AWS re:Invent

DAT308

Replicate & Manage Data Using Managed Databases & Serverless Technologies

Prahlad Rao
Solutions Architect
Amazon Web Services

Shree Kenghe
Solutions Architect
Amazon Web Services

Rob Robinson
Senior Manager
Solutions Architecture
Amazon Web Services

AWS re:Invent

aws

Agenda

- Workshop intro and outcomes
- Use case and proposed solution
- Lab setup and environment
- AWS service introductions—AWS Database Migration Service (AWS DMS), Amazon Aurora Serverless, AWS Glue, Amazon Athena, Amazon QuickSight
- Hands-on lab

Workshop details

- Workshop duration
  - Proposed solution and AWS services presentation—20 minutes
  - Hands-on workshop—1.5 hours

- Workshop team
  - Prahlad Rao, AWS Solutions Architect
  - Shree Kenghe, AWS Solutions Architect
  - Sarah Sleyman, AWS Solutions Architect
  - Rob Robinson, Sr. Manager, AWS Solutions Architecture
  - Greg Powell, AWS Solutions Architect
  - Greg Pryzby, AWS Solutions Architect
  - Ankita Verma, Sr. Product Manager DMS
  - Ryan Bachman, Sr. Technical Account Manager

- Requirements and expectations
  - Students use their own AWS accounts to run the lab with AWS Identity and Access Management (IAM) admin permissions
  - Basic knowledge of AWS services—Amazon Aurora, Amazon Simple Storage Service (Amazon S3), AWS DMS, AWS Schema Conversion Tool (AWS SCT), Amazon Athena, Amazon QuickSight, AWS CloudFormation, AWS Glue)
  - Comfortable working on the AWS Management Console and configuring AWS services
  - Working knowledge of relational databases (SQL Server)

Expected outcomes

- Use AWS serverless technologies to replicate, analyze, and visualize data from relational and semi-structured datasets to gain insight and drive business outcomes
- Leverage AWS Schema Conversion Tool (AWS SCT) and AWS Database Migration Service (AWS DMS) to convert MS SQL server database to Amazon Aurora Serverless MySQL
- Use AWS Glue to build data catalog & transform data
- Use Amazon Athena to run ad-hoc SQL queries and interactively analyze data stored on Amazon S3 data store
- Use Amazon QuickSight to visualize data and drive additional insights (optional)

Breakout repeats

Monday, November 26

Replicate, Analyze, and Visualize Datasets Using AWS Database Migration Service and Serverless Big Data Technologies

4:45 p.m. PST | Bellagio

Tuesday, November 27

Replicate, Analyze, and Visualize Datasets Using AWS Database Migration Service and Serverless Big Data Technologies

11:30 a.m. PST | Mirage

Friday, November 30

Replicate, Analyze, and Visualize Datasets Using AWS Database Migration Service and Serverless Big Data Technologies

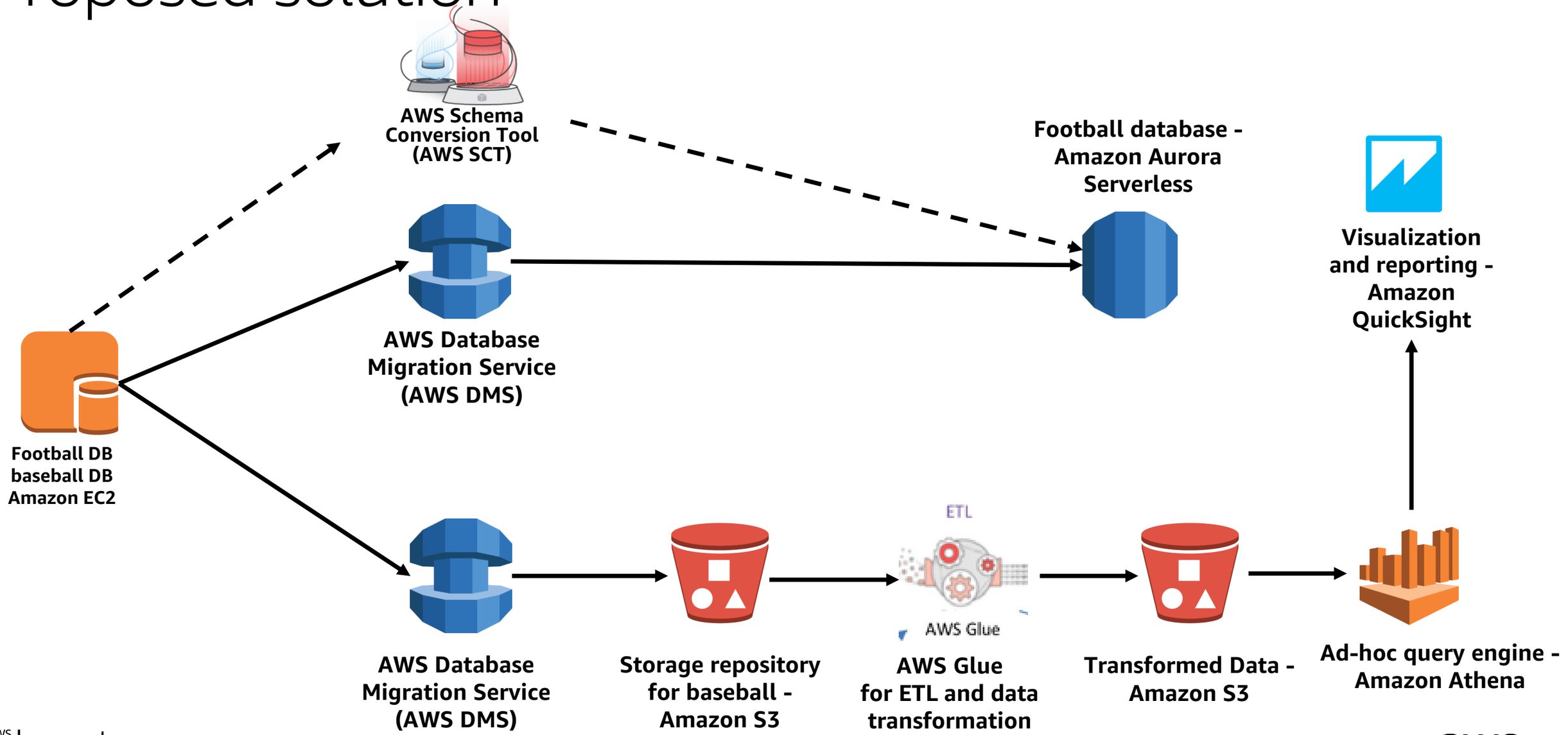
9:15 a.m. PST | Venetian

Use case and proposed solution

Use case

- The dataset—football and baseball data stored on a single source SQL server database
- Football is the current season: Data to be replicated to target serverless relational database on AWS for transactional and complex queries
- Baseball season has ended: Data replicated to target data store on Amazon S3 to transform data using AWS Glue and run ad-hoc queries using Amazon Athena
- Need for data analysis and visualization on target baseball dataset using Amazon QuickSight (optional lab)

Proposed solution



Solution steps

1. Launch AWS CloudFormation template to setup Amazon Virtual Private Cloud (Amazon VPC), source, and target database environment
2. Use AWS Schema Conversion Tool (AWS SCT) to convert schema from source SQL server to target Amazon Aurora Serverless MySQL for football DB
3. Use AWS Database Migration Service (AWS DMS) to replicate source SQL server data to target Amazon Aurora Serverless MySQL for football DB
4. Use AWS Database Migration Service (AWS DMS) to replicate source SQL server data to target Amazon S3 for baseball DB
5. Use AWS Glue to build data catalog and transform data to Parquet
6. Set up and configure Amazon Athena to run ad-hoc queries on baseball data stored on Amazon S3
7. Set up and configure Amazon QuickSight to build visualization for baseball data stored on Amazon Aurora Serverless MySQL

AWS service overview

AWS re:Invent

aws

AWS DMS and AWS SCT

AWS Database Migration Service (AWS DMS) easily and securely migrates and/or replicates your databases *and* data warehouses to AWS

AWS Schema Conversion Tool (AWS SCT) converts your commercial database and data warehouse schemas to open-source engines or AWS-native services, such as Amazon Aurora and Amazon Redshift

Migrated over 30,000 databases. And counting …

When to use AWS DMS and AWS SCT?

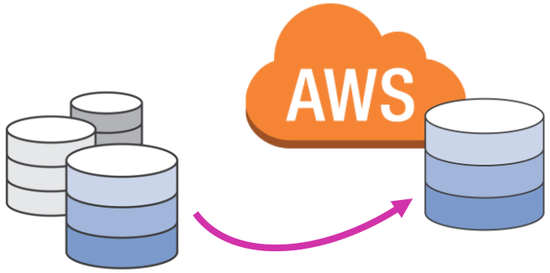
Modernize

Heterogeneous database migrations

Modernize your database tier –
- Commercial to open-source
- Commercial to Amazon Aurora

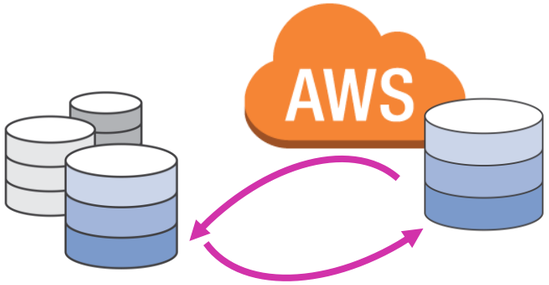
Modernize your data warehouse –
- Commercial to Amazon Redshift

Migrate



- Migrate business-critical applications
- Migrate from classic to Amazon VPC
- Migrate data warehouse to Amazon Redshift
- Upgrade to a minor version
- Consolidate shards into Aurora

Replicate



- Create cross-region read replicas
- Run your analytics in the cloud
- Keep your dev/test and production environment sync

Serverless computing + relational databases

What if you could have a relational database that:

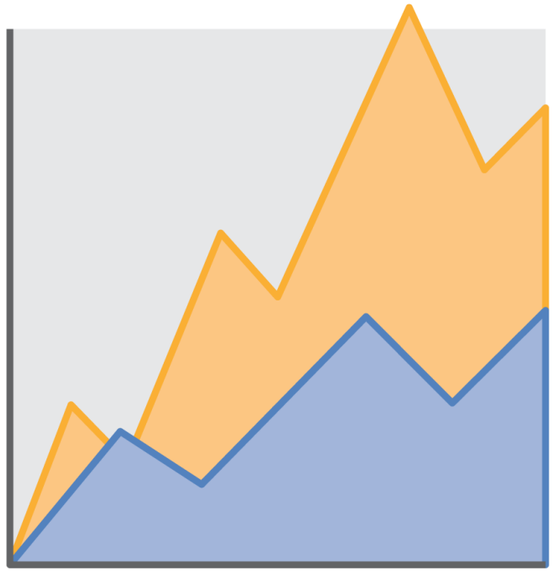
- Starts up on demand, shuts down when not in use?
- Seamlessly scales with no instances to manage?
- Is billed per second for the database capacity you use?
- Is compatible with existing database applications?

Introducing Amazon Aurora Serverless:

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

Aurora Serverless use cases

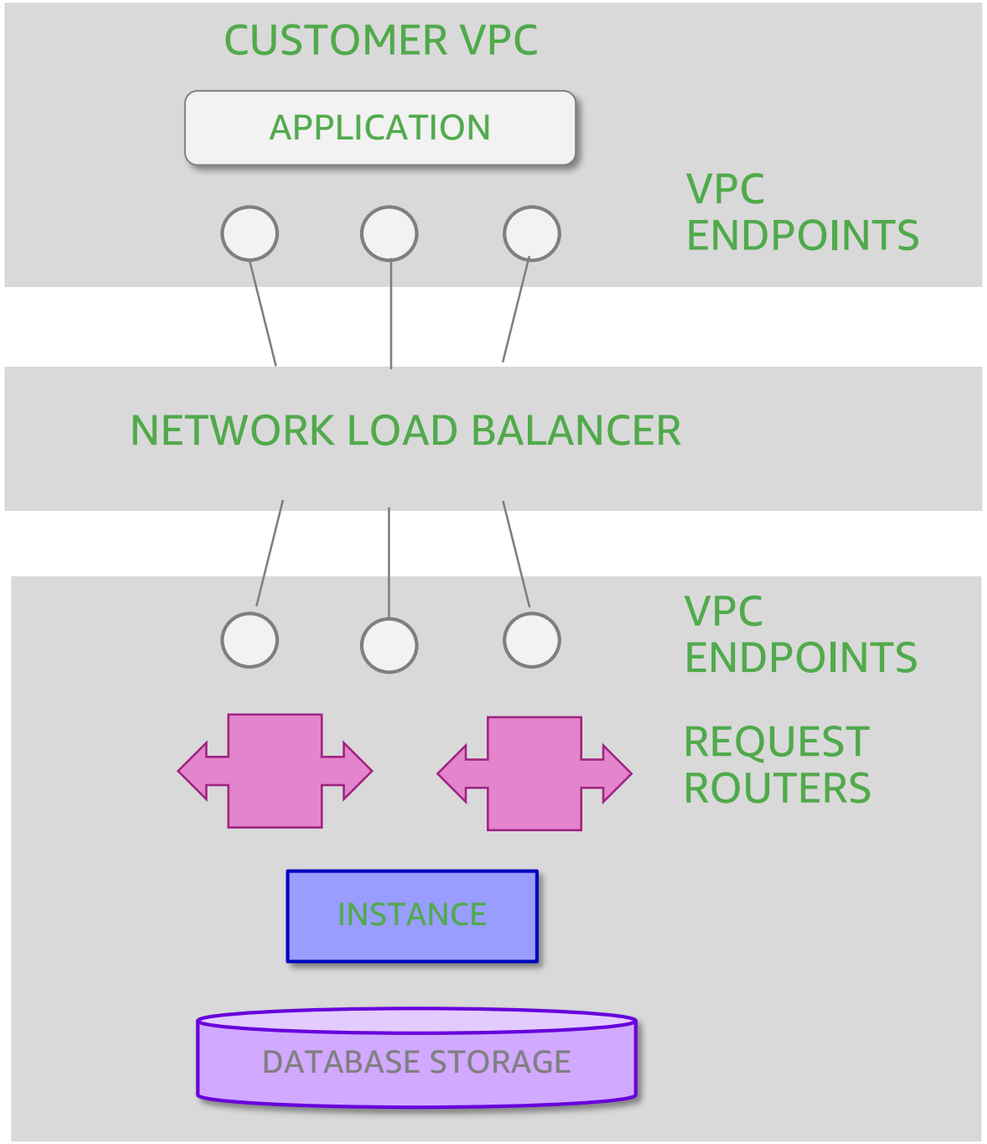
- Infrequently used applications (for example, low-volume blog site)
- Applications with variable load—peaks of activity that are hard to predict (for example, news site)
- Development or test databases not needed on nights or weekends



Aurora Serverless provisioning

When you provision a database, Aurora Serverless:

- Creates an Aurora storage volume
- Creates endpoint in your VPC for the application connection
- Configures network load balancer behind endpoint
- Configures request routers (multi-tenant) to route database traffic
- Provisions initial instance capacity



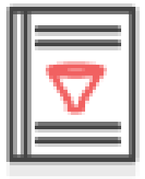
AWS Glue automates the undifferentiated heavy lifting of ETL

Discover Automatically discover and categorize your data making it immediately searchable and queryable across data sources

Develop Generate code to clean, enrich, and reliably move data between various data sources; you can also use their favorite tools to build ETL jobs

Deploy Run your jobs on a serverless, fully managed, scale-out environment. No compute resources to provision or manage.

AWS Glue: Components



Data catalog

- Hive metastore compatible with enhanced functionality
- Crawlers automatically extracts metadata and creates tables
- Integrated with Amazon Athena, Amazon Redshift Spectrum

Job authoring

- Auto-generates ETL code
- Build on open frameworks – Python and Spark
- Developer-centric – editing, debugging, sharing

Job execution

- Run jobs on a serverless Spark platform
- Provides flexible scheduling
- Handles dependency resolution, monitoring, and alerting

Amazon Athena

Amazon Athena is an interactive query service that makes it easy to analyze data directly from Amazon S3 using Standard SQL

Query data directly from Amazon S3

- No loading of data
- Query data in its raw format
  - Athena supports multiple data formats
  - Text, CSV, TSV, JSON, weblogs, AWS service logs
  - Convert to an optimized form like **ORC** or **Parquet** for the **performance and cost**
- No ETL required
- Stream data directly from Amazon S3
- Take advantage of Amazon S3 durability and availability

Amazon QuickSight is a business analytics service that lets business users quickly and easily visualize, explore, and share insights from their data.

aws re:Invent

aws

Deep integration with AWS data sources

Amazon QuickSight is **deeply integrated** with AWS data sources like Amazon Redshift, Amazon Relational Database Service (Amazon RDS), Amazon S3, Athena and others, as well as third-party sources like Excel, Salesforce, as well as on-premises databases

Lab setup and environment

AWS re:Invent

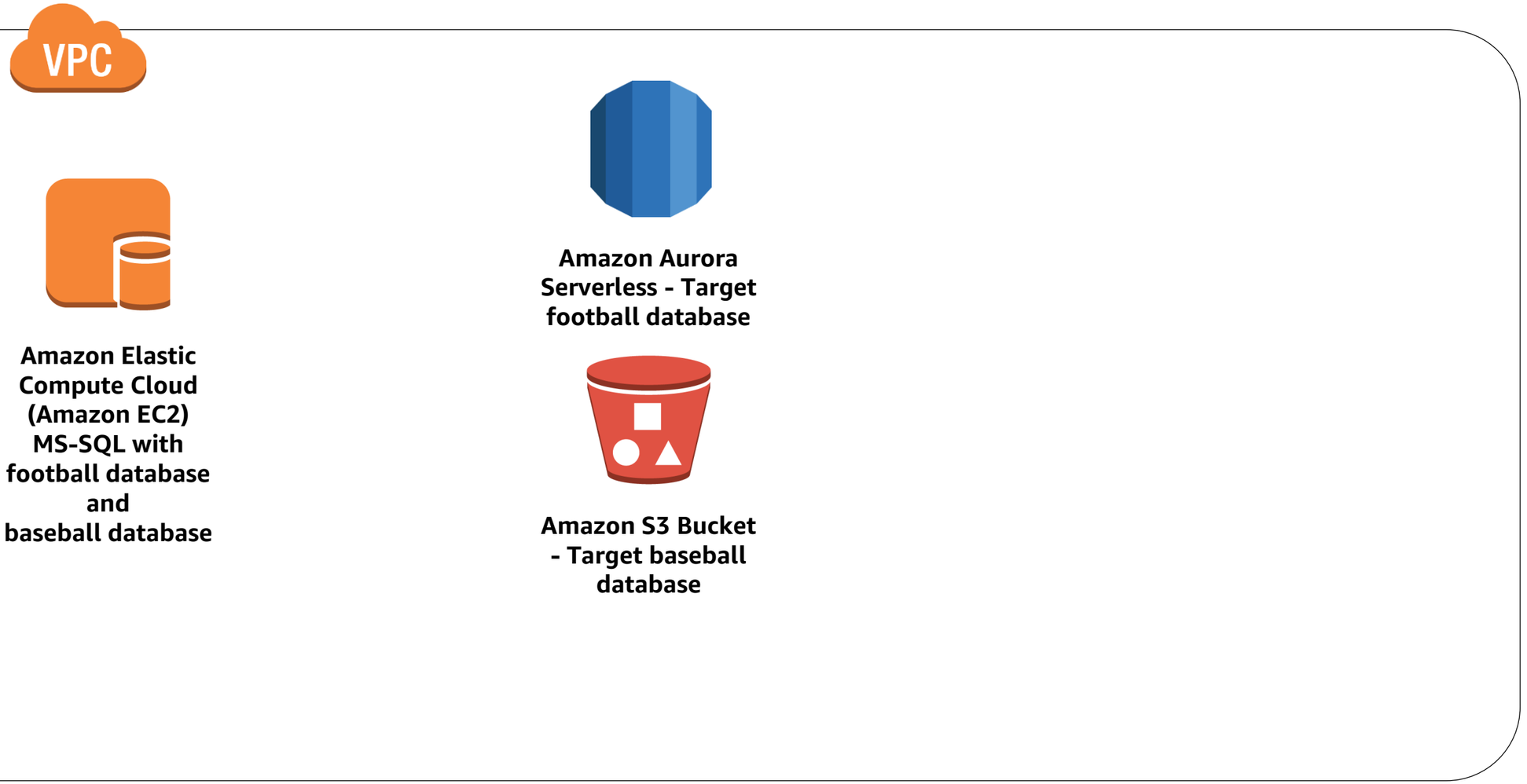
aws

Hands-on lab overview

- AWS Region for the lab—eu-west-1 (Ireland)
- Download zip file from this location - http://bit.ly/2RCppPx
- Follow instructions in lab guide
    - **Lab guide - Begin from here**
    - Exercise 1 - Region verification, launch CF stack, verify VPC, source and target database, Amazon S3 bucket
    - Exercise 2 - Source SQL server configuration, AWS SCT schema conversion steps
    - Exercise 3 - AWS DMS configuration and data migration to Amazon S3 and Amazon Aurora Serverless
    - Exercise 4 - AWS Glue to build data catalog and transform data to Parquet
    - Exercise 5 - Amazon Athena configuration and queries
    - Exercise 6 - Amazon QuickSight configuration and visualization (optional lab)

- Zip file extract contains
    - AWS CloudFormation template
    - ConfigureSQLServer.sql—source SQL Server configuration script
    - AthenaQueries—queries for Athena text file
    - Lab guide and exercises

AWS CloudFormation launch environment



Thank you!

Prahlad Rao
Solutions Architect
Amazon Web Services

Shree Kenghe
Solutions Architect
Amazon Web Services

Rob Robinson
Senior Manager
Solutions Architecture
Amazon Web Services

AWS re:Invent

aws

Please complete the session survey in the mobile app.

AWS re:Invent

aws