AWS re:INVENT

ABD324 - Migrating Your Oracle Data Warehouse to Amazon Redshift Using AWS DMS and AWS SCT

Workshop – November 28, 2017

AWS re:Invent

aws

Agenda

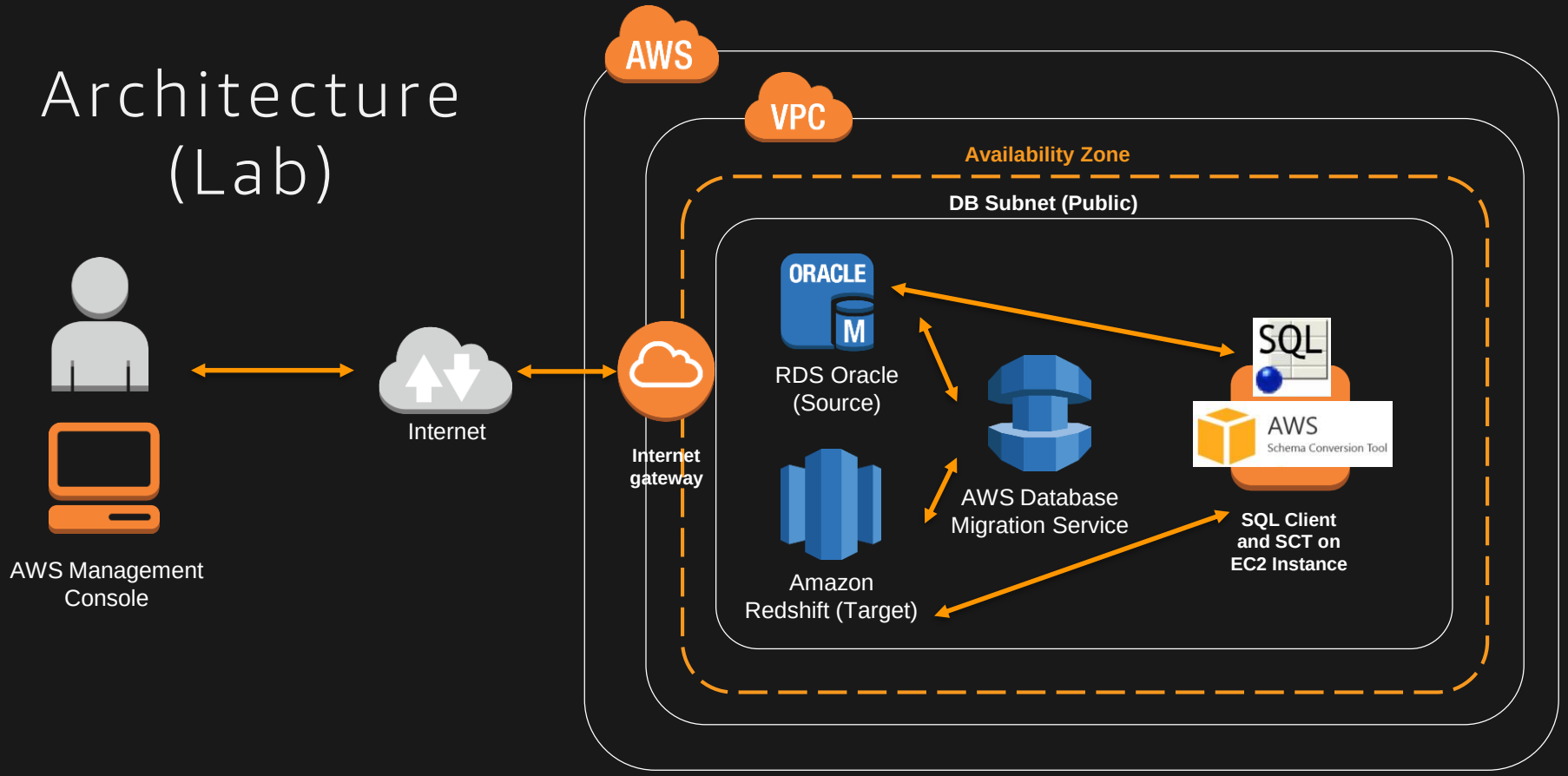
- Lab architecture and environment setup
- Amazon Redshift overview and best practices
- Migrating data warehouse to Amazon Redshift
- AWS DMS & AWS SCT overview and best practices
- Workshop lab

Session Details

- Workshop Duration – 2 1/2 hours
    - Proposed solution, AWS Services and Best Practices Overview
    - Hands-on workshop
- Prerequisites
    - Students use their own AWS accounts to run the lab with IAM Admin permissions
    - Sufficient limits for AWS resources (Amazon RDS, Amazon EC2, Amazon VPC, Amazon Redshift)
    - Comfortable working on the AWS console and configure AWS services
    - Working knowledge of relational databases (Oracle) and Amazon Redshift
- Workshop Team
    - Shree Kenghe, AWS Solutions Architect
    - Wesley Wilk, AWS Solutions Architect
    - Ramya Kaushik, Database Engineer DMS/SCT
    - Sarah Sleyman, AWS Solutions Architect

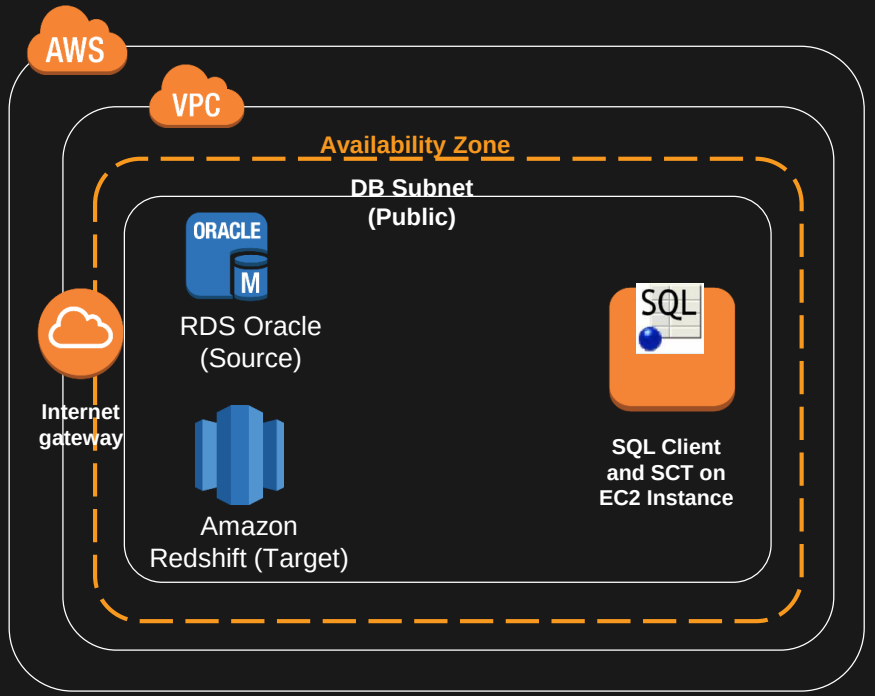
Lab Architecture and Environment Setup

Architecture (Lab)



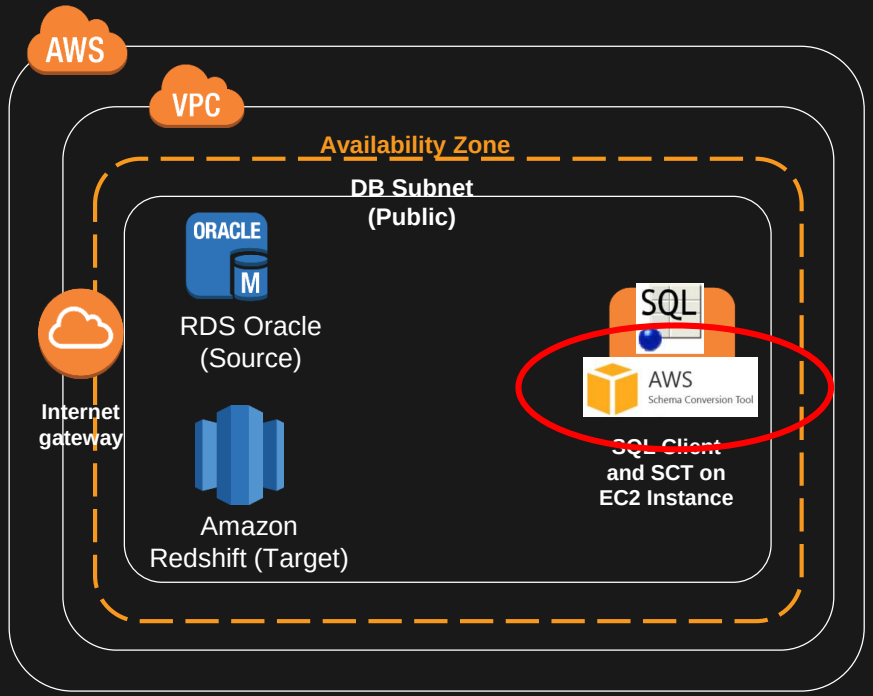
What you will accomplish

1. Launch AWS CloudFormation template to set up environment in Ireland



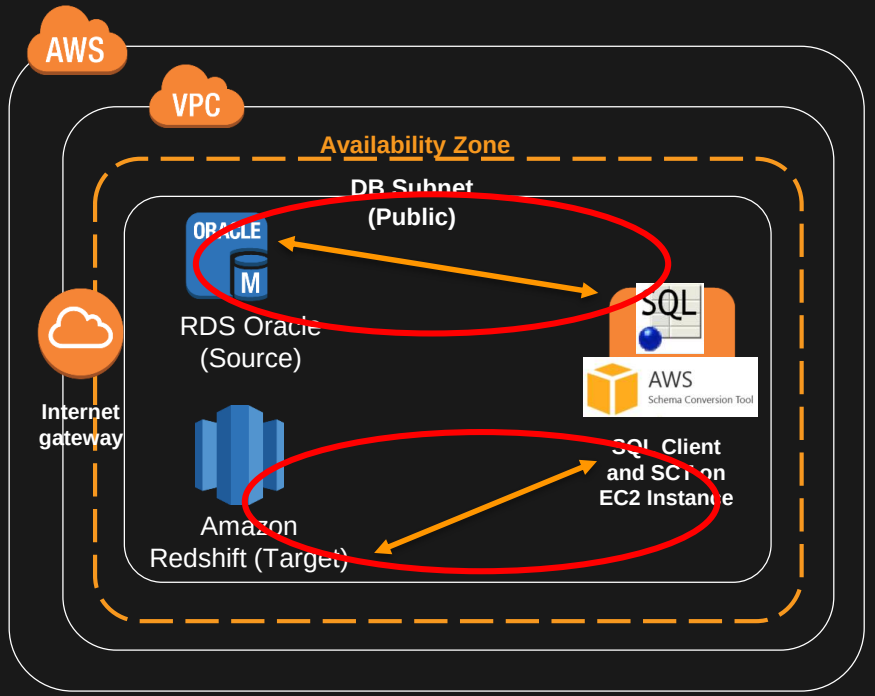
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)



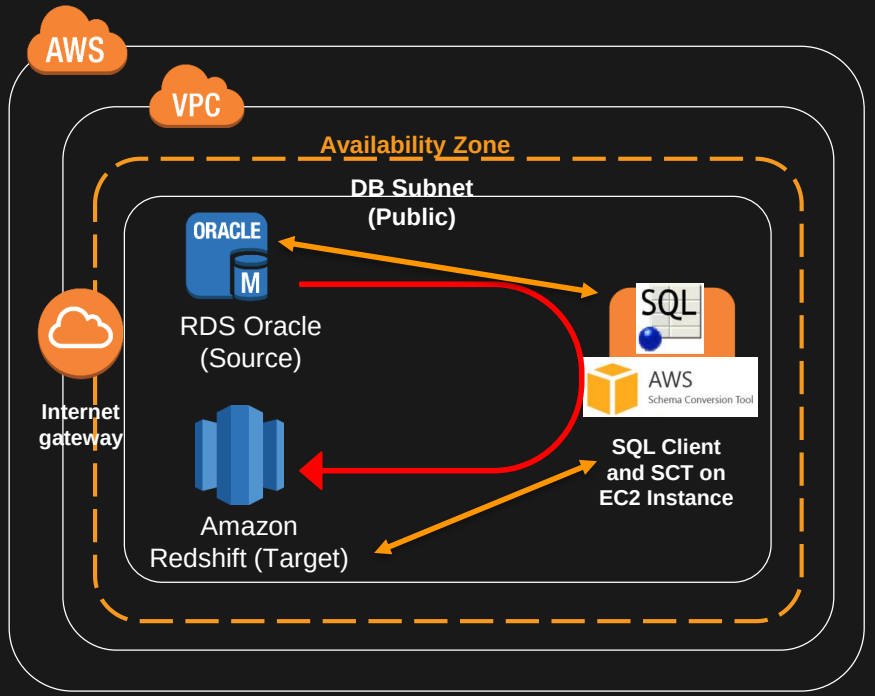
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT



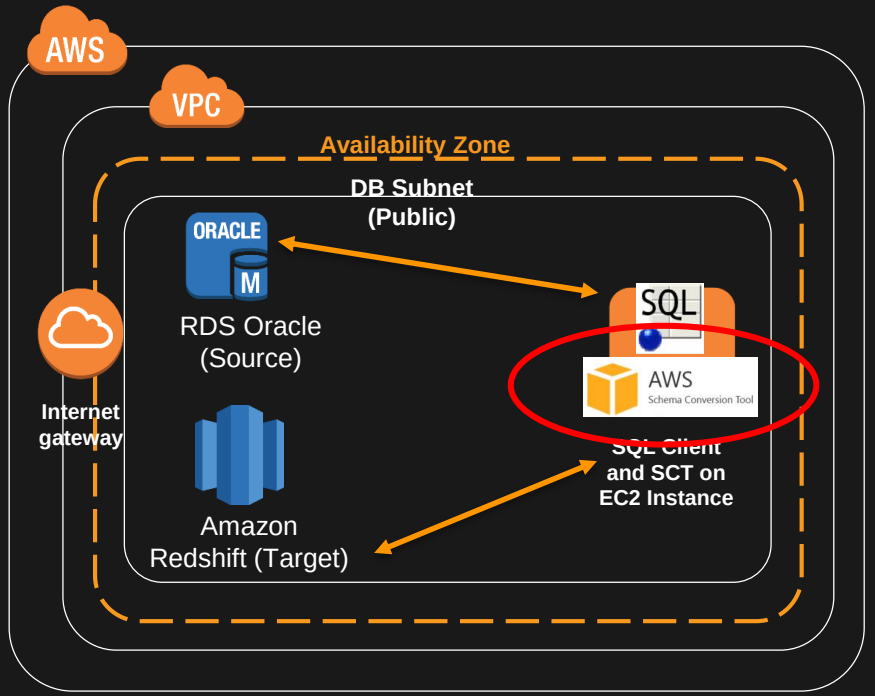
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT
4. Use AWS SCT to convert schema from source Oracle to target Amazon Redshift



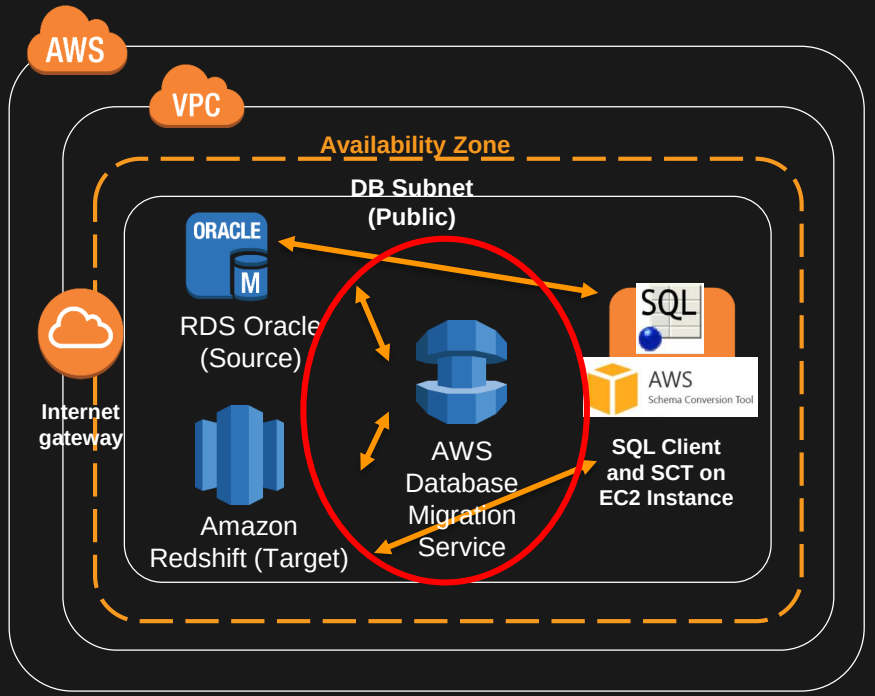
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT
4. Use AWS SCT to convert schema from source Oracle to target Amazon Redshift
5. Validate schema conversion and work through conversion issues



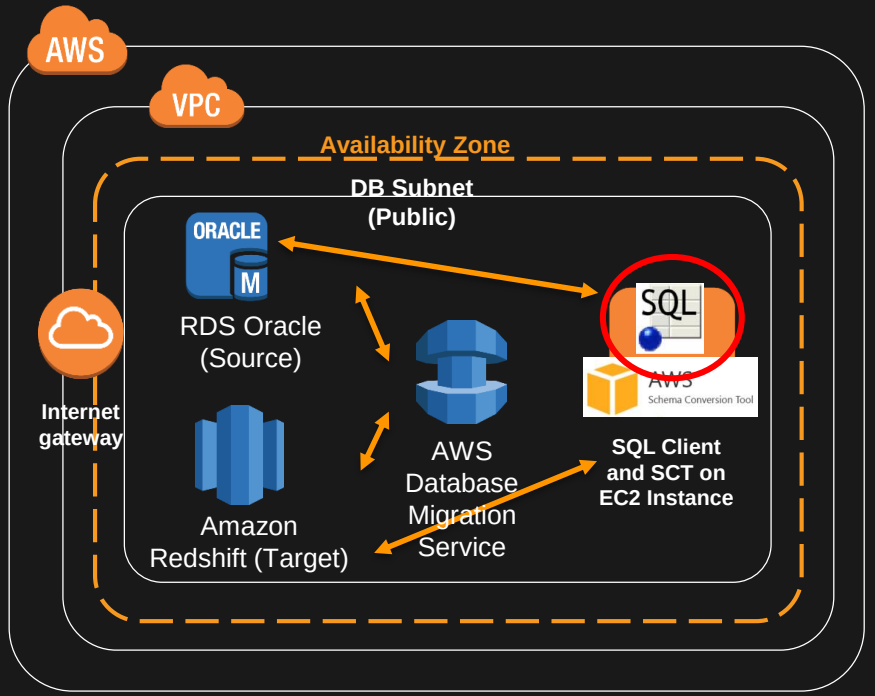
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT
4. Use AWS SCT to convert schema from source Oracle to target Amazon Redshift
5. Validate schema conversion and work through conversion issues
6. Use AWS Database Migration Service (DMS) to migrate data from source Oracle to Amazon Redshift



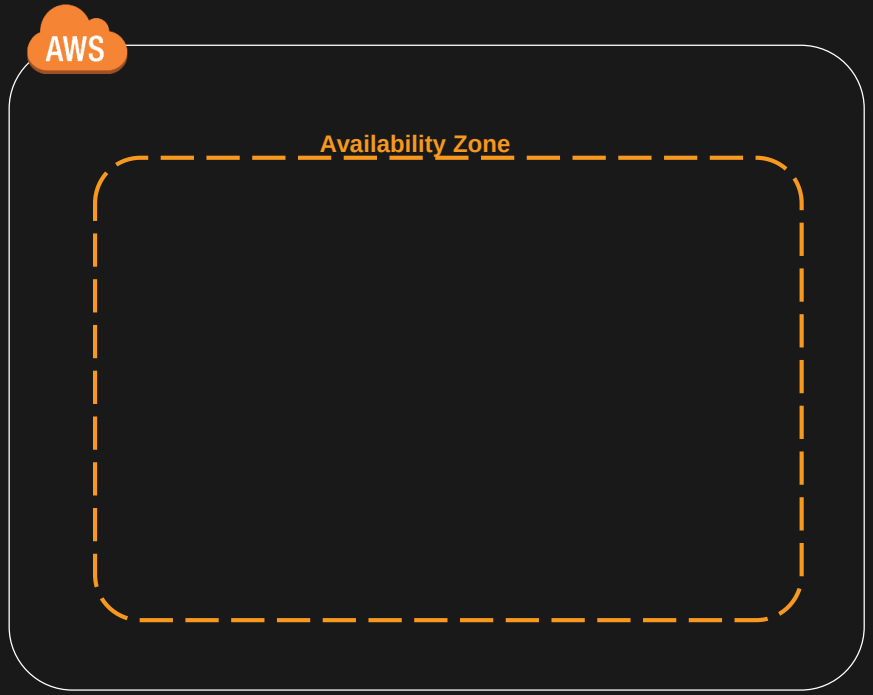
What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT
4. Use AWS SCT to convert schema from source Oracle to target Amazon Redshift
5. Validate schema conversion and work through conversion issues
6. Use AWS Database Migration Service (DMS) to migrate data from source Oracle to Amazon Redshift
7. Verify data migration completed successfully



What you will accomplish

1. Launch CloudFormation template to setup environment in Ireland
2. Connect to your environment and install AWS Schema Conversion Tool (SCT)
3. Test connectivity to Oracle RDS and Amazon Redshift from AWS SCT
4. Use AWS SCT to convert schema from source Oracle to target Amazon Redshift
5. Validate schema conversion and work through conversion issues
6. Use AWS Database Migration Service (DMS) to migrate data from source Oracle to Amazon Redshift
7. Verify data migration completed successfully
8. Delete all resources



Lab Setup and Environment

- Download zip file: **http://amzn.to/2zHGHpG**
  - CloudFormation template
  - Lab guide
  - SQL file
  - PowerPoint presentation

- In AWS Management Console choose AWS Region **eu-west-1** (Ireland)

- Follow lab guide Step 1 - Launch CloudFormation template

Amazon Redshift Overview and Best Practices

AWS re:Invent

aws

Amazon Redshift

Amazon
Redshift

a lot faster
a lot cheaper
a whole lot simpler

Petabyte scale

Fast, Massively parallel

Relational Data Warehouse

Inexpensive

Fully managed; zero admin

Amazon Redshift Architecture

Leader Node

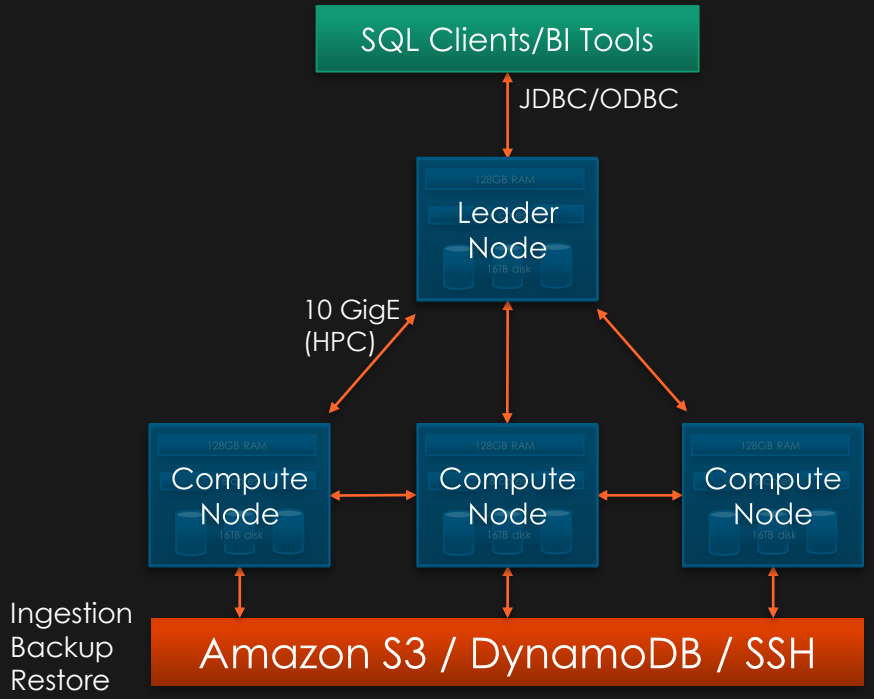
- SQL endpoint, JDBC/ODBC
- Stores metadata
- Coordinates query execution

Compute Nodes

- Local, columnar storage
- Execute queries in parallel
- Load, backup, restore via Amazon S3
- Load from Amazon DynamoDB or SSH
- Fault Tolerant

Two hardware platforms

- Optimized for data processing
- DS2: HDD; scale from 2TB to 2PB
- DC2: SSD; scale from 160GB to 326TB



Amazon Redshift – Columnar Storage

Row Storage

SSN	Name	Age	Addr	City	St
101259797	SMITH	88	899 FIRST ST	JUNO	AL
892375862	CHIN	37	16137 MAIN ST	POMONA	CA
318370701	HANDU	12	42 JUNE ST	CHICAGO	IL

Columnar Storage

SSN	Name	Age	Addr	City	St
101259797	SMITH	88	899 FIRST ST	JUNO	AL
892375862	CHIN	37	16137 MAIN ST	POMONA	CA
318370701	HANDU	12	42 JUNE ST	CHICAGO	IL

Amazon Redshift – Zone Maps and Sort Keys

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

- Single Column
- Compound
- Interleaved

Choose the Best Distribution Style

- Distribute the fact table and one dimension table on their common columns.
- Choose the largest dimension based on the size of the filtered data set.
- Choose a column with high cardinality in the filtered result set.
- Change some dimension tables to use ALL distribution.

Choose the Best Sort Key

- If recent data is queried most frequently, specify the timestamp column as the leading column for the sort key.
- If you do frequent range filtering or equality filtering on one column, specify that column as the sort key.
- If you frequently join a table, specify the join column as both the sort key and the distribution key.

Amazon Redshift – Best Practices & Pointers

- Maximize Load Performance
  - COPY multiple files
  - COPY to multiple nodes
  - Compress source data
  - Use a manifest file
- RedShift does not enforce primary key constraints
  - If you load data multiple times, Amazon Redshift will not complain
- After Loading
  - Data all added at end of columns for speed
  - Fully functional, but not set for optimum performance
- VACUUM command
  - Massages data to optimum disk organization for performance

Migrating Data Warehouses to AWS

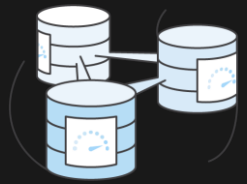
AWS re:Invent

aws

Why Migrate to Amazon Redshift?

Transactional database

- Amazon Redshift is 100x faster
- Scales from GBs to PBs
- Analyze data without storage constraints



MPP database

- Amazon Redshift is 10x cheaper
- Easy to provision and operate
- Higher productivity

Apache Hadoop

- Amazon Redshift is 10x faster
- No programming
- Standard interfaces and integration to leverage BI tools, machine learning, streaming

AWS DMS and SCT Overview & Best Practices

AWS re:Invent

aws

What are DMS and SCT?

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

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

When to use DMS and SCT?

Modernize

Modernize your database tier –

- Commercial to open-source
- Commercial to Amazon Aurora

Modernize your Data Warehouse –

- Commercial to Amazon Redshift

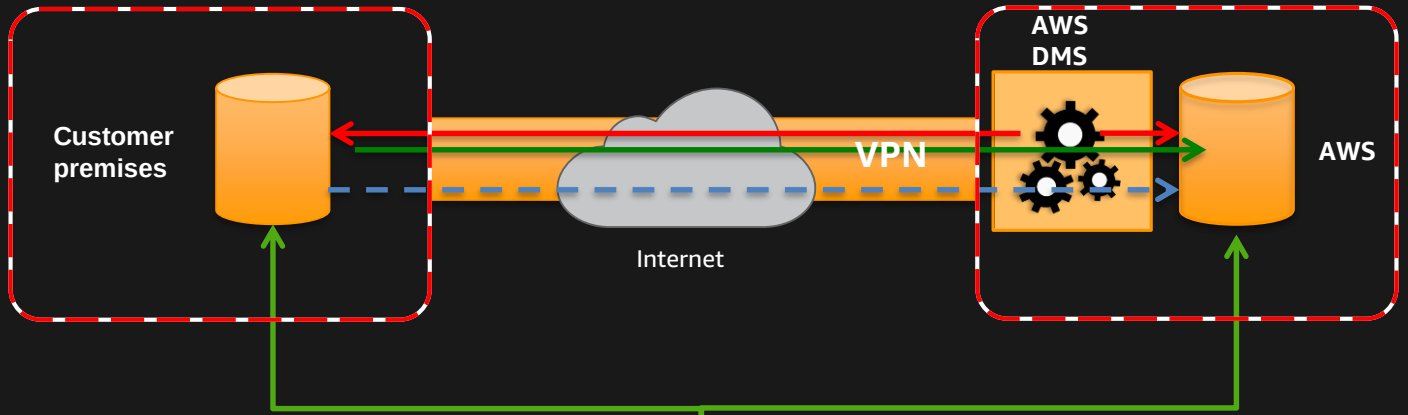
Migrate

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

Replicate

- Create cross-regions Read Replicas
- Run your analytics in the cloud
- Keep your dev/test and production environment sync

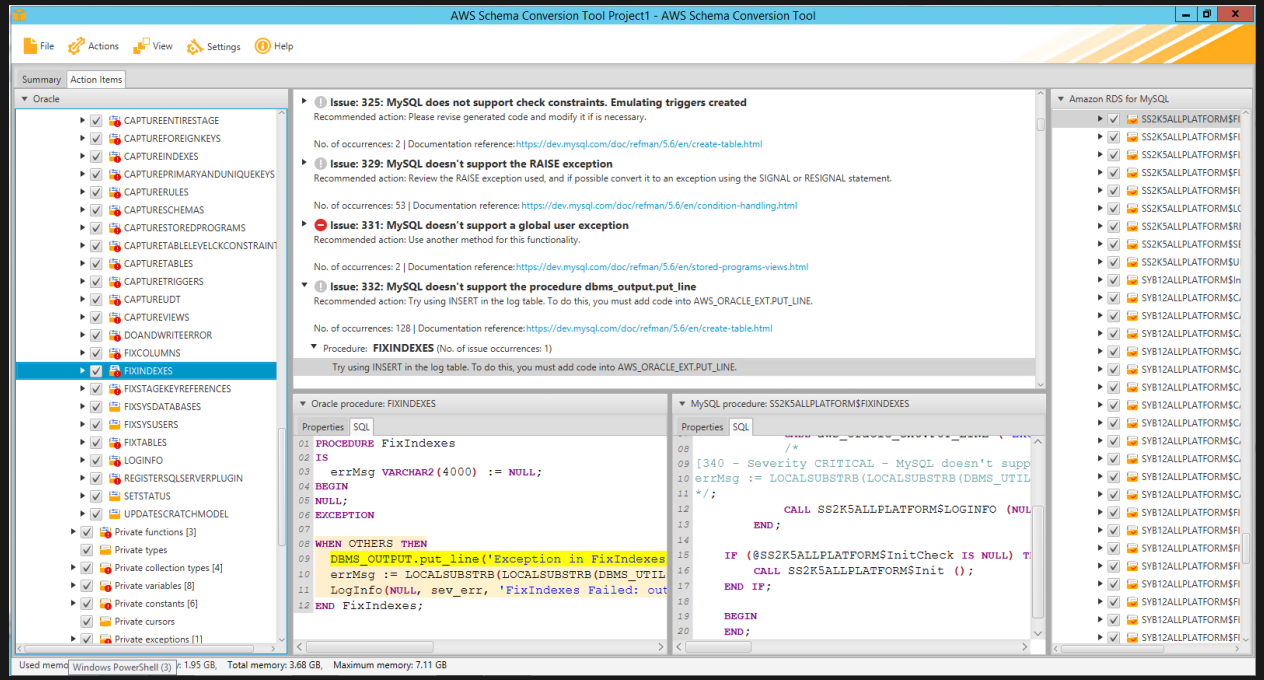
Keep your apps running during the migration



- Start a replication instance
- Connect to source and target databases
- Select tables, schemas, or databases

- Let AWS DMS create tables, load data, and keep them in sync
- Switch applications over to the target at your convenience

SCT helps with converting tables, views, and code

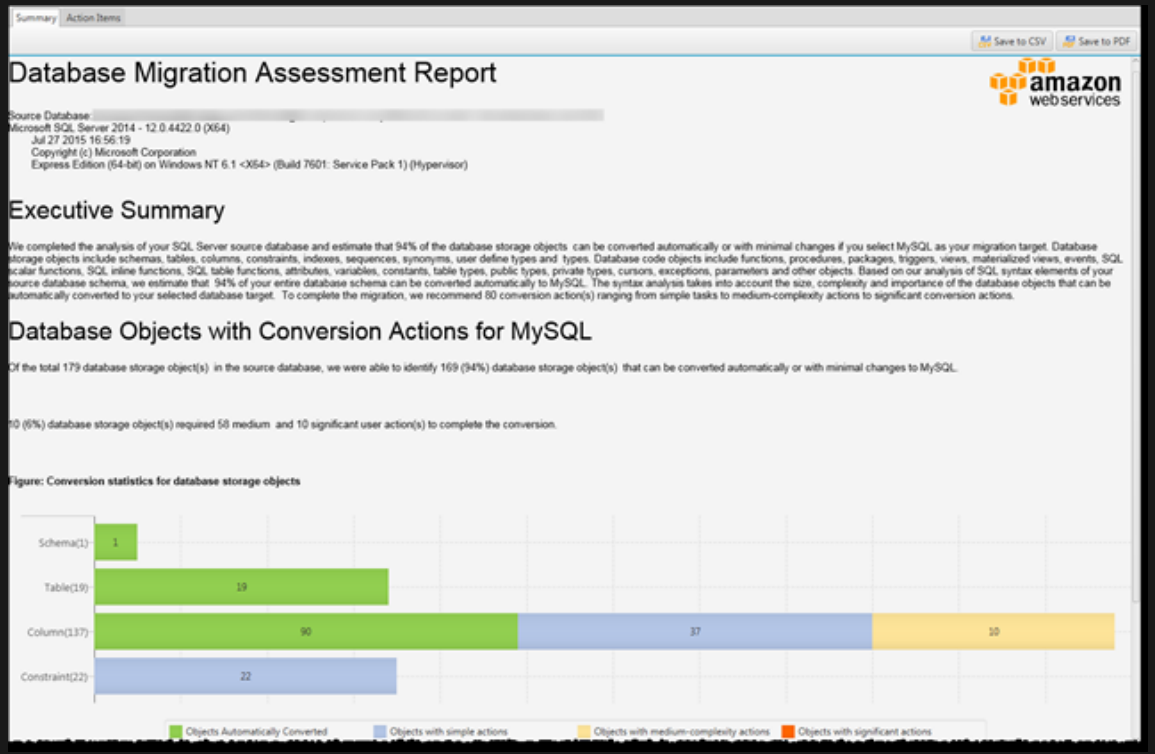


- Sequences
- User-defined types
- Synonyms
- Packages
- Stored procedures
- Functions
- Triggers
- Schemas
- Tables
- Indexes
- Views
- Sort and distribution keys

AWS re:Invent

aws

SCT Migration Assessment Report



- Assessment of migration compatibility of source databases with open-source database engines – RDS MySQL, RDS PostgreSQL, and Aurora
- Recommends best target engine
- Provides details level of efforts to complete migration

Amazon Redshift as a Target for DMS

- Amazon Redshift cluster must be in the same AWS account and same AWS Region as the replication instance.
- AWS DMS first moves data to an S3 bucket.
- Once the files are in an S3 bucket, AWS DMS then transfers them to the proper tables in the Amazon Redshift data warehouse
- The Amazon Redshift endpoint provides full automation for
    - Schema generation and data type mapping
    - Full load of source database tables
    - Incremental load of changes made to source tables
    - Application of schema changes in data definition language (DDL) made to the source tables
    - Synchronization between full load and change data capture (CDC) processes

DMS Best Practices

- Improving performance of an AWS DMS Migration
    - Load multiple tables in parallel
    - Remove bottlenecks on the target
    - Use multiple tasks
    - Improving LOB performance
    - Optimizing change processing
- Determining the optimum size for a replication instance
- Reducing load on your source database
- Ongoing replication

LAB

http://amzn.to/2zHGHpG

Resources/Blogs

- How to Migrate Your Oracle Data Warehouse to Amazon Redshift Using AWS SCT and AWS DMS
- How to Migrate Your Oracle Database to PostgreSQL
- Migrating Oracle Database from On-Premises or Amazon EC2 Instances to Amazon Redshift
- Top 10 performance Tuning Techniques for Amazon Redshift

AWS re:Invent

THANK YOU!