

# AWS re:INVENT

## Database & Analytics State of the Union

Raju Gulabani

Vice President, Databases, Analytics & Machine Learning

AWS  
re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



In this session, we discuss the evolution of database and analytics services in AWS, the new database and analytics services and features we launched this year, and our vision for continued innovation in this space. We are witnessing an unprecedented growth in the amount of data collected, in many different forms. Storage, management, and analysis of this data require database services that scale and perform in ways not possible before. AWS offers a collection of database and other data services—including Amazon Aurora, Amazon DynamoDB, Amazon RDS, Amazon Redshift, Amazon ElastiCache, Amazon Kinesis, and Amazon EMR—to process, store, manage, and analyze data. In this session, we provide an overview of AWS database and analytics services and discuss how customers are using these services today.

## What to Expect

1

### Understand

the portfolio of  
AWS Database and  
Analytics services

2

### Get a sense

of the volume  
and scale at which  
customers are using  
AWS services

3

### Learn

about common  
customer use cases  
and architectures

4

### Decide

when to use  
which services



## Purpose-built

Purpose-built  
databases & analytic  
engines. Right tool  
for the right job.



AWS has the right database service type for different workload use cases that customers might have with easy cloud use and scale

# AWS Databases and Analytics

Broad and deep portfolio, purpose-built for builders

## Business Intelligence & Machine Learning

 QuickSight

 SageMaker

 Comprehend

### Relational Databases

 Aurora

 RDS


### Non-Relational Databases

 DynamoDB

 ElastiCache  
(Redis, Memcached)

 Neptune  
(Graph)

### Data Lake

 S3/Glacier

 AWS Glue  
(ETL & Data Catalog)

### Data Movement

Database Migration Service | Snowball | Snowmobile | Kinesis Data Firehose | Kinesis Data Streams

### Analytics

DW | Big Data Processing | Interactive

 Redshift

 EMR

 Athena

### Real-time

 Elasticsearch  
Service

 Kinesis  
Analytics

**AWS**  
**re:Invent**

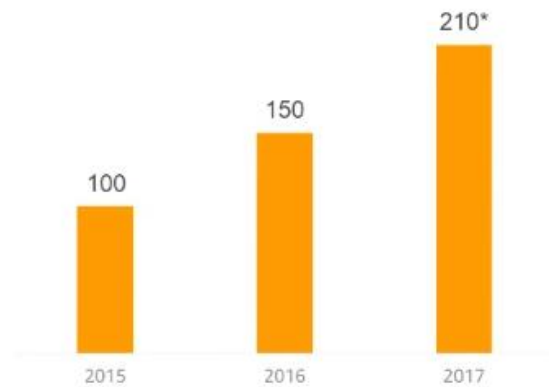
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Accelerating the Pace of Innovation

**210** features in 2017

**6** new database and analytics services GA in 2017



\*Projected number of releases to year-end

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



## Delivering Capabilities to Meet Customer Needs

RDS supports sharing encrypted databases.

**AWS Glue**

**Database Migration Service**

Schema conversion tool supports Oracle and Teradata

Aurora database cloning

**Aurora PostgreSQL**

Kinesis JSON support

Glue per second billing

Deploy Seibel on RDS

Redshift data compression

ElasticSearch supports 100 nodes

EMR fine-grained access controls

DynamoDB auto scaling

**Neptune**

Redshift SQL Scale UDFs

**460<sup>+</sup>**

RDS supports read replicas of encrypted DBs

**Kinesis Analytics**

Redshift query monitoring rules

Kinesis Streams encryption

Redshift federated authentication

Kinesis Firehose management console

Aurora advanced auditing

**Athena**

QuickSight supports Redshift

Redshift now 2x improved throughput

Elasticsearch service supports ElasticSearch 5

ElasticCache supports MemCached 1.4.34

features released since 2015

Amazon RDS for Aurora

QuickSight supports Presto and Spark

Aurora & RDS HIPAA eligible

EMR supports encryption for Spark, Tez, MapReduce

EMR per second billing

**Redshift Spectrum**

DMS support for Redshift

QuickSight Machine Generated Insights

RDS supports latest MySQL

**Amazon Machine Learning**

EMR support Spark 2.1

**DynamoDB Accelerator**

**Kinesis Video Streams**

EMR auto scaling

**RDS for MariaDB**

RDS supports Oracle Enterprise Manager

RDS supports Oracle Label Security

**QuickSight**

Aurora support encryption of globally distributed database

EMR instance fleets for Spot instances/blocks

RDS Supports SQL Server 2016

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



\*Projected number of launches to year-end 2017



# Most Database & Analytics Cloud Customers



## Recognized by Customers and Analysts

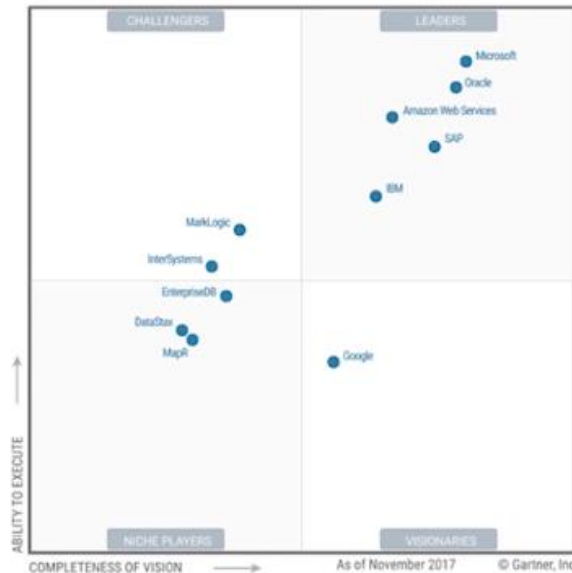
The Forrester Wave: Database-As-A-Service, Q2 2017



The Forrester Wave is copyrighted by Forrester Research, Inc. Forrester and Forrester Wave are trademarks of Forrester Research, Inc. The Forrester Wave is a graphical representation of Forrester's opinion on a market and is plotted using a detailed spreadsheet with explicit scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave. Information is based on best-available resources. Opinions reflect judgment at the time and are subject to change.

# Recognized by Customers and Analysts

Gartner Magic Quadrant for Operational Database Management Systems, November 2017



This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from AWS. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



## AWS Database Services

### Relational



Aurora



RDS



MariaDB

ORACLE

Microsoft SQL Server



AWS Database Migration Service

# Amazon RDS

Managed relational database service with a choice of six popular database engines

Amazon  
**Aurora**

**MySQL**

**PostgreSQL**

**MariaDB**

**Microsoft SQL Server**

**ORACLE**

**Easy to administer**



No need for infrastructure provisioning, installing and maintaining DB software

**Available & Durable**



Automatic Multi-AZ data replication; automated backup, snapshots, failover

**Highly scalable**



Scale database compute and storage with a few clicks with no application downtime

**Fast & Secure**



SSD storage and guaranteed provisioned I/O; data encryption at rest and in transit

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



## RDS Customers—Commercial Databases

**verizon**



**intuit**

**vodafone**

**Agero**

**MIDAS**

**HESS**

**Liberty Mutual**

**TATA MOTORS**

**neo@Ogilvy**

**Expedia**

**3M**

**TREND MICRO**

**MERCK**

**Sierra-Cedar**

**rackspace**

**NORDSTROM**

**ESRI**

**INFORMATICA**

**Capital One**

**Johnson-Johnson**

**NASDAQ**

**SOCIETE GENERALE**

**COMCAST**

**aptos**

**ticketmaster**

**Smartronix**

**Unilever**

**docomo**

**SAMSUNG**

**Cerner**



**DAIKIN**

**ORION HEALTH**

**SK planet**

**Decisiv**

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



AWS probably runs more Oracle databases than Oracle runs in their cloud, same with SQL Server and Azure

# Old World Commercial Databases



Very expensive



Proprietary



Lock-in



Punitive  
licensing



You've  
got mail

## Moving to Open Database Engines



Enterprise-grade performance and reliability

## Amazon Aurora

MySQL and PostgreSQL compatible relational database built for the cloud

Performance and availability of commercial-grade databases at 1/10th the cost

### Performance & scalability



5x throughput of standard MySQL and 3x of standard PostgreSQL; scale-out up to 15 read replicas

### Availability & durability



Fault-tolerant, self-healing storage; six copies of data across three AZs; continuous backup to S3

### Highly Secure



Network isolation, encryption at rest/transit

### Fully managed

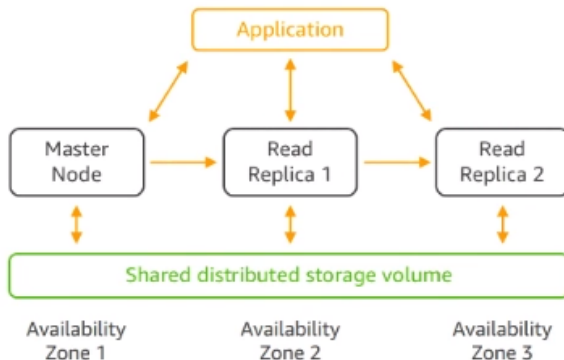


Managed by RDS: no hardware provisioning, software patching, setup, configuration or backups

# Amazon Aurora—High Performance

Scale out to millions of reads per second

Scale out read performance



Up to 15 read replicas across three AZs

Auto-scale new read replicas

Seamless recovery from read replica failures

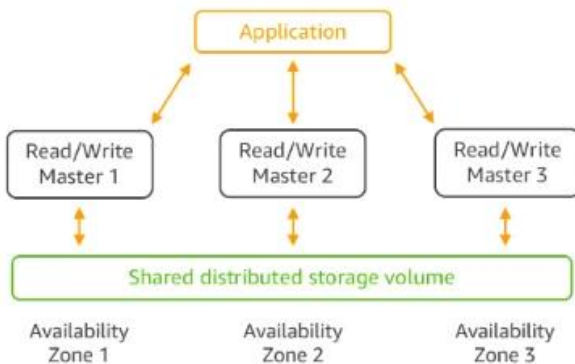
This is a shared database architecture where we do replication at the database level

## Aurora Multi-Master (Preview)

First relational DB service to scale out reads and writes, across multiple data centers

NEW!

Scale out both reads **and** writes



Zero application downtime from ANY instance failure

Zero application downtime from ANY AZ failure

Faster write performance and higher scale

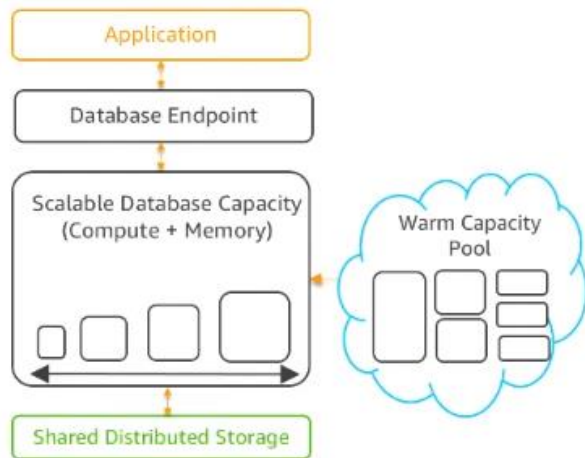
Sign up for single-region multi-master preview today; multi-region multi-master coming in 2018



# Aurora Serverless (Preview)

NEW!

On-demand, auto-scaling database for applications with variable workloads



Starts up on demand, shuts down when not in use

Automatically scales with no instances to manage

Pay per second for the database capacity you use

You simply get a database endpoint that you can use when you need it like for Dev and Test databases that your workers use during work hours and you want to shut down at night. This can also be used for cyclical or seasonal workloads.

## Aurora is the Fastest Growing Service in AWS History

NASDAQ

SAMSUNG

ticketmaster

bmc

URBAN AIRSHIP

EA  
ELECTRONIC ARTS

hulu

nielsen

Pearson

ancestry

RavenPack

zynga

ZUMBA  
fitness

Blackboard

THOMSON REUTERS

SAFE SOFTWARE

INRIX

Expedia

DOW JONES

INTERCOM

Alfresco

SAFE SOFTWARE

FICO

jenny

AstraZeneca

CBS Interactive

workfront

gumi

Experian

delight

Zillow

Lookout

docomo

PeopleAdmin

rackspace

FUNNY  
GAMES

SysAid

Bristol-Myers Squibb

毎日新聞

CyberAgent

SRA OSS

TalentBin  
SOURCE THE WEB.

AWS re:Invent

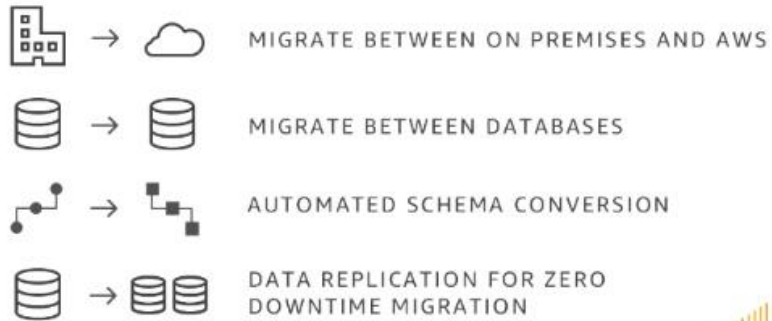
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

# AWS Database Migration Service

**MIGRATING  
DATABASES  
TO AWS**

**45,000+**  
DATABASES MIGRATED



## >45,000 Databases Migrated with DMS



# DHI Migrated from Oracle to Amazon Aurora



*"AWS gave us the opportunity to build an enterprise-class, cost-effective, open source-based database strategy with Amazon Aurora. We have been able to **easily migrate our database from Oracle to Amazon Aurora**, achieving the **database freedom** our team needs to deliver specialized insights and relevant connections to our customers."*

Brian Hostetter, Director, DevOps and Global Technology Architecture, Dice Holdings, Inc.

DHI Group was using Oracle to support Dice.com

Oracle had restrictive licensing that was overpriced

DHI Group migrated to Aurora

Aurora gave them an enterprise-class, cost effective DB

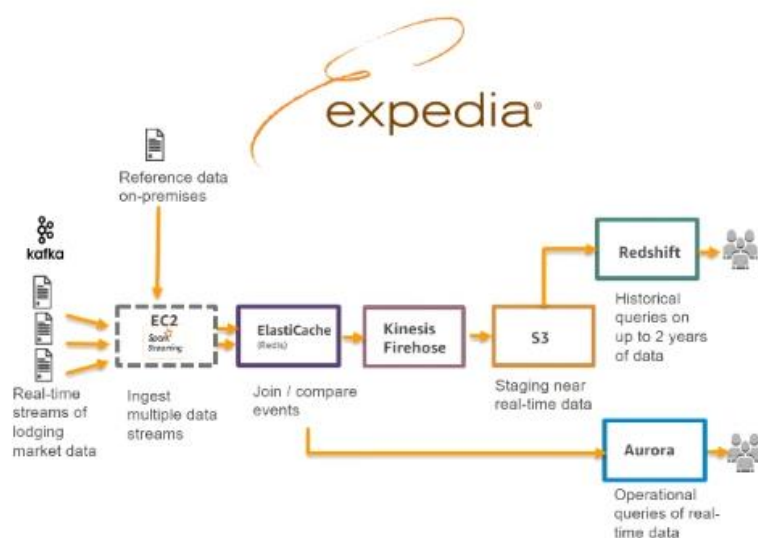
Now delivers specialized insights to customers

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Expedia Migrated from SQL Server to AWS



Needed real-time analysis of lodging market pricing

Migrated from Microsoft SQL Server

Use Aurora, Amazon Redshift, Kinesis, and ElasticCache

Process high-volume pricing and availability data

Query execution times reduced 80–95%

Database has >15B rows and continues to grow

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# AWS Database Services

## Non-relational



DynamoDB

Key value | Document



ElastiCache



redis



memcached



Neptune

Graph



Database Migration Service

## Amazon DynamoDB

We needed to adapt to power Amazon.com

### Dynamo: Amazon's Highly Available Key-value Store

Giuseppe DeCandia, Deniz Hastorun, Madan Jambani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall and Werner Vogels

Amazon.com

#### ABSTRACT

Reliability at massive scale is one of the biggest challenges we face at Amazon.com, one of the largest e-commerce operations in the world, even the slightest outage has significant business consequences and impacts customer trust. The Amazon.com platform, which provides services for some of the world's largest retailers, is implemented on top of an infrastructure of hundreds of thousands of servers and network components located in multiple data centers around the world. At this scale, small failures are inevitable, and the way persistent state is managed is critical to the success of these failures. This paper describes the design and implementation of Dynamo, a highly available key-value storage system that powers Amazon's core services and is designed to provide an always-on experience. To achieve this level of availability, Dynamo maintains consistency under certain failure scenarios. It makes extensive use of object replication and application-assisted conflict resolution as a means that provides a novel interface for developers to use.

One of the biggest challenges we face at Amazon.com is that the reliability and availability of a system is directly tied to how its application state is managed. Amazon.com is a highly decentralized, loosely coupled, service-oriented architecture that consists of hundreds of services. In this paper, we describe a particular need for storage technologies that are highly available. For example, customers should be able to browse and add items to their shopping cart even if links are broken, network routes are flapping, or data centers are being decommissioned by Amazon. Therefore, the service responsible for managing shopping carts requires that it can always write to and read from its data store, and that its data needs to be available across multiple data centers.

Dealing with failures in an infrastructure composed of millions of components is no standard task of operation; there are always a small but significant number of server and network components that are failing at any given time. As such, Amazon's software systems need to be constructed in a manner that treats failure handling as the normal case without impacting availability or performance.

Needed to power Amazon.com

Required massive scalability and reliability

DynamoDB designed to meet this need



## Amazon DynamoDB

Fast and flexible NoSQL database service for any scale

### Highly scalable



Auto-scaling to hundreds of terabytes of data that serve millions of requests per second

### Fast, consistent performance



Consistent single-digit millisecond latency; DAX in-memory performance reduces response times to microseconds

### Fully managed



Automatic provisioning, infrastructure management, scaling, and configuration with zero downtime

### Business critical reliability



Data is replicated across fault tolerant Availability Zones, with fine-grained access control



# Amazon DynamoDB

Delivering on customer needs

February 2017



Time to Live (TTL)

April 2017



VPC Endpoints

April 2017



DynamoDB Accelerator (DAX)

June 2017



Auto Scaling



Global Tables (GA)



On-demand Backup (GA)

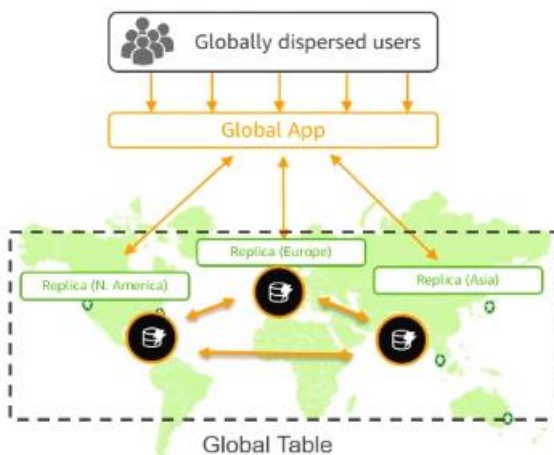


Encryption at rest (Coming Soon)

## DynamoDB Global Tables (GA)

NEW!

First fully managed, multi-master, multi-region database



Build high performance, globally distributed applications

Low latency reads & writes to locally available tables

Disaster proof with multi-region redundancy

Easy to set up and no application re-writes required

You simply create a DynamoDB table and tell us the regions you want to have it running in and we will replicate it to all those regions, this is very good for applications that have a global scale and need low latency in all the regions for the user applications.

# DynamoDB—Backup and Restore

NEW!

First NoSQL database to automate on-demand and continuous backups



On-demand backups for long-term data archival and compliance (GA)



Point in time restore for short-term retention and data corruption protection (coming soon)



Back up hundreds of TB instantaneously with NO performance impact

## Amazon DynamoDB Powers **SAMSUNG**



Backup and restore on mobile application for 300M users

300+ PBs in AWS, 850 TBs in DynamoDB, 130M daily API requests

Migrated from Cassandra to DynamoDB

Consistent performance and 70% cost savings (TCO)

“DynamoDB provided **consistent high performance** at a **drastically lower cost** than Cassandra.”

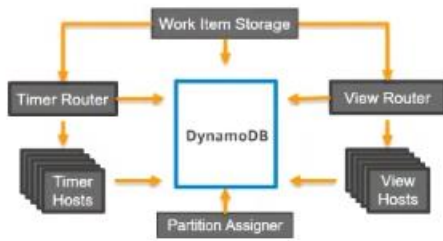
Seongkyu Kim, Server Engineer, Samsung Electronics

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Amazon DynamoDB Powers **amazon.com**



Herd is the database system powering customer purchases

Migrated from Oracle to DynamoDB

Extreme scale to handle millions of requests per second

Workflow processing dropped from 1s to 100ms



**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

**Prime Day 2017:** DynamoDB handled **peak of 12.9M requests per second** (an increase from 6.3M) with no increase in latency

" Herd is a **mission-critical system** for Amazon, and we are **extremely confident** in DynamoDB as the technology on which to run it. "

Mike Thomas, Software Development Manager, Amazon Herd



## Amazon ElastiCache

Managed, in-memory data store service

Redis or Memcached to power real-time apps with sub-millisecond latency

### Extreme performance



In-memory data store and cache using optimized stack to deliver sub-millisecond response times

### Secure & hardened



VPC for cluster isolation, encryption at rest/transit, and HIPAA compliance

### Easily scalable



Read scaling with replicas; write and memory scaling with sharding; non-disruptive scaling

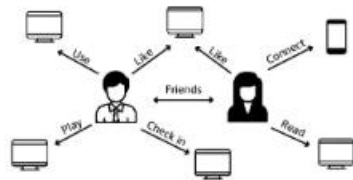
### Highly available & reliable



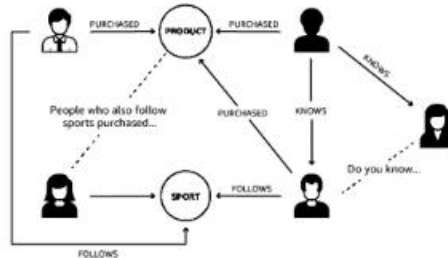
Multi-AZ with automatic failover

# Graph Use Cases

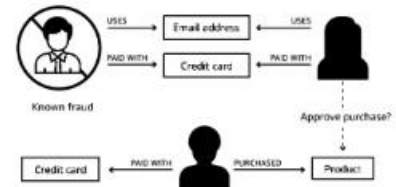
## Social News Feed



## Recommendations



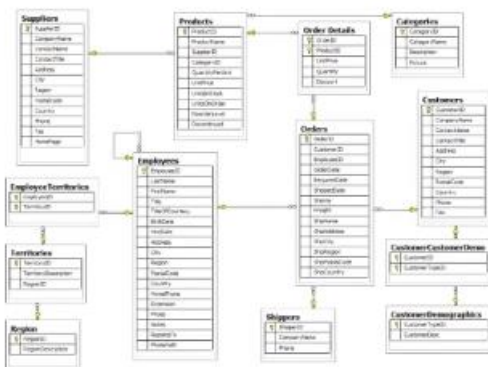
## Retail Fraud Detection



## Highly Connected Data Best Represented in a Graph

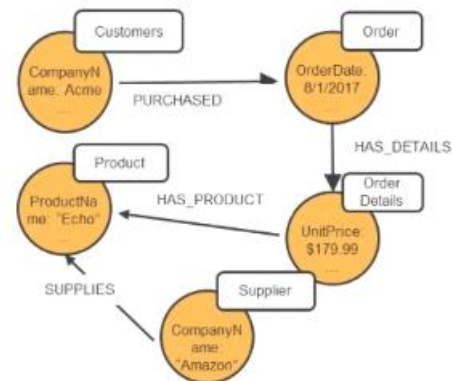
### Relational model

Foreign keys used to represent relationships  
Queries can involve nesting & complex joins  
Performance can degrade as datasets grow



### Graph model

Relationships are first-order citizens  
Easy to write queries that navigate the graph  
Results returned quickly, even on large datasets



**AWS re:Invent**

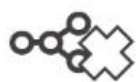
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.





# Challenges Building Apps with Highly Connected Data

## Relational Databases



Unnatural for  
querying  
graph

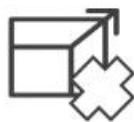


Inefficient  
graph  
processing



Rigid schema  
inflexible for  
changing graphs

## Existing Graph Databases



Difficult to  
scale



Difficult to  
maintain high  
availability



Too  
expensive



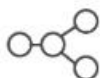
Limited support  
for open  
standards

## Amazon Neptune (Preview)

NEW!

Fully managed graph database

### Open



Supports Apache  
TinkerPop & W3C  
RDF graph models

### Fast & Scalable



Store billions of  
relationships; query with  
millisecond latency

### Reliable



Six replicas of your  
data across three AZs  
with full backup and  
restore

### Easy



Gremlin  
SPARQL

Build powerful  
queries easily with  
Gremlin and SPARQL

# Thomson Reuters Uses Neptune

Navigate a web of global tax policies

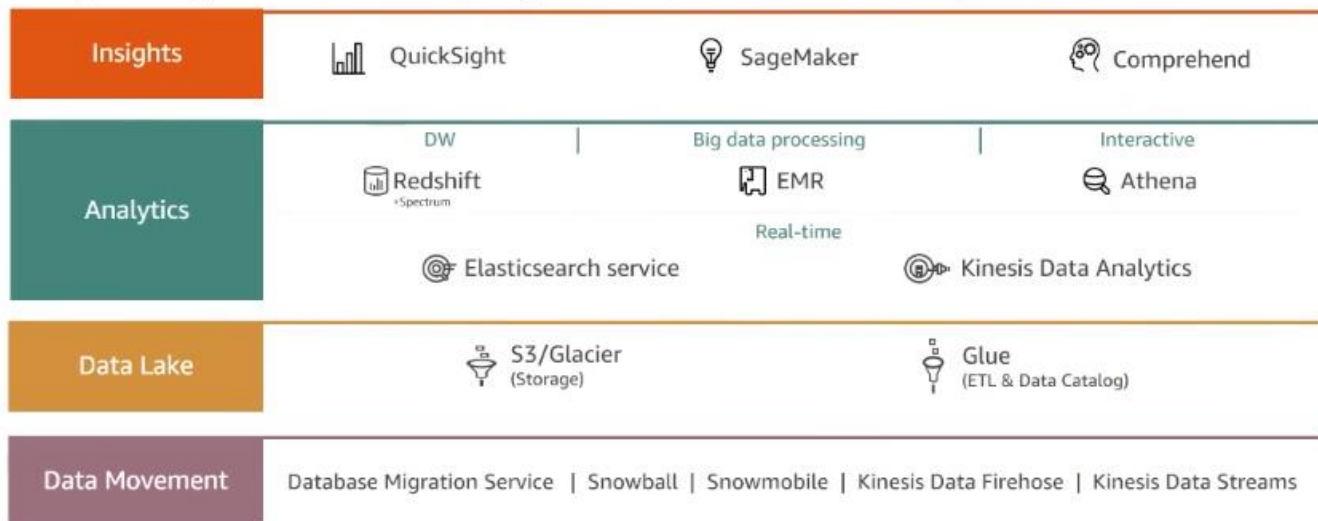


" Our customers are increasingly required to navigate a complex web of global tax policies and regulations. We need an approach **to model the sophisticated corporate structures** of our largest clients and deliver an end-to-end tax solution. We use a microservices architecture approach for our platforms and are beginning to **leverage Amazon Neptune as a graph-based system** to quickly create links within the data. "

*Tim Vanderham, chief technology officer*

## AWS Big Data and Analytic Services

Any analytic workload, any scale, at the lowest possible cost



## Traditionally, Analytics Used to Look Like This



Relational data

---

TBs-PBs scale

---

Schema defined prior to data load

---

Operational reporting and ad hoc

---

Large initial capex + \$10k-\$50k / TB / Year

## Data Lakes Extend the Traditional Approach



Relational and non-relational data

---

TBs-EBs scale

---

Schema defined during analysis

---

Diverse analytical engines to gain insights

---

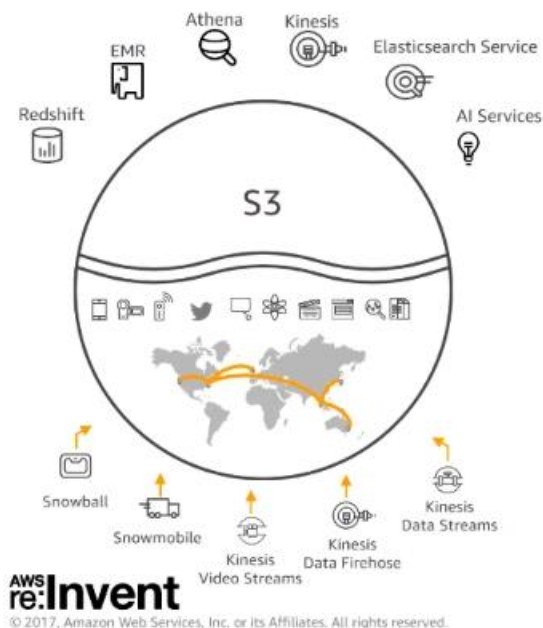
Designed for low-cost storage and analytics

**AWS**  
**re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Data Lakes on AWS



Most ways to bring data in

Unmatched durability and availability at EB scale

Best security, compliance, and audit capabilities

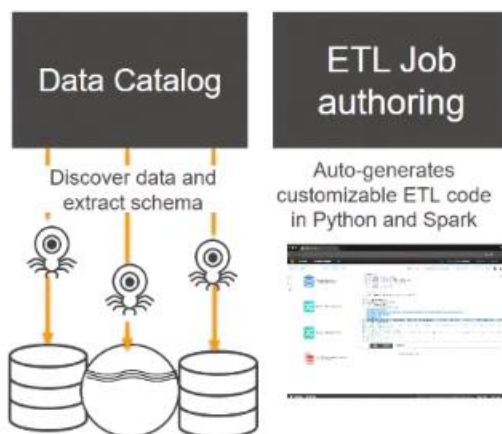
Run any analytics on the same data without movement

Scale storage and compute independently

Store data at \$0.023 / month; Query for \$0.05/GB scanned



## AWS Glue—Serverless Data catalog & ETL service



Automatically discovers data and stores schema

Data searchable, and available for ETL

Generates customizable code

Schedules and runs your ETL jobs

Serverless



# Amazon Redshift—Data Warehousing

Fast, powerful, simple, and fully managed data warehouse at 1/10 the cost

Massively parallel, scale from gigabytes to petabytes

## Fast at scale



Columnar storage technology to improve I/O efficiency and scale query performance

## Open file formats



Analyze optimized data formats on the latest SSD, and all open data formats in Amazon S3

## Secure



Audit everything; encrypt data end-to-end; extensive certification and compliance

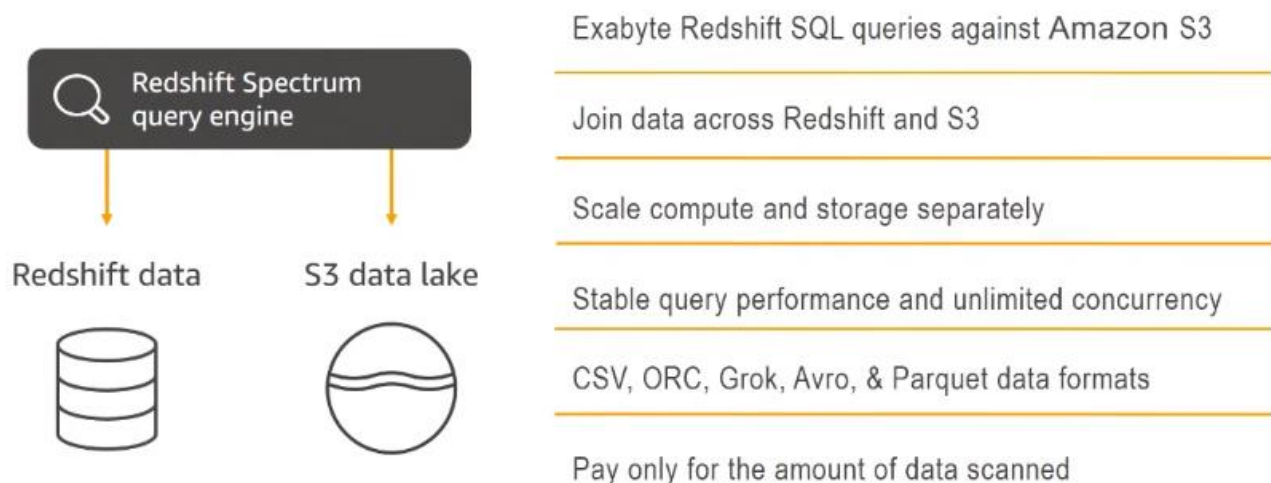
## Inexpensive



As low as \$1,000 per terabyte per year, 1/10th the cost of traditional data warehouse solutions; start at \$0.25 per hour

# Amazon Redshift Spectrum

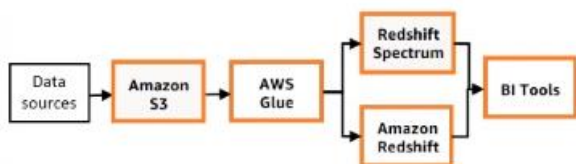
Extend the data warehouse to exabytes of data in S3 data lake



Spectrum is about taking compute to storage location and do the analysis there.

# NUVIAD —Data Lake Analytics with Redshift Spectrum

NUVIAD is a marketing platform that helps media buyers [optimize](#) their mobile [bidding](#)



Use AWS for marketing campaign and bidding analytics

Scale S3 storage for unlimited data capacity

Use Spectrum for unlimited scale and query concurrency

80% performance gain using Parquet data format

**"Amazon Redshift Spectrum is a game changer for us. Reports that took minutes to produce are now delivered in seconds. We like the ability scale compute on-demand to query Petabytes of data in S3 in various open file formats."**

**Rafi Ton, CEO, NUVIAD**

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



They are directly running queries using Spectrum on S3 data to get this done.

## Amazon EMR—Big Data Processing

Analytics and ML at scale

Nineteen open-source projects: Apache Hadoop, Spark, HBase, Presto, and more

Enterprise-grade security

### Latest versions



Updated with the latest open source frameworks within 30 days of release

### Low cost



Flexible billing with per-second billing, EC2 spot, Reserved Instances and auto-scaling to reduce costs 50–80%

### Use S3 storage



Process data directly in the S3 data lake securely with high performance using the EMRFS connector

### Easy



Launch fully managed Hadoop & Spark in minutes; no cluster setup, node provisioning, cluster tuning

EMR is our managed Hadoop service that uses S3 as the data lake. We use the notion of spot market for running those batch jobs for you

# Amazon Elasticsearch Service

Easy to deploy, secure, operate, and scale Elasticsearch

Customers use Elasticsearch for log analytics, full-text search & application monitoring

## Easy to Use



Fully-managed.  
Deploy production-ready  
clusters in minutes

## Open



Direct access to  
Elasticsearch open-source  
APIs; supports Logstash  
and Kibana

## Secure



Secure access with VPC  
to keep all traffic within  
AWS network

## Available



Zone awareness replicates  
data between two AZs;  
automatically monitors &  
replaces failed nodes

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Amazon Kinesis—Real Time

Easily collect, process, and analyze video and data streams in real time

**New**



## Kinesis Video Streams

Capture, process,  
and store video  
streams for analytics



## Kinesis Data Streams

Build custom  
applications that  
analyze data streams



## Kinesis Data Firehose

Load data streams  
into AWS data stores



## Kinesis Data Analytics

Analyze data streams  
with SQL

# Amazon Athena—Interactive Analysis

Interactive query service to analyze data in Amazon S3 using standard SQL

No infrastructure to set up or manage and no data to load

Ability to run SQL queries on data archived in Glacier (Coming soon)

## Query Instantly



Zero setup cost; just point to S3 and start querying

## Pay per query



Pay only for queries run; save 30–90% on per-query costs through compression

## Open



ANSI SQL interface, JDBC/ODBC drivers, multiple formats, compression types, and complex joins and data types

## Easy



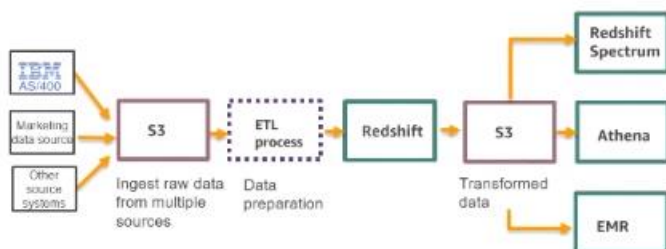
Serverless: zero infrastructure, zero administration  
Integrated with Amazon QuickSight

**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Sysco—Analytics on the Data Lake



Sysco is the leader in selling, marketing, & distributing food

Challenge: large volumes of data in multiple systems

Consolidated data into a single S3 data lake

Data scientists use Amazon EMR notebooks, Athena, & Redshift Spectrum used by business users for reporting



# Amazon QuickSight

Fast, easy to use, serverless analytics at 1/10<sup>th</sup> the cost of traditional BI



Empower everyone



Seamless connectivity



Fast analysis



Serverless



## Serverless Analytics

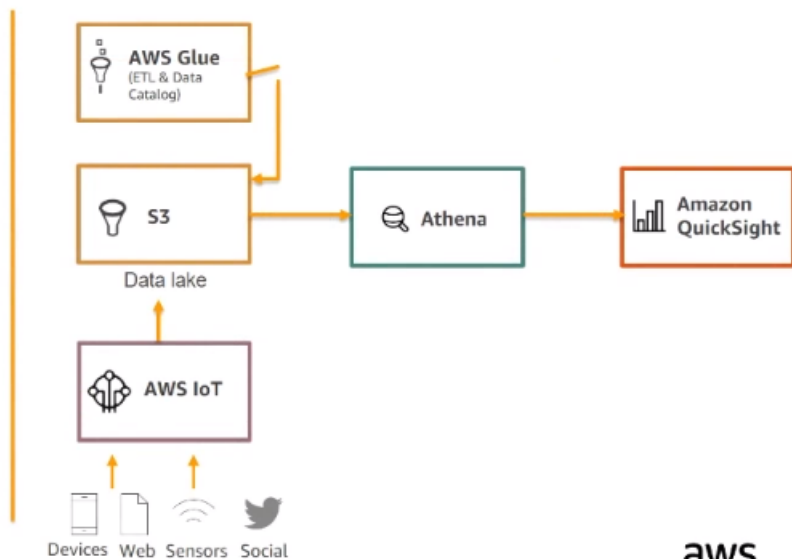
Deliver cost-effective analytic solutions faster

  
Serverless;  
zero infrastructure;  
zero administration

  
Never pay for  
idle resources

  
Automatically  
scales resources  
with usage

  
Availability and  
fault tolerance  
built in



**AWS re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



# Enabling All Types of Data-Driven Analytics



**Retrospective**  
analysis and  
reporting



**Here-and-now**  
real-time processing  
and dashboards



**Predictions**  
to enable smart  
applications

NEW!

## Introducing: Amazon SageMaker (GA)

A managed service

that provides **the quickest and easiest way** for  
your data scientists and developers to get

**ML models from idea to production.**

NEW!

## Amazon SageMaker (GA)

The quickest and easiest way to get ML models from idea to production



End-to-End  
Machine Learning  
Platform



Zero setup



Flexible Model  
Training

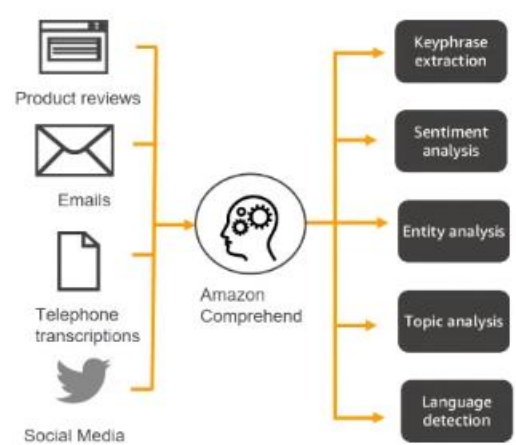


Pay by the  
second

NEW!

# Introducing Amazon Comprehend

Natural Language Processing to discover insights from text



Classify language, extract key phrases, understand sentiment, identify / organize documents by topic

Continuously trained and constantly improving

Integrated with Amazon S3 and AWS Glue

## When to Use Which Services

Situation	Solution
Existing application	Use your existing engine on RDS <ul style="list-style-type: none"><li>MySQL → Amazon Aurora, RDS for MySQL</li><li>PostgreSQL → Amazon Aurora, RDS for PostgreSQL</li><li>Oracle → Amazon Aurora, RDS for Oracle</li><li>SQL Server → Amazon Aurora, RDS for SQL Server</li><li>MariaDB → Amazon Aurora, RDS for MariaDB</li></ul>
New application	<ul style="list-style-type: none"><li>If you can avoid relational features → DynamoDB</li><li>If you need relational features → Amazon Aurora</li></ul>
In-memory store/cache	<ul style="list-style-type: none"><li>Amazon ElastiCache</li></ul>
Data Warehouse & BI	<ul style="list-style-type: none"><li>Amazon Redshift, Amazon Spectrum, and Amazon QuickSight</li></ul>
Interactive, serverless analysis of data in S3	<ul style="list-style-type: none"><li>Amazon Athena and Amazon QuickSight</li></ul>
Apache Spark, Apache Hadoop, Apache Hbase	<ul style="list-style-type: none"><li>Amazon EMR</li></ul>
Log analytics, operational monitoring and search	<ul style="list-style-type: none"><li>Amazon Elasticsearch Service and Amazon Kinesis</li></ul>
Natural language processing	<ul style="list-style-type: none"><li>Amazon Comprehend</li></ul>
Machine Learning	<ul style="list-style-type: none"><li>Amazon SageMaker</li><li>Apache Spark → Amazon EMR</li></ul>

CLICK TO ADD TEXT

**AWS**  
**re:Invent**

**THANK YOU!**

CLICK TO ADD TEXT

**AWS**  
**re:Invent**

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

