Getting Started with Amazon Redshift

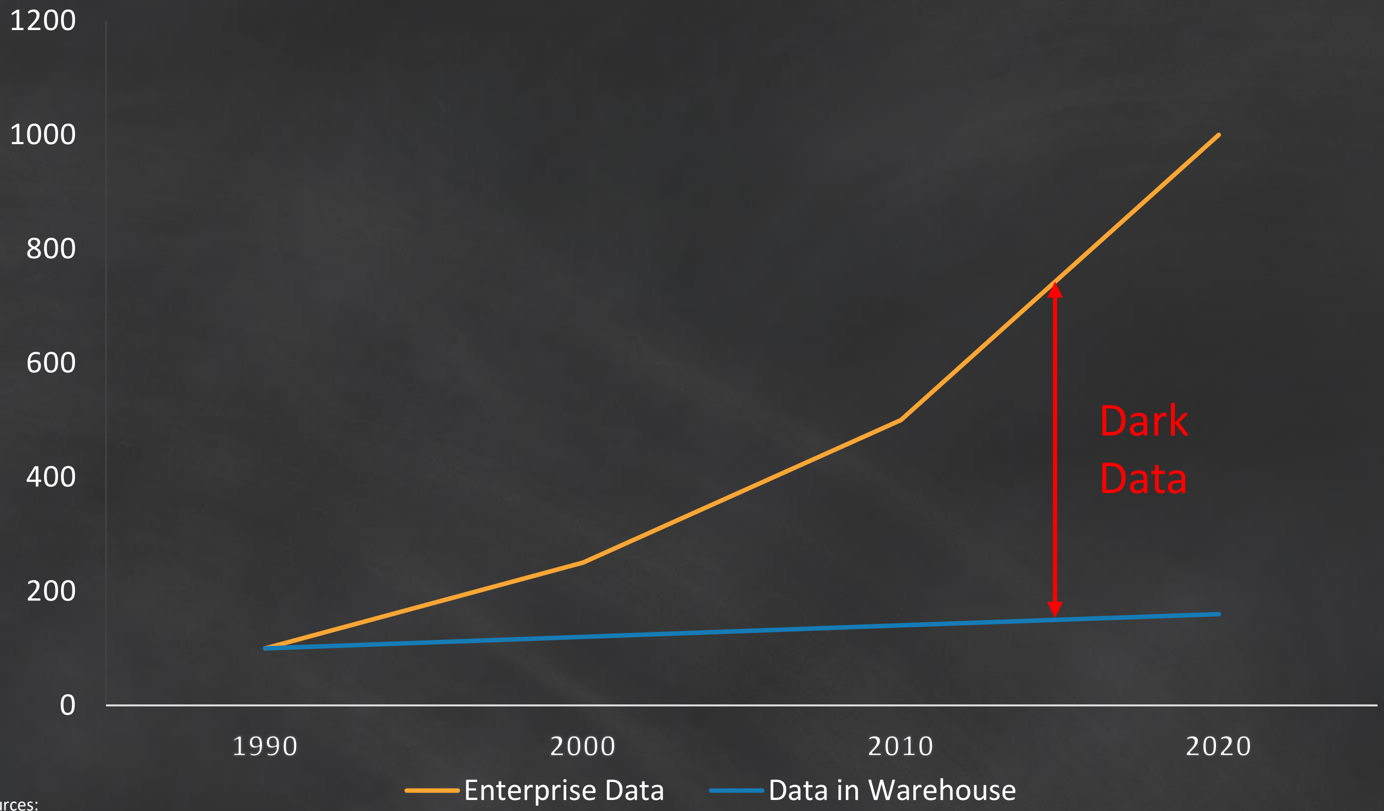
Avinash Nidumbur
Sr Business Development Manager
Amazon Redshift
avinid@amazon.com

aws | Pop-up Loft

Agenda

- Introduction
- Benefits
- Use cases
- Getting started
- Q&A

Legacy architectural models lead to Dark Data



Very Expensive

Inflexible licensing

Proprietary

Lock-In

Sources:
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

AWS Big Data Portfolio

Collect

- Amazon Kinesis Firehose
- AWS Direct Connect
- Amazon Kinesis Streams
- Amazon Snowball

Store

- Amazon S3
- Amazon Glacier
- Amazon Dynamo DB
- Amazon RDS, Amazon Aurora
- Amazon CloudSearch
- Amazon Elasticsearch

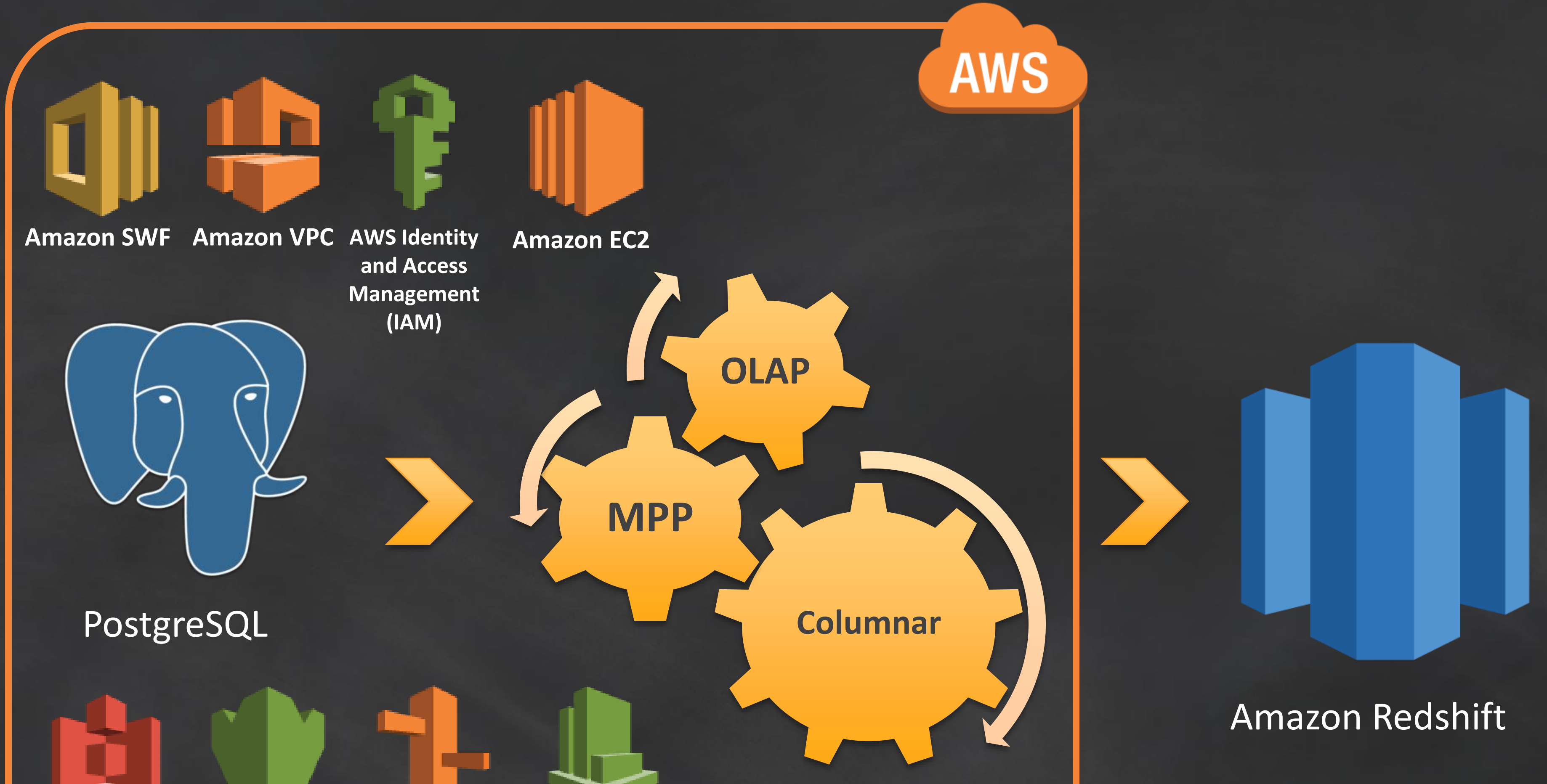
Analyze

- Amazon EMR
- Amazon EC2
- Amazon Redshift
- Amazon Machine Learning
- Amazon Kinesis Analytics
- Amazon QuickSight
- Amazon Athena

AWS Database Migration Service

AWS Data Pipeline

AWS Glue



AWS

Amazon SWF

Amazon VPC

AWS Identity and Access Management (IAM)

Amazon EC2

PostgreSQL

OLAP

MPP

Columnar

Amazon Redshift

Amazon S3

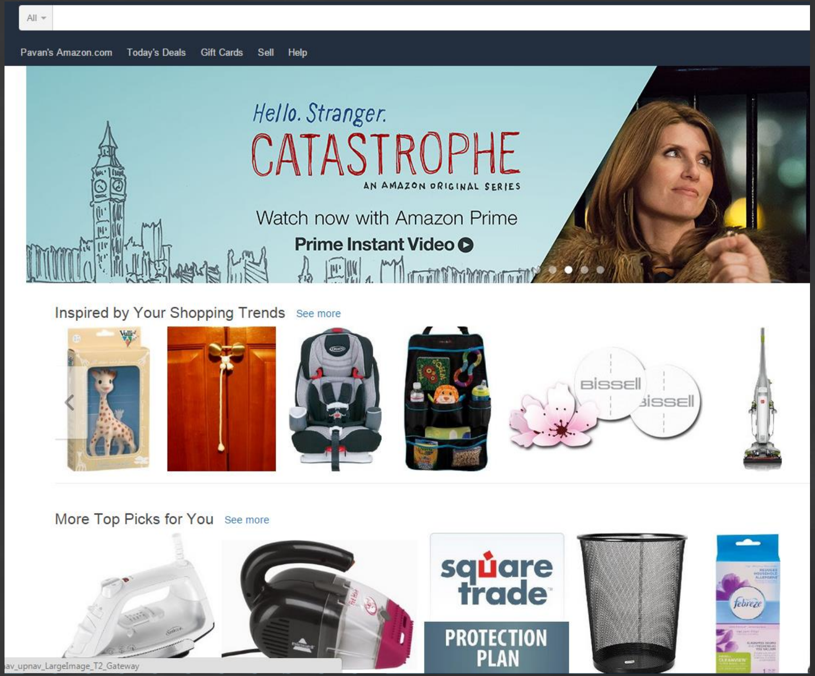
AWS KMS

Amazon Route 53

Amazon CloudWatch

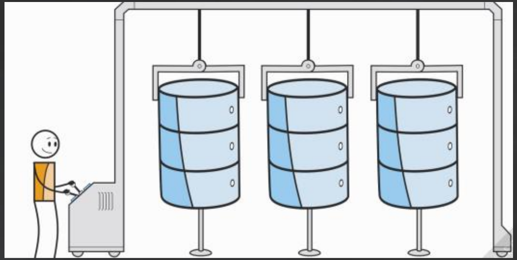
Amazon.com clickstream analytics

- Web log analysis for Amazon.com
  - PBs workload, 2TB/day@67% YoY
  - Largest table: 400 TB
- Understand customer behavior
- Previous solution
  - Legacy DW (Oracle)—query across 1 week/hr
  - Hadoop—query across 1 month/hr



Results with Amazon Redshift

- Query 15 months **in 14 min**
- Load 5B rows in **10 min**
- 21B w/ 10B rows: **3 days to 2 hrs** (Hive → Redshift)
- Load pipeline: **90 hrs to 8 hrs** (Oracle → Redshift)



- 100 node DS2.8XL clusters
- Easy resizing
- Managed backups and restore
- Failure tolerance and recovery

- 20% time of one DBA
- Increased productivity

Amazon Redshift

a lot faster
a lot simpler
a lot cheaper

Relational data warehouse

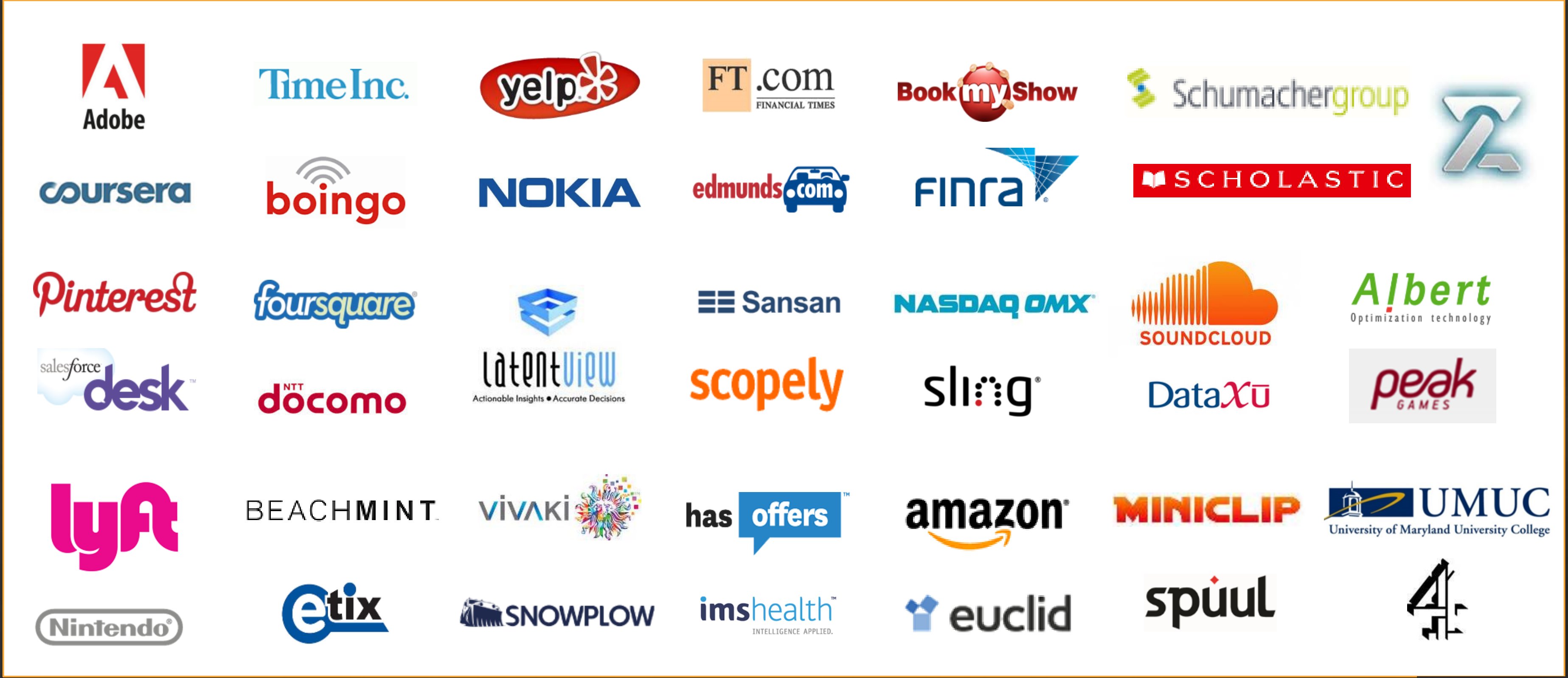
Massively parallel

Fully managed

HDD and SSD platforms

$1,000/TB/year; starts at $0.25/hour

Selected Amazon Redshift Customers



Use Case: Traditional Data Warehousing

Business Reporting

Advanced pipelines and queries

Secure and Compliant

Bulk Loads and Updates

Easy Migration – Point & Click using AWS Database Migration Service

Secure & Compliant – End-to-End Encryption. SOC 1/2/3, PCI-DSS, HIPAA and FedRAMP compliant

Large Ecosystem – Variety of cloud and on-premises BI and ETL tools

NTT docomo

Japanese Mobile Phone Provider

SCHOLASTIC

World's Largest Children's Book Publisher

Nasdaq

Powering 100 marketplaces in 50 countries

aws

Use Case: Log Analysis

Log & Machine IOT Data

Clickstream Events Data

Time-Series Data

Cheap – Analyze large volumes of data cost-effectively

Fast – Massively Parallel Processing (MPP) and columnar architecture for fast queries and parallel loads

Near real-time – Micro-batch loading and Amazon Kinesis Firehose for near-real time analytics

Pinterest

Interactive data analysis and recommendation engine

lyft

Ride analytics for pricing and product development

yelp

Ad prediction and on-demand analytics

aws

Use Case: Business Applications

Multi-Tenant BI Applications

Back-end services

Analytics as a Service

Fully Managed – Provisioning, backups, upgrades, security, compression all come built-in so you can focus on your business applications

Ease of Chargeback – Pay as you go, add clusters as needed. A few big common clusters & data marts

Service Oriented Architecture – Integrated with other AWS services. Easy to plug into your pipeline

Infosys

Infosys Information Platform (IIP)

accenture

Analytics-as-a-Service

Amplitude

Product and Consumer Analytics

Amazon Redshift architecture

- **Leader node**

  Simple SQL endpoint

  Stores metadata

  Optimizes query plan

  Coordinates query execution

- **Compute nodes**

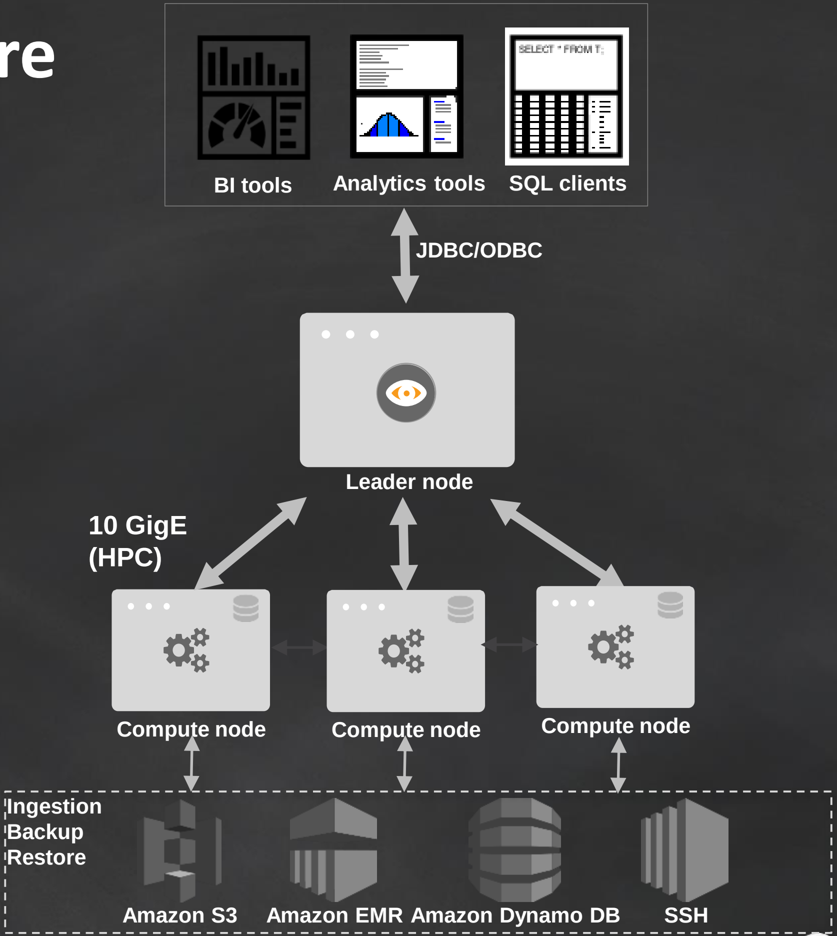
  Local columnar storage

  Parallel/distributed execution of all queries, loads, backups, restores, resizes

- **Start at just $0.25/hour, grow to 2 PB (compressed)**

  DC1/DC2: SSD; scale from 160 GB to 326 TB

  DS2: HDD; scale from 2 TB to 2 PB

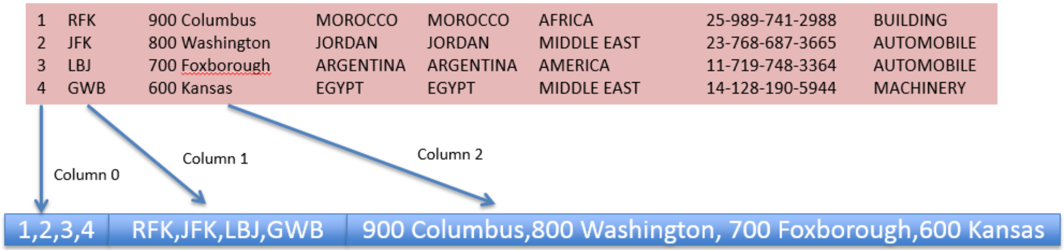


Benefit #1: Amazon Redshift is fast

- **Dramatically less I/O**
  - Column storage
  - Data compression
  - Zone maps
  - Direct-attached storage
  - Large data block sizes

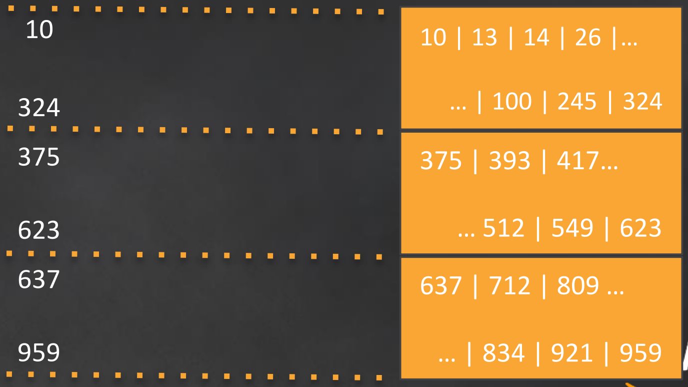
```
analyze compression listing;

  Table  |     Column     | Encoding
---------+----------------+----------
 listing | listid         | delta
 listing | sellerid       | delta32k
 listing | eventid        | delta32k
 listing | dateid         | bytedict
 listing | numtickets     | bytedict
 listing | priceperticket | delta32k
 listing | totalprice     | mostly32
 listing | listtime       | raw
```



1	RFK	900 Columbus	MOROCCO	MOROCCO	AFRICA	25-989-741-2988	BUILDING
2	JFK	800 Washington	JORDAN	JORDAN	MIDDLE EAST	23-768-687-3665	AUTOMOBILE
3	LBJ	700 Foxborough	ARGENTINA	ARGENTINA	AMERICA	11-719-748-3364	AUTOMOBILE
4	GWB	600 Kansas	EGYPT	EGYPT	MIDDLE EAST	14-128-190-5944	MACHINERY

1,2,3,4	RFK,JFK,LBJ,GWB	900 Columbus,800 Washington, 700 Foxborough,600 Kansas



Benefit #2: Amazon Redshift is inexpensive

DS2 (HDD)	Price per hour for DS2.XL single node	Effective annual price per TB compressed
On-demand	$ 0.850	$ 3,725
1 year reservation	$ 0.500	$ 2,190
3 year reservation	$ 0.228	$ 999

DC1 (SSD)	Price per hour for DC1.L single node	Effective annual price per TB compressed
On-demand	$ 0.250	$ 13,690
1 year reservation	$ 0.161	$ 8,795
3 year reservation	$ 0.100	$ 5,500

Pricing is simple

Number of nodes x price/hour

No charge for leader node

No upfront costs

Pay as you go

Benefit #3 : Amazon Redshift is easy to use

Provisioning in minutes

Automatic patching

SQL - Data loading

Backups are built-in

Security is built-in

Compression is built-in

Benefit #4: Amazon Redshift is fully managed

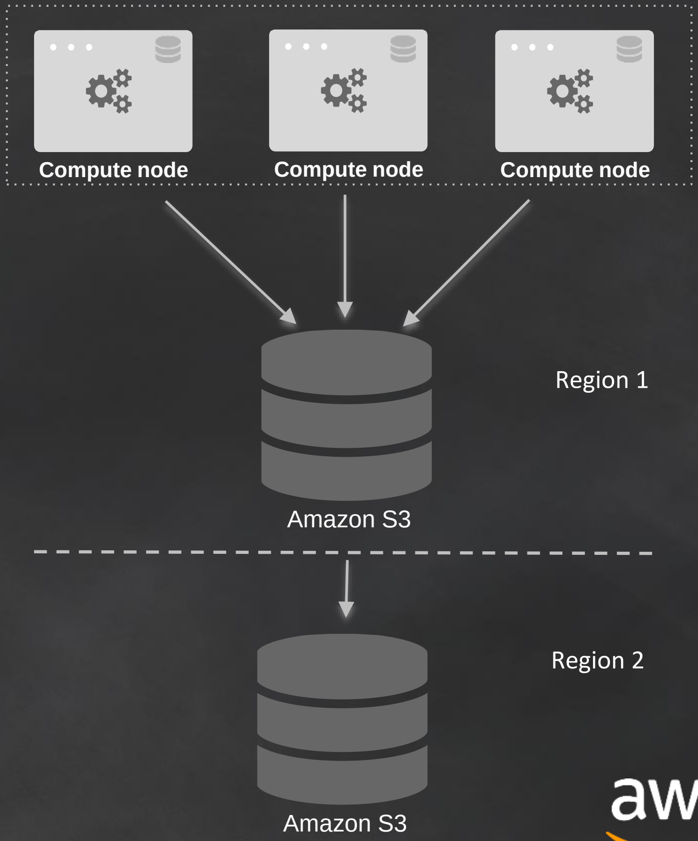
Continuous/incremental backups

Multiple copies within cluster

Continuous and incremental backups to Amazon S3

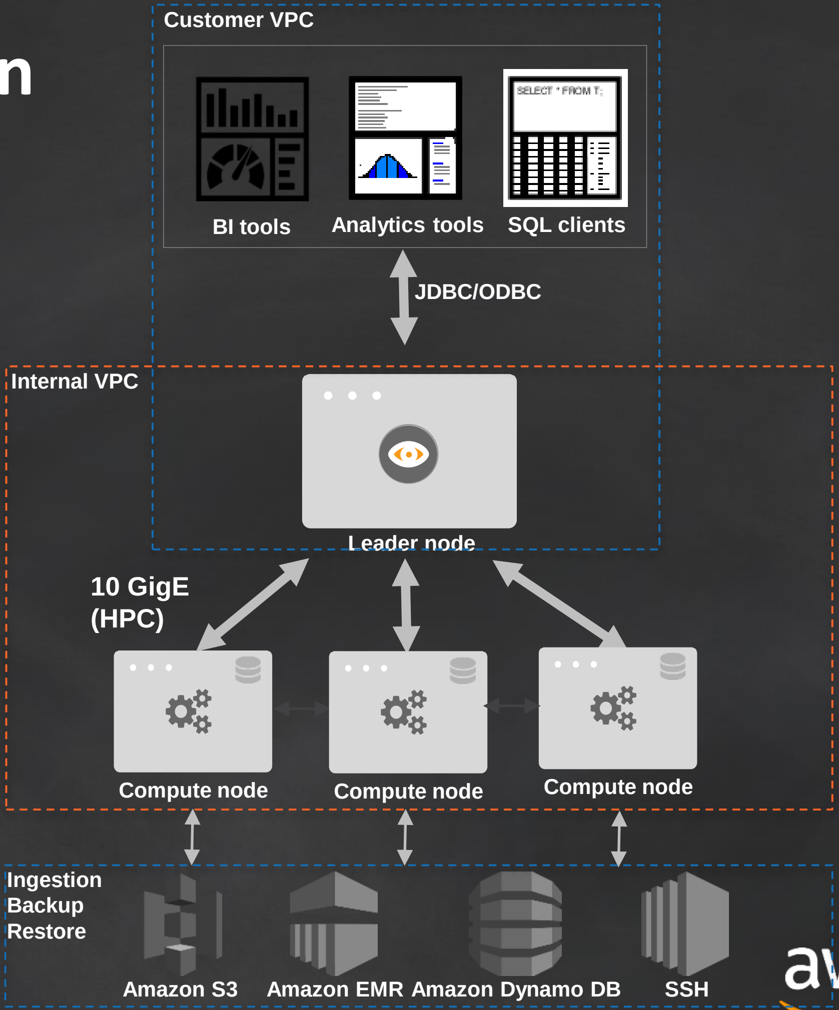
Continuous and incremental backups across regions

Streaming restore



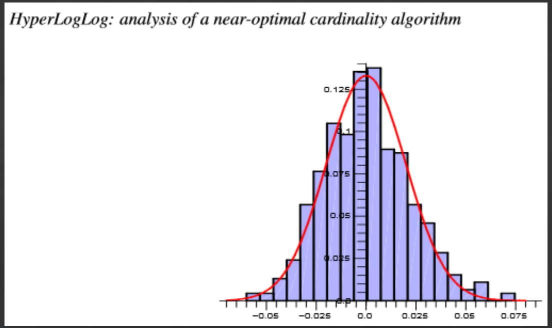
Benefit #5: Security is built-in

- Load encrypted from S3
- SSL to secure data in transit
  - ECDHE perfect forward secrecy
- Amazon VPC for network isolation
- Encryption to secure data at rest
  - All blocks on disks and in S3 encrypted
  - Block key, cluster key, master key (AES-256)
  - On-premises HSM & AWS CloudHSM support
- Audit logging and AWS CloudTrail integration
- SOC 1/2/3, PCI-DSS, FedRAMP, BAA



Benefit #6: Amazon Redshift is powerful

- Approximate functions
- User defined functions
- Machine learning
- Data science


HyperLogLog: analysis of a near-optimal cardinality algorithm

Benefit #7: Amazon Redshift has a large ecosystem

Data integration

ATTUNITY CloudBeam

bryte Systems

FlyData

INFORMATICA

Matillion

pentaho

Segment

snapLogic

talend integration at any scale

Qubole

xplenty

Business intelligence

Actian

infor

periscope

alteryx

JReport JINFONET SOFTWARE

Qlik

ACTUATE The BIRT Company™

Logi ANALYTICS

SiSense

birst

REDROCK BI

looker

ClearStory DATA Now You See It™

MicroStrategy

tableau SOFTWARE

Chart.io Turn data to insight

MODE

TIBCO®

Dundas

PANORAMA

Business Partner IBM

pentaho A Hitachi Data Systems Company

Yellowfin

Systems integrators

2ND WATCH

accenture

BEEVA

BRIGHTLIGHT Ask more of your data

Capgemini CONSULTING.TECHNOLOGY.OUTSOURCING

cloudpack

Cognizant

FULL 360

iOLAP

isobar

MAGNUS DATA

NorthBay solutions

SLALOM CONSULTING

aws

Amazon Redshift Spectrum

Amazon Redshift Spectrum

Run SQL queries directly against data in S3 using thousands of nodes

Fast @ exabyte scale

Elastic & highly available

On-demand, pay-per-query

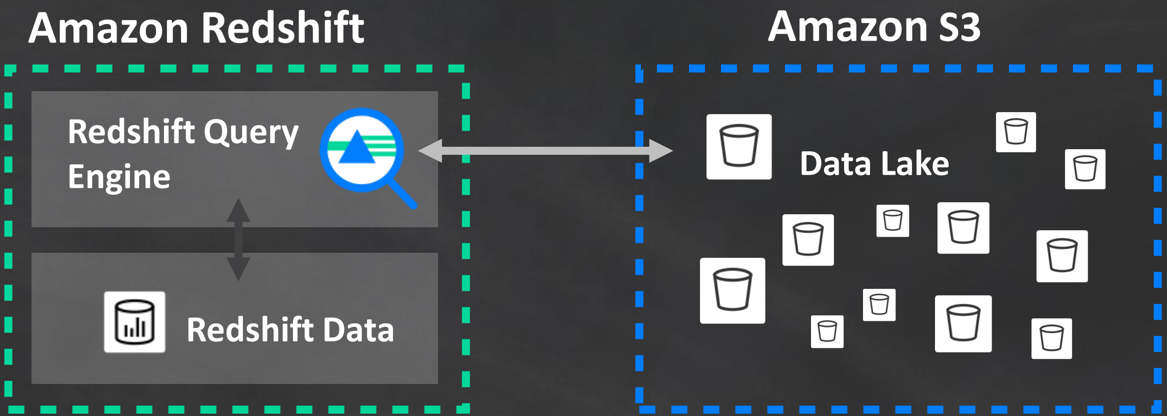
High concurrency: Multiple clusters access same data

No ETL: Query data in-place using open file formats

I ♥ SQL

Full Amazon Redshift SQL support

Extend Redshift Queries To Amazon S3

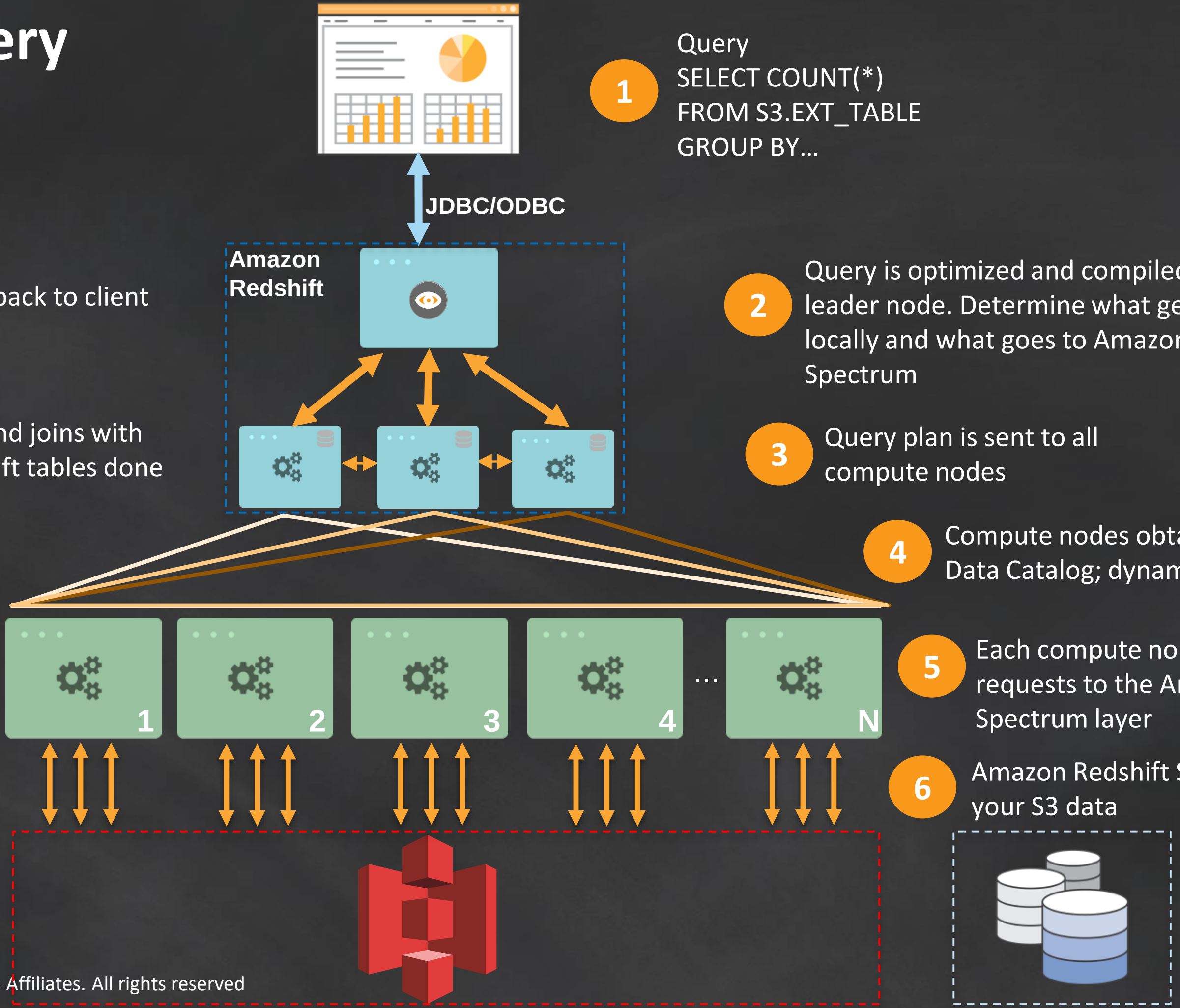


Query S3 directly or join data across Redshift and S3

Scale Redshift compute and storage separately

Support for CSV, Parquet, ORC, Grok, Avro and more formats

Life of a query



1 Query
SELECT COUNT(*)
FROM S3.EXT_TABLE
GROUP BY…

2 Query is optimized and compiled at the leader node. Determine what gets run locally and what goes to Amazon Redshift Spectrum

3 Query plan is sent to all compute nodes

4 Compute nodes obtain partition info from Data Catalog; dynamically prune partitions

5 Each compute node issues multiple requests to the Amazon Redshift Spectrum layer

6 Amazon Redshift Spectrum nodes scan your S3 data

7 Amazon Redshift Spectrum projects, filters, joins and aggregates

8 Final aggregations and joins with local Amazon Redshift tables done in-cluster

9 Result is sent back to client

Amazon S3

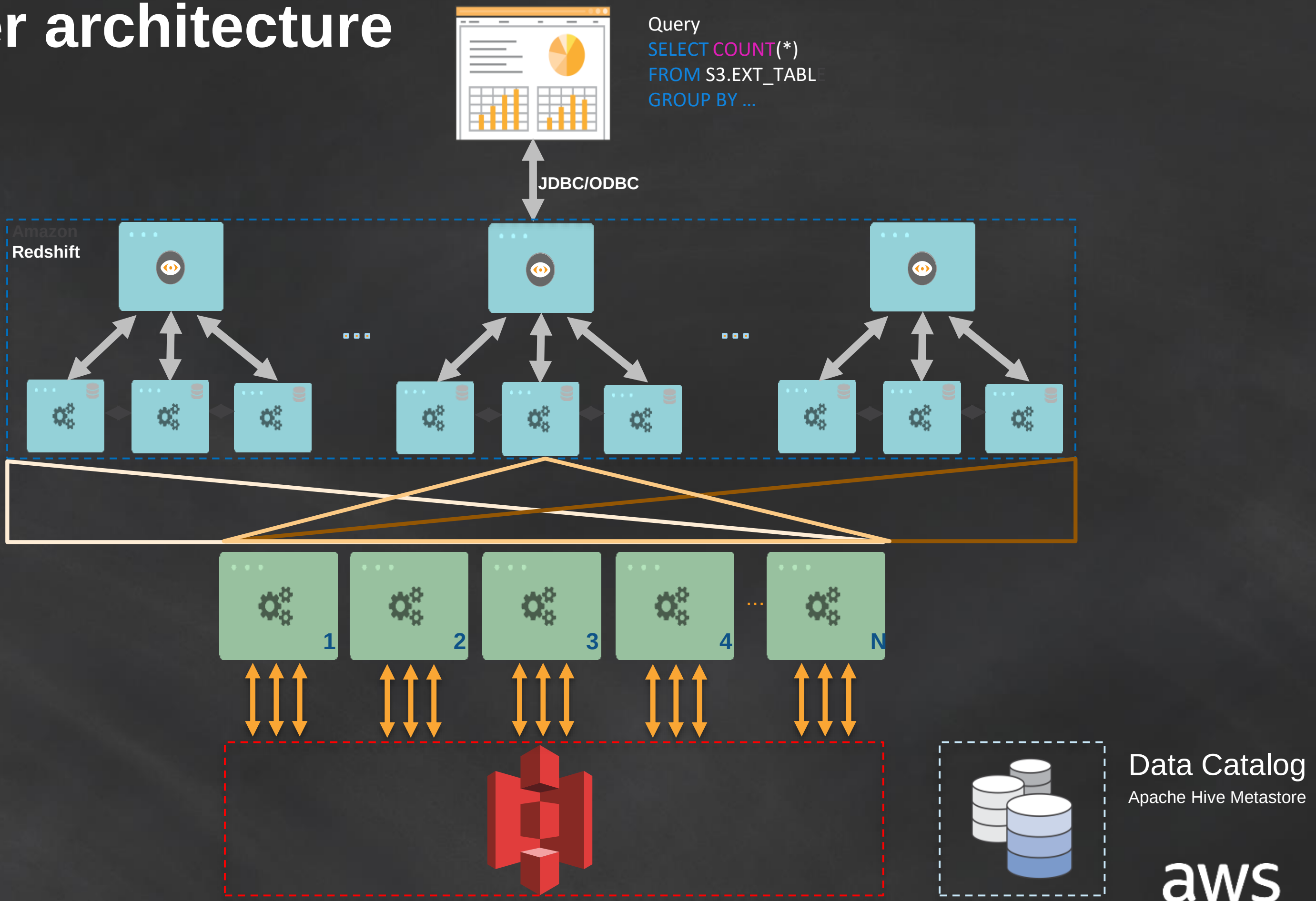
Exabyte-scale object storage

Use cases

- Extend data warehousing to data lake in Amazon S3
- Cost reduction for less frequently accessed data
- Better query performance and minimal time to insight at large scale
- Query many open formats and large data sets that can’t be loaded into cluster
- Improved concurrency with multiple Redshift clusters querying common data
- Elastically scale compute resources separately from the storage layer in S3
- Use familiar SQL to cleanse, transform and load data from S3 to Redshift

Multiple cluster architecture

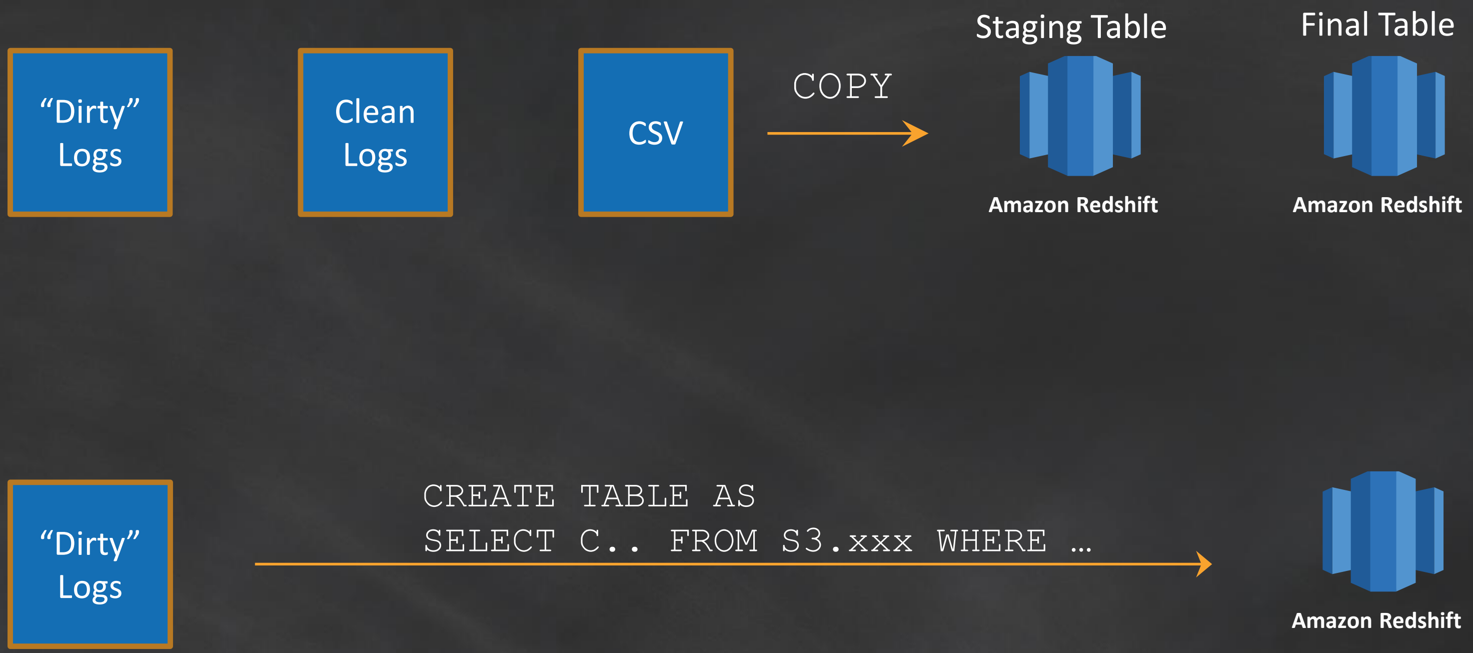
- Use multiple Redshift clusters to query same copy of data in s3



Query

```
SELECT COUNT(*)
FROM S3.EXT_TABLE
GROUP BY …
```

Simplify and accelerate your ETL pipelines



“Dirty” Logs

Clean Logs

CSV

COPY

Staging Table

Amazon Redshift

Final Table

Amazon Redshift

“Dirty” Logs

```
CREATE TABLE AS
SELECT C.. FROM S3.xxx WHERE …
```

Amazon Redshift

Customer Use Cases

Nasdaq: powering 100 marketplaces in 50 countries

Orders, quotes, trade executions, market “tick” data from 7 exchanges

7 billion rows/day

Analyze market share, client activity, surveillance, billing, and so on

Microsoft SQL Server on-premises

Expensive legacy DW ($1.16 M/yr.)

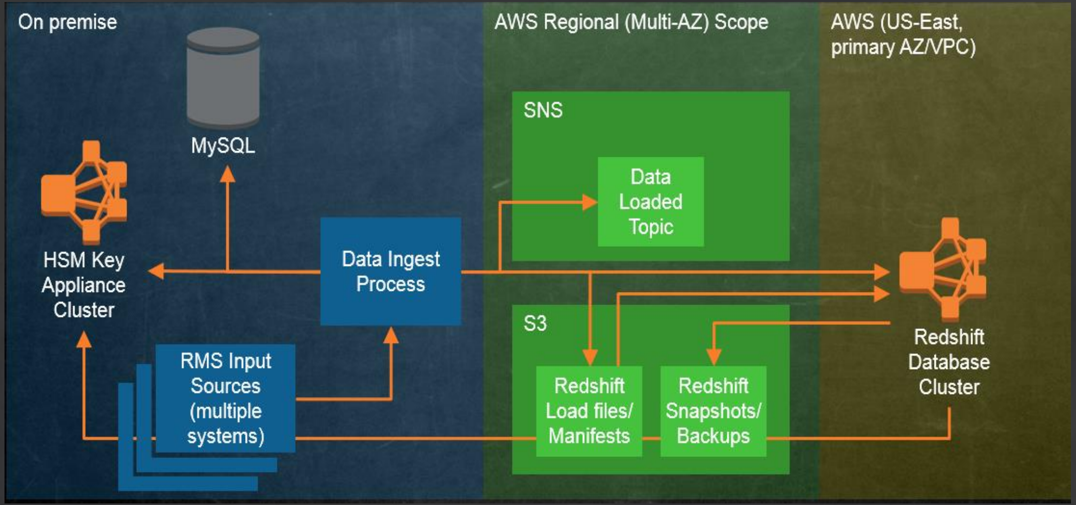
Limited capacity (1 yr. of data online)

Needed lower TCO

Must satisfy multiple security and regulatory requirements

Similar performance

Nasdaq: powering 100 marketplaces in 50 countries


On premise
MySQL
HSM Key Appliance Cluster
Data Ingest Process
RMS Input Sources (multiple systems)
AWS Regional (Multi-AZ) Scope
SNS
Data Loaded Topic
S3
Redshift Load files/ Manifests
Redshift Snapshots/ Backups
AWS (US-East, primary AZ/VPC)
Redshift Database Cluster

23 node DS2.8XL cluster

828 vCPUs, 5 TB RAM

368 TB compressed

2.7 T rows, 900 B derived

8 tables with 100 B rows

7 man-month migration

¼ the cost, 2x storage, room to grow

Faster performance, very secure

Customers love Amazon Redshift Spectrum

Time Inc.

“Redshift Spectrum enables us to directly operate on our data in its native format in Amazon S3 with no preprocessing or transformation.”

NTT docomo

“Redshift Spectrum will let us expand the universe of the data we analyze to 100s of petabytes over time. This is truly a game changer, and we can think of no other system in the world that can get us there.”

edmunds

“Redshift Spectrum’s fast performance across massive data sets is unprecedented.”

REDFIN

“Our data science team using Amazon EMR can now collaborate with our marketing and product teams using Redshift Spectrum to analyze the same Amazon S3 data sets.”

yelp

“Multiple teams can now query the same Amazon S3 data sets using both Redshift and EMR.”

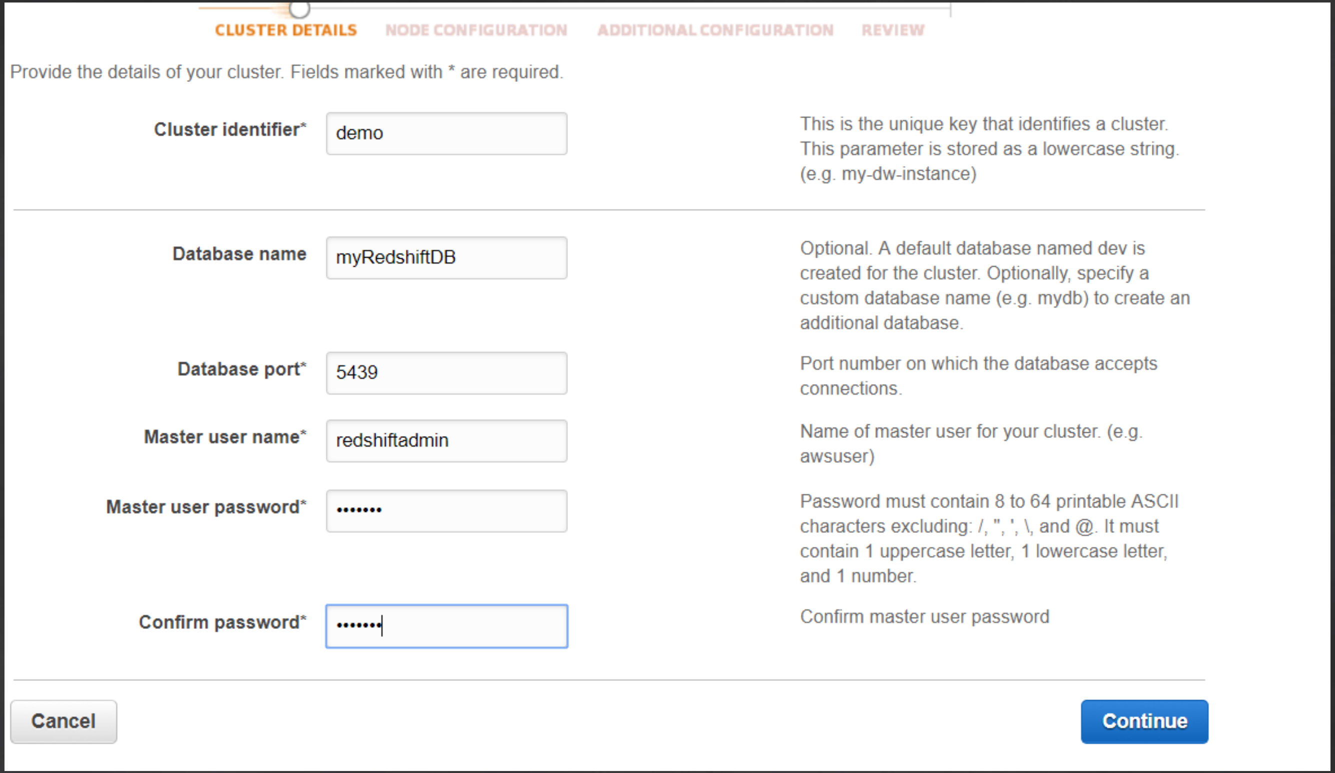
RECRUIT
Recruit Technologies Co.,Ltd.

“Redshift Spectrum will help us scale yet further while also lowering our costs.”

Getting Started

Provisioning

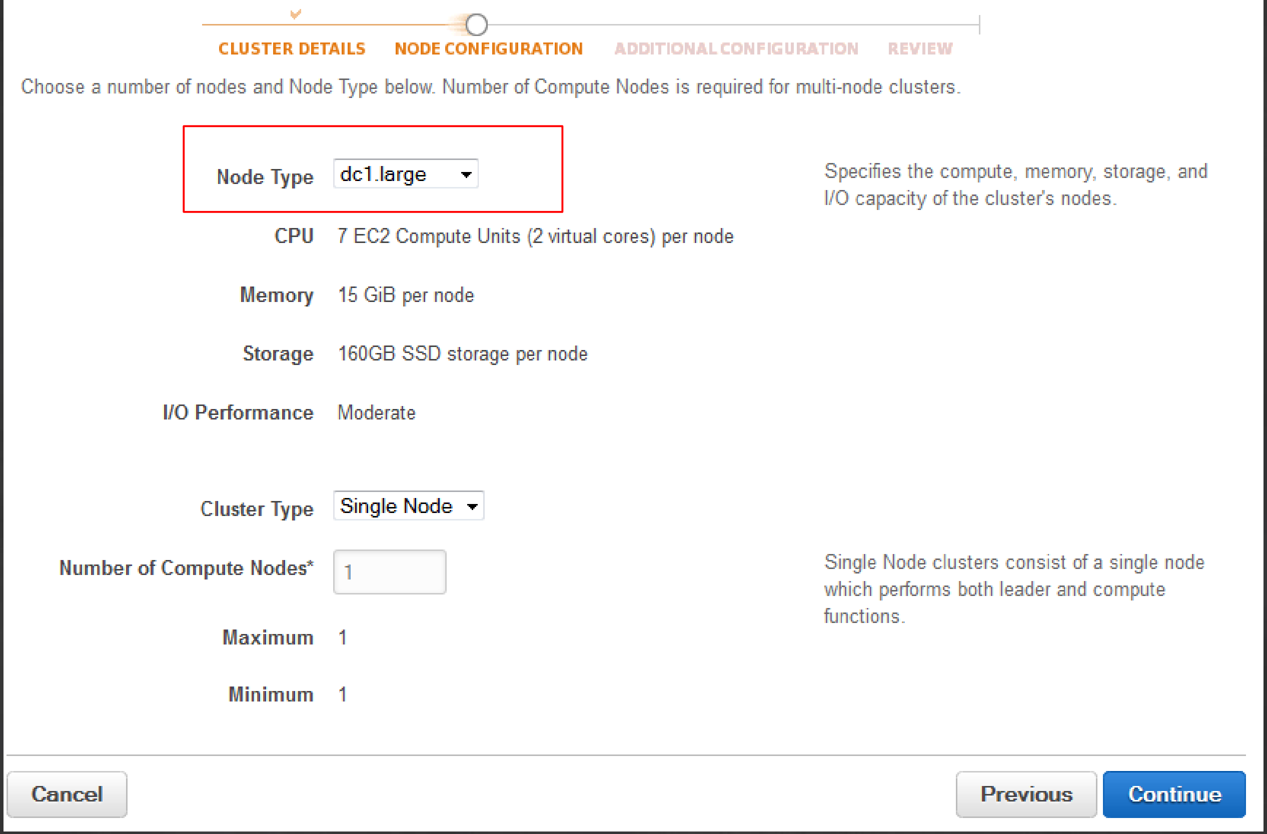
Enter cluster details



Provide the details of your cluster. Fields marked with * are required.

Cluster identifier*	demo	This is the unique key that identifies a cluster. This parameter is stored as a lowercase string. (e.g. my-dw-instance)
Database name	myRedshiftDB	Optional. A default database named dev is created for the cluster. Optionally, specify a custom database name (e.g. mydb) to create an additional database.
Database port*	5439	Port number on which the database accepts connections.
Master user name*	redshiftadmin	Name of master user for your cluster. (e.g. awsuser)
Master user password*	•••••••	Password must contain 8 to 64 printable ASCII characters excluding: /, ", ', \, and @. It must contain 1 uppercase letter, 1 lowercase letter, and 1 number.
Confirm password*	•••••••	Confirm master user password

Select node configuration



Choose a number of nodes and Node Type below. Number of Compute Nodes is required for multi-node clusters.

Node Type: dc1.large

Specifies the compute, memory, storage, and I/O capacity of the cluster's nodes.

CPU: 7 EC2 Compute Units (2 virtual cores) per node

Memory: 15 GiB per node

Storage: 160GB SSD storage per node

I/O Performance: Moderate

Cluster Type: Single Node

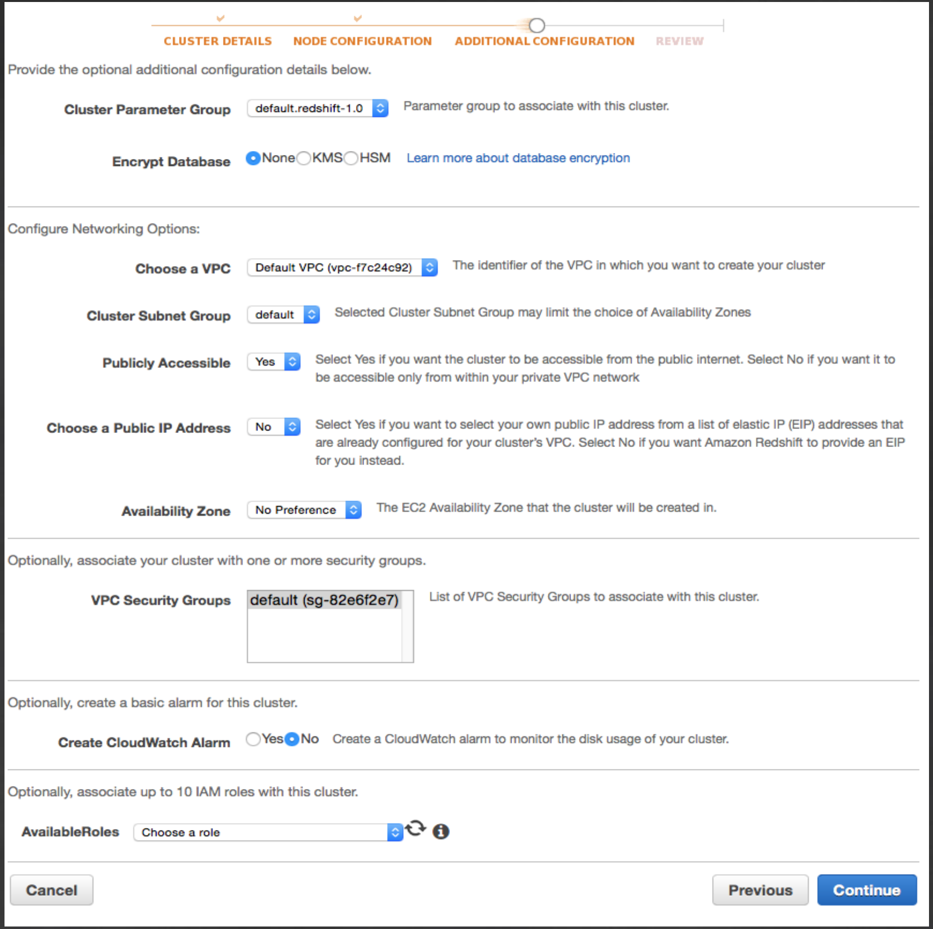
Number of Compute Nodes*: 1

Single Node clusters consist of a single node which performs both leader and compute functions.

Maximum: 1

Minimum: 1

Select security settings and provision



Provide the optional additional configuration details below.

Cluster Parameter Group default.redshift-1.0 — Parameter group to associate with this cluster.

Encrypt Database ● None ○ KMS ○ HSM — Learn more about database encryption

Configure Networking Options:

Choose a VPC Default VPC (vpc-f7c24c92) — The identifier of the VPC in which you want to create your cluster

Cluster Subnet Group default — Selected Cluster Subnet Group may limit the choice of Availability Zones

Publicly Accessible Yes — Select Yes if you want the cluster to be accessible from the public internet. Select No if you want it to be accessible only from within your private VPC network

Choose a Public IP Address No — Select Yes if you want to select your own public IP address from a list of elastic IP (EIP) addresses that are already configured for your cluster's VPC. Select No if you want Amazon Redshift to provide an EIP for you instead.

Availability Zone No Preference — The EC2 Availability Zone that the cluster will be created in.

Optionally, associate your cluster with one or more security groups.

VPC Security Groups default (sg-82e6f2e7) — List of VPC Security Groups to associate with this cluster.

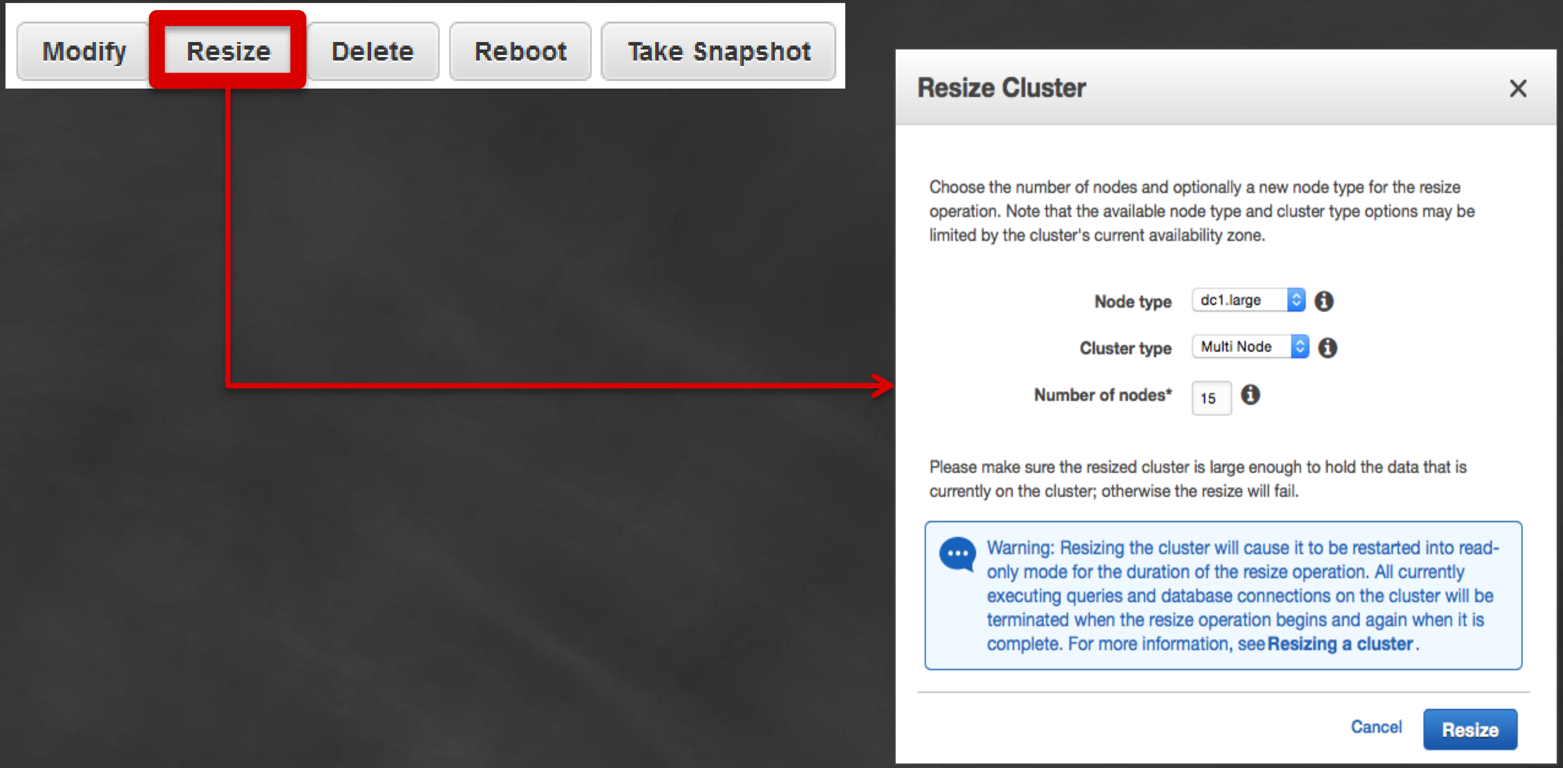
Optionally, create a basic alarm for this cluster.

Create CloudWatch Alarm ○ Yes ● No — Create a CloudWatch alarm to monitor the disk usage of your cluster.

Optionally, associate up to 10 IAM roles with this cluster.

AvailableRoles Choose a role

Point-and-click resize



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	dc1.large
Cluster type	Multi Node
Number of nodes*	15

Please make sure the resized cluster is large enough to hold the data that is currently on the cluster; otherwise the resize will fail.

Warning: 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. For more information, see **Resizing a cluster**.

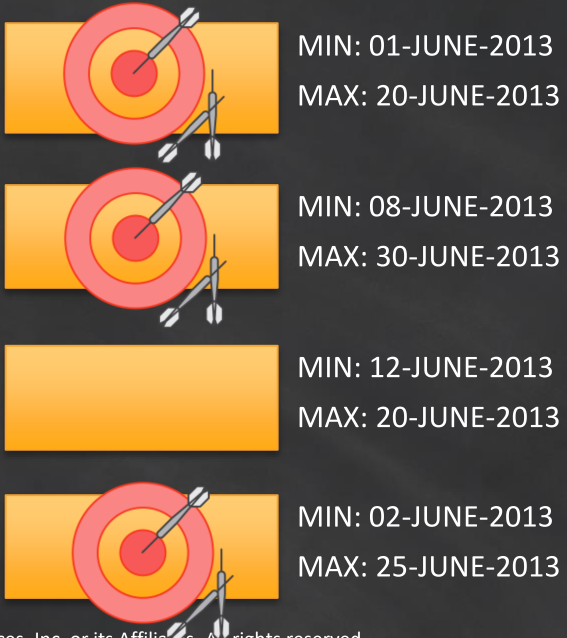
Data Modeling

Sort Keys

Zone maps

SELECT COUNT(*) FROM LOGS WHERE DATE = ‘09-JUNE-2013’

Unsorted table



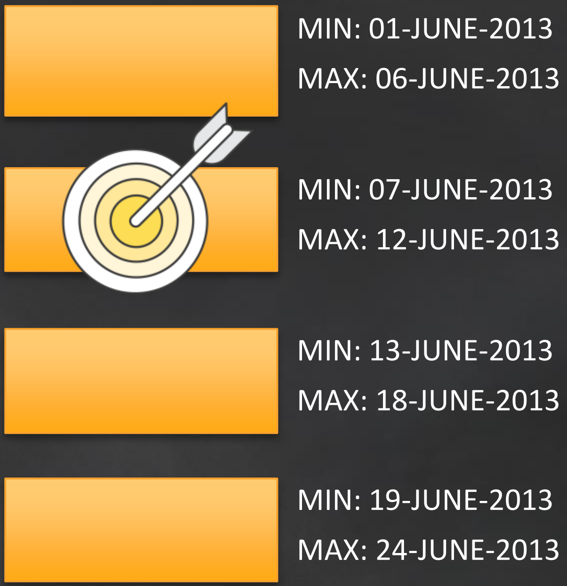
MIN: 01-JUNE-2013
MAX: 20-JUNE-2013

MIN: 08-JUNE-2013
MAX: 30-JUNE-2013

MIN: 12-JUNE-2013
MAX: 20-JUNE-2013

MIN: 02-JUNE-2013
MAX: 25-JUNE-2013

Sorted by date



MIN: 01-JUNE-2013
MAX: 06-JUNE-2013

MIN: 07-JUNE-2013
MAX: 12-JUNE-2013

MIN: 13-JUNE-2013
MAX: 18-JUNE-2013

MIN: 19-JUNE-2013
MAX: 24-JUNE-2013

Sort Keys

- Single column
- Compound
- Interleaved

Sort Keys

Single Column

- Table is sorted by 1 column

```
[ SORTKEY ( date ) ]
```

Date	Region	Country
2-JUN-2015	Oceania	New Zealand
2-JUN-2015	Asia	Singapore
2-JUN-2015	Africa	Zaire
2-JUN-2015	Asia	Hong Kong
3-JUN-2015	Europe	Germany
3-JUN-2015	Asia	Korea

- Best for:
  - Queries that use 1st column (i.e. *date)* as primary filter
  - Can speed up joins and group bys
  - Quickest to VACUUM

Sort Keys

Compound

- Table is sorted by 1st column , then 2nd column etc.

```
[ SORTKEY COMPOUND ( date, region, country) ]
```

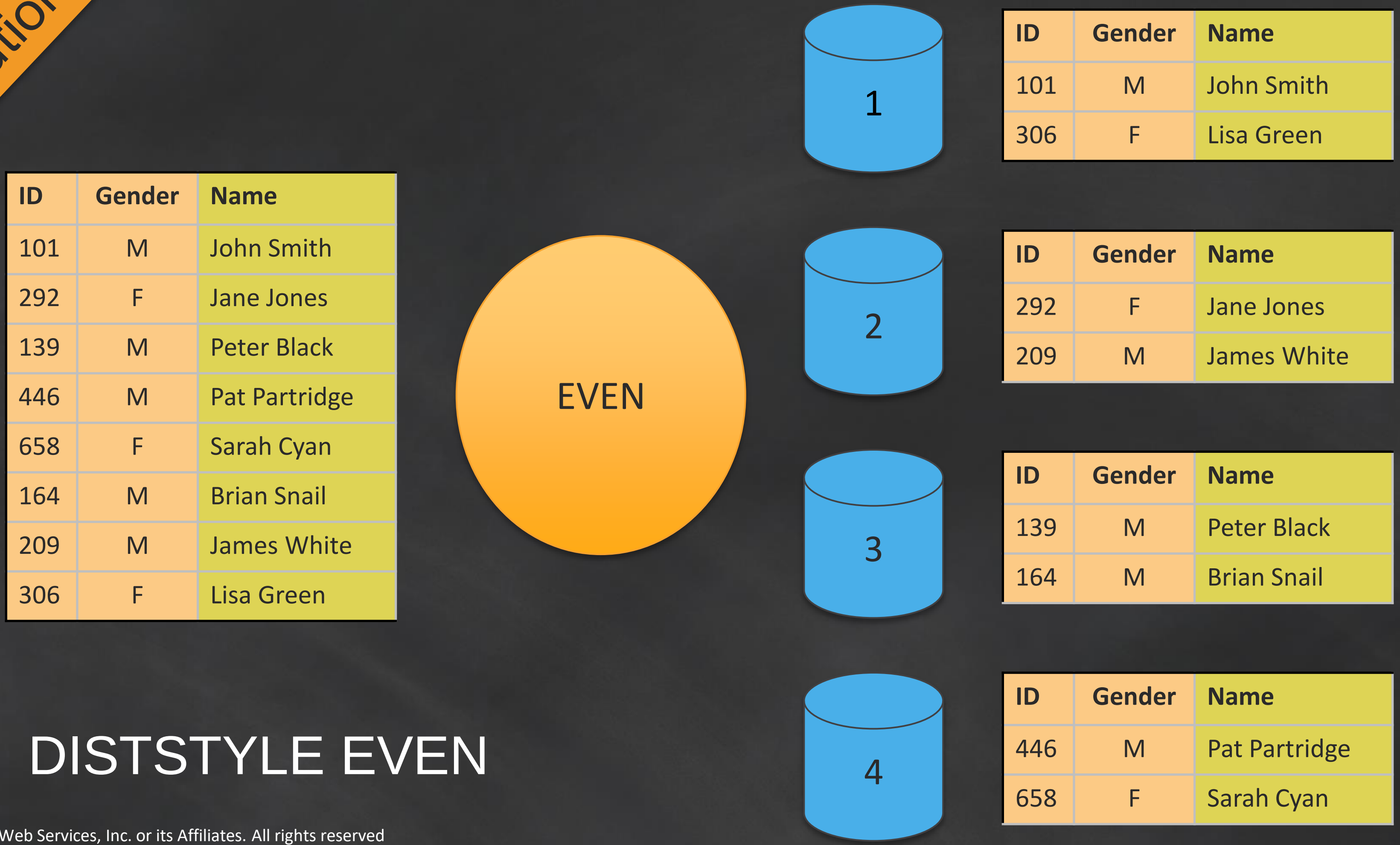
Date	Region	Country
2-JUN-2015	Africa	Zaire
2-JUN-2015	Asia	Korea
2-JUN-2015	Asia	Singapore
2-JUN-2015	Europe	Germany
3-JUN-2015	Asia	Hong Kong
3-JUN-2015	Asia	Korea

- Best for:
  - Queries that use 1st column as primary filter, then other cols
  - Can speed up joins and group bys
  - Slower to VACUUM

Distribution

- EVEN
- KEY
- ALL

Distribution



ID	Gender	Name
101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	James White
306	F	Lisa Green

ID	Gender	Name
101	M	John Smith
306	F	Lisa Green

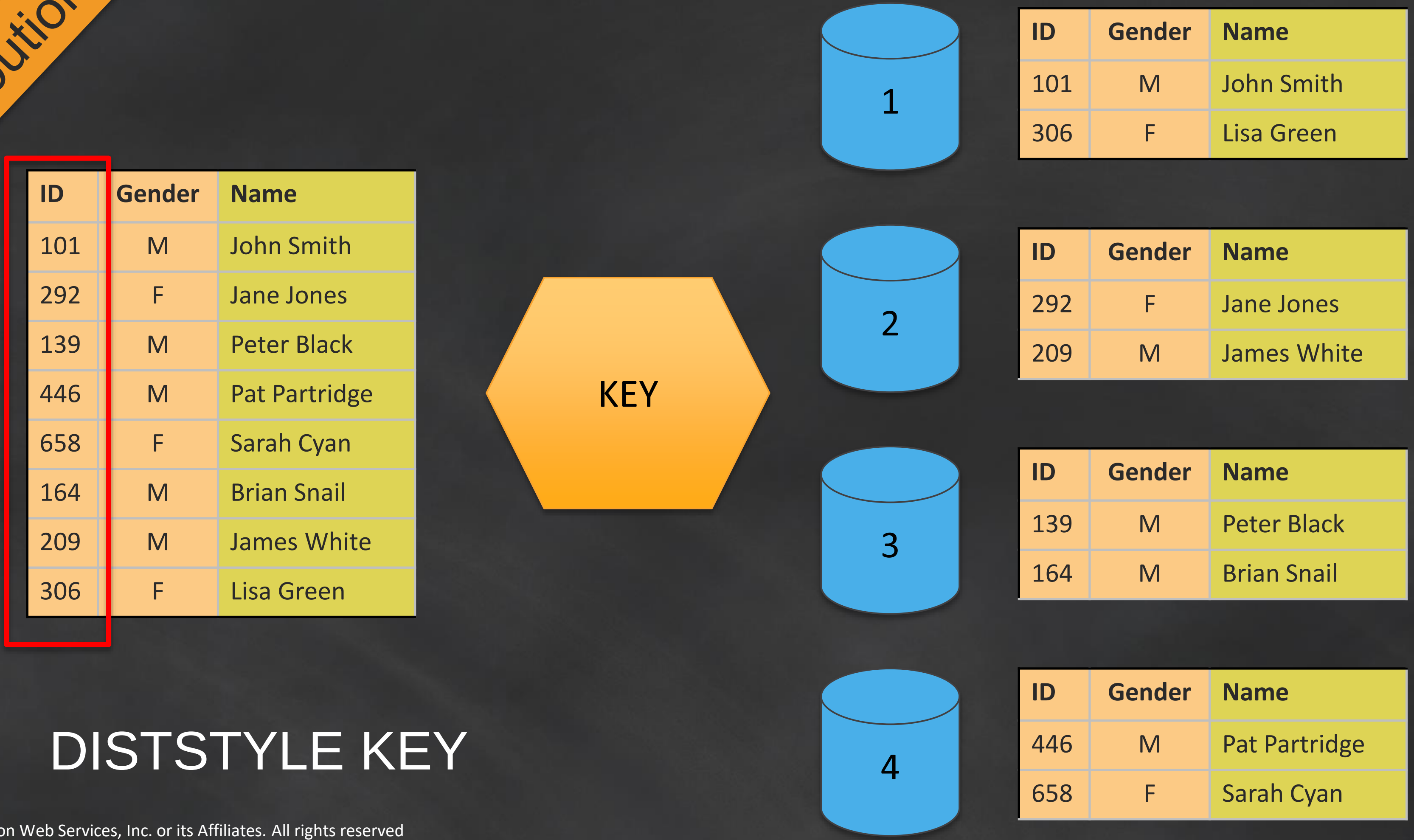
ID	Gender	Name
292	F	Jane Jones
209	M	James White

ID	Gender	Name
139	M	Peter Black
164	M	Brian Snail

ID	Gender	Name
446	M	Pat Partridge
658	F	Sarah Cyan

DISTSTYLE EVEN

Distribution



ID	Gender	Name
101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	James White
306	F	Lisa Green

ID	Gender	Name
101	M	John Smith
306	F	Lisa Green

ID	Gender	Name
292	F	Jane Jones
209	M	James White

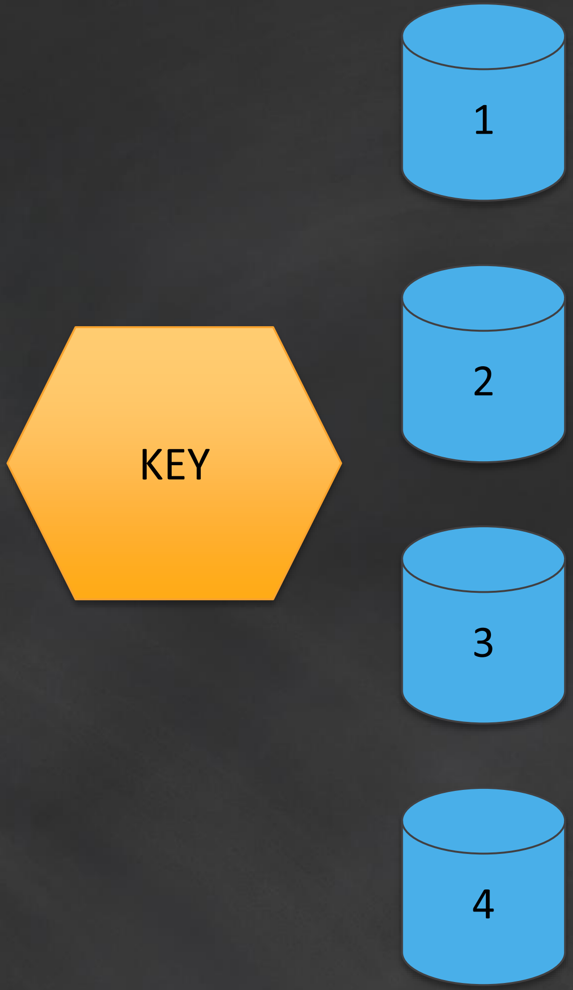
ID	Gender	Name
139	M	Peter Black
164	M	Brian Snail

ID	Gender	Name
446	M	Pat Partridge
658	F	Sarah Cyan

DISTSTYLE KEY

Distribution

ID	Gender	Name
101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	James White
306	F	Lisa Green



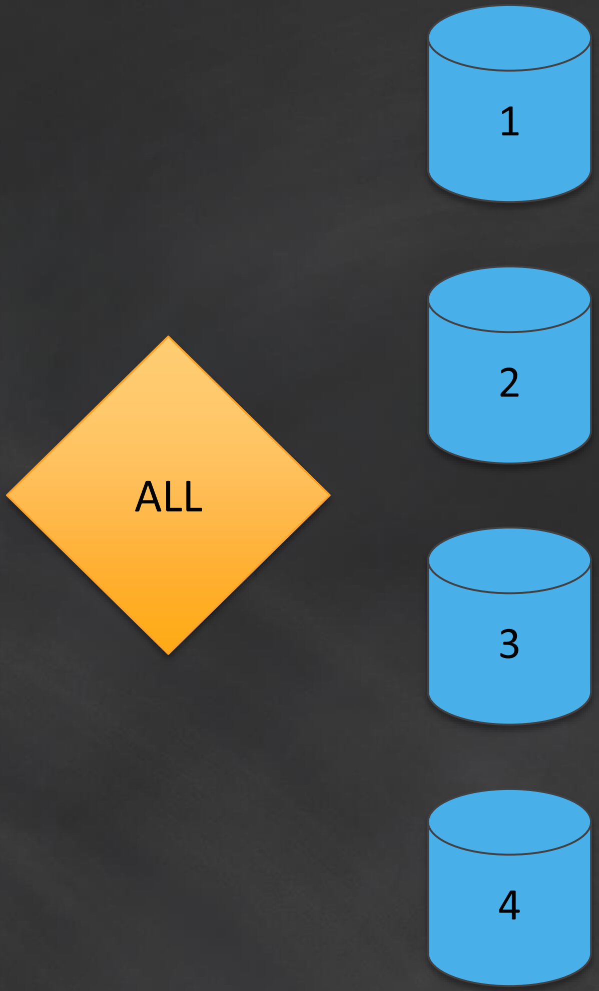
ID	Gender	Name
101	M	John Smith
139	M	Peter Black
446	M	Pat Partridge
164	M	Brian Snail
209	M	James White

ID	Gender	Name
292	F	Jane Jones
658	F	Sarah Cyan
306	F	Lisa Green

DISTSTYLE KEY

Distribution

ID	Gender	Name
101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	James White
306	F	Lisa Green



101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	Lisa Green
306	F	James White

101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	Lisa Green
306	F	James White

101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	Lisa Green
306	F	James White

101	M	John Smith
292	F	Jane Jones
139	M	Peter Black
446	M	Pat Partridge
658	F	Sarah Cyan
164	M	Brian Snail
209	M	Lisa Green
306	F	James White

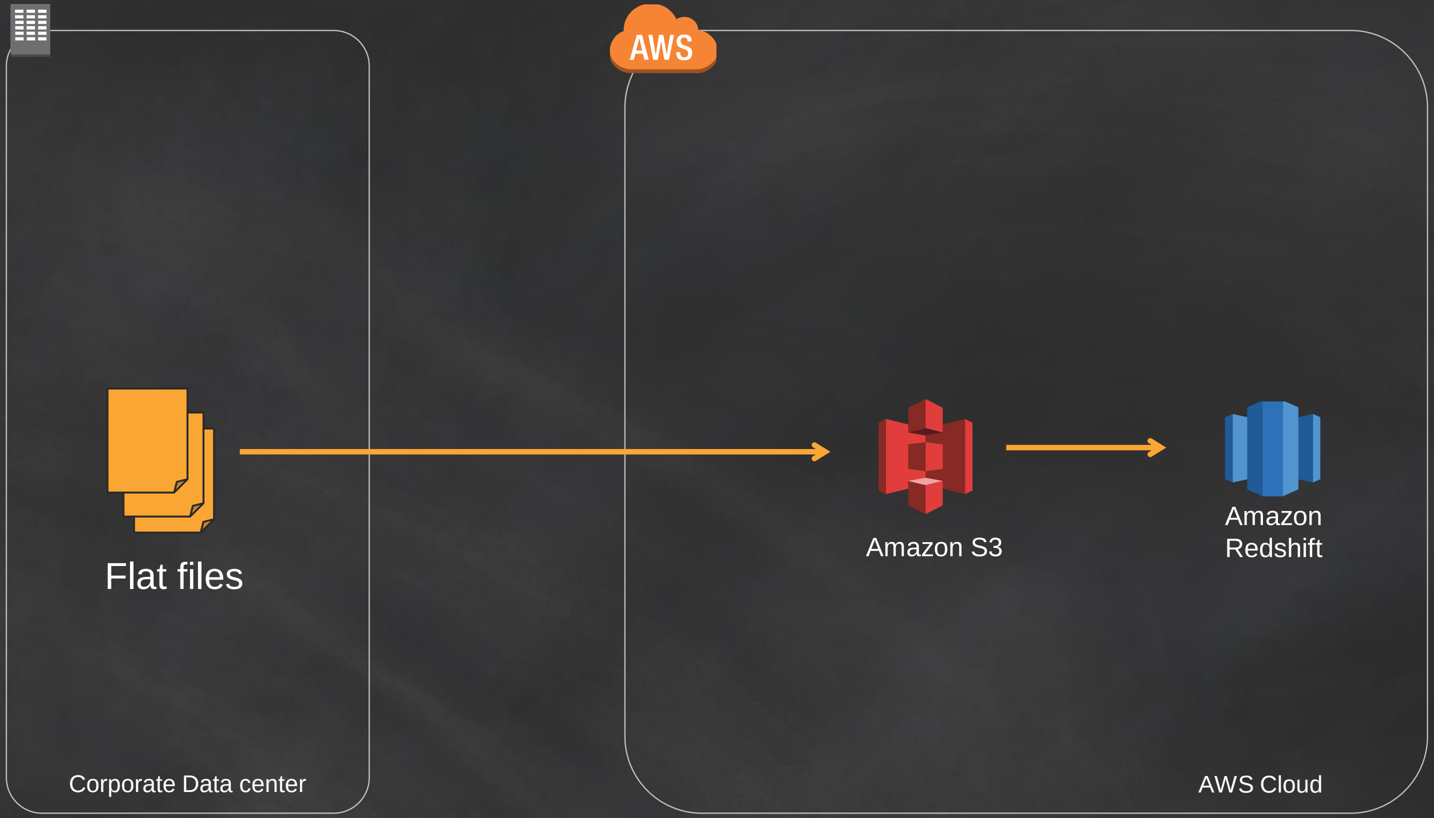
DISTSTYLE ALL

Distribution

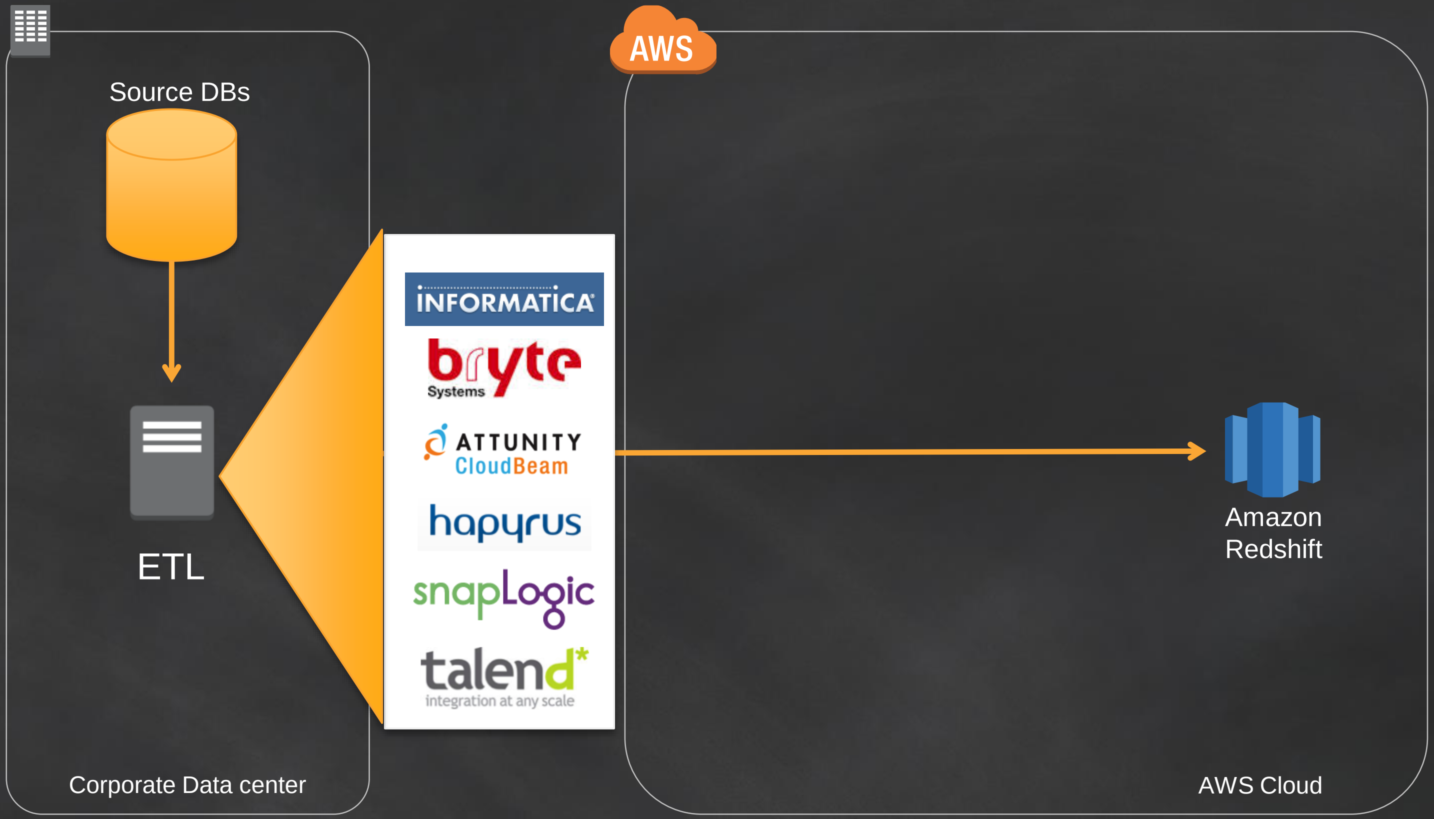
- EVEN
  - Tables with no joins or group by
- KEY
  - Large Fact tables
  - Large dimension tables
- ALL
  - Medium dimension tables (1K – 2M)
  - Small dimension tables

Loading Data

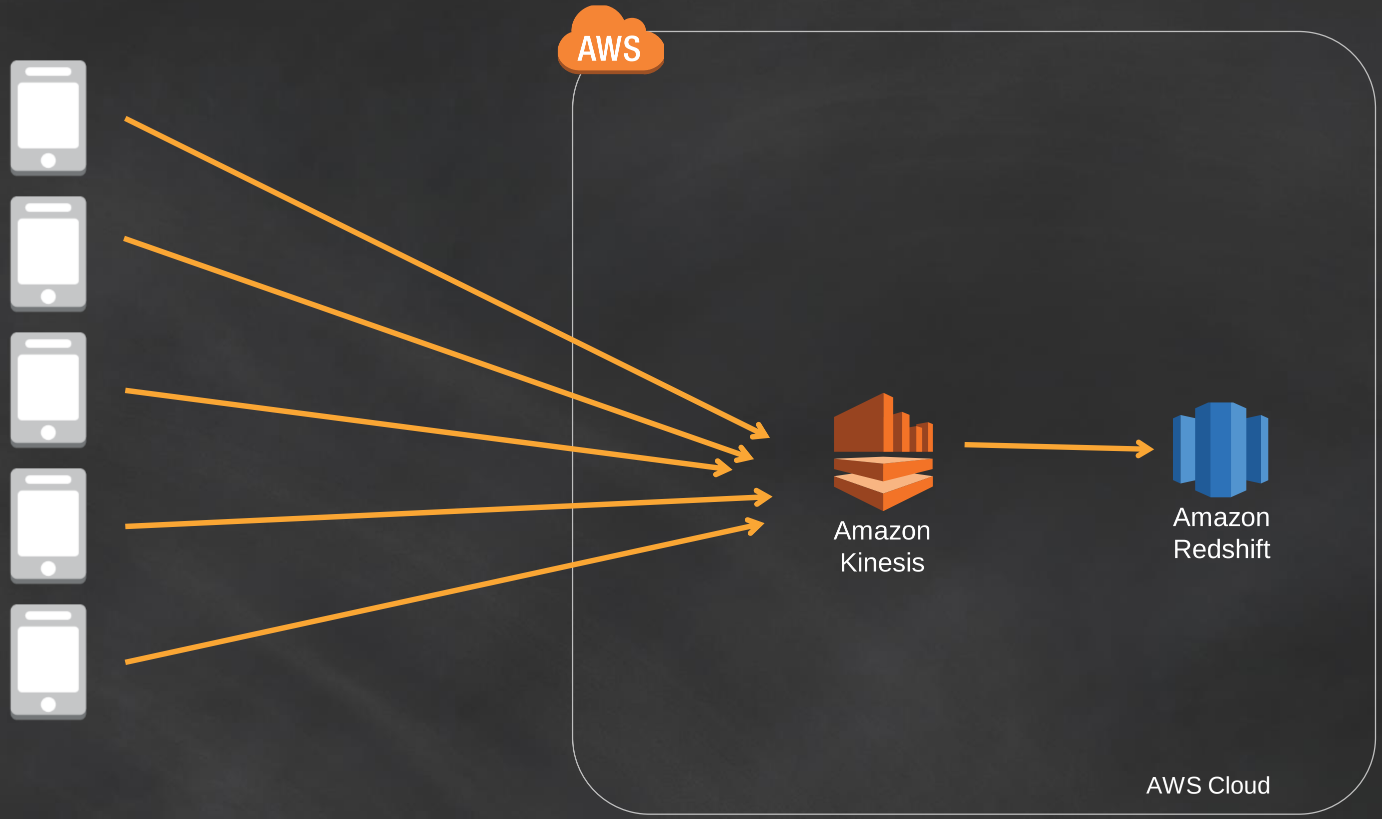
Data loading options



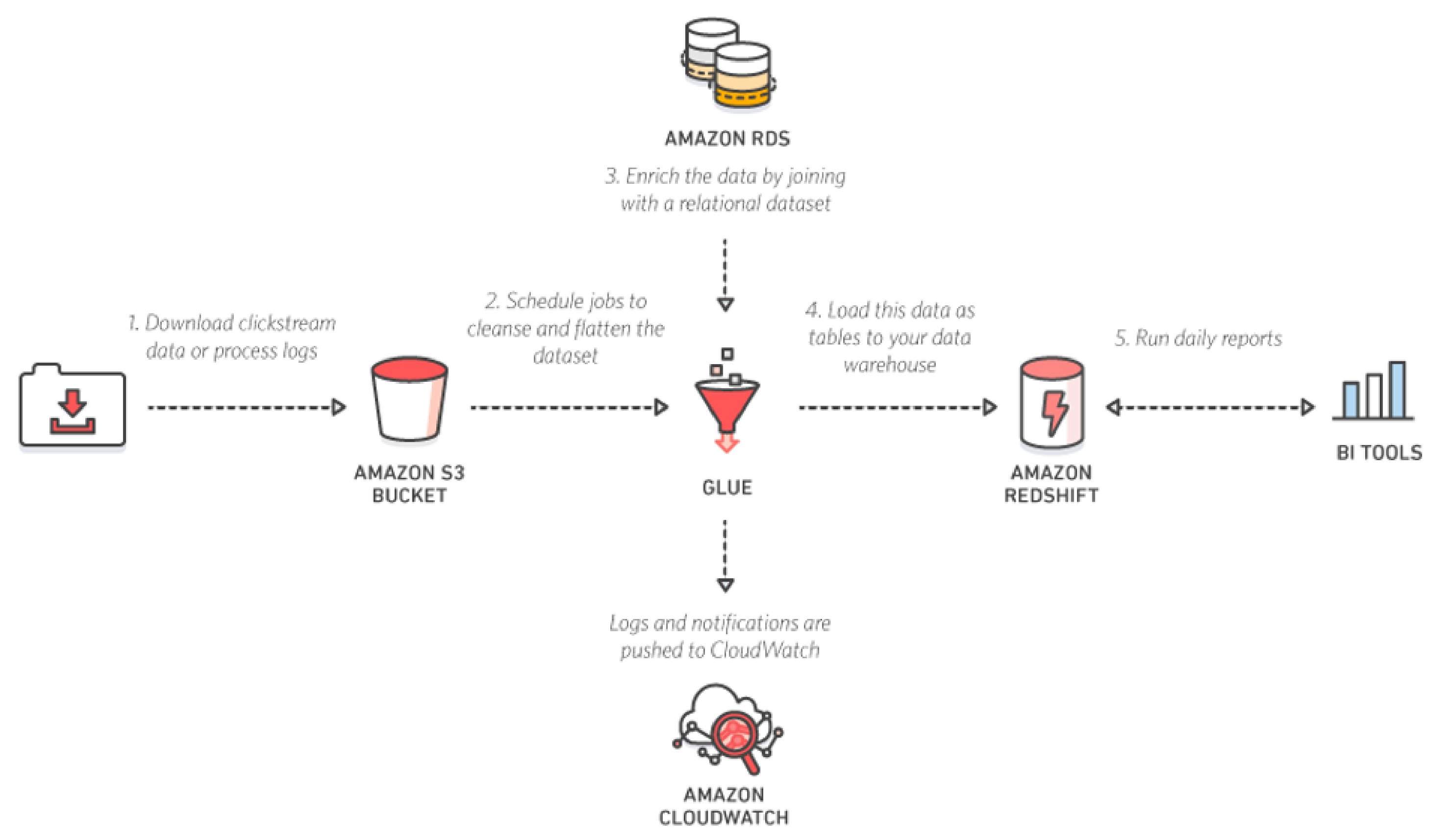
Data loading options



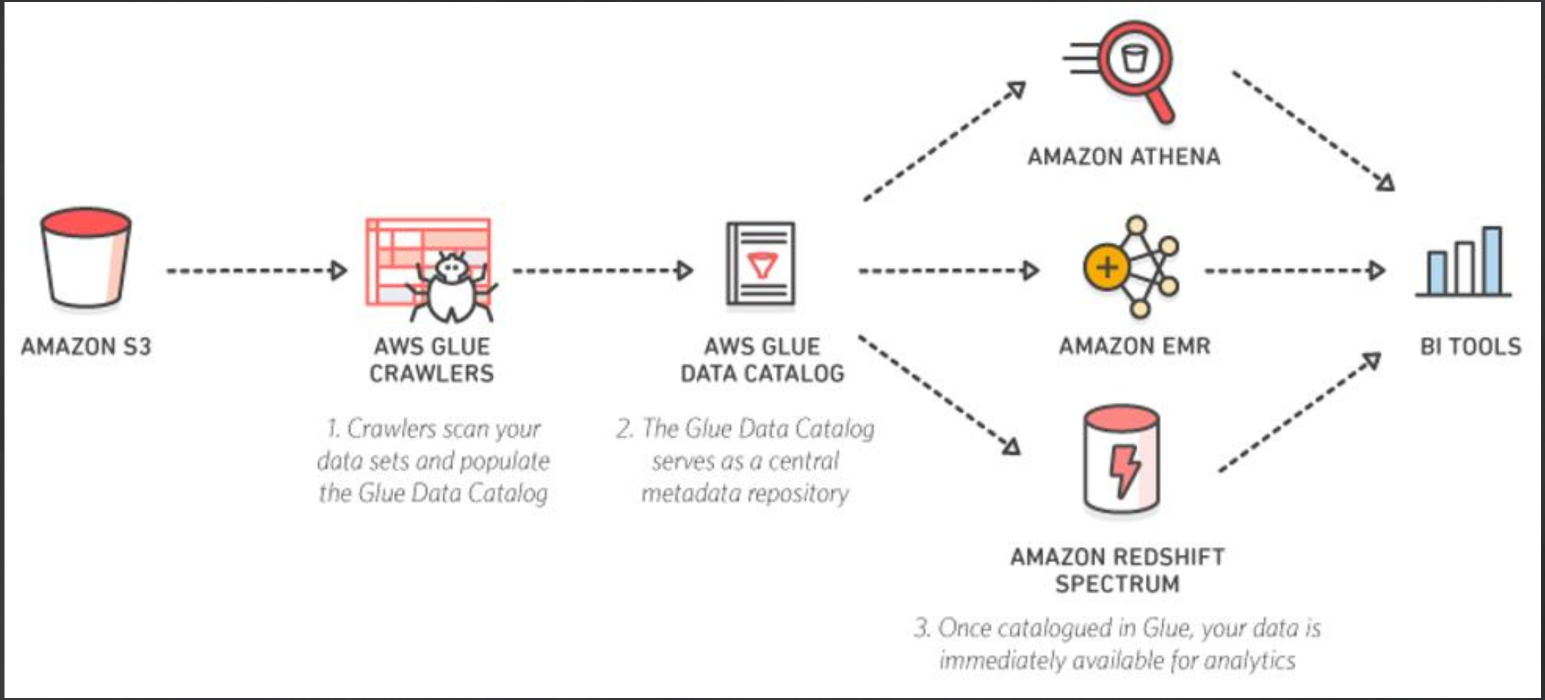
Data loading options



ETL data into your data warehouse


AMAZON RDS
3. Enrich the data by joining with a relational dataset
1. Download clickstream data or process logs
2. Schedule jobs to cleanse and flatten the dataset
4. Load this data as tables to your data warehouse
5. Run daily reports
AMAZON S3 BUCKET
GLUE
AMAZON REDSHIFT
BI TOOLS
Logs and notifications are pushed to CloudWatch
AMAZON CLOUDWATCH

Instantly query your data lake on Amazon S3



1. Crawlers scan your data sets and populate the Glue Data Catalog

2. The Glue Data Catalog serves as a central metadata repository

3. Once catalogued in Glue, your data is immediately available for analytics

Redshift Spectrum: Defining External Schema and External Tables

1. Define an external schema in Amazon Redshift using the AWS Glue Data Catalog or your own Apache Hive Metastore

```
CREATE EXTERNAL SCHEMA <schema_name>
FROM { [ DATA CATALOG ] | HIVE METASTORE }
DATABASE 'database_name'
IAM_ROLE 'iam-role-arn'
```

2. Register external tables using Amazon Athena, your Hive Metastore client, or from Amazon Redshift

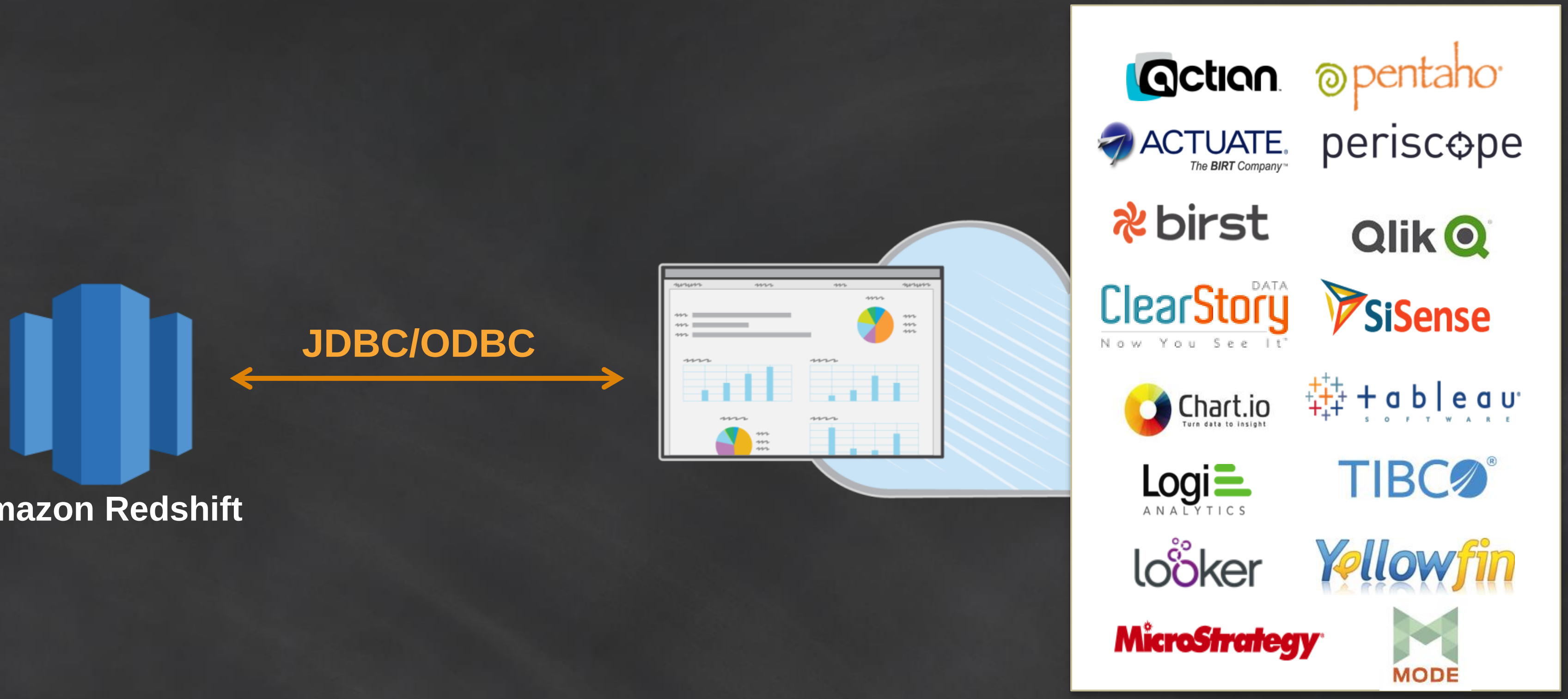
```
CREATE EXTERNAL TABLE <schema_name>.<table_name>
[PARTITIONED BY <column_name, data_type, …>]
STORED AS file_format
LOCATION s3_location
[TABLE PROPERTIES property_name=property_value, …];
```

3. Query external tables

```
SELECT … FROM <schema_name>.<table_name> …
```

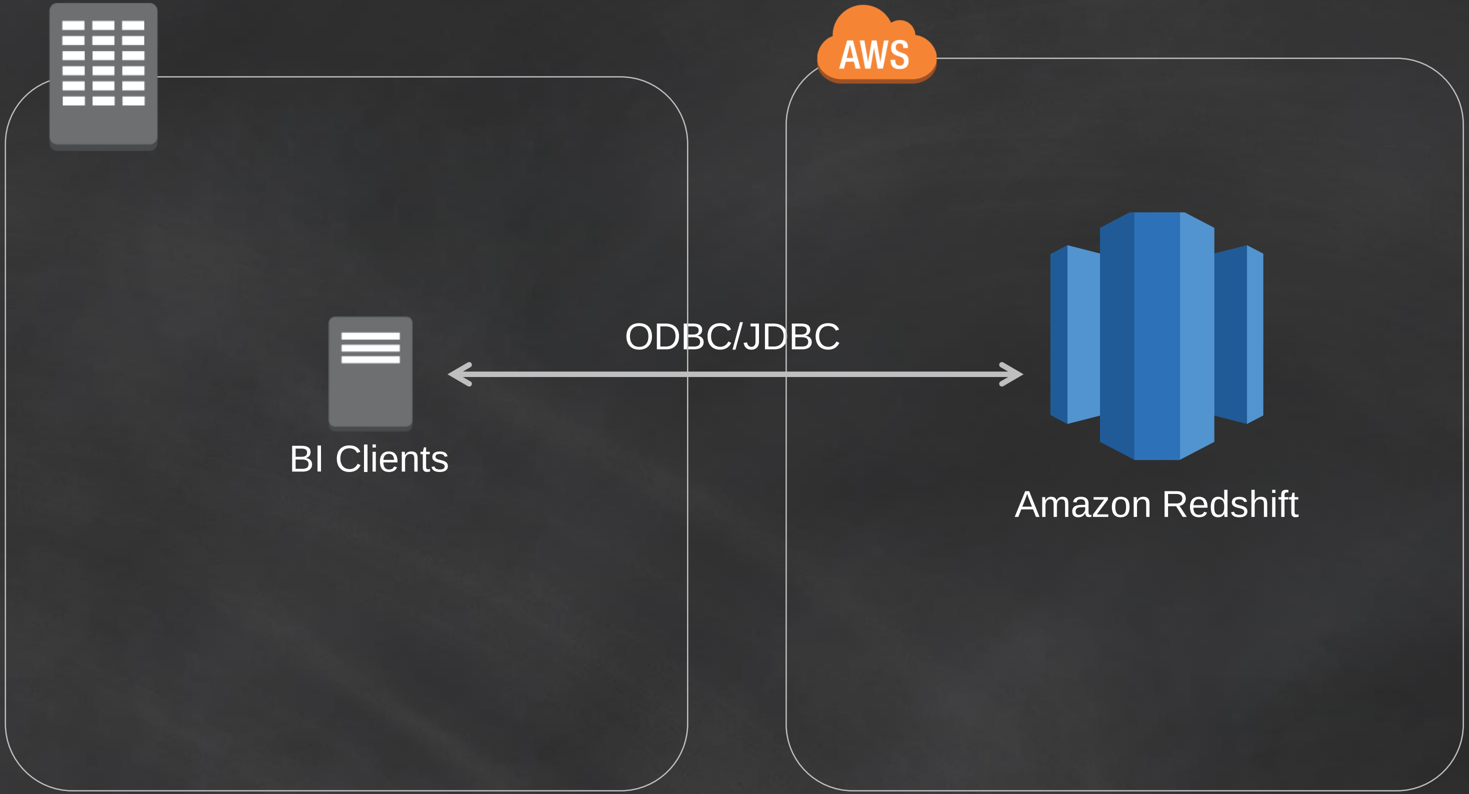
Querying

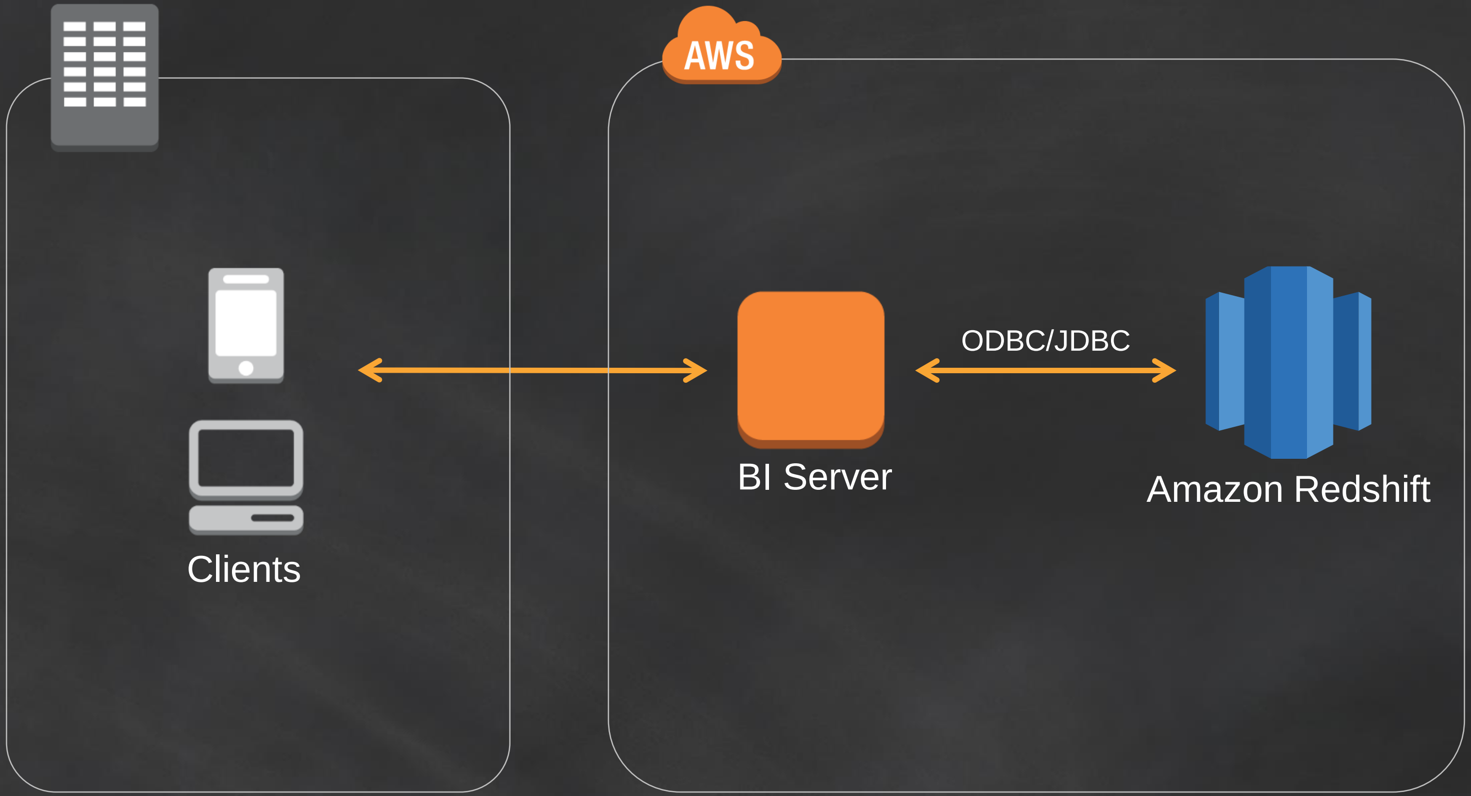
Amazon Redshift works with your existing BI tools



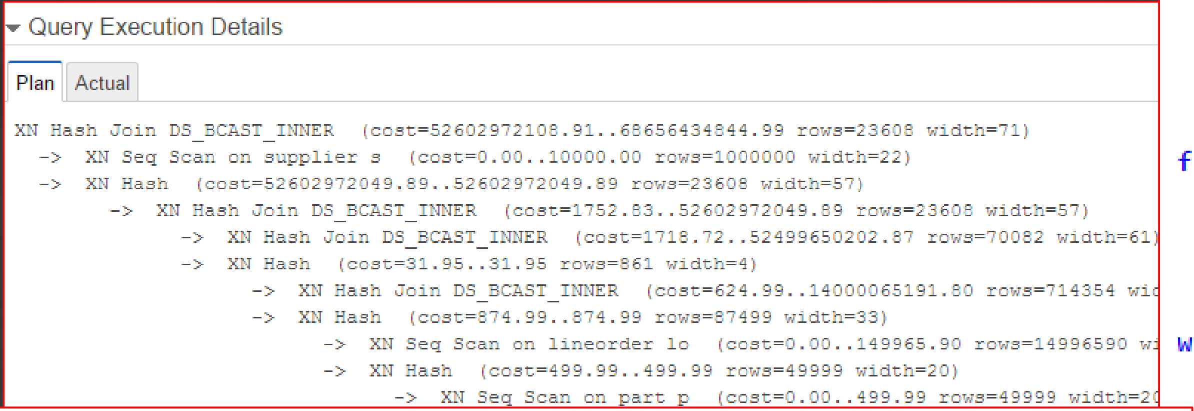
Amazon Redshift

JDBC/ODBC


AWS
ODBC/JDBC
BI Clients
Amazon Redshift


AWS
Clients
BI Server
ODBC/JDBC
Amazon Redshift

View explain plans

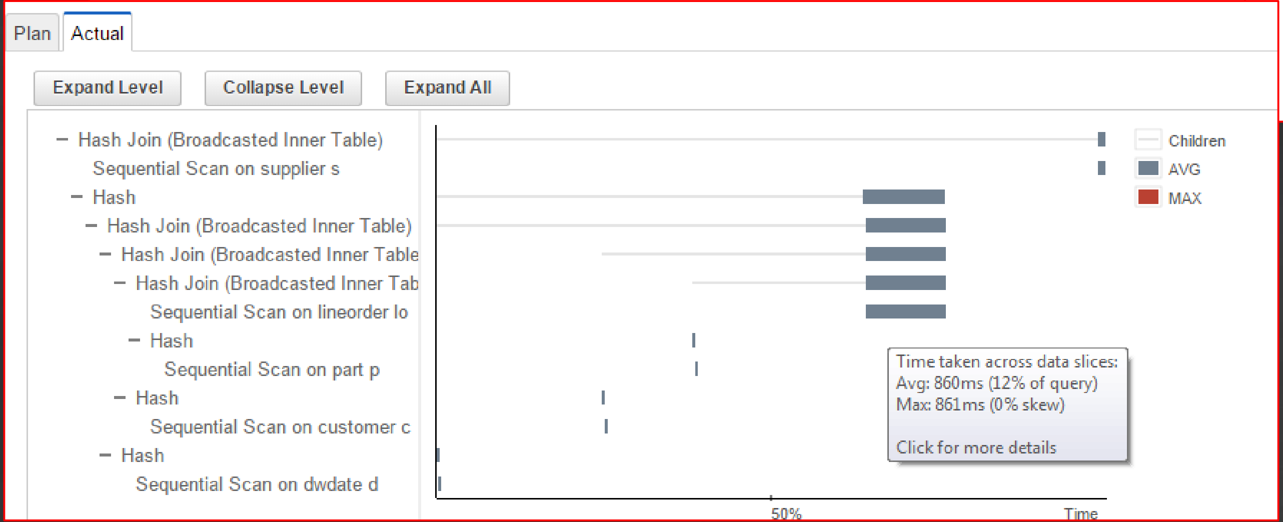


Query Execution Details

Plan | Actual

```
XN Hash Join DS_BCAST_INNER  (cost=52602972108.91..68656434844.99 rows=23608 width=71)
  ->  XN Seq Scan on supplier s  (cost=0.00..10000.00 rows=1000000 width=22)
  ->  XN Hash  (cost=52602972049.89..52602972049.89 rows=23608 width=57)
        ->  XN Hash Join DS_BCAST_INNER  (cost=1752.83..52602972049.89 rows=23608 width=57)
              ->  XN Hash Join DS_BCAST_INNER  (cost=1718.72..52499650202.87 rows=70082 width=61)
              ->  XN Hash  (cost=31.95..31.95 rows=861 width=4)
                    ->  XN Hash Join DS_BCAST_INNER  (cost=624.99..14000065191.80 rows=714354 wid
                    ->  XN Hash  (cost=874.99..874.99 rows=87499 width=33)
                          ->  XN Seq Scan on lineorder lo  (cost=0.00..149965.90 rows=14996590 wi
                          ->  XN Hash  (cost=499.99..499.99 rows=49999 width=20)
                                ->  XN Seq Scan on part p  (cost=0.00..499.99 rows=49999 width=2
```

```sql
select
    lo_orderkey,
    p_name,
    c_name,
    s_address,
    lo_quantity
from
    lineorder lo,
    part p,
    supplier s,
    customer c,
    dwdate d
where
    lo_custkey = c_custkey
    and     lo_partkey = p_partkey
    and     lo_suppkey = s_suppkey
    and     lo_orderdate = d_datekey
    and     d_sellingseason = 'Summer'
```

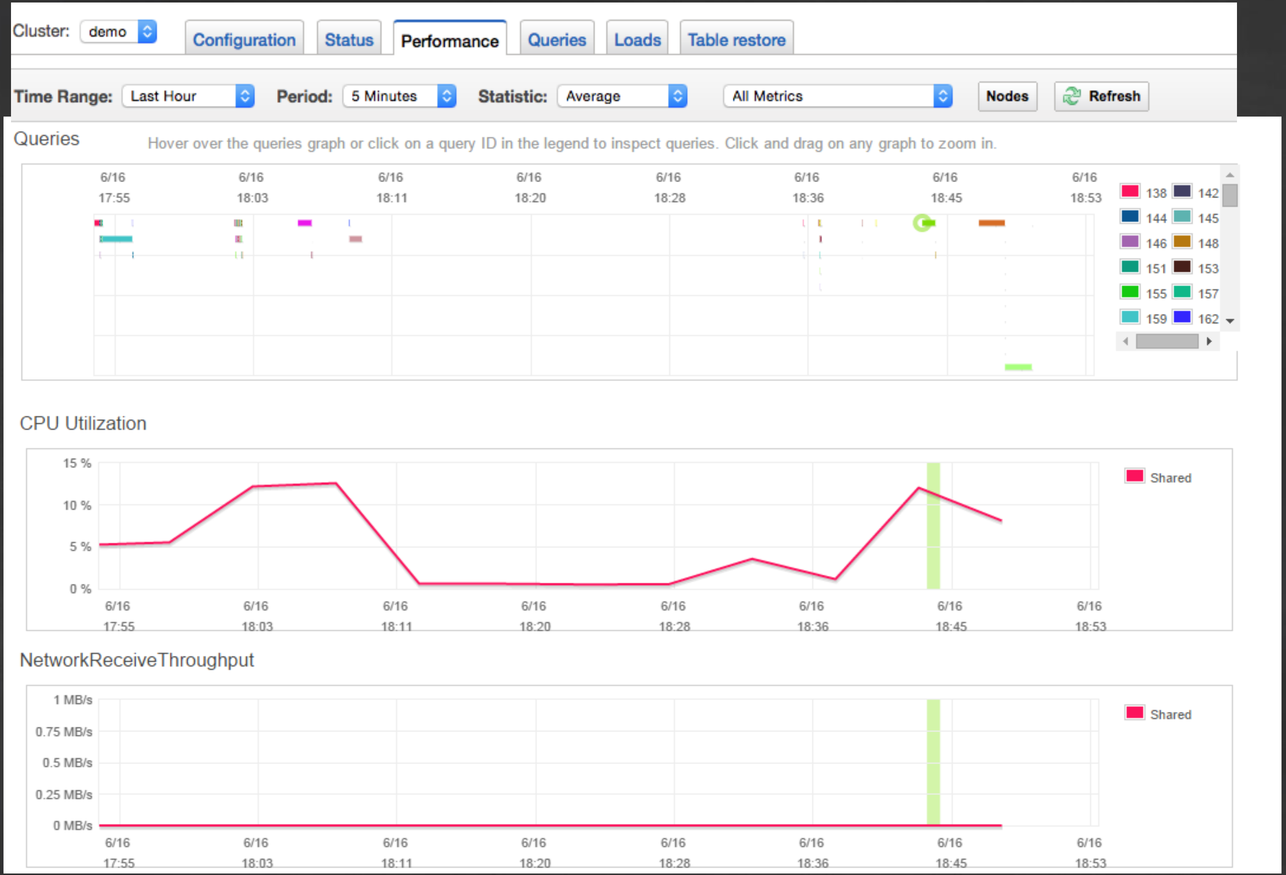


```
− Hash Join (Broadcasted Inner Table)
   Sequential Scan on supplier s
 − Hash
   − Hash Join (Broadcasted Inner Table)
     − Hash Join (Broadcasted Inner Table
       − Hash Join (Broadcasted Inner Tab
           Sequential Scan on lineorder lo
         − Hash
             Sequential Scan on part p
       − Hash
           Sequential Scan on customer c
     − Hash
         Sequential Scan on dwdate d
```

Time taken across data slices:
Avg: 860ms (12% of query)
Max: 861ms (0% skew)

Click for more details

Monitor query performance



Cluster: demo

Time Range: Last Hour

Period: 5 Minutes

Statistic: Average

All Metrics

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.

CPU Utilization

NetworkReceiveThroughput

Resources

- **Detail Pages**
  - http://aws.amazon.com/redshift
  - https://aws.amazon.com/redshift/spectrum
  - https://aws.amazon.com/marketplace/redshift/
  - https://aws.amazon.com/redshift/developer-resources/
  - Amazon Redshift Utilities - GitHub

- **Best Practices**
  - http://docs.aws.amazon.com/redshift/latest/dg/c_loading-data-best-practices.html
  - http://docs.aws.amazon.com/redshift/latest/dg/c_designing-tables-best-practices.html
  - http://docs.aws.amazon.com/redshift/latest/dg/c-optimizing-query-performance.html

Pop-up Loft

Everything and Anything Startups Need to Get Started on AWS

aws.amazon.com/activate