
AWS
re:Invent

ANT304

Migrating a legacy data warehouse to Amazon Redshift

Gal Barnea

Partner Engineering Manager

Amazon Redshift

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

intuit

Intuit Inc. is a business and financial software company that develops and sells financial, accounting, and tax preparation software

Challenge

Previous solutions based on Microsoft SQL Server couldn't scale with explosive growth of data; scaling required significant manual effort and cost, which resulted in delayed insights

Solution

Intuit migrated to a solution based on Amazon Redshift; it scales to more than 7x the data volume with virtually zero effort and delivers 20x the performance, which has led to a 90% reduction in time-to-insight and a 66% reduction in cost

intuit powering prosperity

turbotax

quickbooks

proconnect

mint

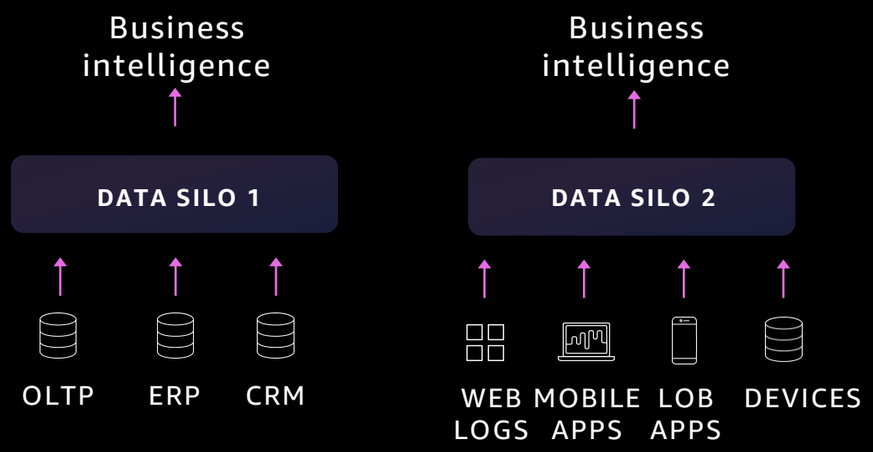


Customers who have migrated from on-premises data warehouses to Amazon Redshift



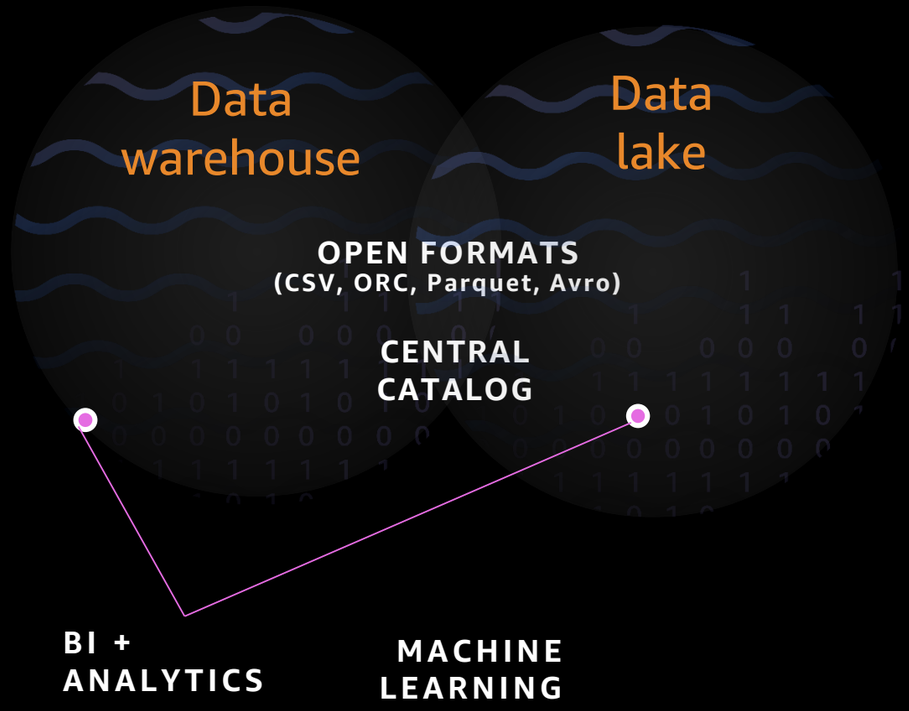
Customers look for a modern data platform

Old-guard data patterns



TO

Modern data architecture



Modern data platform requirements

Use – support exploratory data analysis and ML


Embedded analytics


Notebook automation


Predictive analytics


Operational analytics


Ad hoc query

Manage – data discovery, search, and collaboration


Catalog and search


Governance share data

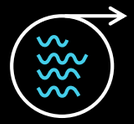

Master data management


Data quality

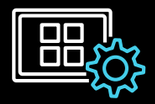

Security and management

Run – data processing and platform frameworks


Data ingestion


Data transformation


Databases and storage


Code and infrastructure automation

Migrating to Amazon Redshift



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer migration journey

Analyze and plan → **Workshop and pilot** → **Continuous workload migrations**

Discovery & planning

Migration business case

Migration expertise

Migration plan

Workload analysis and pilot

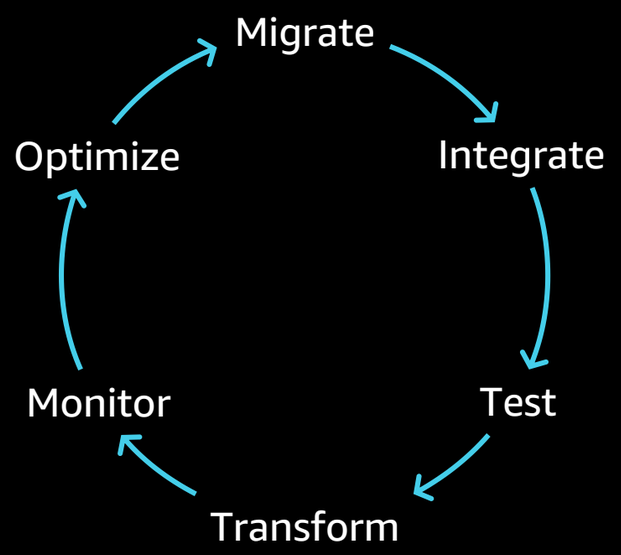
Create new target

Modify / develop BI and apps to dual target

Modify / develop ETL to dual target

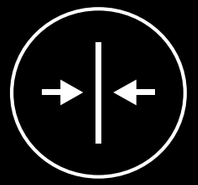
Skills / CoE

Migration



AWS data migrations: Broadest toolkit

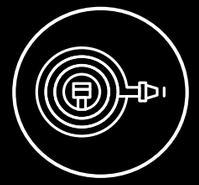
AWS PROVIDES THE BROADEST RANGE OF TOOLS FOR EASY, FAST, AND SECURE DATA MOVEMENT TO AND FROM THE AWS CLOUD



AWS Direct Connect



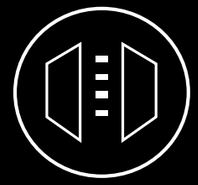
AWS Snowball



Amazon Kinesis Data Firehose



Amazon S3 Transfer Acceleration



AWS Storage Gateway

AWS Schema Conversion Tool (AWS SCT)



AWS Database Migration Service (AWS DMS)



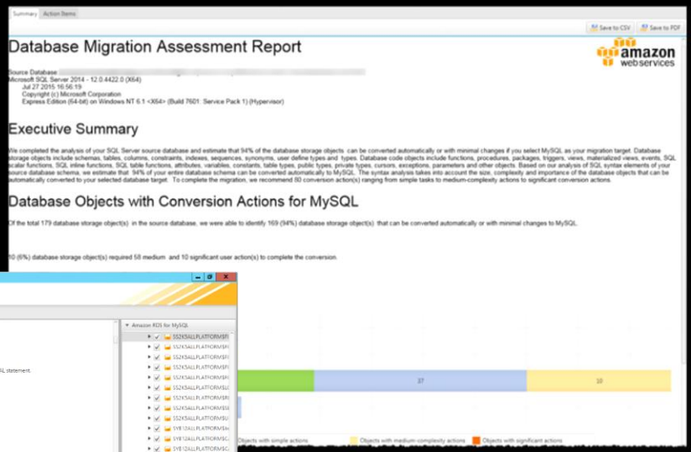
AWS Schema Conversion Tool (AWS SCT)

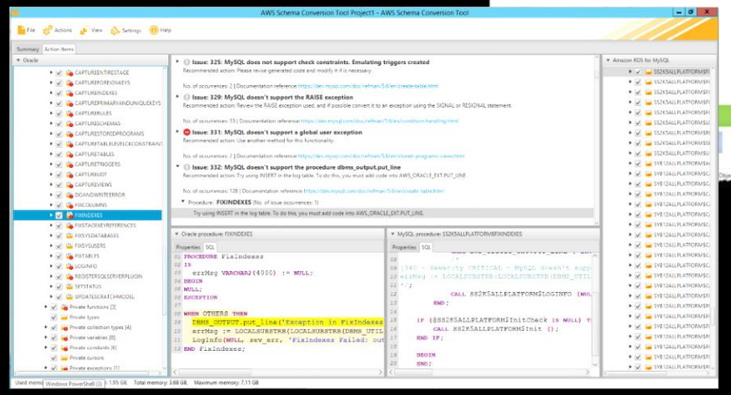
ACCELERATE MIGRATIONS TO AMAZON REDSHIFT

AWS SCT converts your commercial database and data warehouse schemas to Amazon Redshift and other native services, such as Amazon RDS and Amazon Aurora

- Support a number of sources including Oracle, Teradata, Greenplum, IBM Netezza, HPE Vertica, and MS SQL Server
- Generates a detailed migration assessment report
- Converts source tables, views, stored procedures, functions, and application SQL code
- Automatic schema optimization
- AWS SCT data migration agents can extract, prepare, optimize, and upload data securely and in parallel from source data warehouse to Amazon Redshift







AWS Database Migration Service (AWS DMS)

ACCELERATE MIGRATIONS TO AMAZON REDSHIFT

1. Relational databases

PostgreSQL Microsoft Azure SQL

MariaDB MySQL Oracle

Microsoft SQL Server

2. Non-relational databases

3. Other sources



Amazon S3

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



AWS DMS



Amazon Redshift

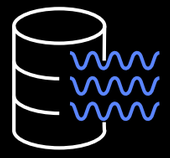
Migration-focused features



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift innovates to meet your needs



Analyze all your data

Lake house with AWS integration

NEW!

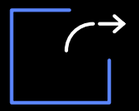
Amazon Redshift ML

NEW!

Data sharing

NEW!

Super data type with JSON support

UPDATED!

Federated Query

NEW!

Lambda UDF

NEW!

Partner console integration


Amazon Redshift Spectrum + Lake Formation


Data Lake Export



Performance & scale

Fast and self-tuning

UPDATED!

RA3 nodes & managed storage

PREVIEW!

AQUA

NEW!

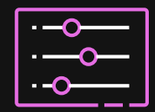
Performance tuning: automated

UPDATED!

Materialized views

NEW!

100K tables

NEW!

HyperLogLog


Concurrency scaling



Low cost & best value

Predictable costs

UPDATED!

Automatic workload manager

NEW!

Cross-AZ cluster recovery

NEW!

Data API


On-demand and RIs


Pause and resume


Cost controls


Built-in security features

Migration-focused features

AMAZON REDSHIFT INNOVATING TO ACCELERATE MIGRATIONS

Stored Procedures support

Enhanced Security Controls (CLP)

Increased Catalog Limits

Enhanced Spatial Functionality

New Data Types: Time & TimeTZ

Support for stored procedures provides the ability to run code where the data is to efficiently run ETL, data validation, and custom business logic

```
CREATE OR REPLACE PROCEDURE test_spl(f1 int, f2 varchar)
AS $$
BEGIN
  RAISE INFO 'f1 = %, f2 = %', f1, f2;
END;
$$ LANGUAGE plpgsql;

call test_spl(5, 'abc');
INFO: f1 = 5, f2 = abc
CALL
```

Migration-focused features

AMAZON REDSHIFT INNOVATING TO ACCELERATE MIGRATIONS

- Stored Procedures support
- Enhanced Security Controls (CLP)
- Increased Catalog Limits
- Enhanced Spatial Functionality
- New Data Types: Time & TimeTZ

Enforce security for sensitive data like PII and PCI with Column level access control for local tables

```
GRANT SELECT(id, ssn, sal) on emp_data to authorized_user;

REVOKE SELECT(ssn) on emp_data from unauthorized_user;

--As authorized_user
select * from emp_data; -- ok
select id, ssn from emp_data; -- ok

-- As unauthorized_user
select * from emp_data; -- fail
ERROR:  permission denied for relation emp_data
select id from emp_data; -- ok, ID column can be accessed
```

Migration-focused features

AMAZON REDSHIFT INNOVATING TO ACCELERATE MIGRATIONS

NEW

Stored Procedures support

Enhanced Security Controls (CLP)

Increased Catalog Limits

Enhanced Spatial Functionality

New Data Types: Time & TimeTZ

- Up to 100,000 tables
- Up to 10,000 stored procedures

Migration-focused features

AMAZON REDSHIFT INNOVATING TO ACCELERATE MIGRATIONS

NEW

Stored Procedures support

Enhanced Security Controls (CLP)

Increased Catalog Limits

Enhanced Spatial Functionality

New Data Types: Time & TimeTZ

- 30 new functions
- Shapefile import support
- ODBC/JDBC driver support
- Spatial join performance improvements

Migration-focused features

AMAZON REDSHIFT INNOVATING TO ACCELERATE MIGRATIONS

NEW

Stored Procedures support

Enhanced Security Controls (CLP)

Increased Catalog Limits

Enhanced Spatial Functions

New Data Types: Time & TimeTZ

Store and process ‘time’ values with/without a timezone
ODBC/JDBC driver support

Data Types

`TIME`
`TIMETZ`

Functions

`EXTRACT()`
`DATEADD()`
`DATEDIFF()`

Operators

+ (Concatenate)
>, <, =, <=, >=, !=, <> **(compare)**

Native semi-structured data support

PREVIEW

New data type: **SUPER**

Easy, efficient, and powerful JSON processing

Fast row-oriented data ingestion

Fast column-oriented analytics with materialized views over SUPER/JSON

Access to schema-less nested data with easy-to-use SQL extensions powered by the PartiQL query language

```
{
"id":1,
"name":{"given":"Jane", "family":"Doe"},
"phone":[{"type":"work", "num": "9252364000"},
         {"type":"cell", "num": 6501234444}]
}
{
"id":2,
"name":{"given":"Graham", "family":"Bell"},
"phone":[{"type":"work", "num": 5106101234}]
}
```

```
SELECT name.given AS firstname, ph.num
FROM customers c, c.phone ph
WHERE ph.type = 'cell';

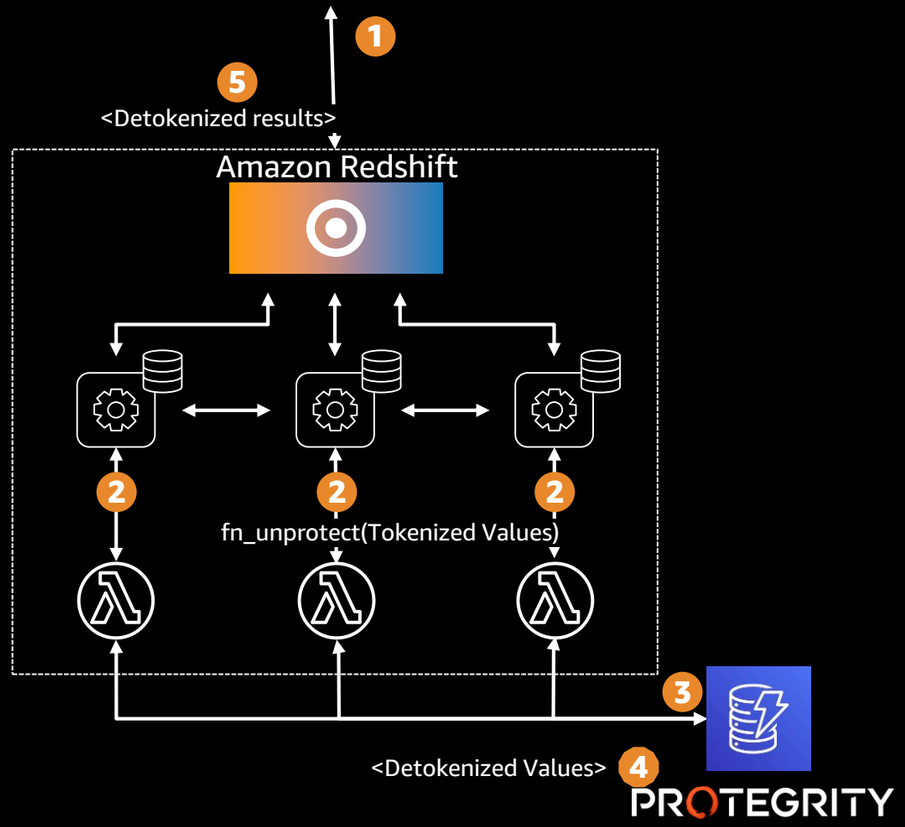
firstname | num
----------+---------------
"Jane"    | 6501234444
```

Tokenization with Lambda UDFs

NEW

```
CREATE OR REPLACE EXTERNAL FUNCTION unprotect(varchar) RETURNS varchar
    VOLATILE lambda 'fn_unprotect'
    iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';

SELECT unprotect(t.tokenized_column) FROM tokenized_table t;
```



Invoke AWS Lambda programs as UDFs in Amazon Redshift SQL queries

Simple integration with external services

- Tokenization with third-party vendors like Protegrity
- More languages runtimes (C++, Java etc.)
- Access DynamoDB, SageMaker, etc.

Concurrent and batch processing

Cost controls and error controls

Amazon Redshift automates performance tuning

ML-BASED OPTIMIZATIONS TO GET STARTED EASILY AND GET THE FASTEST PERFORMANCE QUICKLY

Automates physical data design and optimization

Optimizes for peak performance as data and workloads scale

Leverages machine learning to adapt to shifting workloads

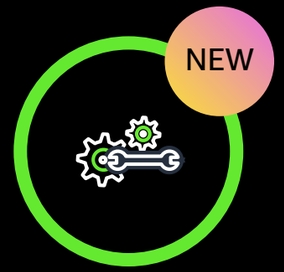
Automated performance tuning



Automatic vacuum delete



Automatic distribution keys



Automatic sort keys



Auto workload manager



Automatic table sort



MV auto-refresh and rewrite

Migration best practices



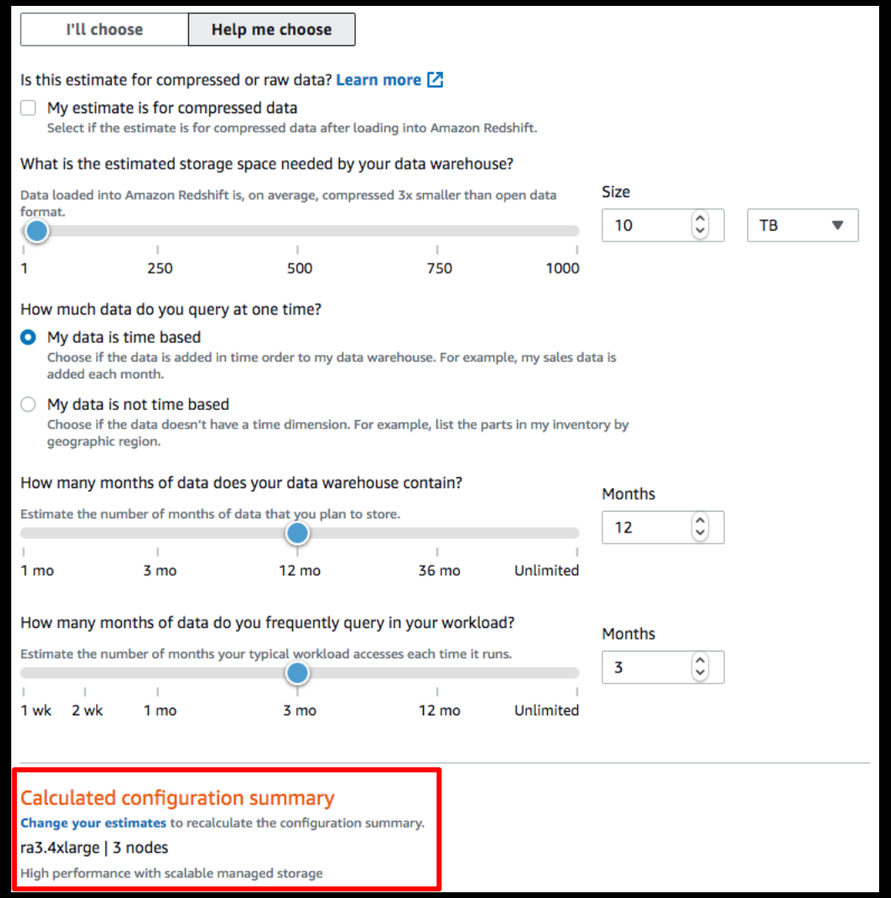
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Determining the target Amazon Redshift cluster size is easy

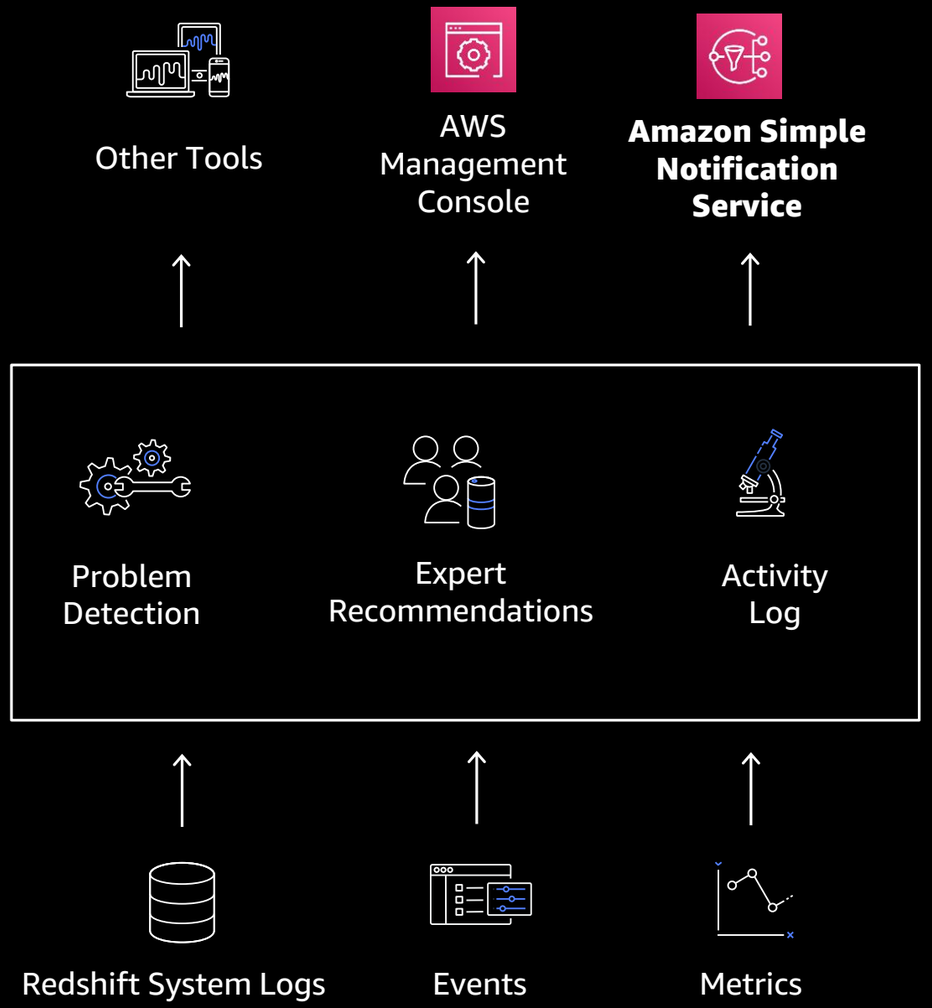
Size a cluster to meet performance needs for steady state workload

Amazon Redshift console helps you determine the size for the steady state workload
https://console.aws.amazon.com/redshift/



Redshift advisor

EXPERT ADVICE PERSONALIZED FOR YOUR CLUSTER AND WORKLOAD



Continuously monitor millions of data points to detect, analyze, and surface issues before they impact your users.

- Expert health checks
- Actionable recommendations to optimize cost and performance
- Machine learning powered

General legacy migration considerations

The majority of the DDL, SPs, and SQL scripts can be automatically converted by AWS SCT, and you can use the assessment reports for a deeper analysis

LOBs (large objects), such as images, pdfs, or other binary data, are not directly supported but can be migrated to Amazon S3

Stored Procedure support makes porting legacy procedures easier

Lambda UDFs extend support for C, C++, and Java UDFs and macros

Materialized Views help migrate complex queries for faster query performance;
for example, BI dashboard queries

Leverage Amazon Redshift Spectrum for external tables and infrequently accessed data

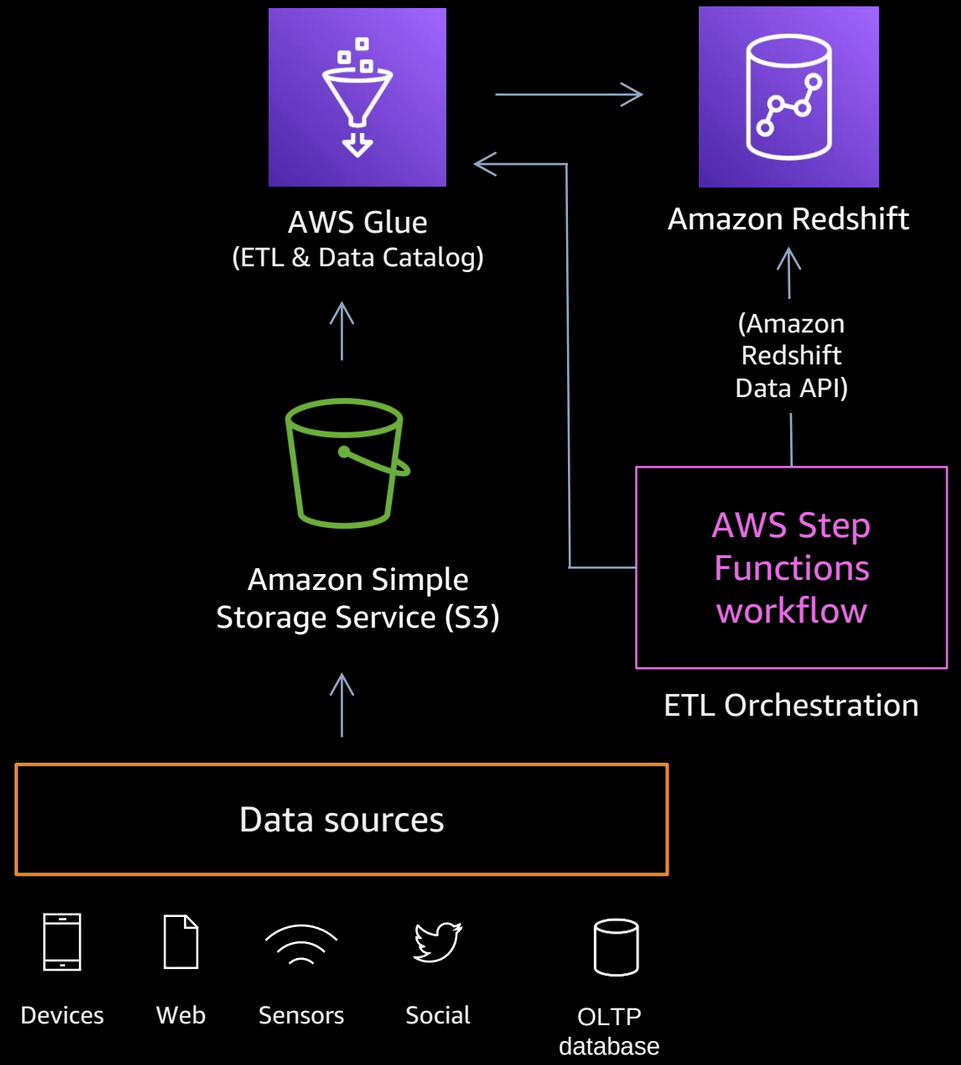
ETL migration

Modernize your data pipelines using AWS Glue/AWS EMR/Amazon Redshift Data API

Oracle ETL and Teradata BTEQ can be converted to AWS Glue using AWS SCT

Leverage your existing 3rd party ETL tools; Informatica, Matillion, Talend, and many others support Amazon Redshift Natively

Custom ETL scripts should be modified to use the Amazon Redshift COPY command to load from Amazon S3



Migration best practices

Vertica

FACTSET RetailMeNot ROVIO

Use materialized views in place of Vertica projections

Vertica 9.x support bulk export to Amazon S3 using S3EXPORT functions

Create User Defined Functions (UDFs) with the same name and parameters as Vertica-specific functions used in ETL queries (e.g. time slice) for ease of migration

Migration best practices



Netezza/ Greenplum

PostGres lineage enables convenient table designs, for example sort/dist keys

Use native bulk-export of the data to flat files for fast data movement to Amazon S3

Migration best practices

Oracle/ SQL Server







Uniqueness, primary key, and foreign key constraints are informational only; they are not enforced by Amazon Redshift

OLTP workloads can be migrated to AWS RDS or Amazon Aurora and queried using Amazon Redshift Federated Query

Sequences are not directly supported, but can be migrated to IDENTITY columns

Kick-start your data warehouse migration today!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Migrate with AWS partners

AWS consulting partners offer a wide range of migration services to help you move your data warehouse to Amazon Redshift. AWS Data Warehouse Migration Partners provide support to accelerate moving a data warehouse to the cloud with proven best practices and resources. The AWS Service Delivery Partners have deep understanding of specific AWS services, follow best practices, and have proven success delivering AWS services to customers.



This is not a complete list; to view all Amazon Redshift partners, visit https://aws.amazon.com/redshift/partners/

Start your data warehouse migrations today

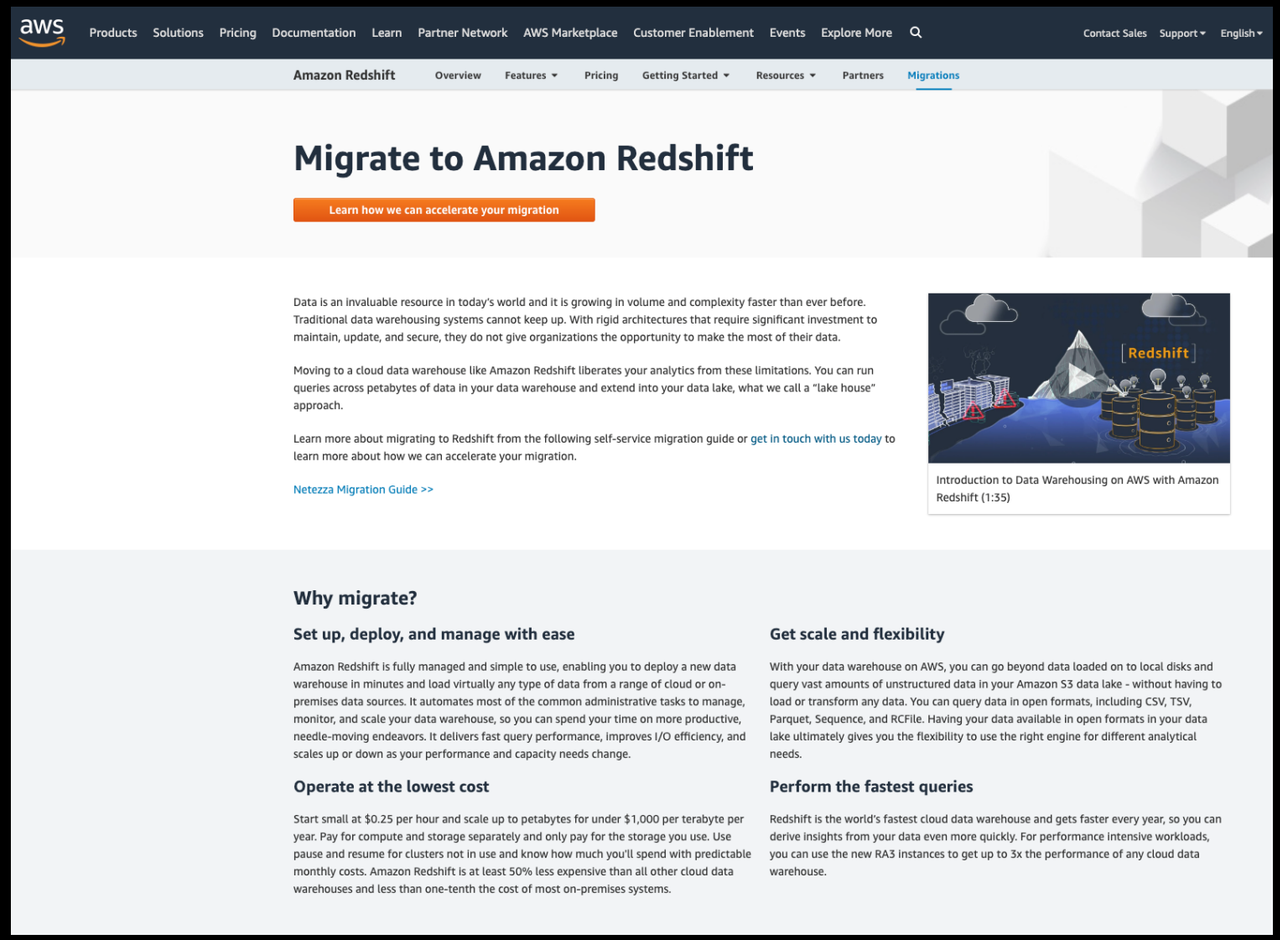
VISIT OUR MIGRATION PAGE ON THE REDSHIFT WEBSITE

Learn more about Amazon Redshift and new features

Hear migration success stories from customers and partners

Talk to a Migration Specialist

https://aws.amazon.com/redshift/data-warehouse-migration



Thank you!

Gal Barnea

galbarn@amazon.com

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.