
AWS
re:Invent

DAT212

New capabilities to build graph apps quickly with Amazon Neptune

Karthik Bharathy

Product Leader, Neptune

AWS

AWS re:Invent

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

Agenda

Graphs help you build awesome new things

What is Amazon Neptune?

Key innovations in Neptune



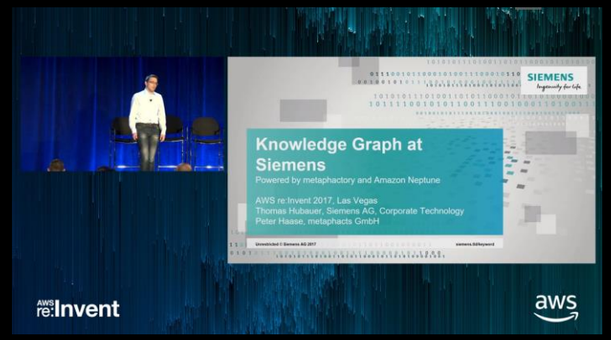
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Neptune journey . . .



2017

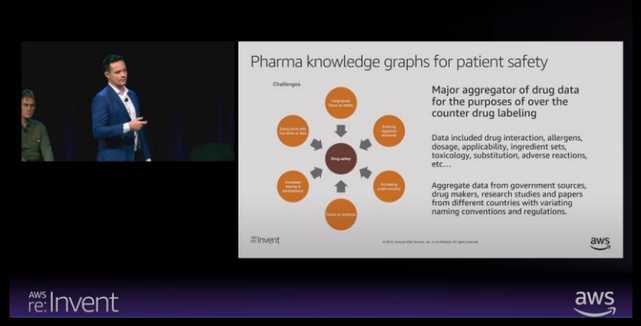
Announced at re:Invent – a fully managed fast, scalable, reliable graph database; Siemens presented their use case

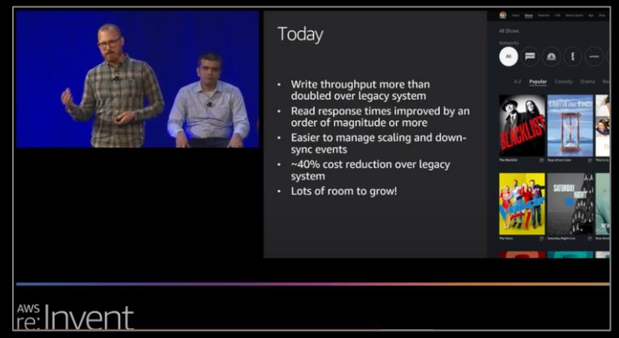




2018

Nike and PwC present their graph use cases at re:Invent





2019

Uber and NBCU present their use cases at re:Invent



2020

ADP and Cox Automotive present their use cases at re:Invent



Amazon Neptune



Fast

Query billions of relationships with millisecond latency



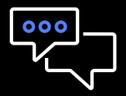
Scalable

Scale compute and storage independently; scale out to 15 read replicas in minutes



Easy

Build powerful queries easily with Gremlin and SPARQL



Open

Supports Apache TinkerPop and W3C RDF graph models



Fully managed

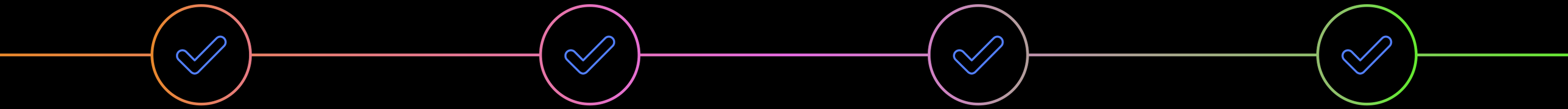
Managed by AWS – no hardware provisioning; auto patching, quick setup, secure, and automatic backups



Secure

VPC only, IAM integrated, encryption in-transit and at rest

Amazon Neptune is globally available and secure by default



21 AWS Regions*

Encryption-at-rest using CMK AWS KMS

Encryption-in-transit using TLS 1.2 only

In scope for PCI-DSS, ISO, SOC compliance, and HIPAA eligibility

**21 AWS Regions – Americas (N. Virginia, Montréal, Ohio, Oregon, N. California, Sao Paulo), Europe (Frankfurt, Ireland, London, Paris, Stockholm), Middle East (Bahrain), Asia-Pacific (Mumbai, Seoul, Singapore, Sydney, Tokyo, Hong Kong), AWS China (Ningxia) operated by NWCD, and AWS GovCloud (US-East) and (US-West)*



aws

ANALYST REPORT

Forrester Names **AWS a Leader in Graph Data Platforms** Wave™

The Forrester Wave™: Graph Data Platforms, Q4 2020

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 call on a market and is plotted using a detailed spreadsheet with exposed 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.

Reference customers



Amazon Neptune

Customers across different verticals and use cases use Amazon Neptune in production today

THOMSON REUTERS · finra · SIEMENS · AstraZeneca · LIFEOMIC · freshworks

amazon alexa · intuit · FACTSET · SAMSUNG · Blackfynn · Rappi

Uber ATG · HUUUGE · NBCUniversal · NETFLIX · asurion · COX AUTOMOTIVE

PaySense · zeta · Pearson · MARINUS ANALYTICS · noonum · audible FOR BUSINESS

Case studies: https://aws.amazon.com/solutions/case-studies/?customer-references-cards.q=neptune

Amazon Neptune: Key innovations

IT'S BEEN A BUSY YEAR FOR US!

Cost and performance

- T3 instances
- Start / stop cluster
- Federated query
- Alternative query engine

Enterprise

- Language enhancements
- DMS support
- Neptune streams
- Database cloning
- Tags in IAM
- Graph visualizations
- Cross-Region snapshot copy
- Delete protection
- Fast DB reset (new)
- Custom endpoints (new)

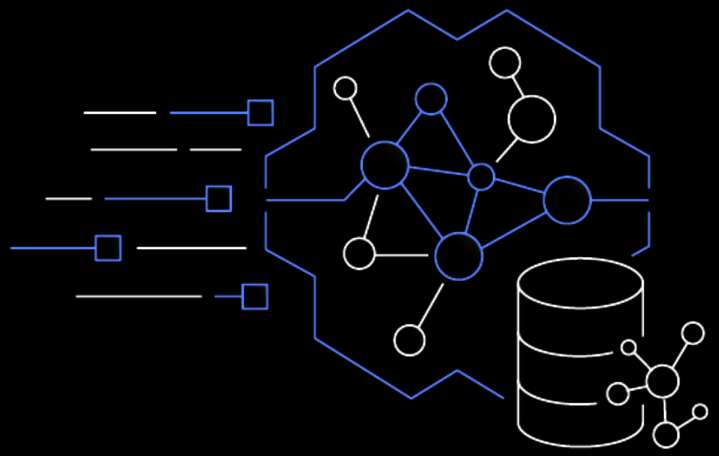
Integrations and ML

- Neptune ML (new)
- Amazon ES integration
- Amazon Kinesis integration
- Event subscriptions (new)
- DMS integration
- Export graph

New

Amazon Neptune ML

Easy, fast, and accurate predictions on graphs with graph neural networks (GNNs), powered by the Deep Graph Library (DGL) and Amazon SageMaker



Make predictions on graph data without ML expertise

Automatically choose and train the best ML model for your workload, enabling you to make ML-based predictions on graph data in hours instead of weeks

Use state-of-the-art graph machine learning

Up to 50% more accurate, GNNs are state-of-the-art ML purpose-built to use the relationships in graphs based on research from Stanford University

Scale to large datasets

For graph applications with billions of relationships in knowledge graphs, fraud detection, or product recommendation

Integrations and ML

AWS Database Blog

How to get started with Neptune ML

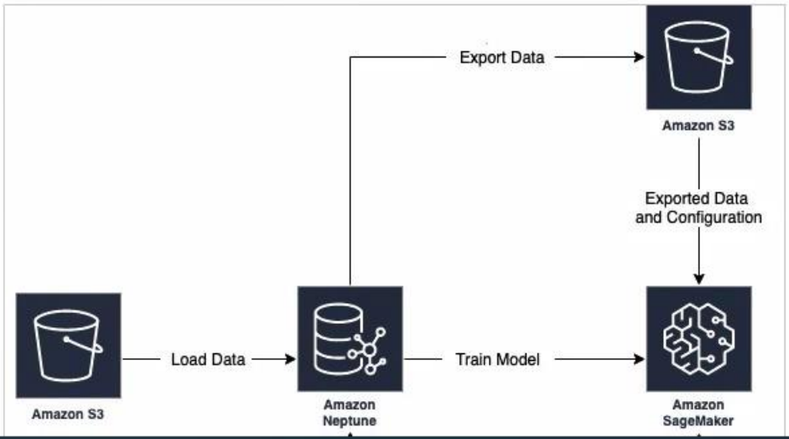
by George Karypis, Dave Bechberger, and Karthik Bharathy | on 08 DEC 2020 | in Amazon Neptune | Permalink | Comments | Share

Amazon Neptune ML is an easy, fast, and accurate approach for predictions on graphs. In this post, we show you how you can easily set up Neptune ML and infer properties of vertices within a graph.

For our use case, we have a movie streaming application and we want to infer the the top genres for a movie using a process known as node classification. To illustrate this node classification use case with Neptune ML, we use the MovieLens 100K Dataset** as an example property graph. The MovieLens dataset contains users, movies, genres, and ratings as vertices, and user-to-rating, rating-to-movie, and movie-to-genre as edges. You can try out other capabilities of Neptune ML using samples provided with Neptune.

The workflow contains six steps:

1. Set up the test environment.
2. Launch the node classification notebook sample.
3. Load the sample data into the cluster.
4. Export the graph.
5. Perform ML training.
6. Run Gremlin queries with Neptune ML.



Resources
Getting Started
What's New
Top Posts
Official AWS Podcast
AWS Case Studies

Follow
Twitter
Facebook
LinkedIn
Twitch
RSS Feed
Email Updates

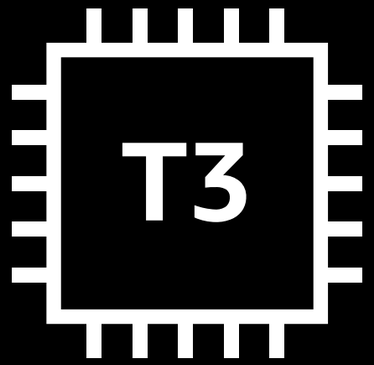
Cost and performance



AWS re:Invent

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

Lower cost – T3 instances



Start as low as $0.09/hr

Upgrade to large instances as workload grows; use with start / stop instances to further lower your cost

Access full Neptune functionality

Each cluster can have 64 TB of storage and up to 15 read replicas

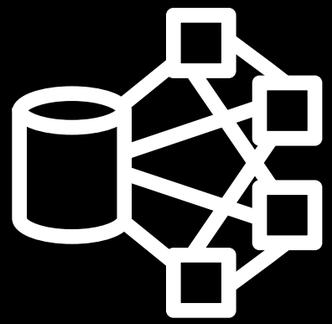
Use for development and test use cases

Burstable instances with CPU credits offer an easy way to get started with Neptune



Cost and performance

Lower cost – Start / stop cluster



Stop instances when not in use

For clusters that need not be up for shorter windows (up to 7 days)

Remove instances for longer periods

For clusters that need not be up for larger windows

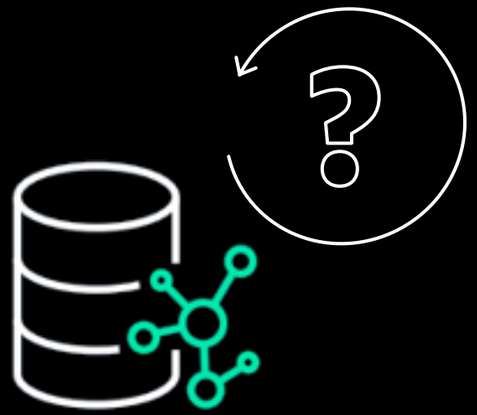
Storage continues to be with the cluster

Decoupled model lets you bring up the cluster anytime and resume processing



Cost and performance

Performance – Query Engine (DFE)



Efficient use of system resources

Optimizes resource allocation and execution for certain queries; DFE usage can be examined in the explain plan

Use statistics for query execution

Autogenerated statistics that are triggered when 10% of your graph changes or updates are older than 10 days

Lab mode for Gremlin and SPARQL

Set DFEQueryEngine parameter to enable for development or test use

Cost and performance

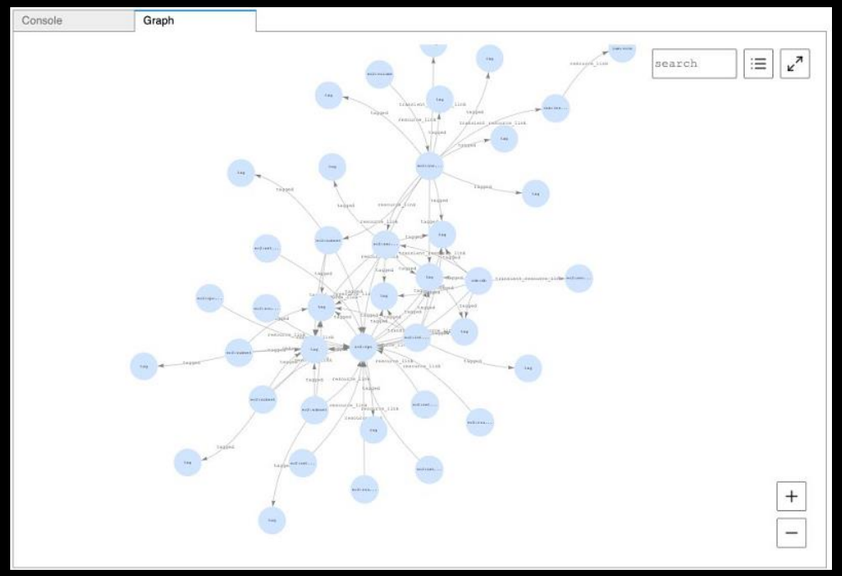
Enterprise capabilities



AWS re:Invent

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

Graph visualization



Graph visualization for Gremlin and SPARQL

Jupyter notebook to run queries and visualize the graph data

Open-source graph notebook

Use graph notebook in Amazon SageMaker, Amazon EC2, or your local desktop; contribute to the code!

Built-in samples – Graph your AWS resources

Getting-started examples to view relationships between AWS Cloud resources as a graph for audit or compliance use cases

Enterprise capabilities

AWS Database Blog

Graph your AWS resources with Amazon Neptune

by Dave Bechberger | on 28 SEP 2020 | in Amazon Neptune, Database | Permalink | Comments | Share

When building graph applications, you want to get started with graph databases and need a way to do it quickly. If you're new to using graph databases, you may find it easy to relate to examples or demos using your own data. Specifically, you want examples that highlight the best practices in Amazon Neptune and applications that showcase the features available. These sample applications can help you right-size your cluster, model your data as a graph, provide best practices in coding, or architect your graph application for scale.

In this post, we walk through an example we released for Neptune with integration with Altimeter. Altimeter is an open-source project (MIT License) from Tableau Software, LLC that scans AWS resources and links these resources into a graph. You can store, query, and visualize the data in Neptune. You can query the graph to examine the AWS resources and their relationships in an account. For example, you can query for resources or pathways that expose a cluster with a public IP address to check for security and compliance.

Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run business critical graph applications. Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying your graph with milliseconds latency.

Solution overview

The AWS CloudFormation template associated with this post automates deploying and configuring Altimeter within a Neptune notebook. The template creates a new Neptune notebook that is associated with the specified Neptune cluster. It also installs Python 3.7, which is required for Altimeter, and installs and configures a version of Altimeter that is optimized for use inside a Neptune notebook environment.

The template interacts with or creates the following AWS resources:

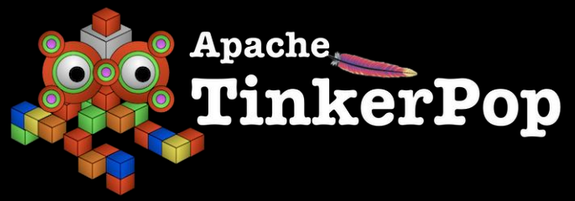
- **Neptune notebook** – The template creates a notebook with Python 3.7 installed and a modified version of Altimeter that automatically configures the account to scan based on the AWS Identity and Access Management (IAM) permissions given to the notebook.
- **Neptune cluster** – The template doesn't create this resource; you can either use an existing cluster or create a new Neptune cluster. For this post, a small Neptune cluster (t3.medium or r5.xlarge) should be sufficient for most accounts.

The following diagram illustrates this architecture.

Resources
Getting Started
What's New
Top Posts
Official AWS Podcast
AWS Case Studies

Follow
Twitter
Facebook
LinkedIn
Twitch
RSS Feed
Email Updates

Query language enhancements







+



Amazon Neptune Gremlin client

Load balance requests using Java client for connection management, retry logic, and exception handling

Support for Apache TinkerPop 3.4.8

New features and improvements, such as the elementMap() step and an improved behavior for working with map instances

New SPARQL and Gremlin query hint

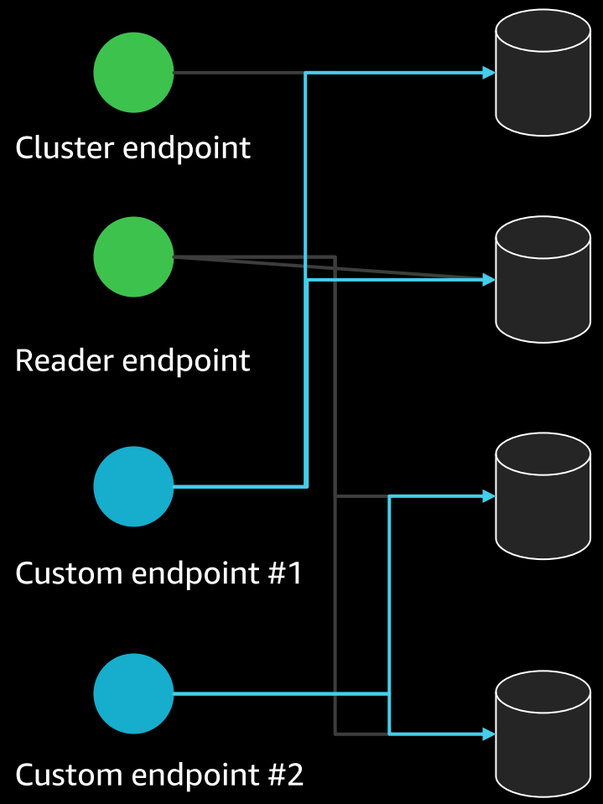
Turn on / off type promotion depending on whether you need to find values of a given type



Enterprise capabilities

Custom endpoints

New



Load balance requests specific to your workload

Create up to 7 custom endpoints to map workload patterns to set of instances; you can continue to use reader and cluster endpoints

Map existing and new instances

Load balance read queries within an endpoint; configure option to add new instances automatically to custom endpoints

Manage using Console, CLI or API

Easy to use interfaces to create and manage custom endpoints

Enterprise capabilities

Fast database reset

New



Reset a graph quickly in minutes

For workloads that load full data periodically or for testing

Optimized to execute at low cost

Fast reset operation does not result in additional I/Os

Execute using Neptune Workbench or API

REST API to reset or use the workbench %db_reset command

Enterprise capabilities

Integrations and ML

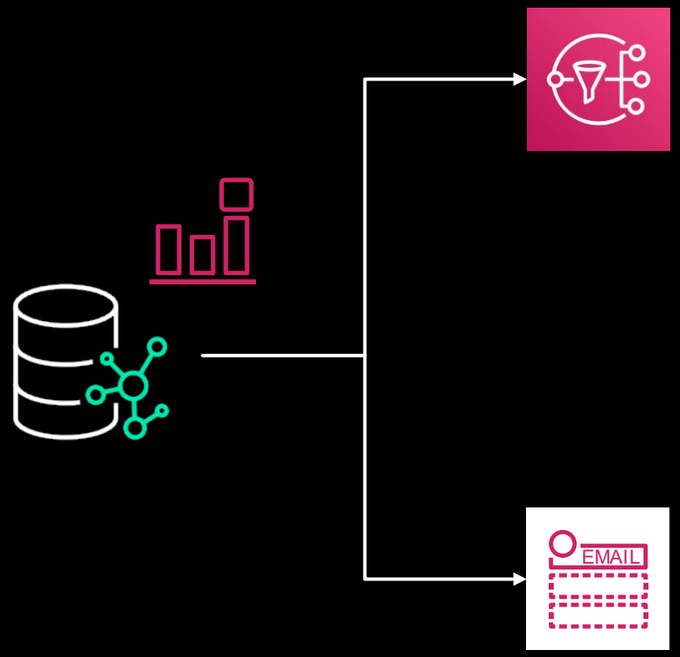


AWS re:Invent

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

Event subscriptions

New



Get notified when interesting events occur

Choose from a host of event types including, maintenance, configuration changes, backup, and so on

Automate using Amazon SNS

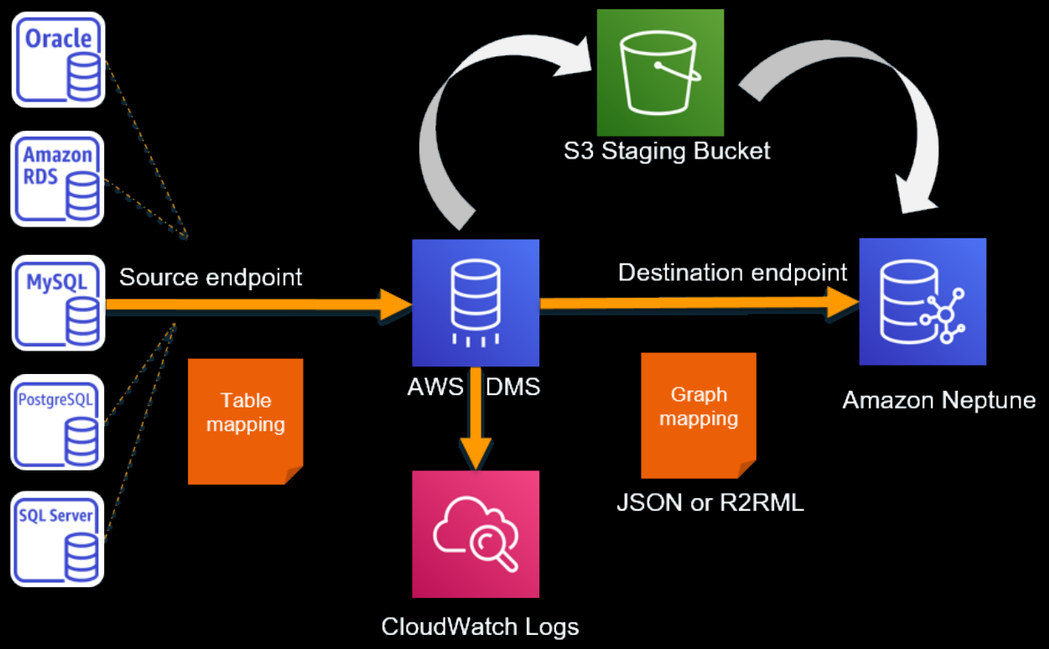
Subscribe to events through email, text, or HTTP endpoint calls

Manage using Console, CLI, or API

Create subscriptions or view events for up to the last 14 days

Integrations and ML

AWS DMS support for Neptune



Copy from relational DB to Neptune

Use relational database as source and Neptune as target

Choose source tables or views

Choose tables and views from the relational database

Flexible configuration to create RDF or property graph model

Use R2RML for SPARQL and JSON configuration for Gremlin



Integrations and ML

Amazon Neptune: Key innovations

IT'S BEEN A BUSY YEAR FOR US!

Cost and performance

- T3 instances
- Start / stop cluster
- Federated query
- Alternative query engine

Enterprise

- Language enhancements
- DMS support
- Neptune streams
- Database cloning
- Tags in IAM
- Graph visualizations
- Cross-Region snapshot copy
- Delete protection
- Fast DB reset (new)
- Custom endpoints (new)

Integrations and ML

- Neptune ML (new)
- Amazon ES integration
- Kinesis integration
- Event subscriptions (new)
- DMS integration
- Export graph

Related sessions

DAT309

Deep dive on Amazon Neptune

DAT315

ADP's next-generation platform powers dynamic teams with Amazon Neptune

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

Thank you!

Karthik Bharathy

@kb_chirps

© 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.