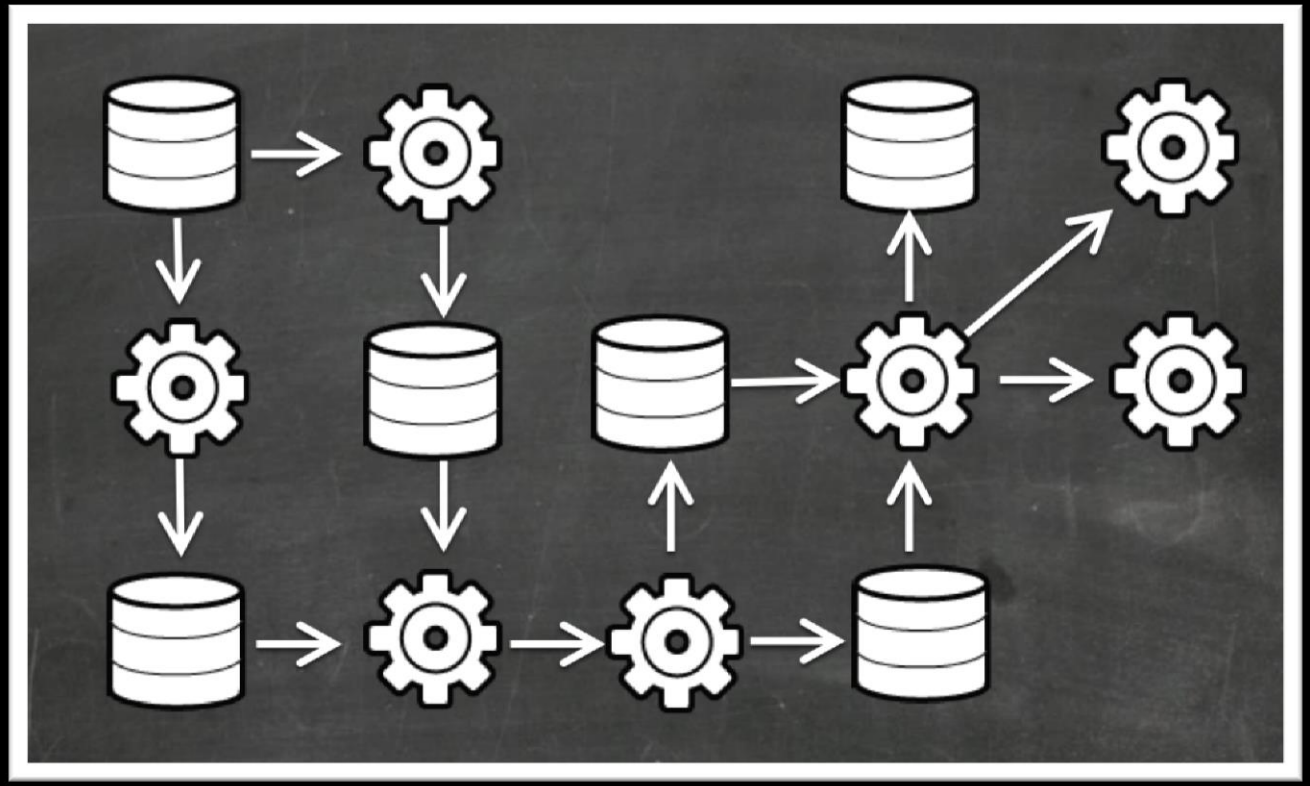
GENERATE ➔ STORE ➔ ANALYZE ➔ SHARE

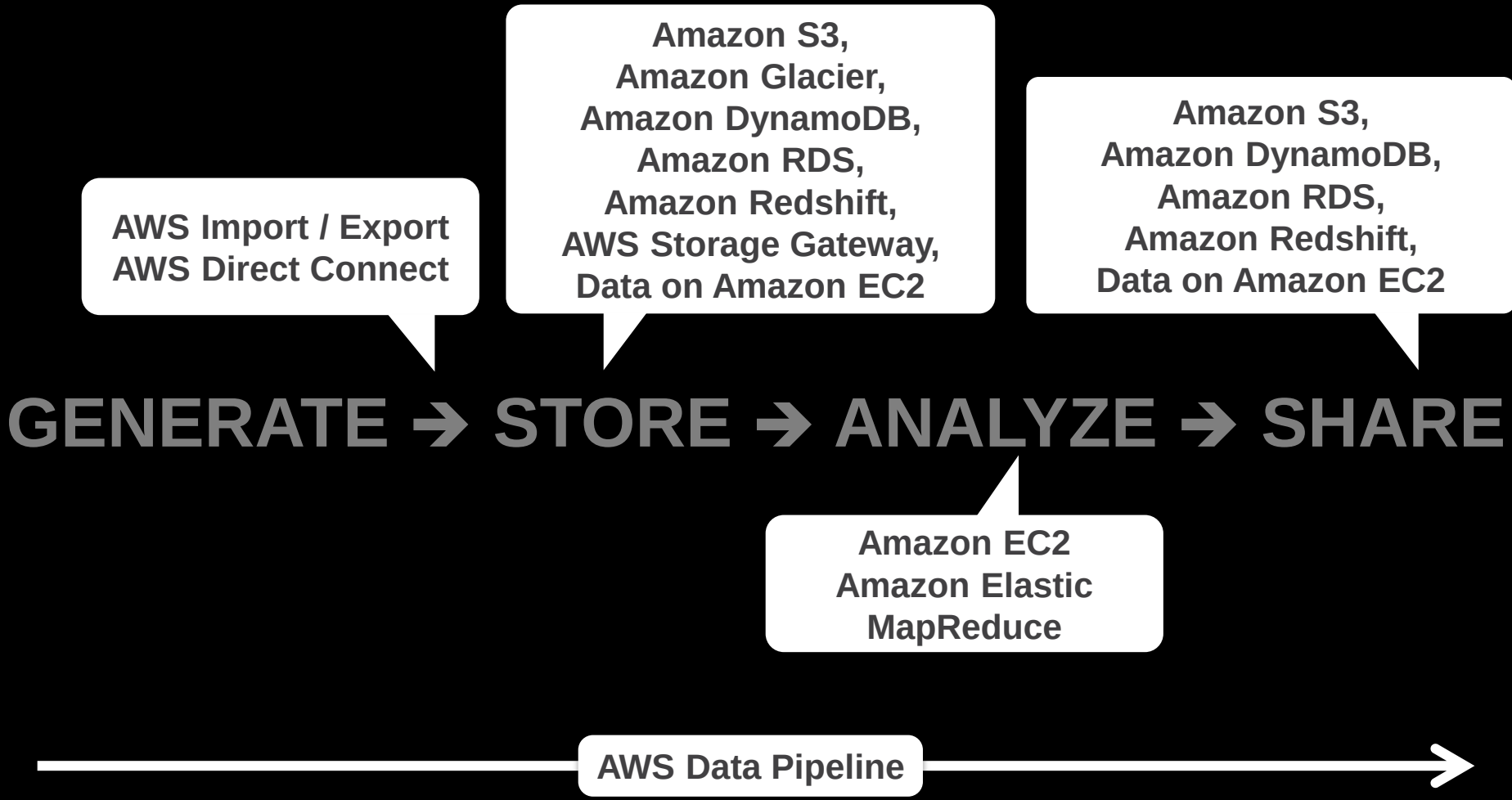
AWS Data Pipeline

- Data-intensive orchestration and automation
- Reliable and scheduled
- Easy to use, drag and drop
- Execution and retry logic
- Map data dependencies
- Create and manage compute resources


Input Datanode
Activity
[Output Datanode]


Input Datanode with precondition check
Activity with failure & delay notifications
Ouput Datanode




AWS Import / Export
AWS Direct Connect
Amazon S3,
Amazon Glacier,
Amazon DynamoDB,
Amazon RDS,
Amazon Redshift,
AWS Storage Gateway,
Data on Amazon EC2
Amazon S3,
Amazon DynamoDB,
Amazon RDS,
Amazon Redshift,
Data on Amazon EC2
GENERATE → STORE → ANALYZE → SHARE
Amazon EC2
Amazon Elastic
MapReduce
AWS Data Pipeline

FROM DATA TO ACTIONABLE INFORMATION


amazon
web services

MUSIXMATCH

Stefano Rodighiero

MXM FACTS

- 7+ million lyrics catalogue in more than 50 distinct languages
- Daily updated with more than 1 million artists and more than 20 million music tracks
- Music Discography Meta Data: Lyrics, Artists, Albums, Songs, Biographies, Worldwide Charts
- Currently musiXmatch is the only lyrics platform allowed for worldwide licensing and has deals with top Music Publishers: Warner Chappell, Universal, BMG, EMI Publishing, Sony ATV, Peer Music, ...
- Synced lyrics!

SYNCED LYRICS

OUR DATA

MUSIC METADATA: RECORDING & PUBLISHING

OUR DATA

CONTENT USAGE

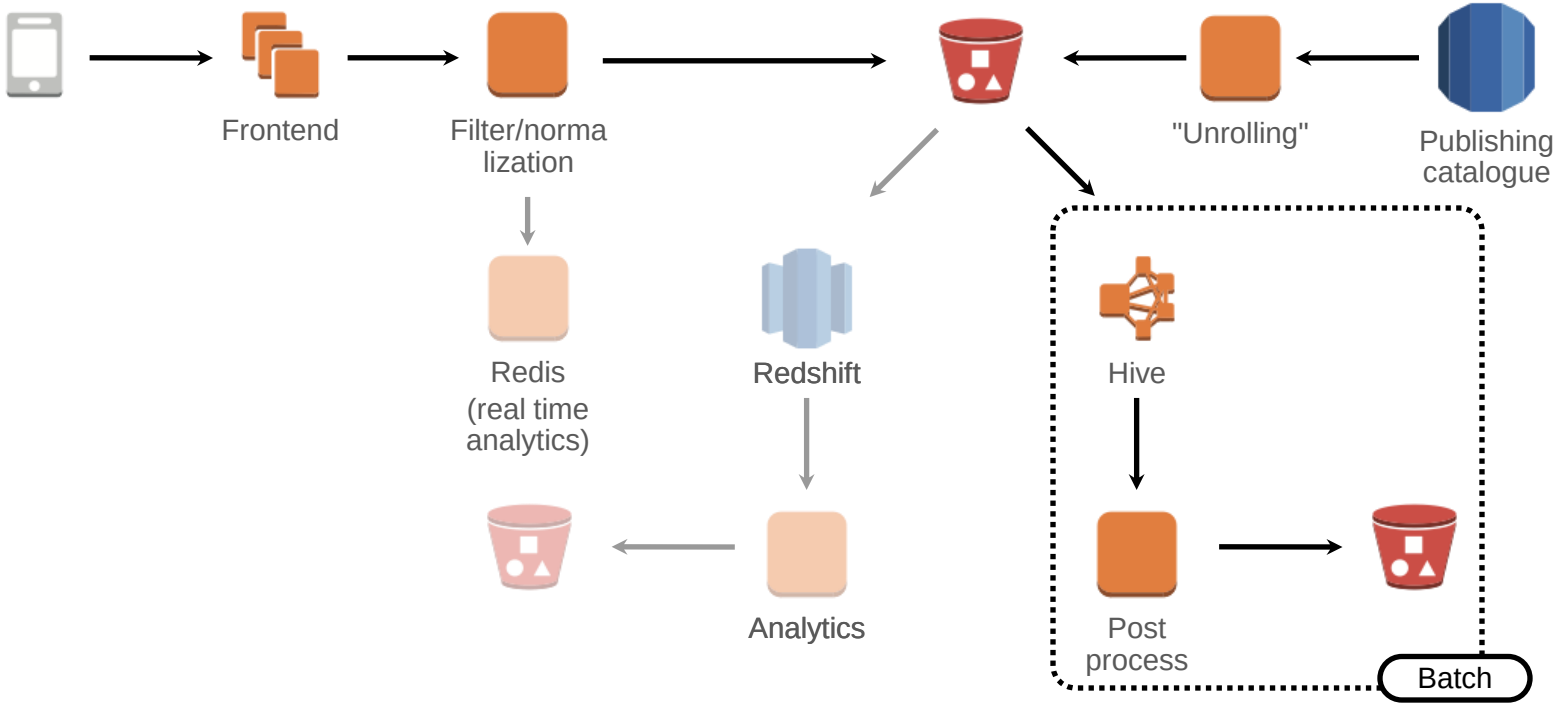
OUR DATA

OTHER SOURCES

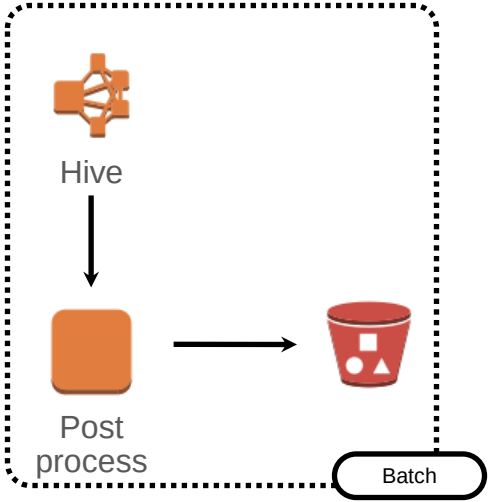
DATA ANALYSIS @ MXM

CONTENT USAGE: REPORTING & ANALYTICS

DATAFLOW

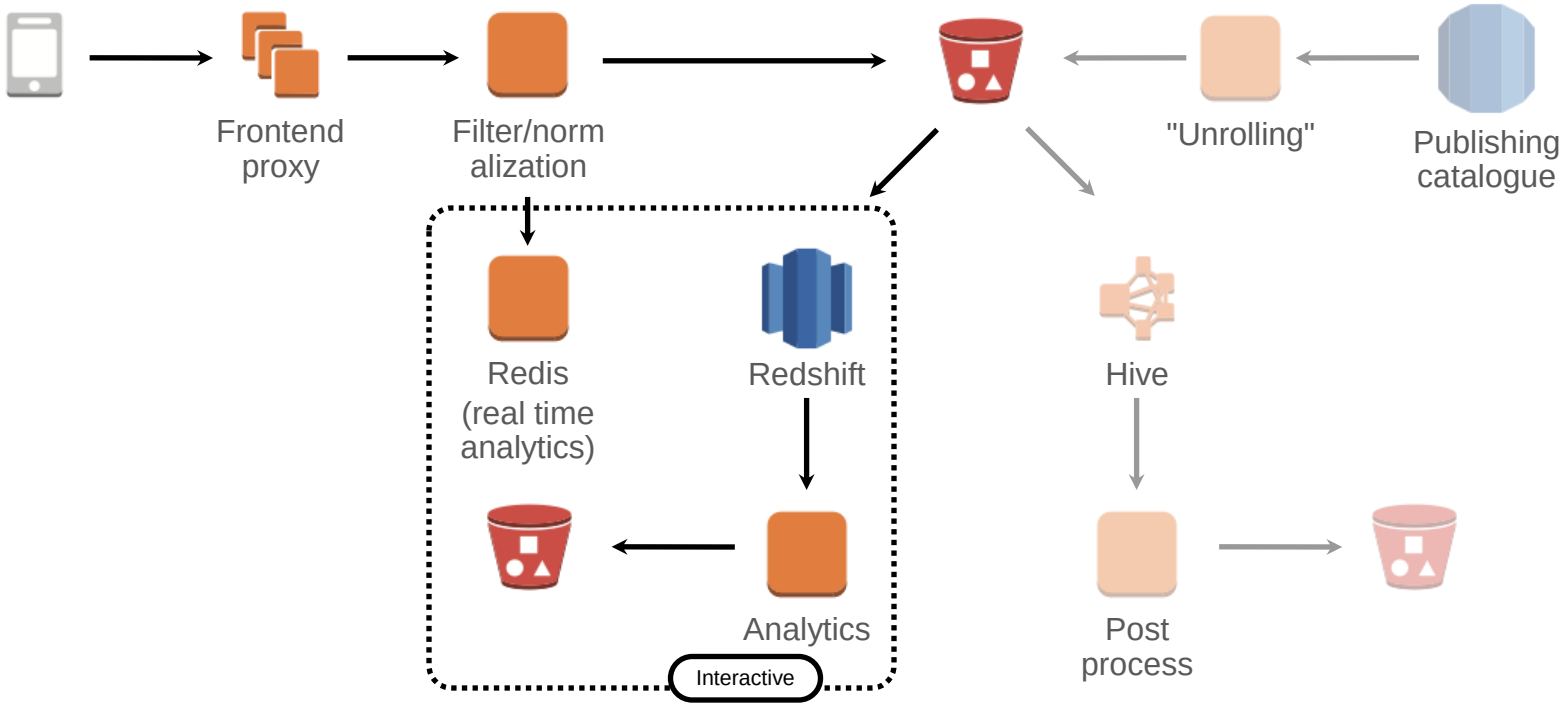

Frontend
Filter/normalization
"Unrolling"
Publishing catalogue
Redis (real time analytics)
Redshift
Hive
Analytics
Post process
Batch

BATCH REPORTING

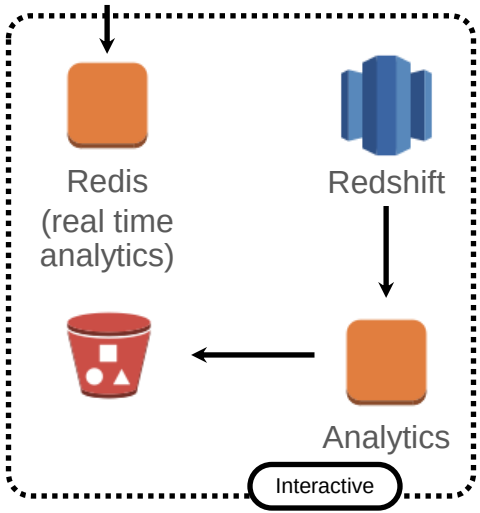


- Step 1. Aggregation of views by country, application and content type
- Step 2. Join with a 500M+ rows table
- It takes approx 1 hour with 5 c1.xlarge instances
- It used to take days with traditional techniques!
- SQL interface makes it easier to review and share the process

DATAFLOW

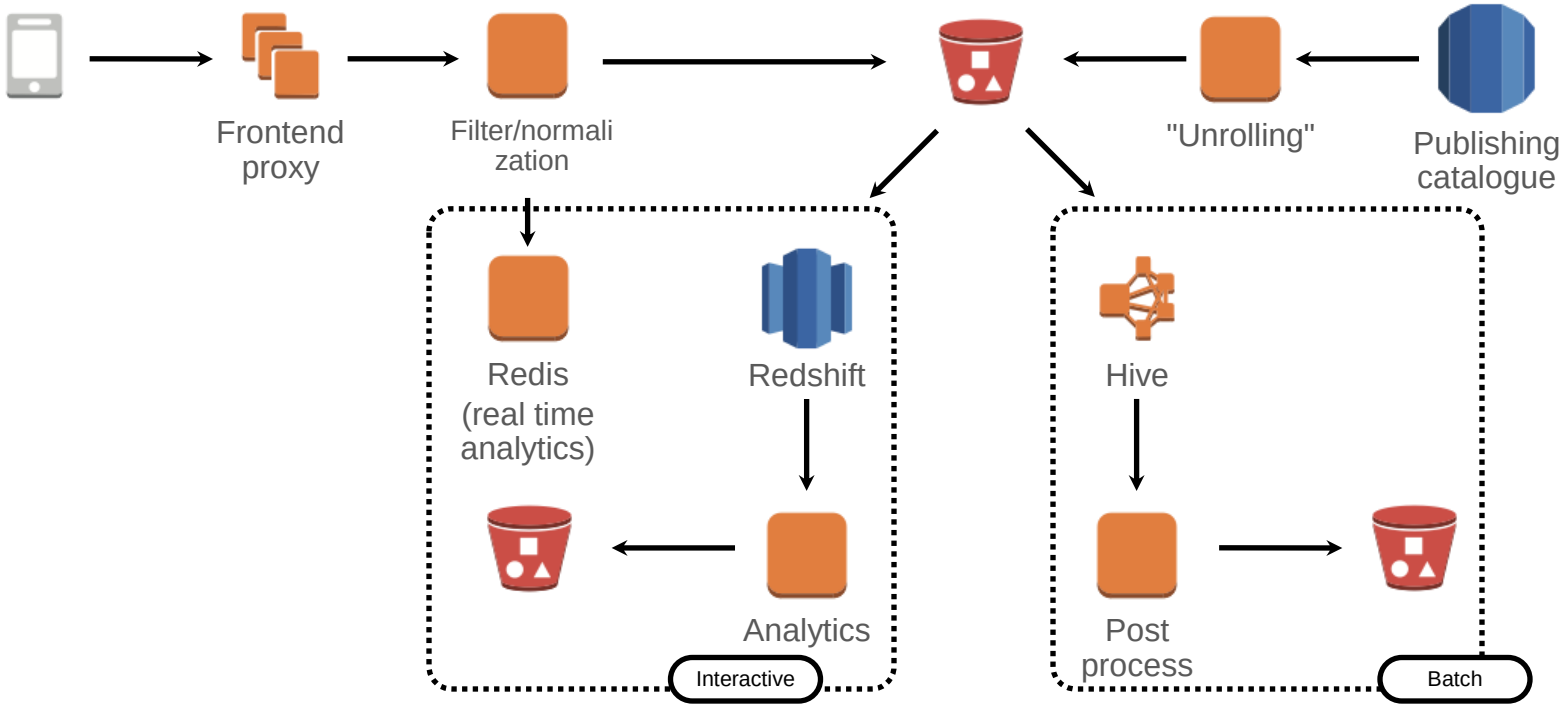

Frontend
proxy
Filter/norm
alization
"Unrolling"
Publishing
catalogue
Redis
(real time
analytics)
Redshift
Hive
Analytics
Interactive
Post
process

INTERACTIVE ANALYTICS



- SQL interface like Hive, accessible with any Postgresql client...
- ...but faster!
- Flexible costs
- With Redshift doing all the heavy lifting, it's easier to build analytics tools

DATAFLOW


Frontend proxy
Filter/normalization
"Unrolling"
Publishing catalogue
Redis (real time analytics)
Redshift
Hive
Analytics
Post process
Interactive
Batch

MUSIXMATCH

Stefano Rodighiero
stefano@musixmatch.com
@larsen

MUSIXMATCH

THANK YOU!

MUSIXMATCH


Available on the
App Store


GET IT ON
Google play


Windows
Store


Spotify

THANK YOU

hakan@amazon.lu