Managing IoT and Time Series Data with Amazon ElastiCache for Redis

Michael Labib, Specialist Solutions Architect

August 2016

Amazon ElastiCache

redis

Learning Objectives

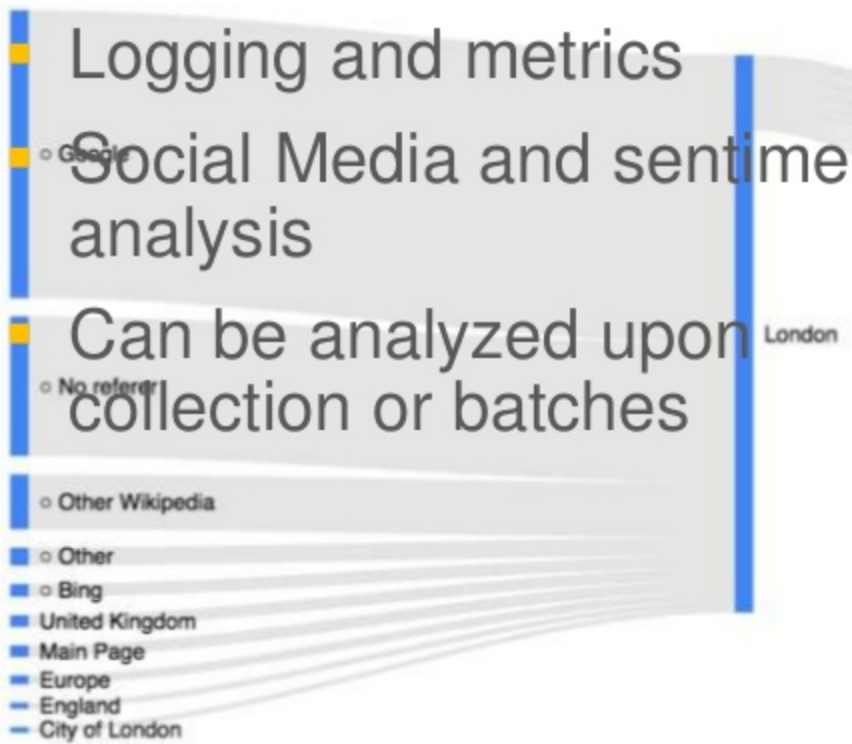
- Understand Time Series Data & Challenges
- Learn about Amazon ElastiCache for Time Series Data
- Learn how to build Time Series Solutions with Amazon ElastiCache
- Explore a Sensor Data Demonstration

2

Time Series Data

Time Series Data – What is it?

- Device / Sensor Data
- Website Clickstream events
- Logging and metrics
- Social Media and sentiment analysis
- Can be analyzed upon collection or batches



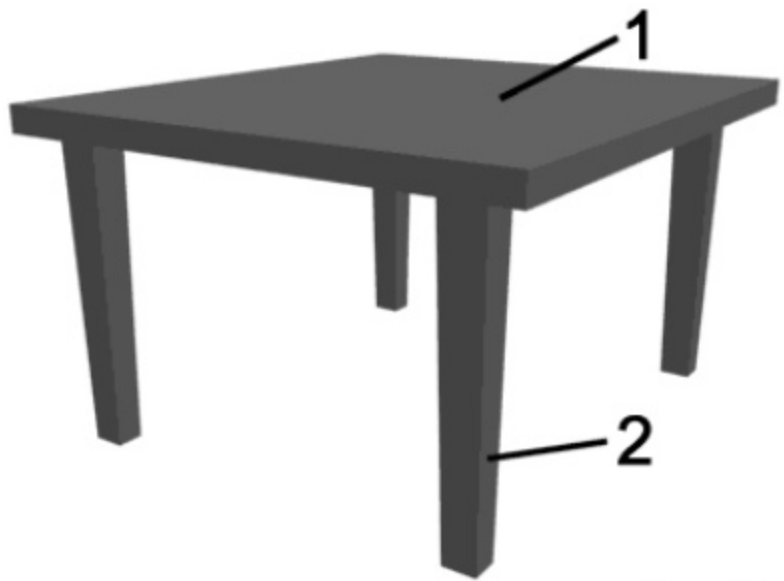
Time Series Data – Challenges

- What type of database should you use?
  - Relational or NoSQL
- How should you model the data to support the queries and analysis needed?
  - What types of aggregations will you need?
  - What type of information would do you need to gather?
- How will the applications access the data?
- How do you build the solution in a cost effective manner?
- How long should you retain the data?
- How do you manage scalability given the amount of data and sensors grow?
- And so on…

Database Families

Relational

+ Mature Technology
+ SQL is widely adopted
+ Durable and Consistent
- Inflexible data model
- Hard to scale
- Performance limitations
  - Updates and other changes can be slow
  - Limited to speed of storage technology

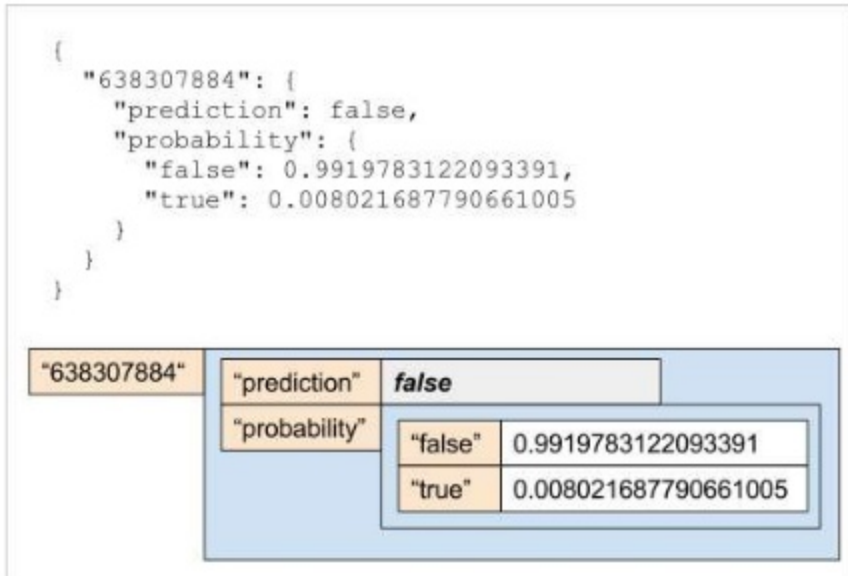


1 = Row
2 = Column

Database Families

NoSQL

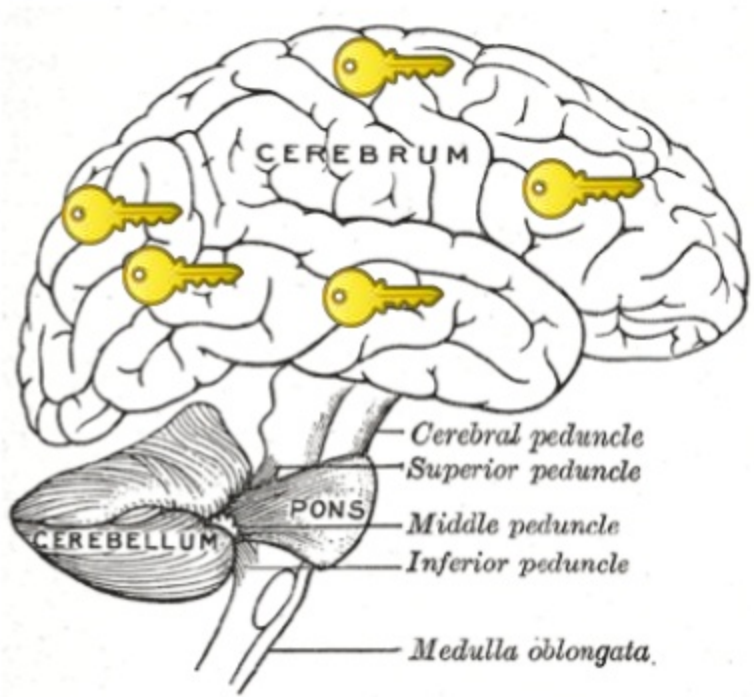
+ 'Schema-less' = flexible data model
+ Quickly adapt to new requirements
+ Easier to scale
+ Performance
+ Generally higher throughput and
  + lower latency than Relational DB
- Can be less durable and consistent
- Query tools less mature

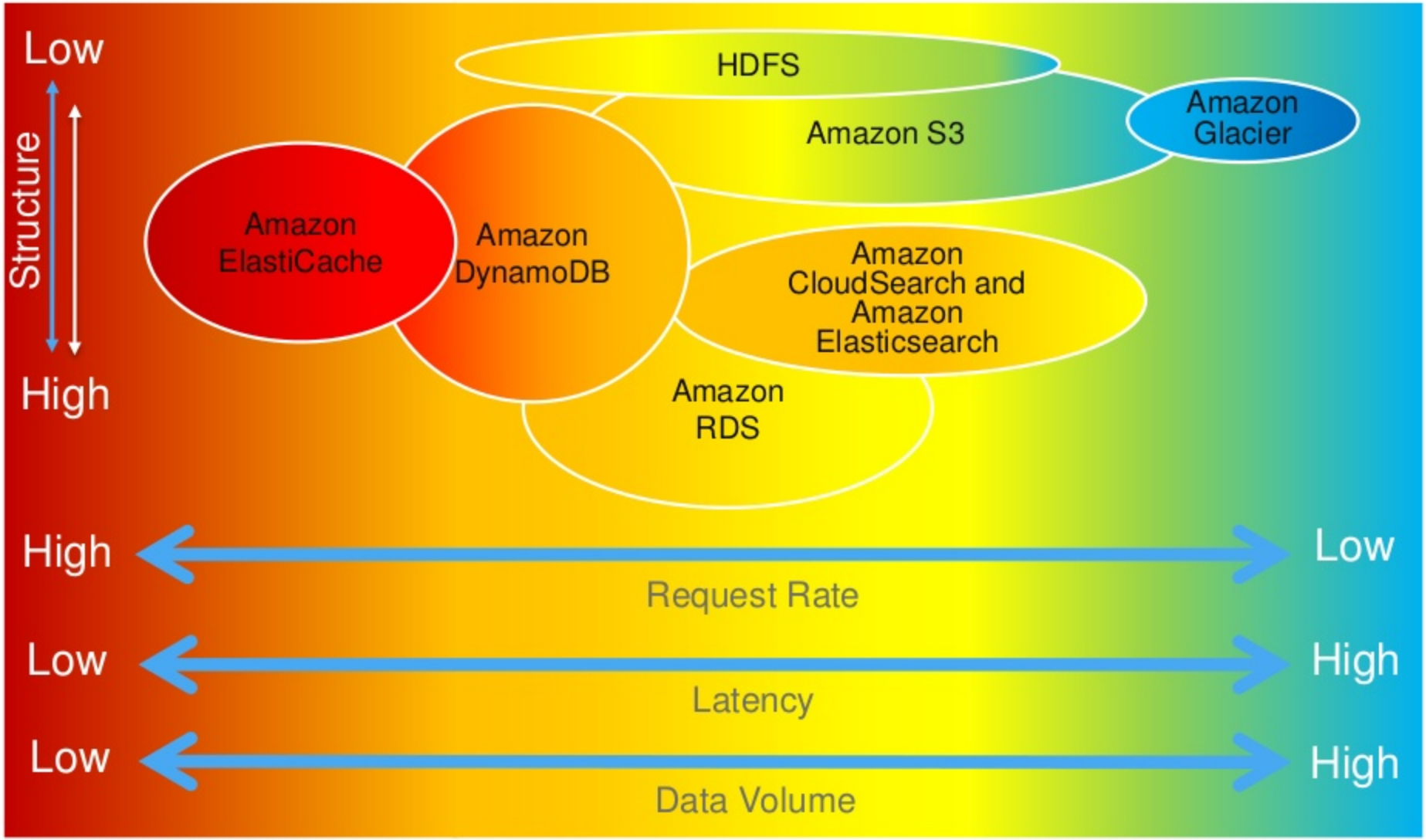


Database Families

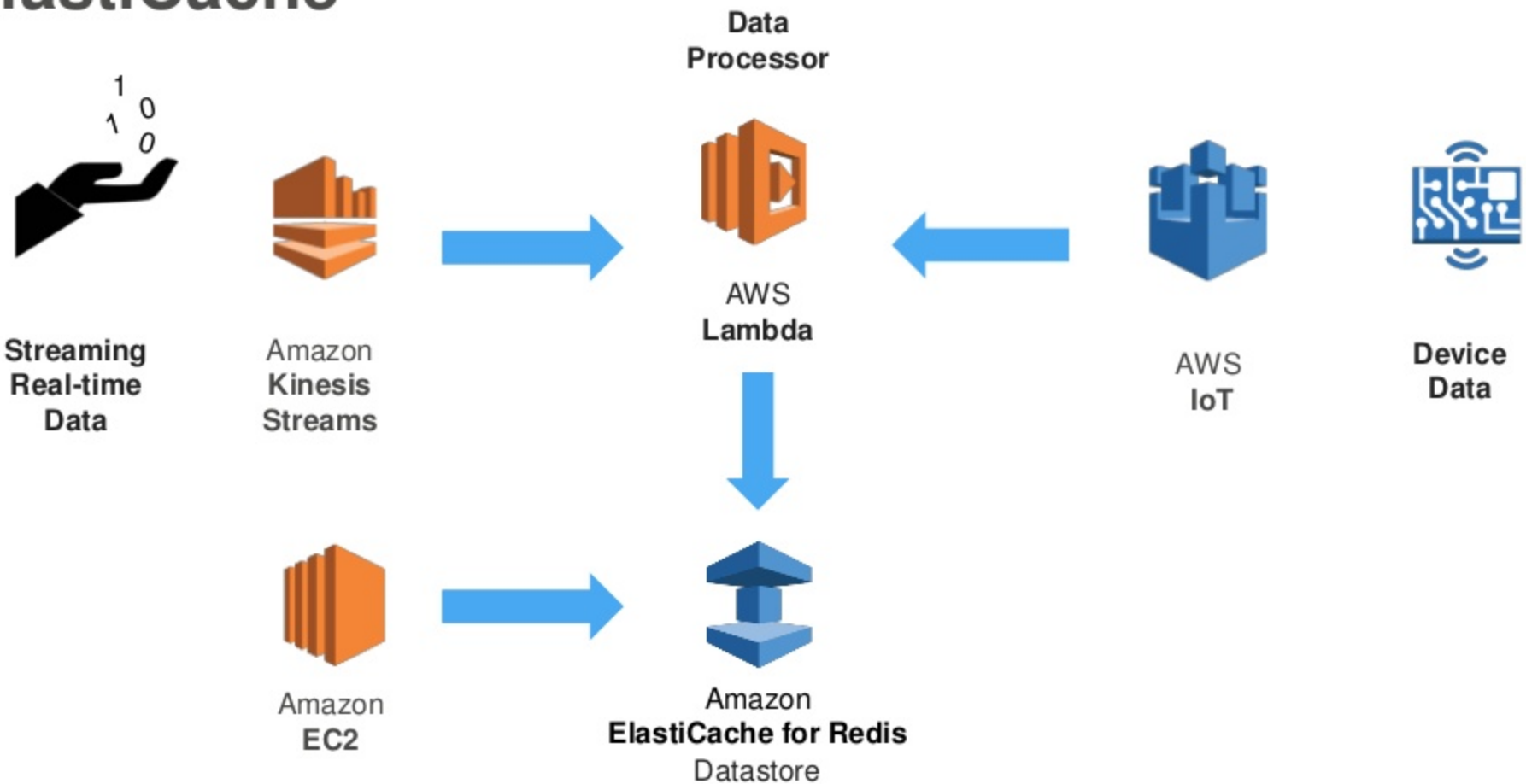
NoSQL: In-Memory Key-Value

+ Very fast performance
  + Sub millisecond reads and writes
  + High throughput
+ Schema-Less – Flexible data model
+ Redis - Advanced Data Structures
- Still emerging technology
  - Query and other tools less mature
  - Less connectors than other database options
- Less durable




Low
Structure
High
HDFS
Amazon S3
Amazon Glacier
Amazon ElastiCache
Amazon DynamoDB
Amazon CloudSearch and Amazon Elasticsearch
Amazon RDS
High
Low
Request Rate
Low
High
Latency
Low
High
Data Volume

Ingesting Time Series Data with Amazon ElastiCache



Amazon ElastiCache Overview

Amazon
ElastiCache

In-Memory Key-Value Store

High-performance

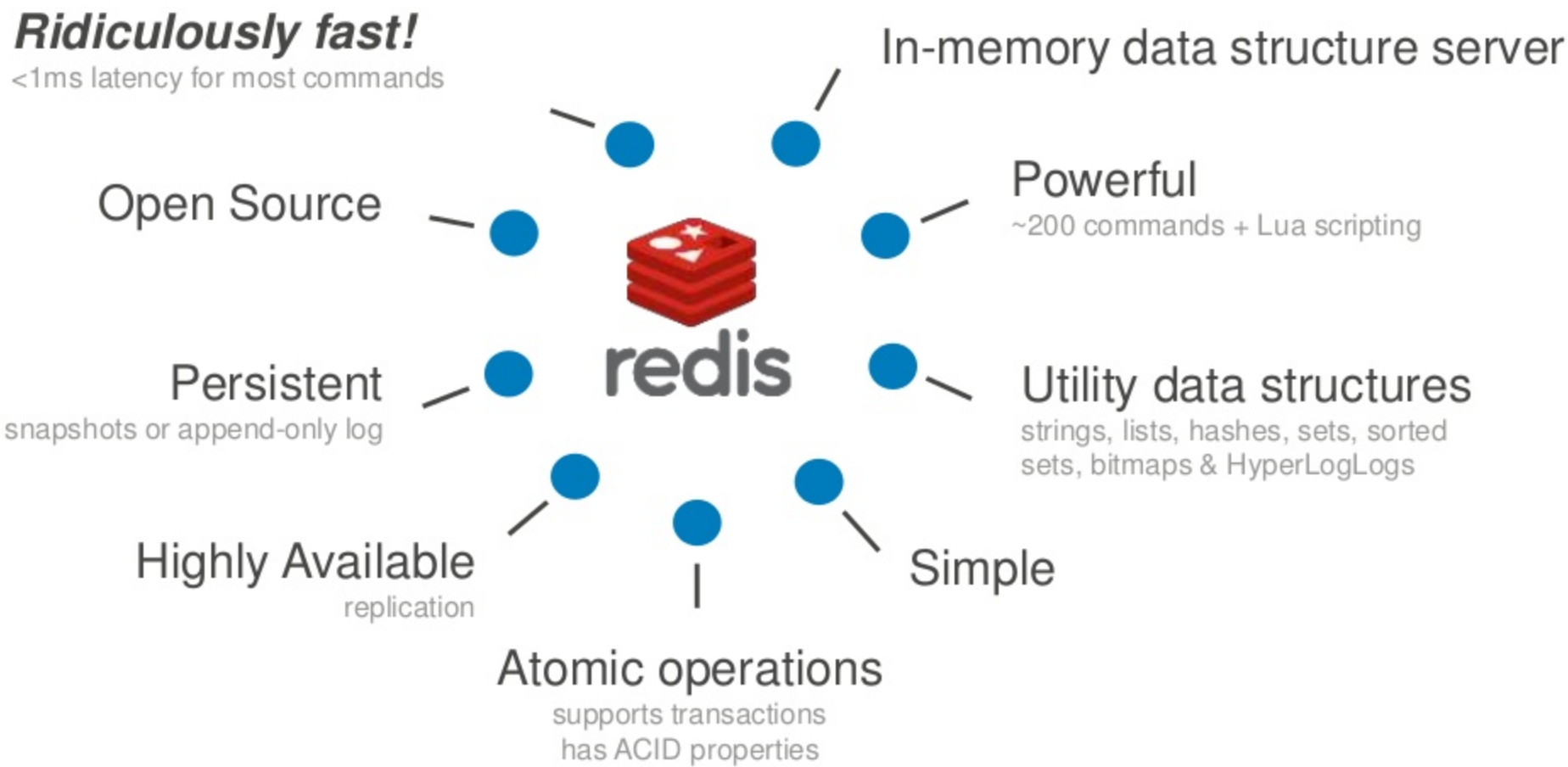
Redis and Memcached

Fully managed; Zero admin

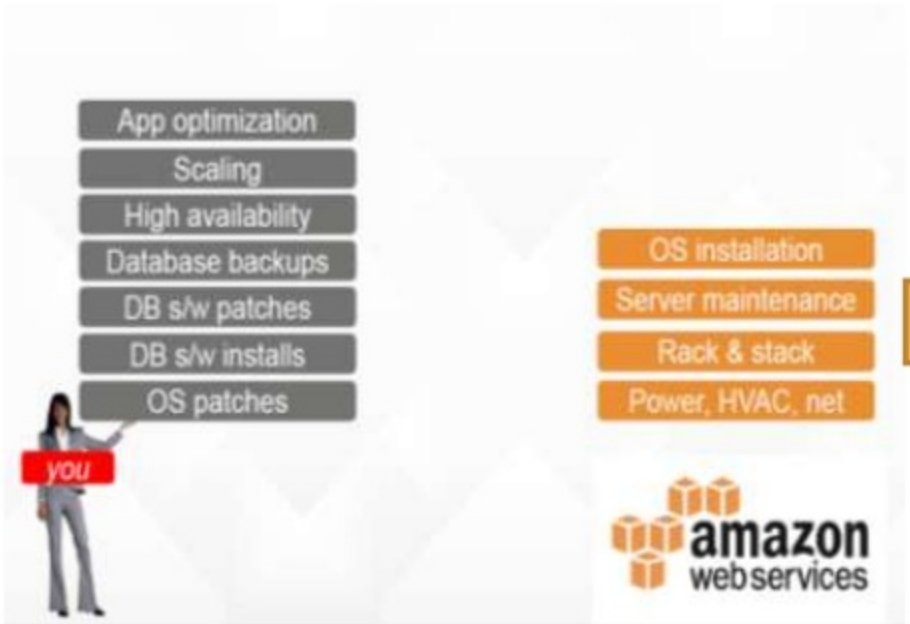
Highly Available and Reliable

Hardened by Amazon

Redis – The In-Memory Leader



Fully managed service = Automated Operations



Redis datastore hosted on Amazon EC2

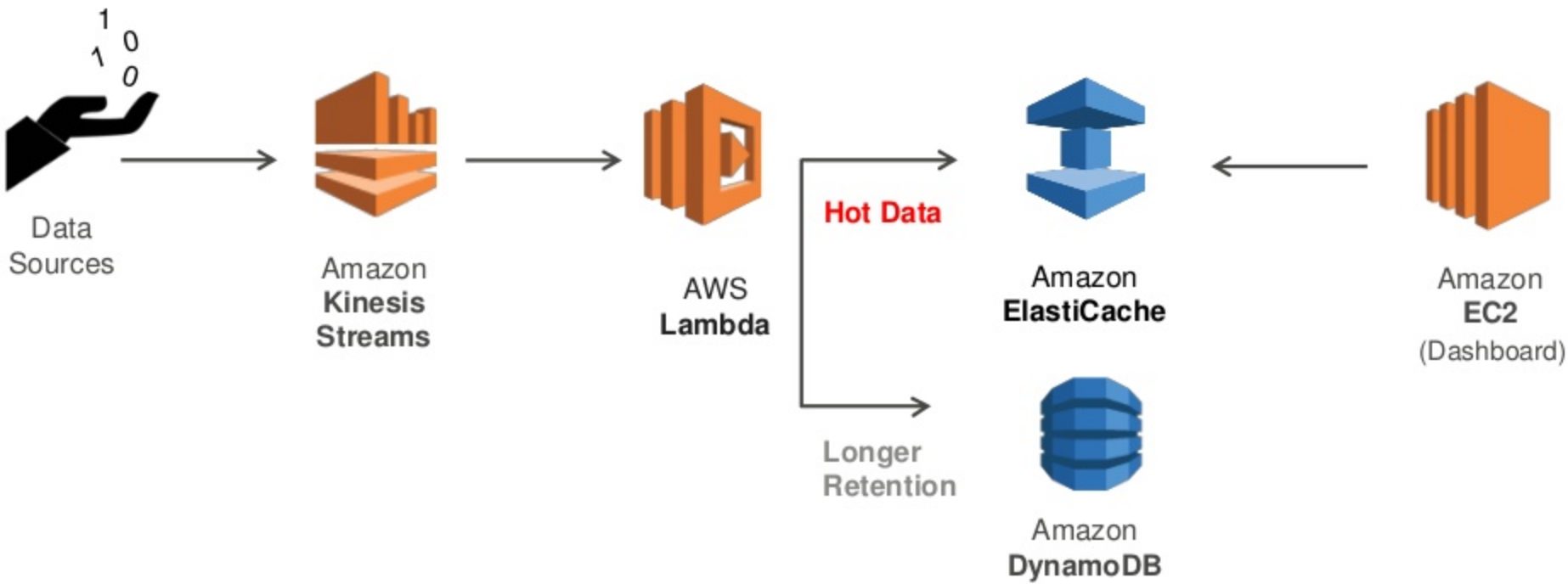
Amazon ElastiCache for Redis

ElastiCache - Customer Value

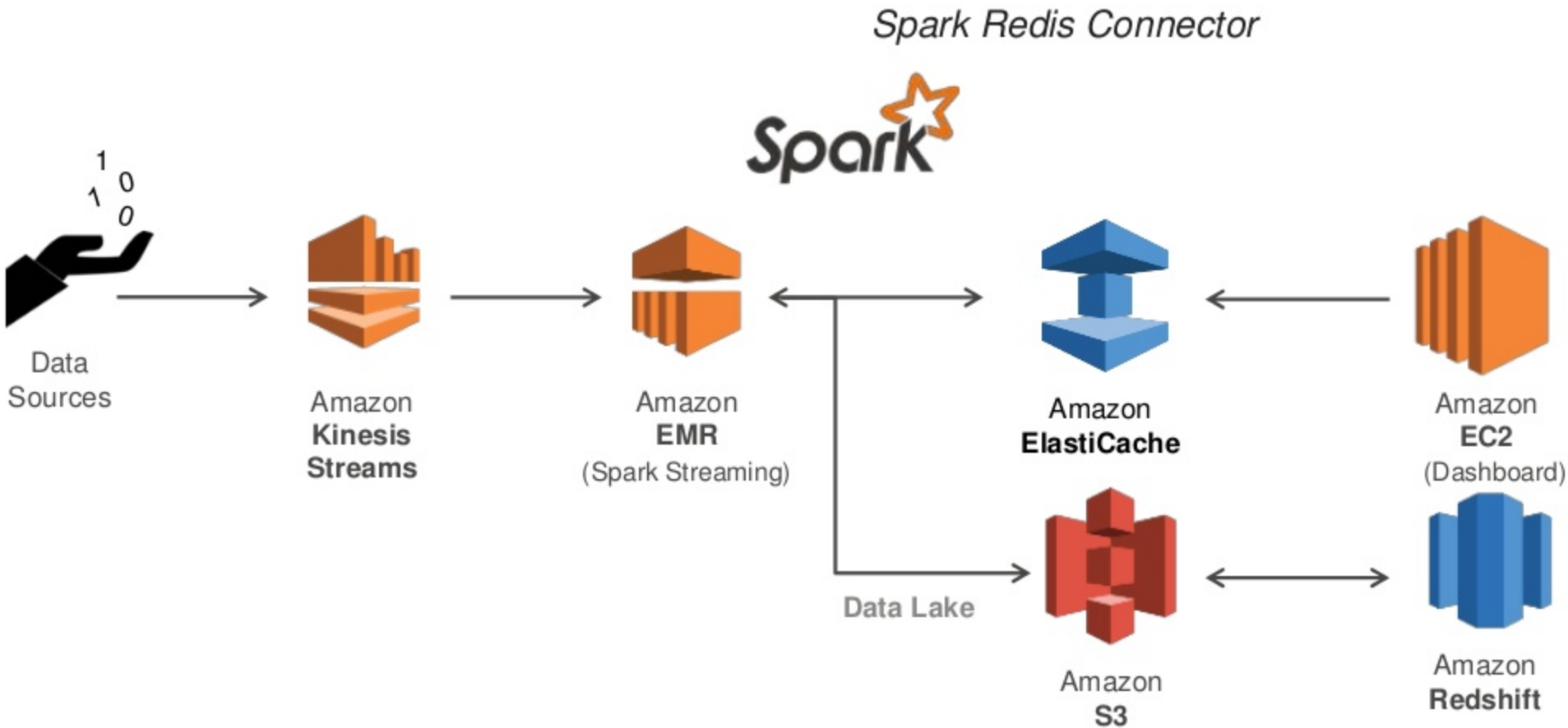
Extreme Performance	Open Source Compatible	Fully Managed	Secure and Hardened	Highly Available and Scalable	Cost Effective
Sub-millisecond access latencies	Compatible with Redis and Memcached	Automates tasks such as failed node replacement, software patching, upgrades and backups	Supports Amazon VPC and IAM for secure and fine grained access.	Multi-AZ with automatic failover to a read replica, no human intervention required.	Pay as low as US$0.017 per hour. Get started with 750 free hours per month of a micro node for a year.
Engineered for Cloud Scale	Existing code will work when you update node end points	CloudWatch enables you to monitor cache performance metrics	Monitors your nodes and applies security patches when necessary	Easily scale your Redis (vertically) and Memcached (horizontally) environments	No cross availability zone data transfer costs.

Building Time Series Solutions with Amazon ElastiCache

Streaming Data



Streaming Data Analytics



Capturing Moving Data

- Equity prices, clickstreams, sensor streams …
- Amazon Kinesis and AWS Lambda are serverless
- Amazon Kinesis + Amazon ElastiCache for Redis ensure order among incoming data
  - Kinesis provides an ordered Stream
  - Redis provides a Sorted Set
- Fault Tolerant
  - Data Retention – Data accessible in Kinesis 24 hours after its been added to the stream and can be extended up to 7 days.
  - ElastiCache provides HA with automatic failover to a read replica
- ElastiCache is fast and flexible
  - Can query and retrieve hot data quickly

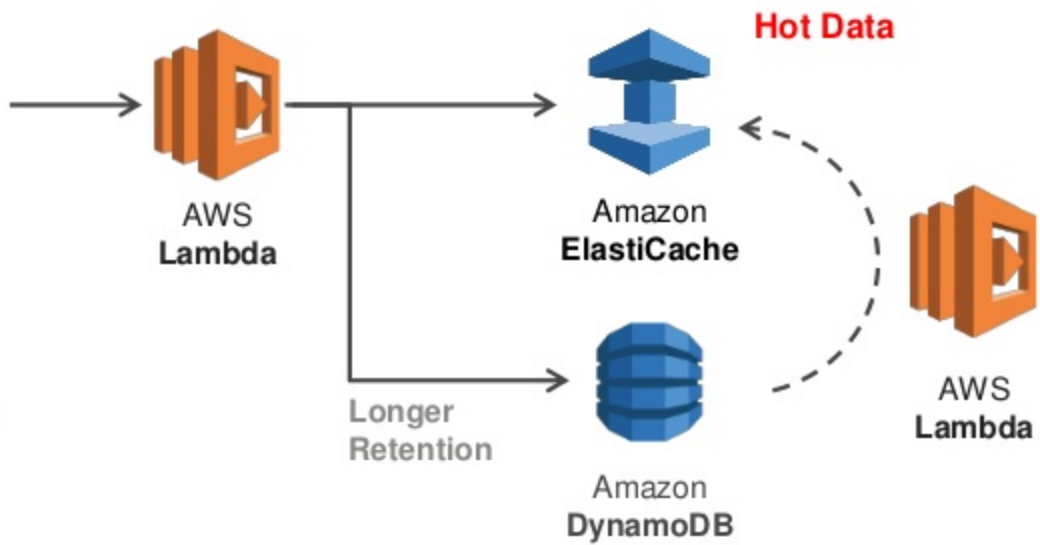
ElastiCache for Streaming Data Analytics

- Access all Redis Data structures directly from Spark
- Significantly accelerate Spark performance over disk-based solutions
- Simplify analytics by leveraging Redis Data Structures reducing processing overhead and complexity
- Improve application performance by accessing aggregations and hot data from Redis

Event Driven Data

Event Sources

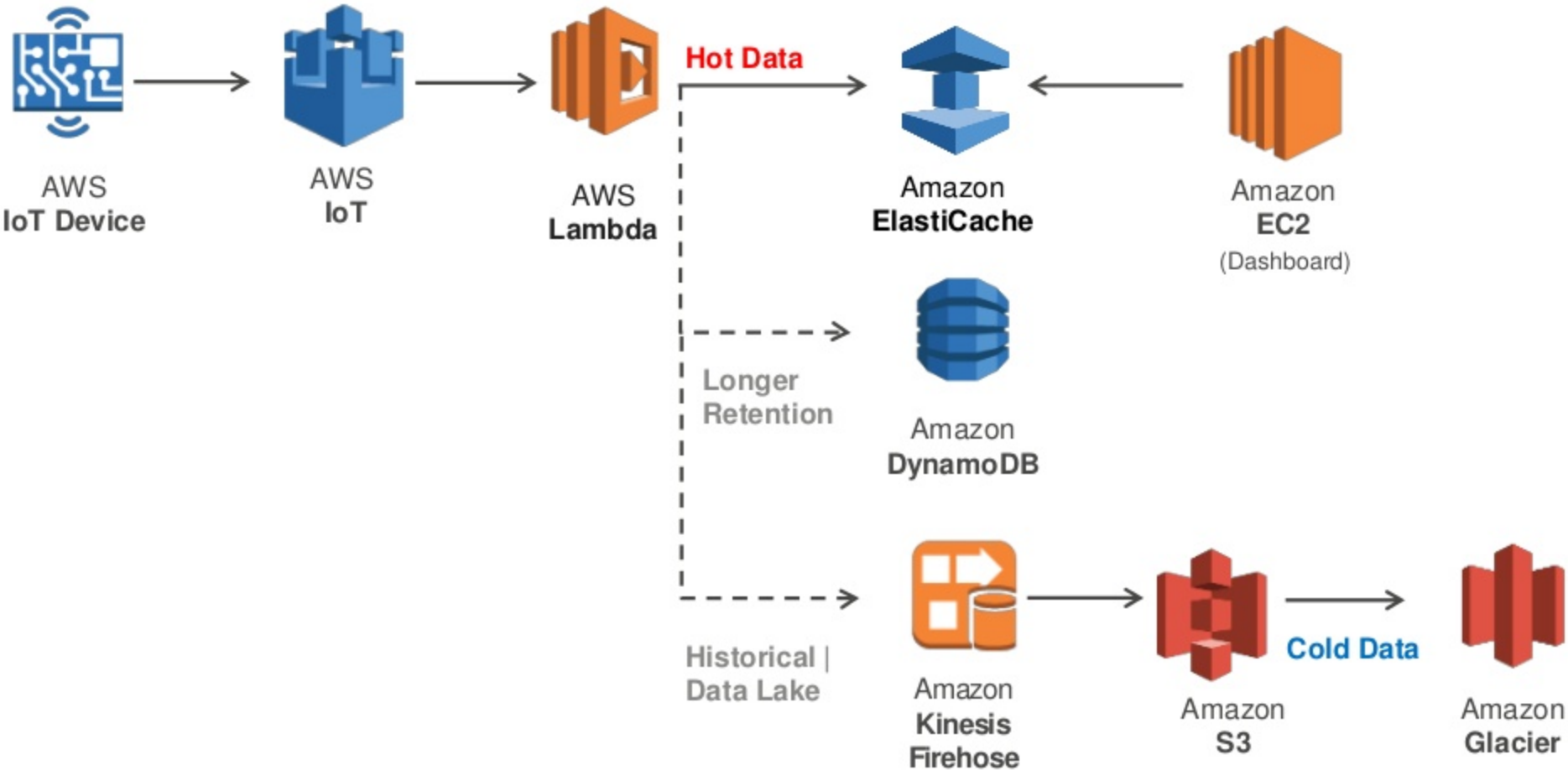
Capture time ordered changes from a DynamoDB table using DynamoDB Streams and propagate to ElastiCache



Responding to Events

- Can cover a huge variety of different data sources
  - Security events, application events, messaging events, user actions…
  - For IoT, often created when a sensor crosses a threshold
- Lambda captures the incoming event data and routes it
  - Hot data to ElastiCache
  - Cold data to slower, disk-based databases
  - Lambda function to keep databases synchronized
- Cost effective
  - Use the speed and flexibility of ElastiCache for hot data queries
    - No incremental cost per query – price is driven by amount of memory available
  - Use the DynamoDB or RDS for longer data retention

Device Data


AWS
IoT Device
AWS
IoT
AWS
Lambda
Hot Data
Amazon
ElastiCache
Amazon
EC2
(Dashboard)
Longer
Retention
Amazon
DynamoDB
Historical |
Data Lake
Amazon
Kinesis
Firehose
Amazon
S3
Cold Data
Amazon
Glacier

Sensors and Data Bursts

- "We are all glorified motion sensors. Some things only become visible to us when they undergo change" – Vera Nazarian, author
- Sensors enable detailed understanding of conditions
  - Quickly perceive and react to changing conditions
- Device data can be streaming (time-based delivery) or event driven (threshold-based delivery)
  - Or both! A temperature sensor might deliver regular readings and also send an alert when it becomes too hot or too cold
- Sensor data solutions must scale to meet loads
  - Sensor clusters can generate large data bursts
  - AWS IoT auto scales
  - ElastiCache for Redis allows flexible sizing to meet changing workloads

Data Modeling for Sensors

- Sensor data is usually of high variety
  - Different sensors, and different versions of the same sensor, collect different data in different formats
  - A single device can have many sensors
- The flexible schema of ElastiCache for Redis allows non-disruptive evolution of data models
  - Quickly test and use new or modified data models with no need to migrate schemas, rebalance shards or other onerous administrative tasks

Sensor Demo

Sensor Demo

Intel Edison –
Grove Temperature Sensor

http://www.seeedstudio.com/wiki/Intel%C2%AE_Edison_and_Grove_IoT_Starter_Kit_Powered_by_AWS

Sensor Demo - Data Model

What were the last N events?

zrevrangebyscore SensorData +inf -inf WITHSCORES LIMIT 0 5

Sorted Set

SensorData

timestamp

deviceId

Time-Series events
Sort by timestamp (score)

Hash

deviceId

deviceIP

temperature

humidity

climate

requestId

timestamp

Strings (INCR/DECR)

climate:hot

climate:warm

climate:cool

climate:cold

Aggregations
For dashboards, analytics, etc.

Which events occurred at a particular time range?

zrangebyscore SensorData (1410000000000 14400000000000

Pub/Sub
If (temperature> 95 F)
publish notification

Event details

What are the totals for each climate type?

get climate:warm
get climate:cool
get climate:cold
get climate:hot

Which devices match a particular pattern?

zscan SensorData 0 MATCH 2* COUNT 100

What is the temperature for a specific device?

hget 2221 temperature

What is all the data for a specific device?

hgetall 2221

Data structures updated within a
transaction block using **Redis MULTI**

Summary

- Time series data can present high volume, velocity and variety challenges
- ElastiCache for Redis is scalable, performant and cost effective service for sensor data and other time series data sets

Amazon
ElastiCache

redis