



---

# Introduction to Database Services

Brian Rice

Product Marketing Manager, Amazon RDS



# Today's agenda

- Why managed database services?
- A non-relational managed database
- A relational managed database
- A managed in-memory cache
- A managed data warehouse
- What to do next



---

# Why managed database services?

# If you host your databases on-premises

App optimization

Scaling

High availability

Database backups

DB software patches

DB software installs

OS patches

OS installation

Server maintenance

Rack and stack

Power, HVAC, net



# If you host your databases on-premises

App optimization

Scaling

High availability

Database backups

DB software patches

DB software installs

OS patches

OS installation

Server maintenance

Rack and stack

Power, HVAC, net



# If you host your databases in Amazon EC2

App optimization

Scaling

High availability

Database backups

DB software patches

DB software installs

OS patches



you

OS installation

Server maintenance

Rack and stack

Power, HVAC, net



# If you host your databases in Amazon EC2

App optimization

Scaling

High availability

Database backups

DB software patches

DB software installs

OS patches



you

OS installation

Server maintenance

Rack and stack

Power, HVAC, net



# If you choose a managed DB service



you

App optimization

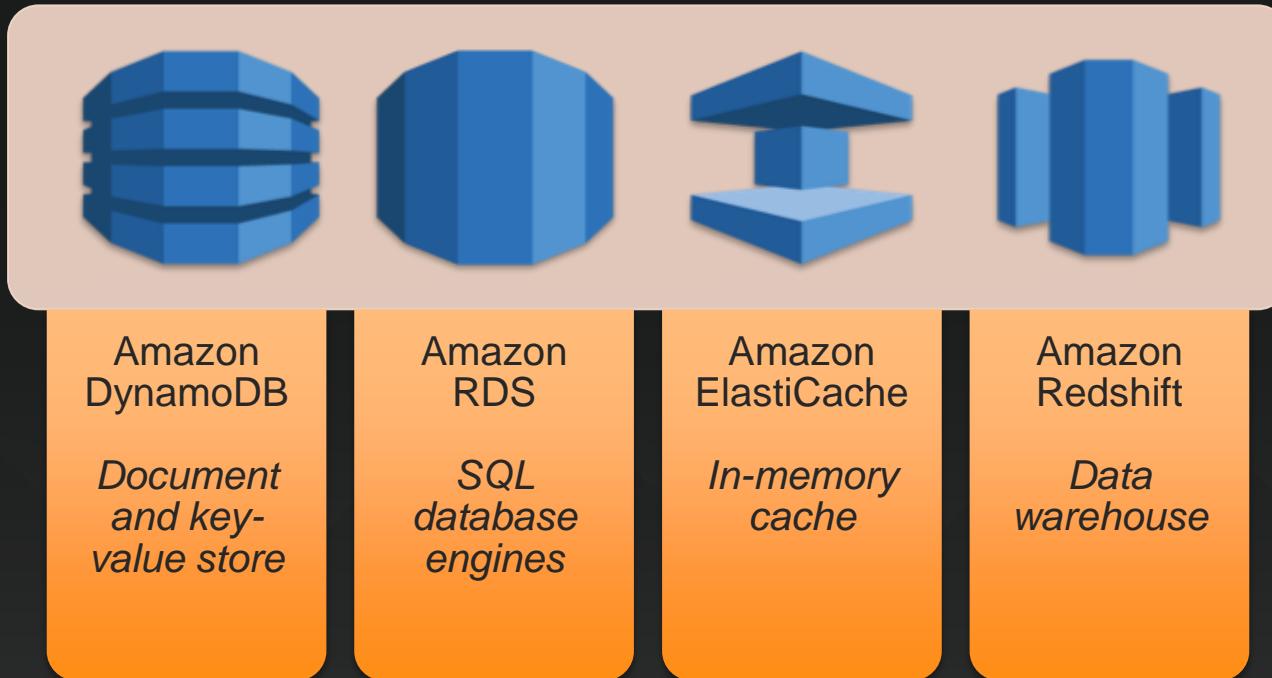
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net



# The self-managed vs. AWS-managed decision

Self-managed database	AWS-managed database
You have full responsibility for upgrades and backup	AWS provides upgrades, backup, and failover as a service
You have full responsibility for security	AWS provides high infrastructure security, certifications; gives you tools to ensure DB security
Full control over parameters of server, OS, and database	Database is a managed appliance, so you can easily automate
Replication is expensive and complex and requires a lot of engineering	AWS provides failover as a packaged service

# A managed service for each major DB type





---

# What is Amazon DynamoDB?

# Amazon DynamoDB: a managed document and key-value store

- Simple and fast to deploy
- Simple and fast to scale
  - To millions of IOPS
- Data is automatically replicated
- Fast, predictable performance
  - Backed by SSD storage
- Secondary indexes offer fast lookups
- No cost to get started; pay only for what you consume



Amazon DynamoDB

# Dropcam relies on Amazon DynamoDB

“By using DynamoDB, we reduced delivery time for video events to less than 50 milliseconds.”

*—Greg Nelson, VP of Software Engineering*

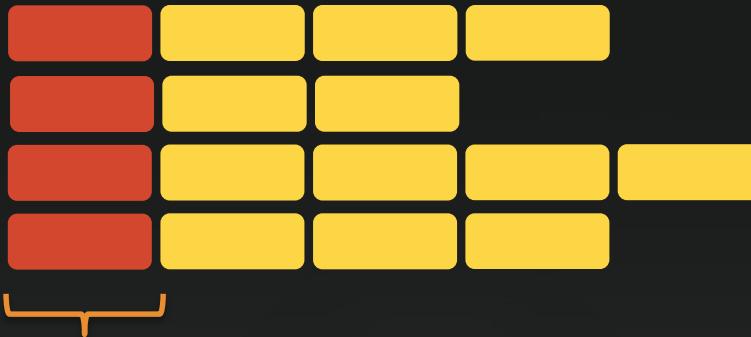


- Dropcam provides video monitoring hardware and software so customers can view high-definition video from iOS and Android devices or the Internet
- Dropcam’s developers report consistent and low latency with DynamoDB

# Amazon DynamoDB is a schemaless database

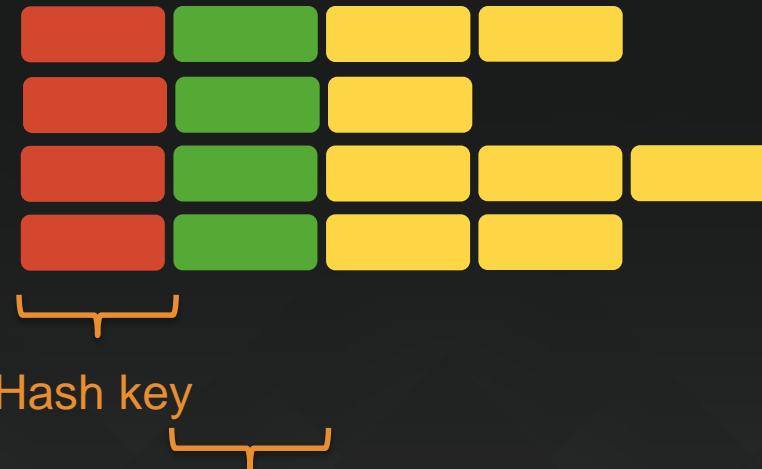


# Each item must include a key



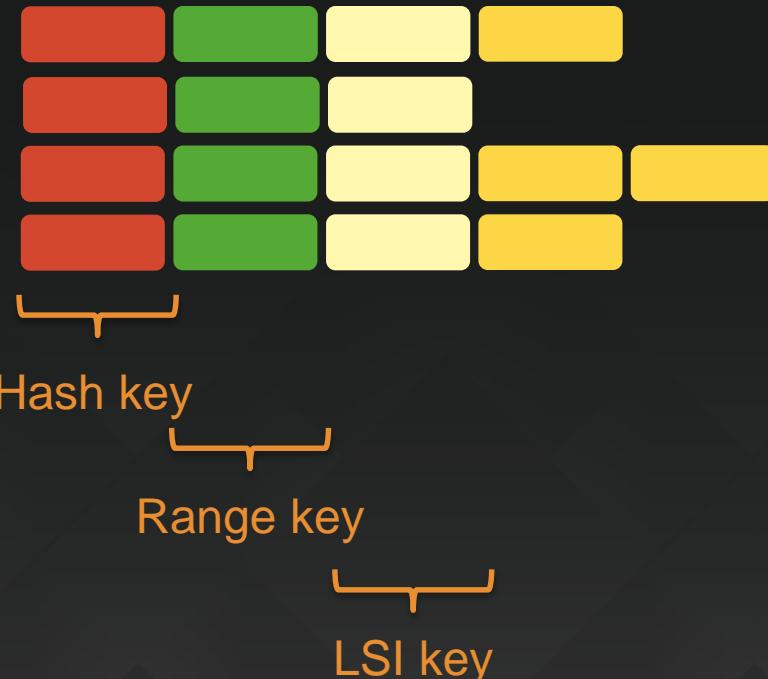
Hash key  
(DynamoDB maintains an  
unordered index)

# Each item must include a key

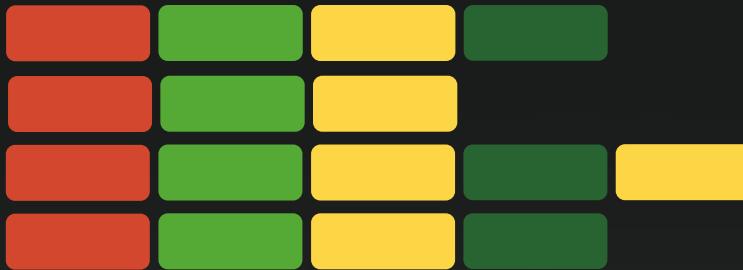


Hash key  
Range key  
(DynamoDB maintains a sorted index)

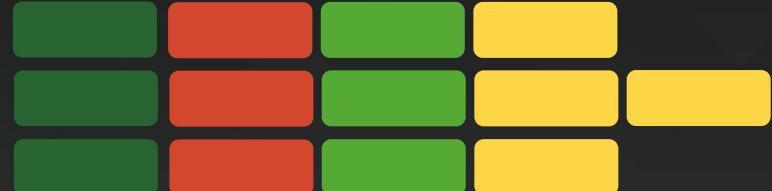
# Local secondary indexes = alternate range keys



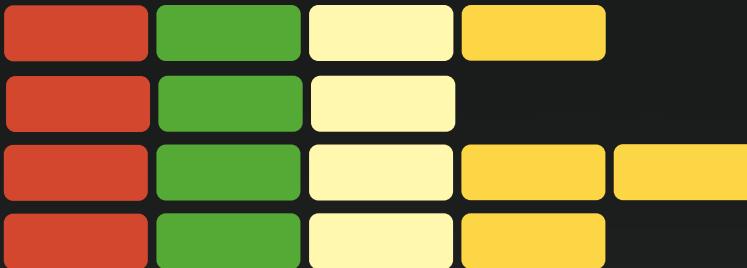
# Global secondary indexes = “pivot charts” for your table



Choose which  
attributes  
to project (if any)



# Amazon DynamoDB: provision throughput



21

Write  
capacity units

35

Read  
capacity units

# DynamoDB: What are capacity units?

1



1

One write capacity unit

Pay to bearer  
on demand  
**1 write per sec  
of up to 1KB**

1

1 Pay to bearer  
on demand  
**1 read per sec  
of up to 4KB**

1

*Eventually consistent reads at 50% off!*

1

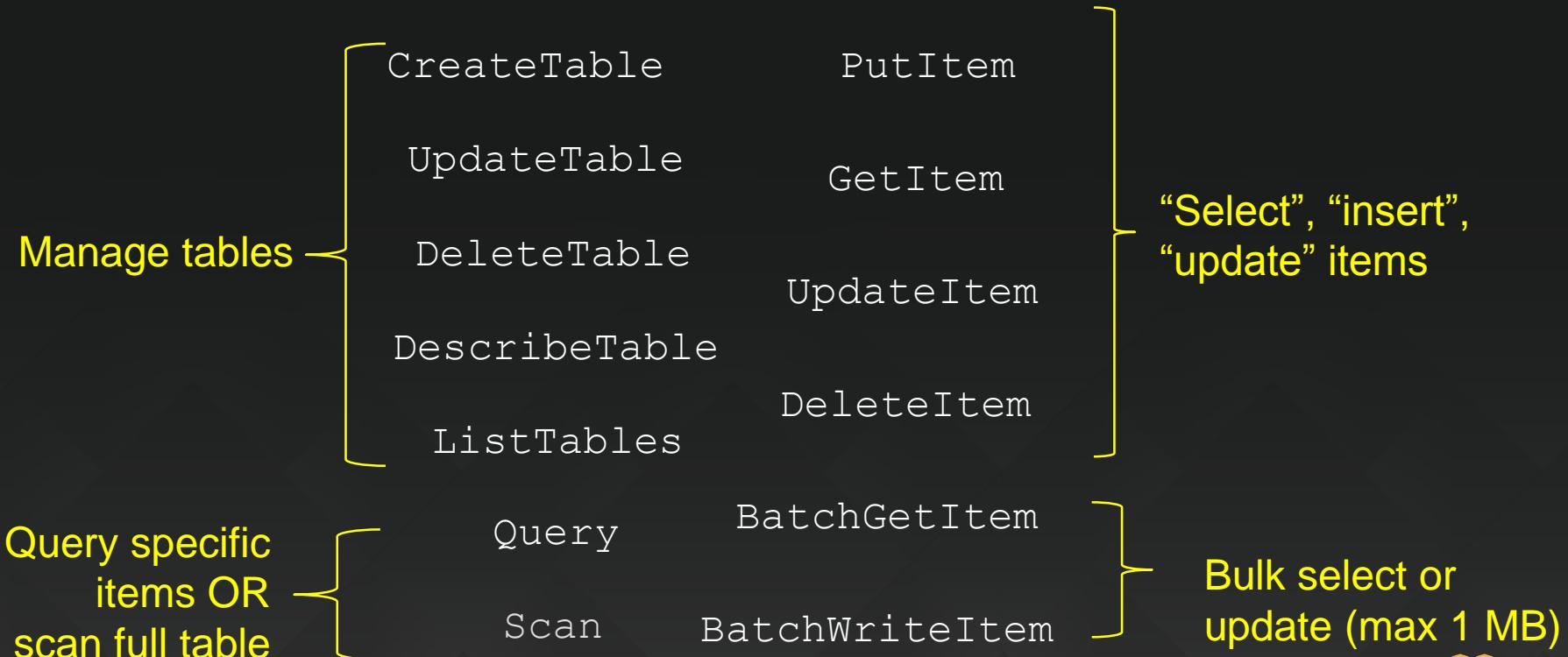


1

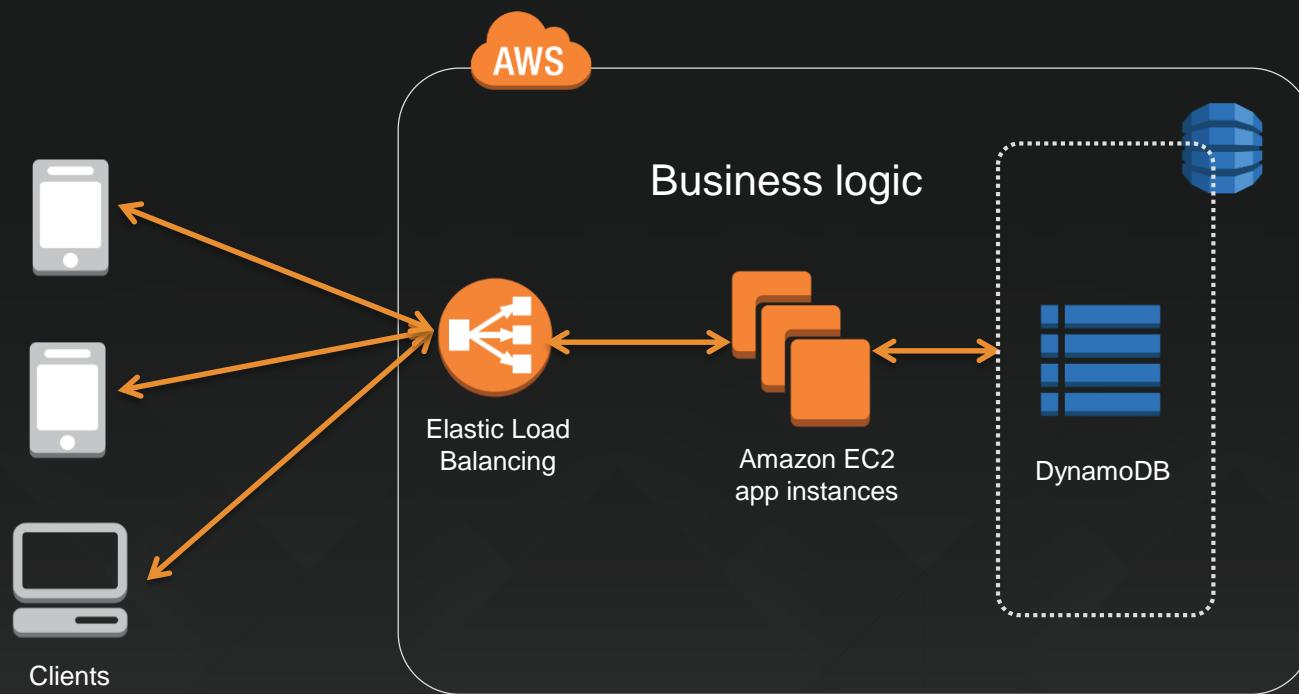
One read capacity unit



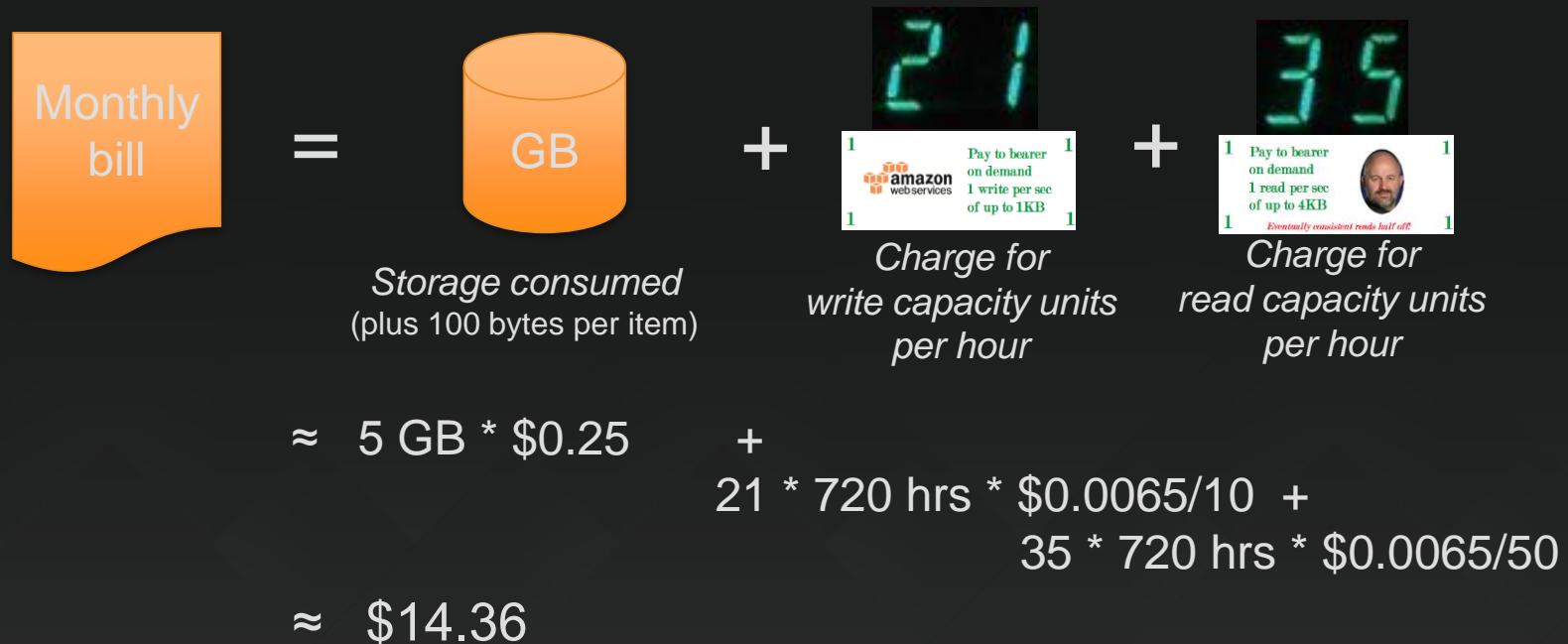
# DynamoDB is optimized for developer productivity



# Simple app architecture with Amazon DynamoDB

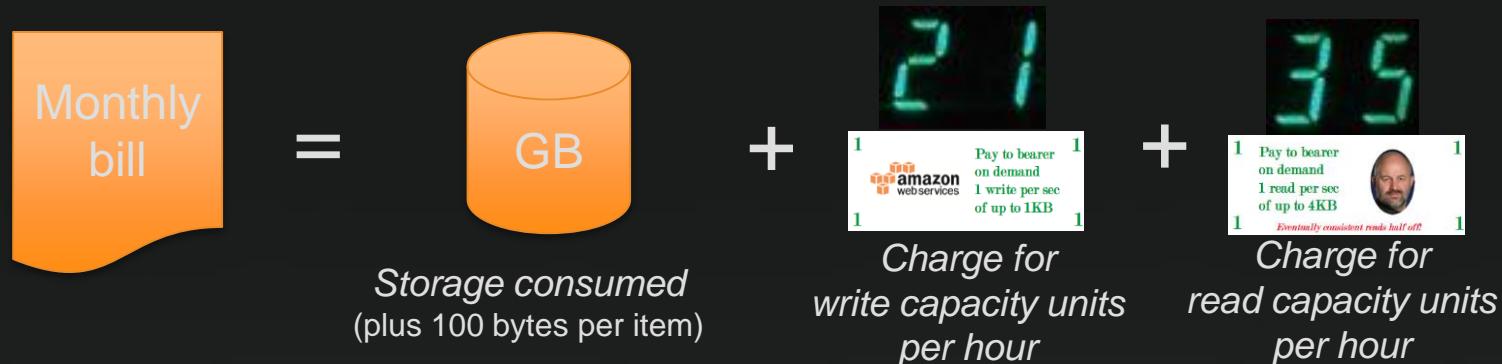


# How DynamoDB billing works



Assumes DB instance accessed only from AWS region  
Further details at <http://aws.amazon.com/dynamodb/pricing/>

# How DynamoDB billing works (with free tier)



$$\approx \text{5--25 GB} * \$0.25 + 21\text{--25} * 720 \text{ hrs} * \$0.0065/10 + 35\text{--25} * 720 \text{ hrs} * \$0.0065/50$$

Assumes DB instance accessed only from AWS region  
Further details at <http://aws.amazon.com/dynamodb/pricing/>

# How DynamoDB billing works (with free tier)





---

# What is Amazon RDS?

# Amazon RDS: a managed SQL service

- Simple and fast to deploy
- Simple and fast to scale
- AWS handles patching, backups, replication
- Compatible with your applications
  - Choose among Amazon Aurora, MySQL, PostgreSQL, Oracle, SQL Server
- Fast, predictable performance
- No cost to get started; pay only for what you consume



Amazon RDS

# Flipboard relies on Amazon RDS

"We were able to go from concept to delivered product in about six months with just a handful of engineers."  
—*Greg Scallan, Chief Architect*



- Flipboard is an online magazine with millions of users and billions of “flips” per month
- Uses Amazon RDS and its Multi-AZ capabilities to store mission critical user data

# How Amazon RDS delivers high performance

- Choose General Purpose (SSD) storage for most workloads
  - 3 IOPS per GB provisioned, with burst capability up to 3,000 IOPS
- Choose Provisioned IOPS (SSD) storage for high, predictable performance
  - Provision up to 3 TB storage and 30 K IOPS per instance
  - Scale IOPS up or down online
- Choose a database instance type with the right amount of CPU and memory

# How Amazon RDS backups work

- Automated backups
  - Restore your database to a point in time
  - Enabled by default
  - Choose a retention period, up to 35 days
- Manual snapshots
  - Initiated by you
  - Persist until you delete them
  - Stored in Amazon S3
  - Build a new database instance from a snapshot when needed

# Choose Multi-AZ for greater availability, durability

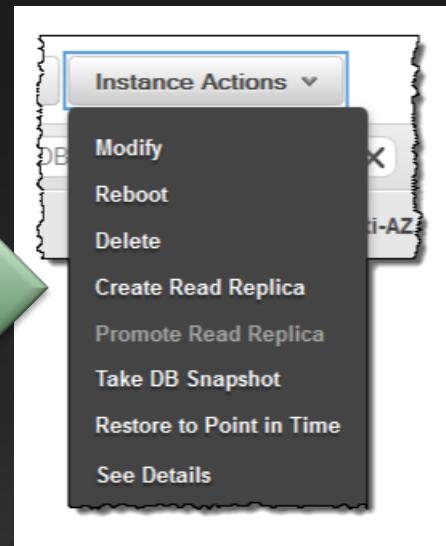
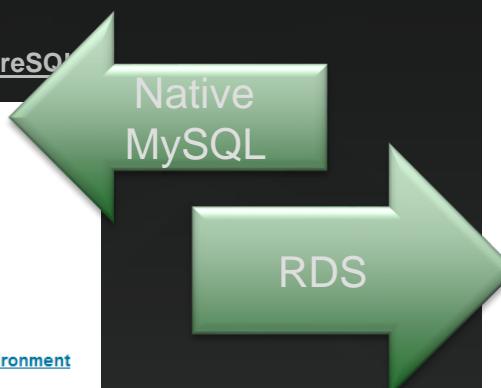
- An *Availability Zone* is a physically distinct, independent infrastructure
- With Multi-AZ operation, your database is synchronously replicated to another zone in the same AWS region
- Failover occurs automatically in response to the most important failure scenarios
- Planned maintenance is applied first to backup

# Choose Read Replicas for greater scalability

- Offload read traffic to an automatically maintained Read Replica
- Create multiple Read Replicas, load-share traffic
- Easy to set up

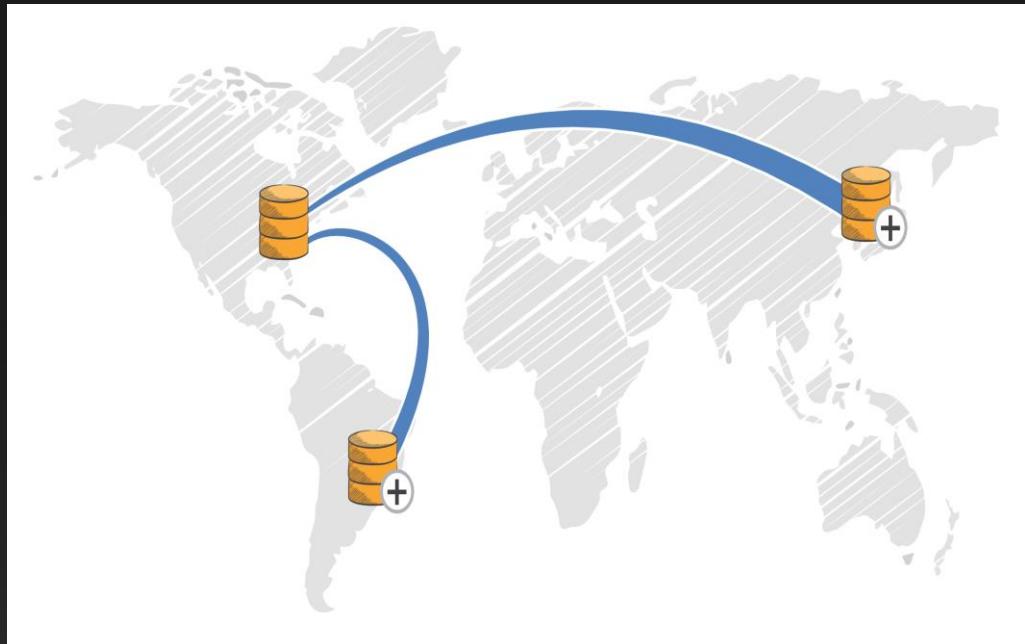
- Available in Amazon RDS for MySQL and PostgreSQL

[16.1.1.1. Setting the Replication Master Configuration](#)  
[16.1.1.2. Setting the Replication Slave Configuration](#)  
[16.1.1.3. Creating a User for Replication](#)  
[16.1.1.4. Obtaining the Replication Master Binary Log Coordinates](#)  
[16.1.1.5. Creating a Data Snapshot Using mysqldump](#)  
[16.1.1.6. Creating a Data Snapshot Using Raw Data Files](#)  
[16.1.1.7. Setting Up Replication with New Master and Slaves](#)  
[16.1.1.8. Setting Up Replication with Existing Data](#)  
[16.1.1.9. Introducing Additional Slaves to an Existing Replication Environment](#)  
[16.1.1.10. Setting the Master Configuration on the Slave](#)

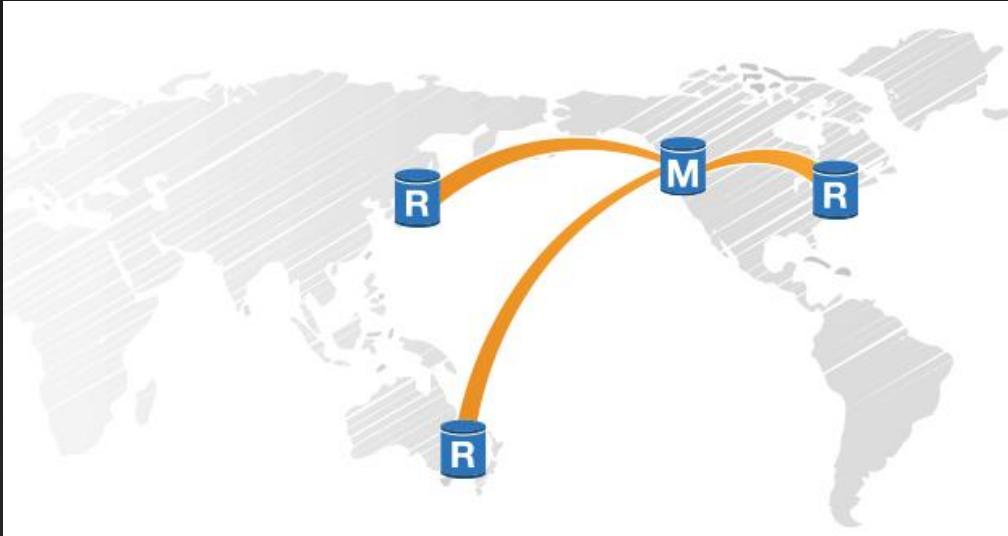


# Choose cross-region snapshot copy for even greater durability, ease of migration

- Copy a database snapshot to a different AWS region
- Warm standby for disaster recovery
- Or use it as a base for migration to a different region



# Choose cross-region Read Replicas for enhanced data locality, even more ease of migration



- Even faster recovery in the event of disaster
- Bring data close to your customers
- Promote to a master for easy migration

# How to scale with Amazon RDS

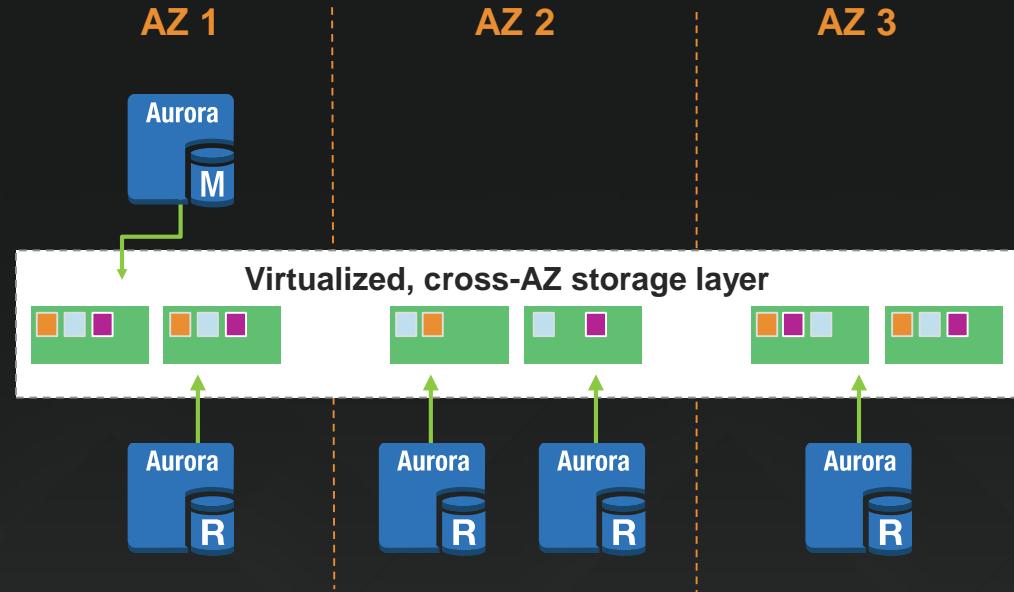
- Scale up or down with resizable instance types
- Scale your storage up with a few clicks while online
- Offload read traffic to read replicas
- Put a cache in front of Amazon RDS
  - Amazon ElastiCache for Memcached or Redis
  - Or your favorite cache, self-managed in Amazon EC2
- Amazon RDS takes some of the pain out of sharding

# Now in preview: Amazon RDS for Aurora

- Amazon Aurora: the relational database reinvented for the cloud
  - Up to five times better performance than MySQL**
  - At a price point 1/10 of a commercial database**
  - Designed for drop-in compatibility with MySQL 5.6**
- Pay only for the storage you use
- Runs in Amazon VPC; offers encryption at rest and in transit
- Amazon RDS handles administrative tasks for Aurora

# Amazon Aurora: high availability by default

- Your data is replicated 6 ways across 3 AZs
- Storage grows up to 64 TB seamlessly
- Up to 15 Aurora Replicas with instant crash recovery



# NoSQL vs. SQL for a new app: How to choose?

- Want simplest possible DB management?
- Want app to manage DB integrity?



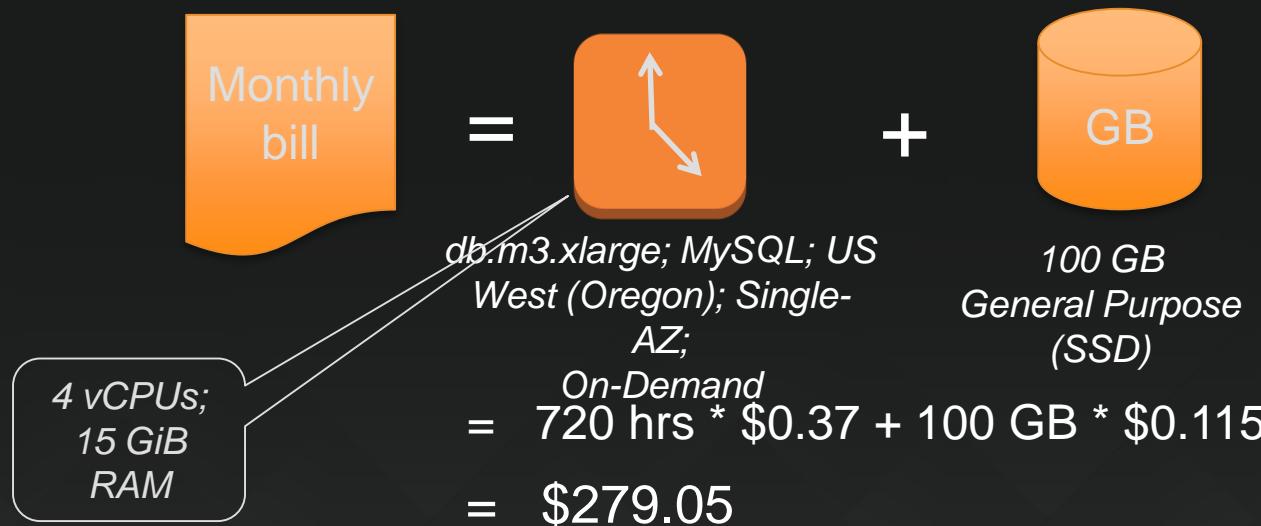
Amazon DynamoDB

- Need joins, transactions, frequent table scans?
- Want DB engine to manage DB integrity?
- Team has SQL skills?



Amazon RDS

# How Amazon RDS billing works



Assumes DB instance accessed only from Amazon EC2  
Further details at <http://aws.amazon.com/rds/pricing/>



---

# What is Amazon ElastiCache?

# Amazon ElastiCache: resizable in-memory cache

- High performance, resizable in-memory caching
- Speed your application by bypassing database access and disk storage
- Compatible with your existing applications
  - Choose between the popular **memcached** and **Redis** engines



redis



# 2U relies on Amazon ElastiCache



“ElastiCache helps us specifically a lot around our social and collaborative tools....It just works. We don’t even know it’s there.”

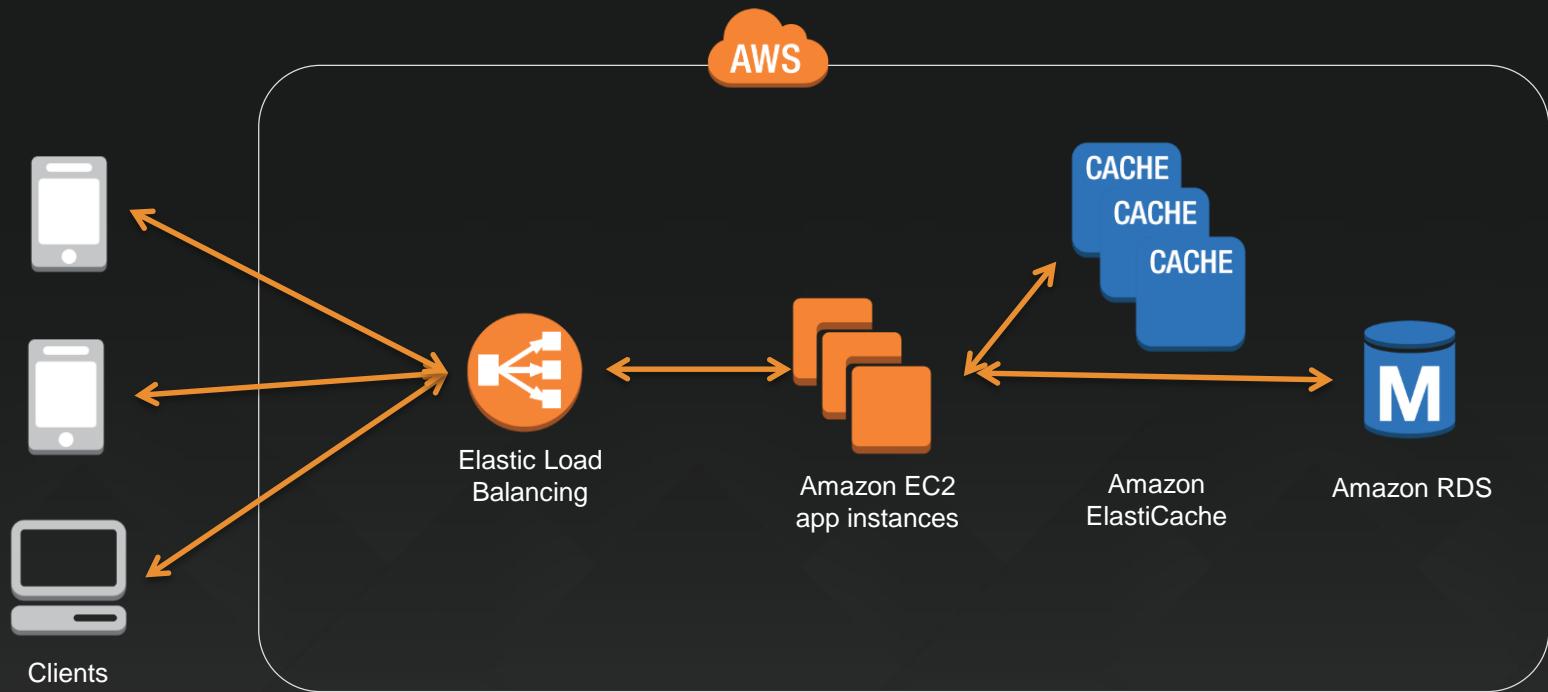
—James Kenigsberg  
Chief Technology Officer

- 2U, Inc., is a “School as a Service” provider that helps universities take their degrees online.
- To support collaboration and learning, the company’s technology platform uses ElastiCache to cache data that grows exponentially as students communicate with instructors and with each other.
- ElastiCache is used to cache news feeds and data from RDS MySQL.

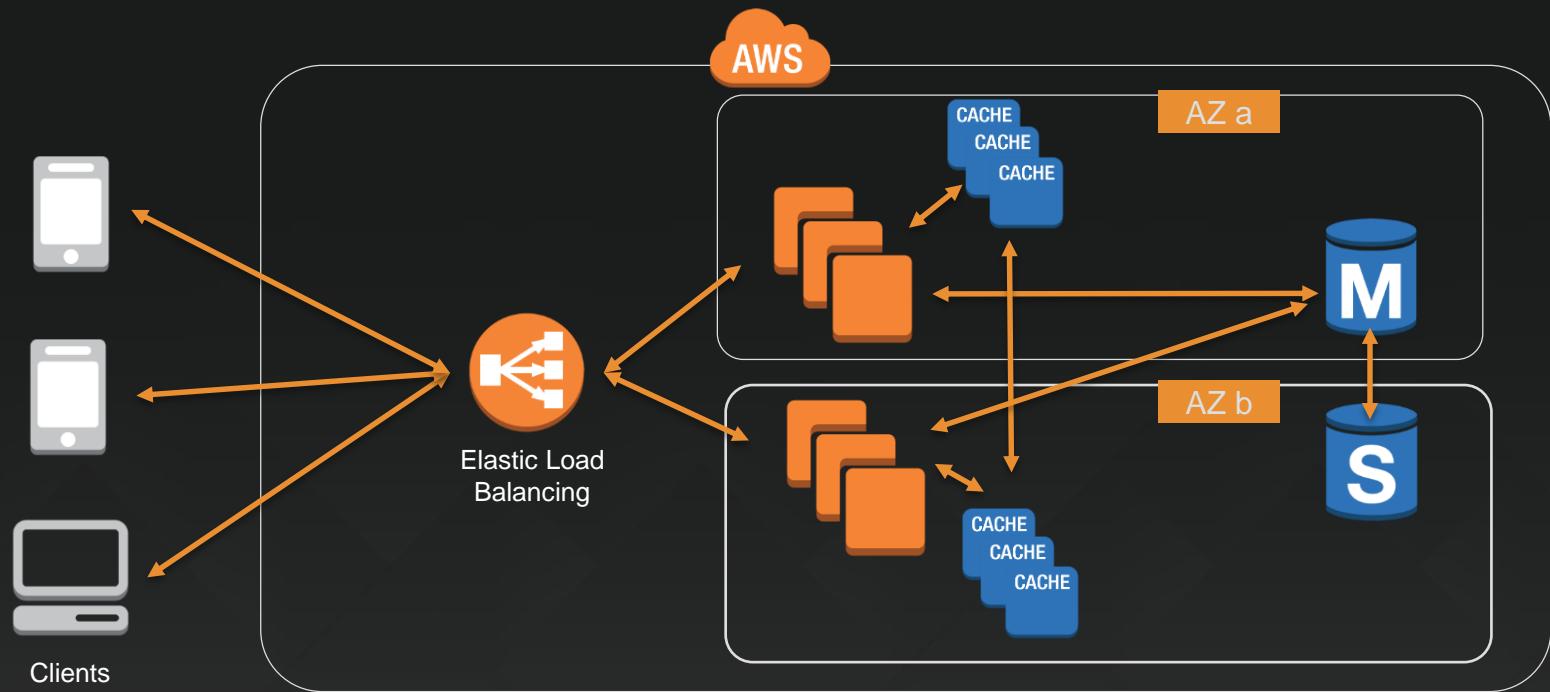
# Use cases for Amazon ElastiCache

- Performance or cost optimization of an underlying database
- Storage of ephemeral key-value data
- High-performance application patterns

# Amazon ElastiCache: simple app architecture



# Amazon ElastiCache: resilient app architecture



# How Amazon ElastiCache billing works

Monthly bill

$$= N \times$$



*m3.large;  
US West (Oregon);  
on demand*

$$= 4 \text{ nodes} * 720 \text{ hrs} * \$0.182$$

$$= \$524.16$$

2 vCPUs;  
7.1 GB  
RAM

*Further details at <http://aws.amazon.com/elasticsearch/pricing/>*



---

# What is Amazon Redshift?

# Amazon Redshift: a managed data warehouse

- Petabyte-scale columnar database
- Fast response time
  - ~10x that of typical relational stores
- Pricing as low as \$1,000 per TB per year



Amazon Redshift

# Foursquare relies on Amazon Redshift

“Amazon Redshift offers the performance we needed while freeing us from the licensing costs of our previous solution.”

—Jon Hoffman  
Software Engineer



- More than 40 million people worldwide use Foursquare to meet up with friends, exchange travel tips, and find money-saving deals
- Foursquare uses AWS to perform analytics across millions of daily check-ins, saving licensing fees and redeploying its dev/ops staff on more strategic work

# Who uses Amazon Redshift?



## Traditional enterprise DW

- Reduce costs by extending DW rather than adding HW
- Migrate completely from existing DW systems
- Respond faster to business; provision in minutes



## Companies with big data

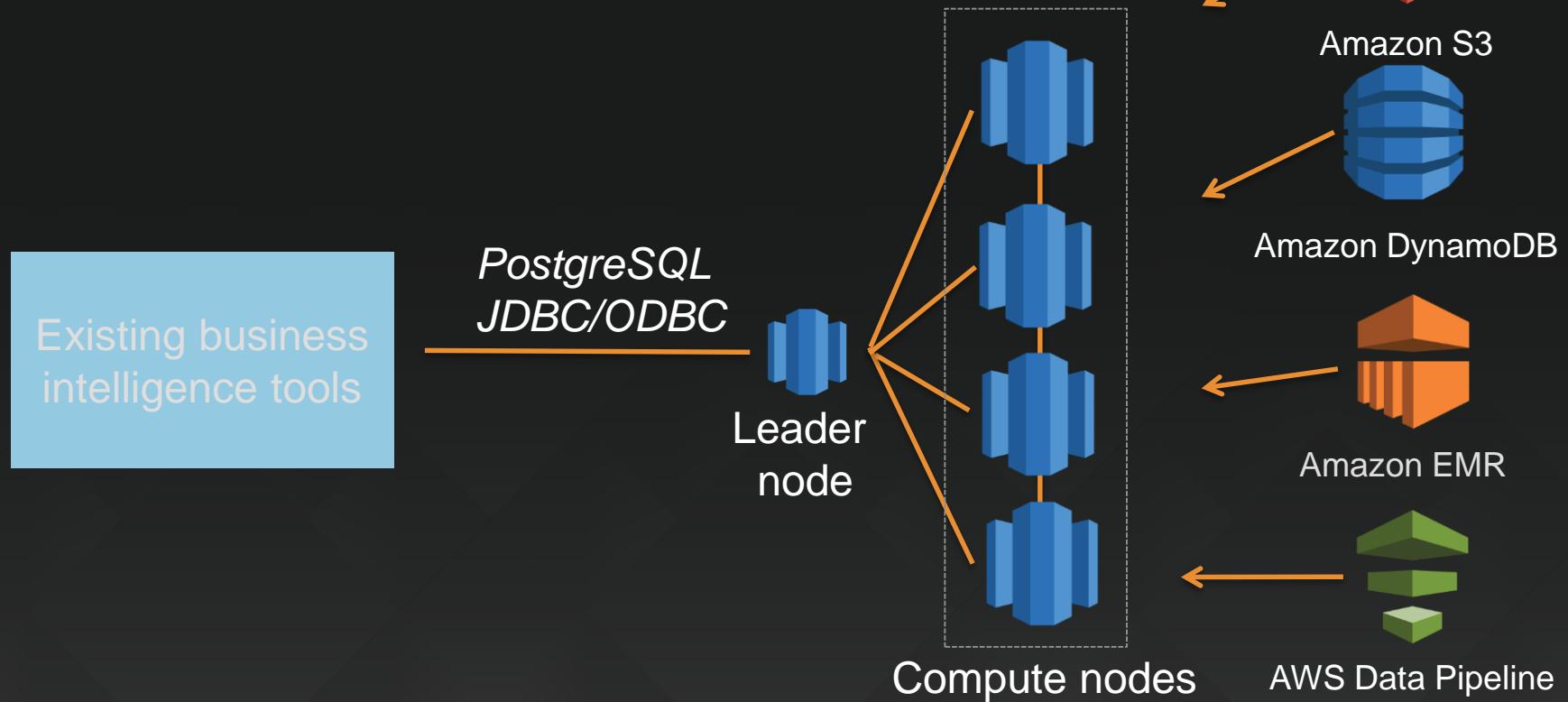
- Improve performance by an order of magnitude
- Make more data available for analysis
- Access business data via standard reporting tools



## SaaS companies

- Add analytic functionality to applications
- Scale DW capacity as demand grows
- Reduce HW and SW costs by an order of magnitude

# Amazon Redshift architecture



# Amazon Redshift dramatically reduces I/O

- **Column storage**

- Data compression
- Zone maps
- Direct-attached storage

ID	Age	State	Amount
123	20	CA	500
345	25	WA	250
678	40	FL	125
957	37	WA	375

- With row storage, you do unnecessary I/O
- To get total amount, you have to read everything



# Amazon Redshift dramatically reduces I/O

- Column storage
- Data compression
- Zone maps
- Direct-attached storage

ID	Age	State	Amount
123	20	CA	500
345	25	WA	250
678	40	FL	125
957	37	WA	375

- With column storage, you only read the data you need

# Amazon Redshift dramatically reduces I/O

- Column storage

- **Data compression**

- Zone maps
- Direct-attached storage

```
analyze compression listing;

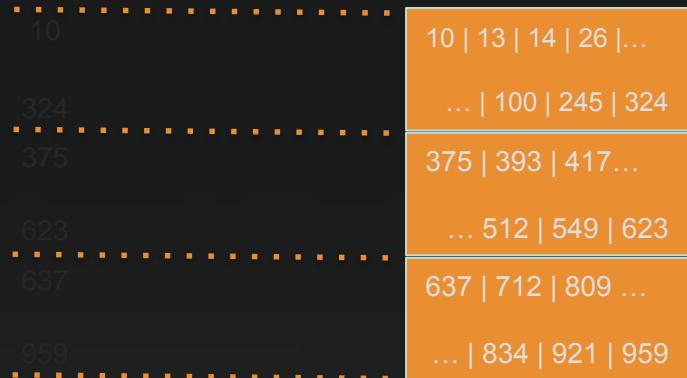
Table | Column      | Encoding
-----+-----+-----+
listing | listid       | delta
listing | sellerid     | delta32k
listing | eventid      | delta32k
listing | dateid       | bytedict
listing | numtickets    | bytedict
listing | priceperticket | delta32k
listing | totalprice    | mostly32
listing | listtime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

- COPY compresses automatically
- You can analyze and override
- More performance, less cost



# Amazon Redshift dramatically reduces I/O

- Column storage
- Data compression
- **Zone maps**
- Direct-attached storage



- Track the minimum and maximum value for each block
- Skip over blocks that don't contain relevant data

# Amazon Redshift dramatically reduces I/O

- Column storage
- Data compression
- Zone maps
- **Direct-attached storage**

DW.HS1.XL:

16 GB RAM
2 cores
2 TB disk

DW.HS1.8XL:

128 GB RAM
16 cores
16 TB disk

- > 2 GB/s scan rate
- Optimized for data processing
- High disk density

# Amazon Redshift: start small and grow big

## Dense Storage Node (dw1.xlarge)

2 TB, 16 GB RAM, 2 cores

## Dense Compute Node (dw2.large)

0.16 TB, 16 GB RAM, 2 cores

## Single Node (2 TB)

XL

## **Cluster 2-32 Nodes (up to 64 TB)**

XL							
XL							
XL							
XL							

Note: Nodes not to scale

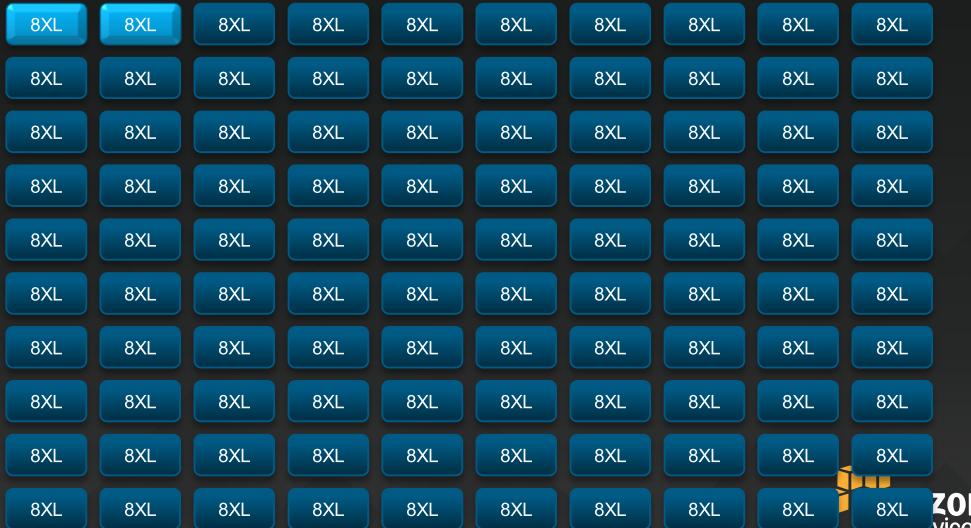
## **8XL Dense Storage Node (dw1.8xlarge)**

16 TB, 128 GB RAM, 16 cores, 10 GigE

# 8XL Dense Compute Node (dw2.8xlarge)

2.56 TB, 128 GB RAM, 16 cores, 10 GigE

## **Cluster 2-100 Nodes (up to 1.6 PB)**



# How Amazon Redshift billing works

$$\text{Monthly bill} = N \times \begin{matrix} \nearrow \\ \searrow \end{matrix}$$

*dw2.large; US West (Oregon);  
on demand*

$$\begin{aligned} &= 4 \text{ nodes} * 720 \text{ hrs} * \$0.25 \\ &= \$720 \end{aligned}$$

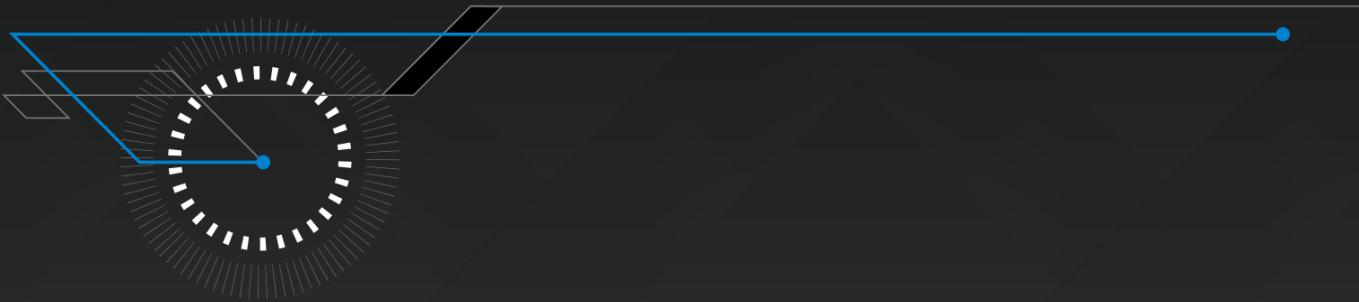
*2 vCPUs;  
15 GiB RAM;  
0.16 TB SSD*

*Further details* <http://aws.amazon.com/rds/pricing/>

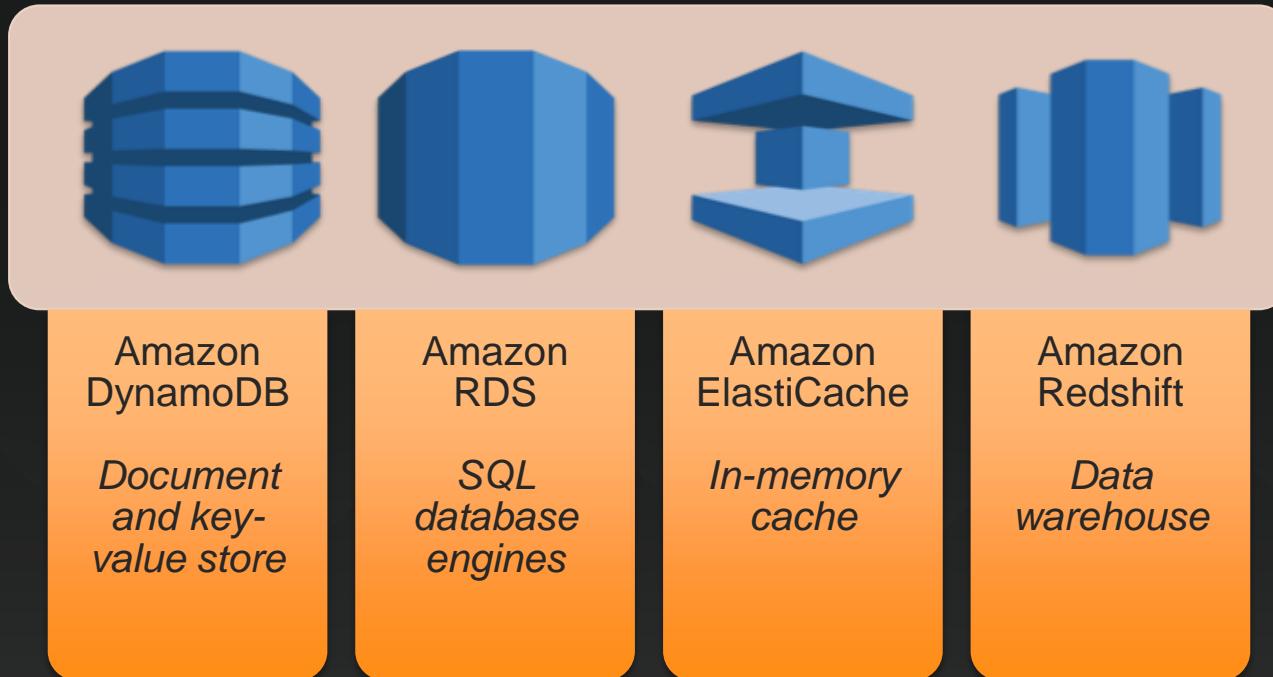
# Try Amazon Redshift with BI and ETL for free!

- <http://aws.amazon.com/redshift/free-trial>
- 2 months, 750 hours/month to try our dw2.large SSD instance with 160 GB of compressed storage per node
- Also try BI and ETL for free from partners at  
<http://aws.amazon.com/redshift/partners/>

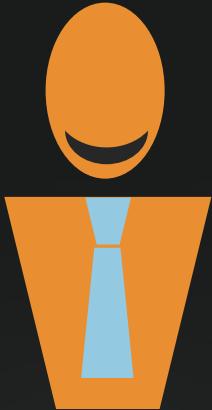
# To sum up...



# Review: AWS managed database services



# Benefits of AWS database services

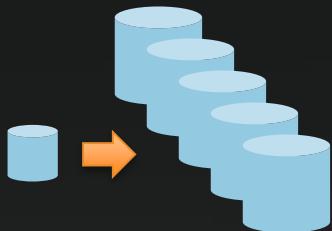


## Managed services

*AWS handles installs,  
patching, restarts*

## Easy to scale

*Grow as you need*



## Pay only for what you use

*No up-front cost*



## Designed for use with other AWS services



# AWS Marketplace

- Find software to use with Amazon RDS, Amazon Redshift, Amazon DynamoDB, and Amazon ElastiCache
- One-click deployments
- Flexible pricing options

The screenshot shows the AWS Marketplace homepage with a search bar at the top. The search term 'Amazon Redshift' has been entered. On the left, there's a sidebar with categories like All Categories, Business Software, Business Intelligence, and Amazon Redshift. Below that are filters for Operating System (Windows releases: All, Linux/UNIX distributions: All). The main content area displays the results for 'Amazon Redshift' with 9 results. Each result includes the provider logo, product name, rating, version, price, and a brief description. For example, 'Jaspersoft Reporting and Analytics for AWS (Hourly)' is listed with a price of '\$0.40 to \$5.54/hr for software'. Other products shown include MicroStrategy, ATTUNITY CloudBeam, and birst.

Product	Provider	Rating	Version	Price	Description
Jaspersoft Reporting and Analytics for AWS (Hourly)	Jaspersoft	★★★★★ (11)	Version 5.5.0	\$0.40 to \$5.54/hr for software	Sold by Jaspersoft. Jaspersoft for AWS is a commercial open source reporting and analytics server built for AWS that can run standalone or be embedded in your application. It is priced very ...
MicroStrategy Free MicroStrategy Suite	MicroStrategy	★★★★★ (3)	Version 9.3.1	\$0.00/hr for software + AWS usage fees	Sold by MicroStrategy. MicroStrategy Suite is a powerful free Mobile and Business Intelligence solution that gives

<http://aws.amazon.com/marketplace>



# Try AWS database services for free

Service	Free every month	Term
Amazon DynamoDB	25 GB of storage 25 units of write capacity 25 units of read capacity	Perpetual
Amazon ElastiCache	750 micro cache node instance hours	First 12 months of AWS account
Amazon RDS	750 micro DB instance hours 20 GB of DB storage 20 GB for backups 10 million I/O operations	First 12 months of AWS account



Thank You  
SAN FRANCISCO