AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

DAT329-R2

Caching for high-volume workloads using Amazon ElastiCache

Michael Labib
Head of Specialist SA, Emerging Databases
AWS

Sashi Varanasi
WW Leader, In-Memory DB Specialist SA
AWS

Roberto Luna Rojas
WW Specialist SA, In-Memory DBs
AWS

Siva Karuturi
WW Specialist SA, In-Memory DBs
AWS

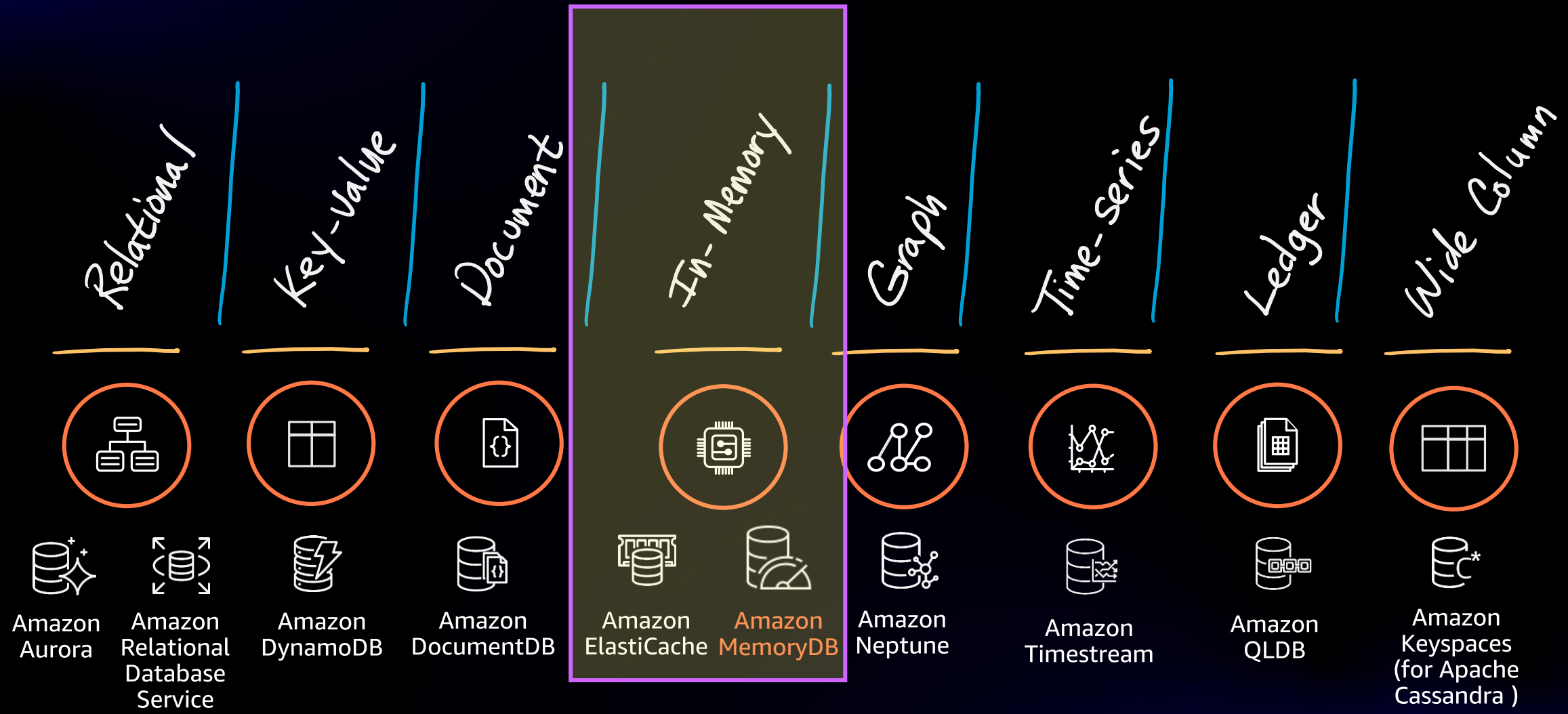
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Introduction
- High-volume workload requirements
- Caching strategies
- Hands-On lab: RDS Aurora + Amazon ElastiCache

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Purpose-built databases



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon ElastiCache for Redis

Fully-managed, Redis compatible, in-memory data store in the cloud



Extreme Performance

In-memory data store and cache for sub-millisecond response times

Fully Managed

AWS manages all hardware and software setup, configuration, monitoring

Easily Scalable

Read scaling with replicas

Write and memory scaling with sharding

Non disruptive scaling

Reliable

Multi-AZ

Deep monitoring

Automatic failover

Secure & Compliant

Amazon VPC

HIPAA, PCI, FedRAMP

Encryption at-rest and in-transit

Authentication

Redis Compatible

Redis 5 Support

Redis clients compatible

AutoScaling

Automatically add shards and replicas to your cluster when certain criteria are met

Data tiering

Save cost by automatically moving data between memory and SSDs by using LRU algorithm

aws

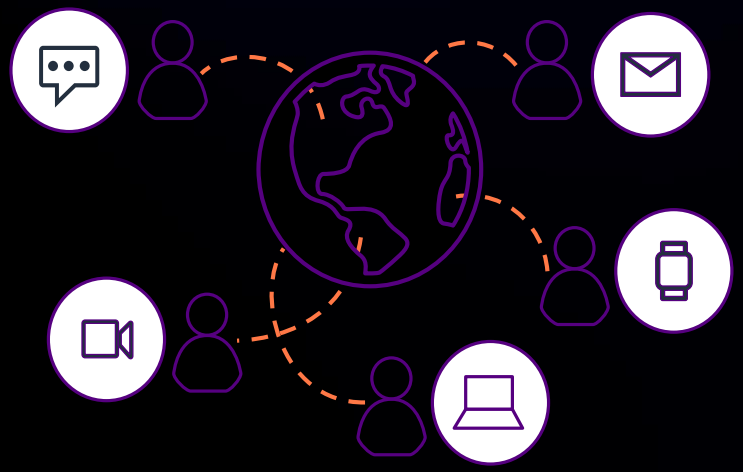
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

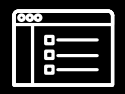
High-volume real-time applications

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern real-time applications require

PERFORMANCE, SCALE & AVAILABILITY





E-Commerce



Media streaming



Social media



Online gaming



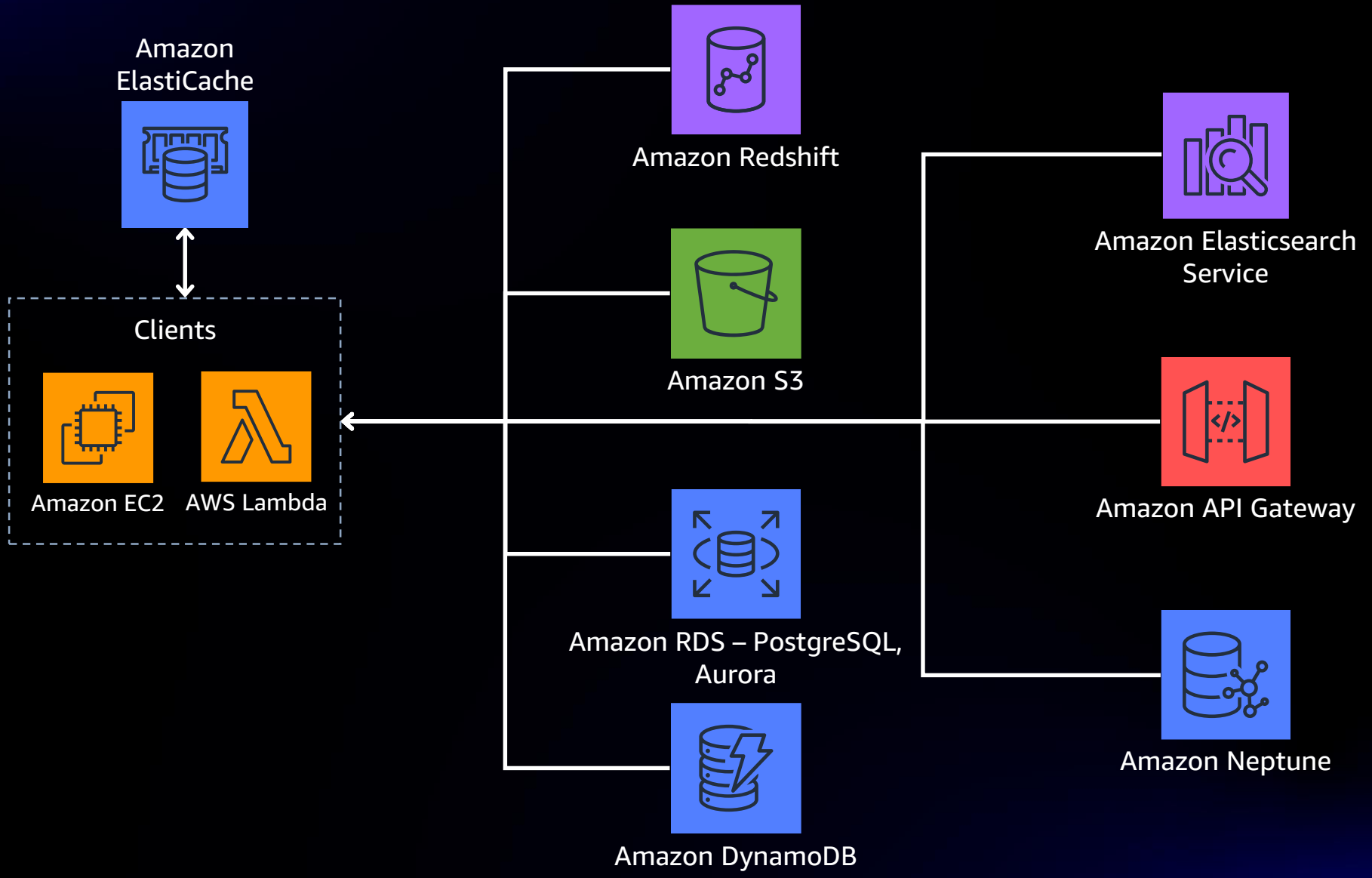
Shared economy

Users	1M+
Data volume	Terabytes—petabytes
Locality	Global expansion
Performance	Microsecond latency
Request rate	Millions per second
Access	Mobile, IoT, devices
Scale	Up-out-in
Economics	Pay-as-you-go
Developer access	Open API

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

In-Memory Data store – storing any data that can be queried



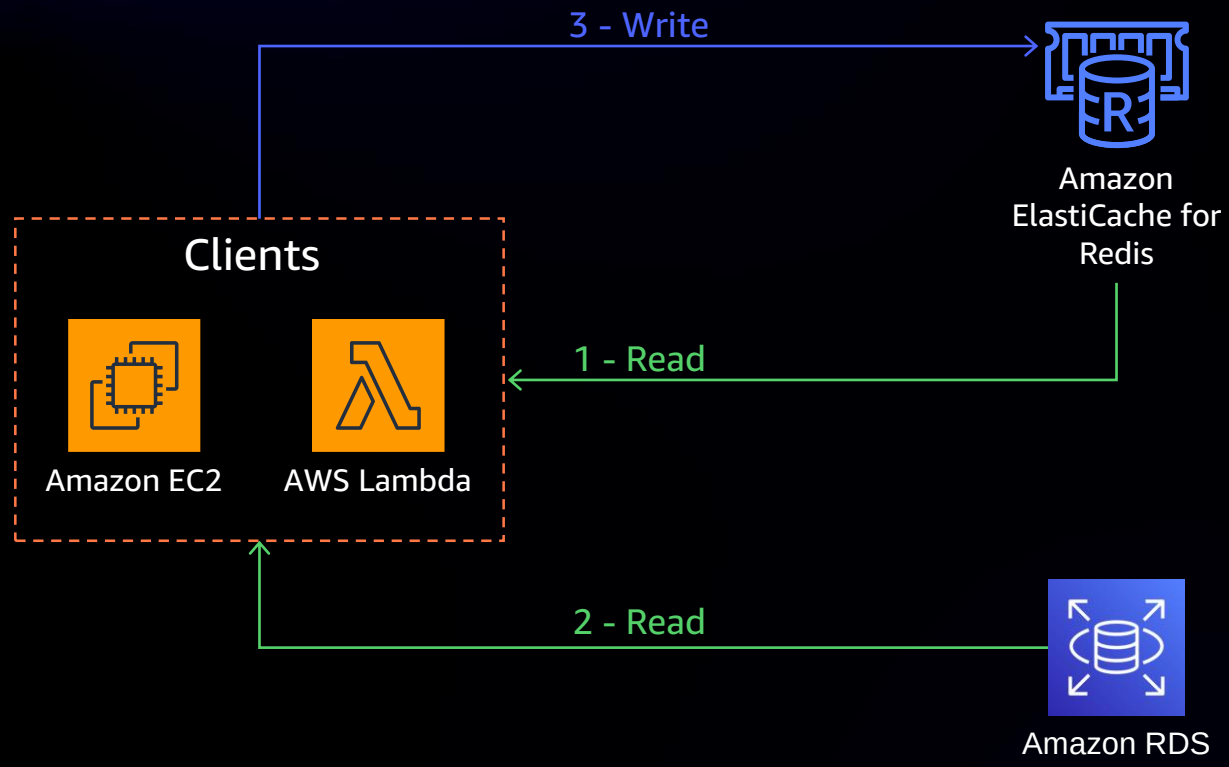
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Populating in-memory data store

Lazy load and write-through patterns

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lazy loading – pattern overview

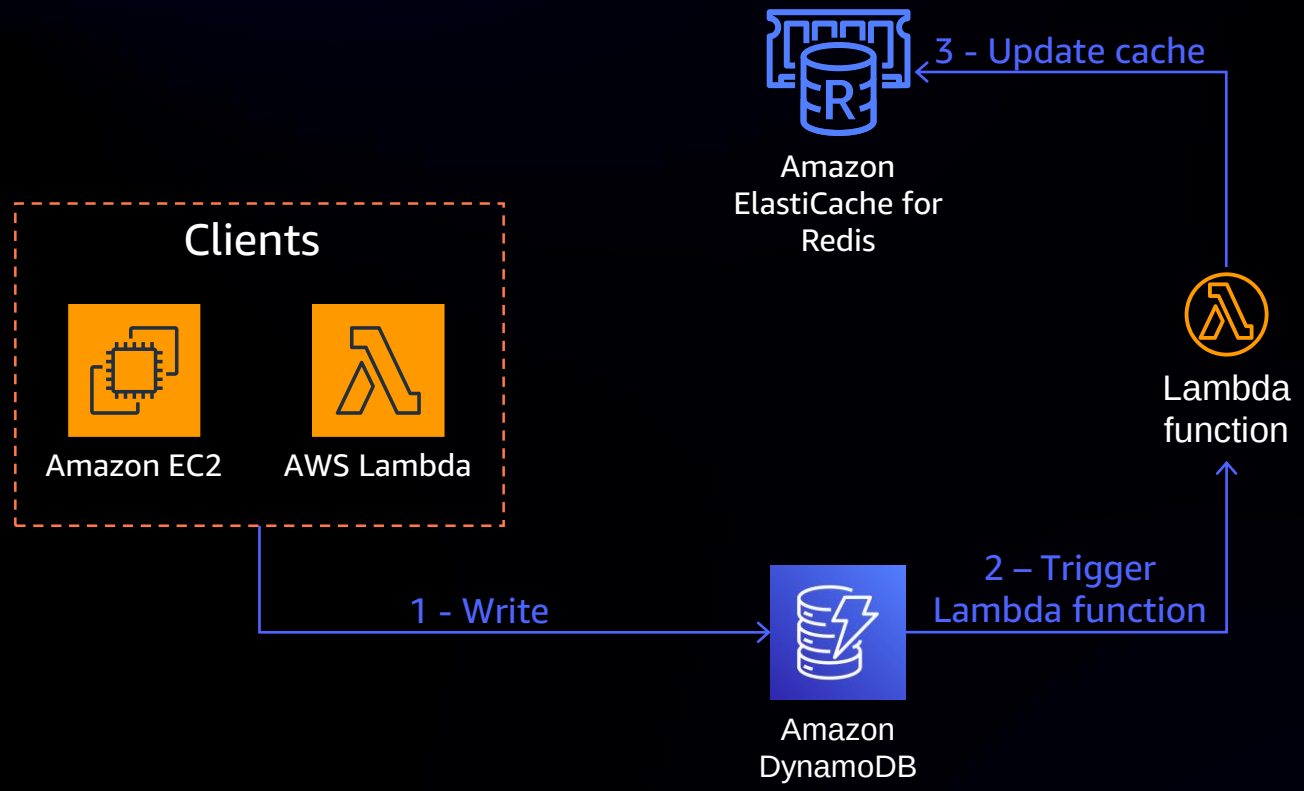


1. Read from cache
2. Read from source (if miss)
3. Write to cache

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Write-through (backend)

Example 1

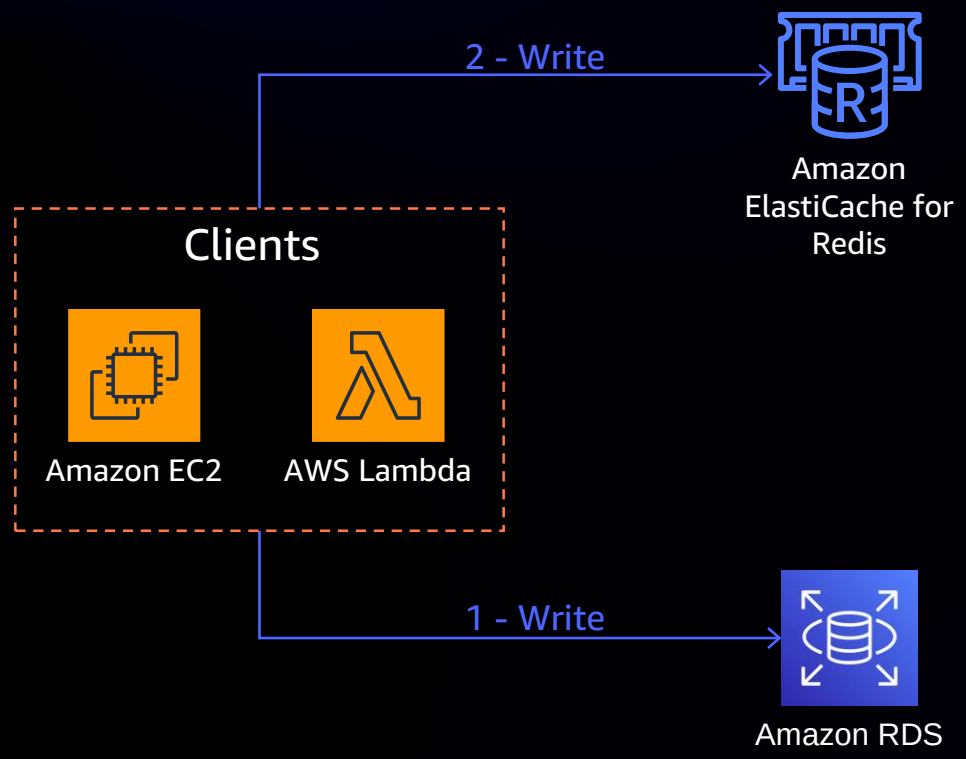


1. Write to source
2. Triggers AWS Lambda function
3. Lambda function updates cache

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Write-through

Example 2



1. Write to source
2. Write to cache

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pre-populating cache

- Use an ETL process that is executed on a regular schedule
- Warm up the cache with the relevant data needed for your application

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon ElastiCache for Redis

Data structures & Use cases

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Redis Data Structures

Data Type	Description	Example Use Cases	Associated Commands
String	Sequence of Bytes	Caching	APPEND,GET,SET,INCR,DECR,GETSET...
List*	A list of Strings	Many	LSET,LLEN,LPUSH,LPOP,LTRIM,RPOP...
Set*	Non-repeating, unordered collection of Strings	Cardinality	SADD,SCARD,SDIFF,SUNION,SINTER,SMEMBERS...
Sorted Set*	Non-repeating, ordered collection of Strings	Leaderboards	ZADD,ZCARD,ZCOUNT,ZRANK,ZSCORE...
Hash*	Map of Key / Value Pairs		HGET,HGETALL,HKEYS,HVALS,HMSET,HMGET...
Streams	Log data structure	Message Queue	XADD,XRANGE,XREAD,XACK,XCLAIM,XLEN...
Geospatial	Longitude / latitude – based entries	Maps, "nearby"	GEOADD,GEODIST,GEOPOS,GEORADIUS...
Bitmaps	Special usage of String type		GETBIT,BITCOUNT,SETBIT,SETRANGE,GETRANGE...
HyperLogLogs	Special usage of String type		PFADD,PFCOUNT,PFMERGE

** These data types are capable of storing over 4 billion sub elements*

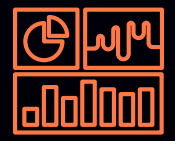
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lots of use cases for real-time apps



Caching



Real-time analytics



Gaming leaderboards



Geospatial



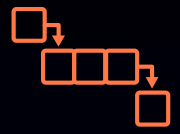
Media streaming



Session store



Chat apps



Message queues



Machine learning

aws

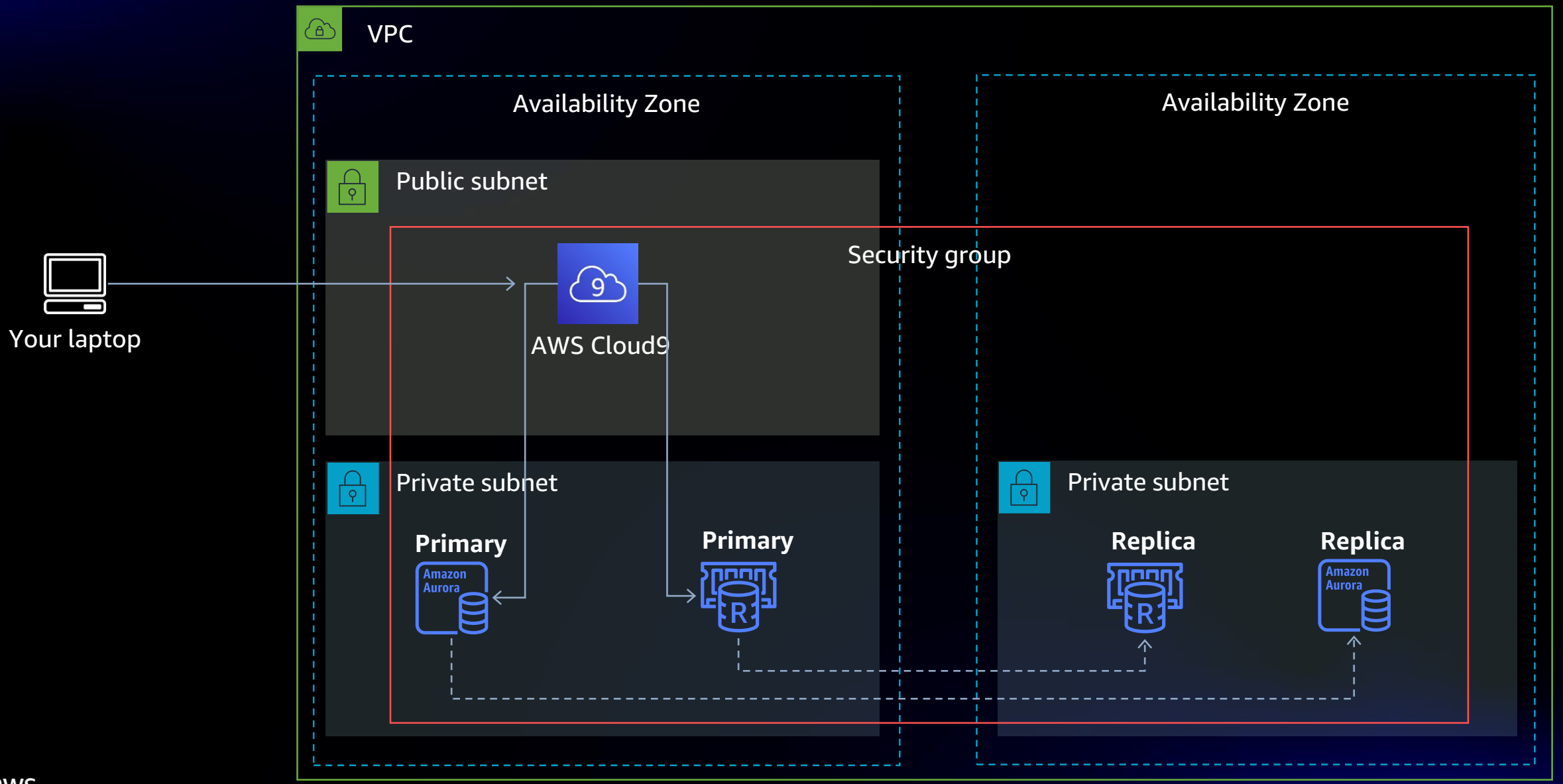
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hands-On lab:

Amazon Aurora + Amazon ElastiCache

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

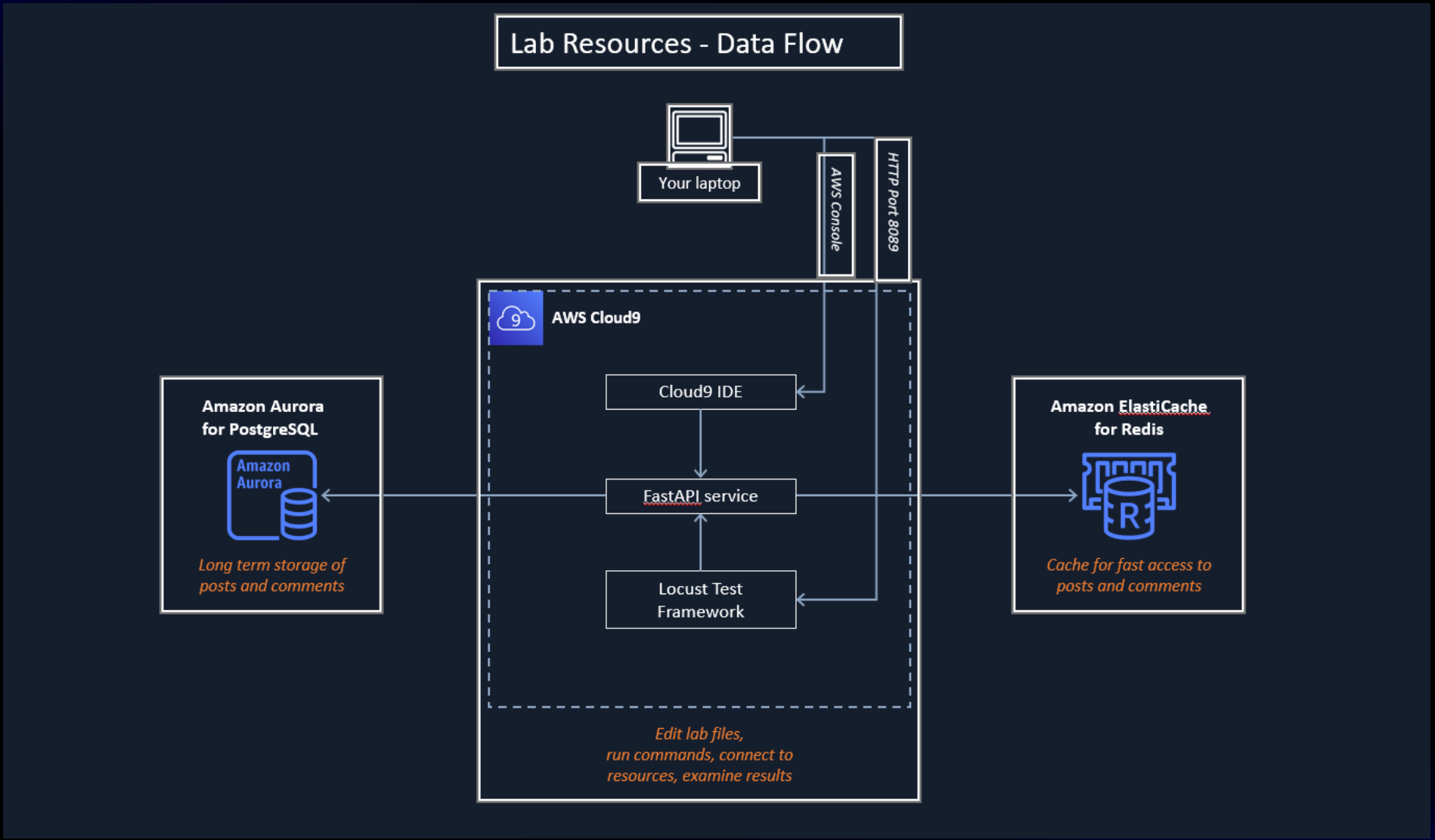
Lab: RDS Aurora + Amazon ElastiCache



Lab Resources - Architecture

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lab resources – Data flow



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

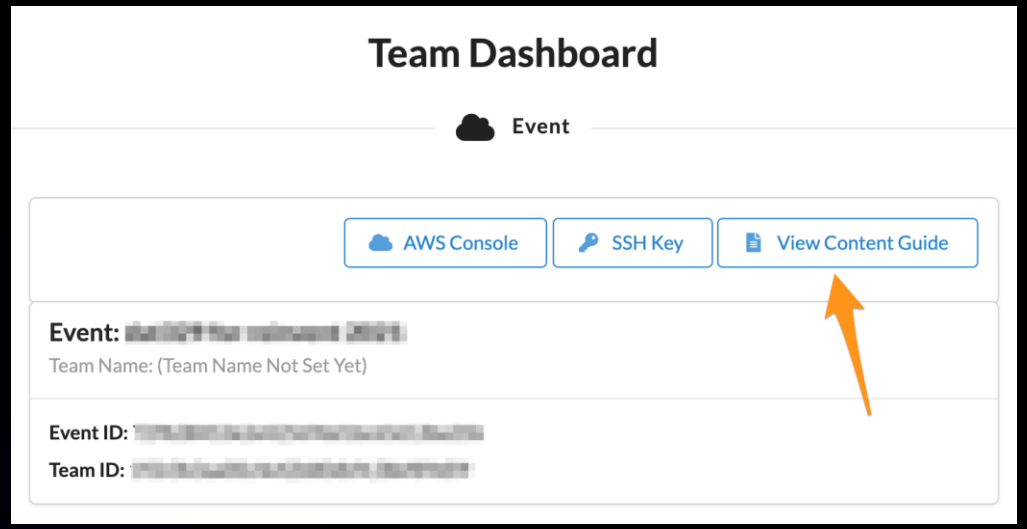
Attendee guide/Instructions for the lab

https://tiny.amazon.com/mnri4c5/dat329

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

https://dashboard.eventengine.run

149e-179e23e6d4-af



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Sashi Varanasi
vsashi@amazon.com

Michael Labib
mlabib@amazon.com

Roberto Luna Rojas
rberoj@amazon.com

Siva Karuturi
maravisk@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.