Understanding Amazon DynamoDB and Redshift

Amazon DynamoDB

Fully managed, highly available and scalable NoSQL

Automatically and synchronously replicates data across three Availability Zones

SSDs and limiting indexing on attributes provides high throughput and low latency
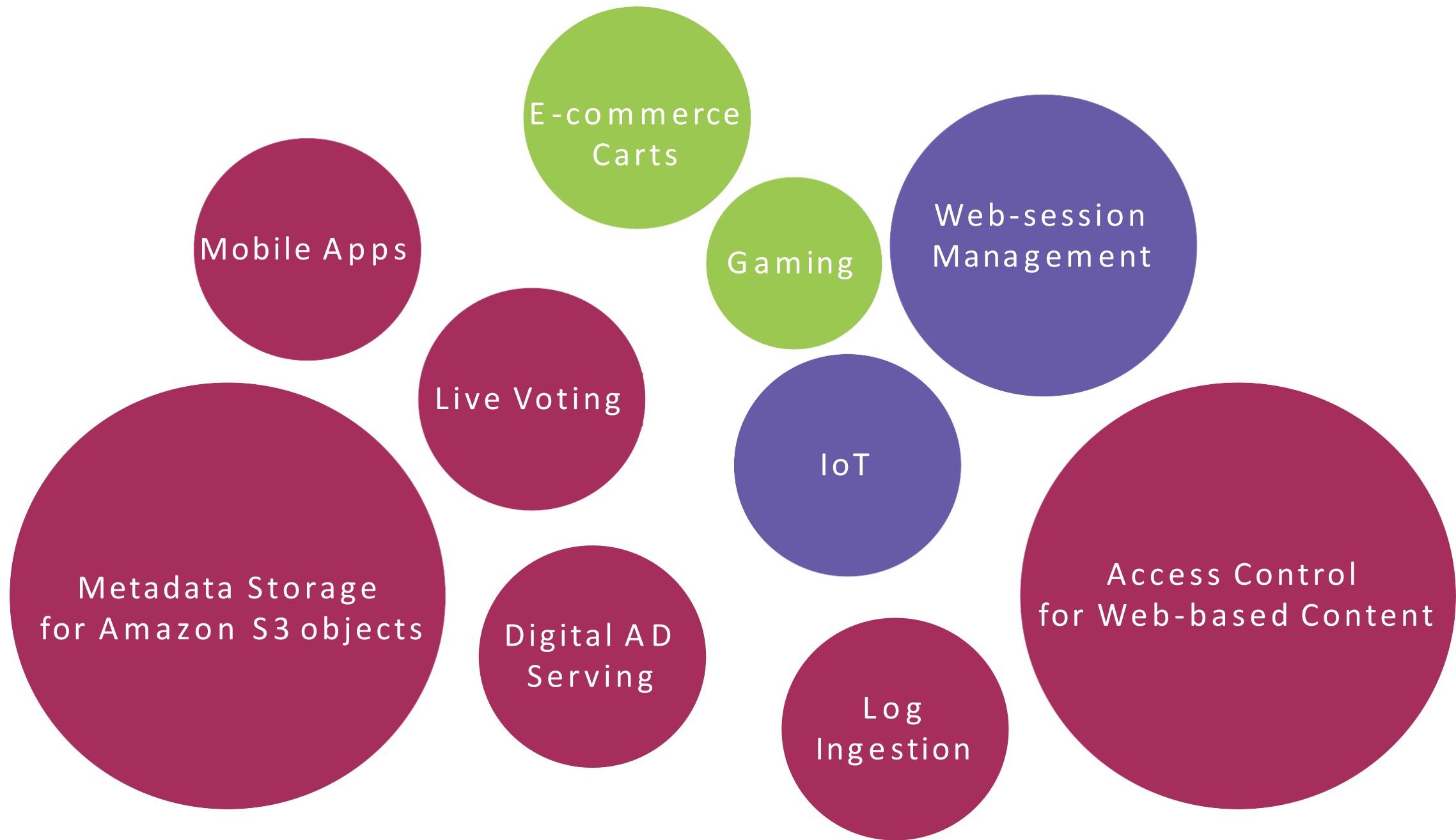
ElastiCache can be used in front of DynamoDB

- Offload high amounts of reads for non-frequently changed data

Amazon DynamoDB

Ideal for existing or new applications that need:

- Flexible NoSQL database with low read and write latencies
- Ability to scale storage and throughput up or down as needed without code changes or downtime

E-commerce Carts

Mobile Apps

Gaming

Web-session Management

Live Voting

IoT

Metadata Storage for Amazon S3 objects

Digital AD Serving

Log Ingestion

Access Control for Web-based Content

Non-ideal DynamoDB Scenarios

Prewritten application tied to a traditional relational database

Joins and/or complex transactions

BLOB data

Large data with low I/O rate

DynamoDB Integration

Amazon Elastic MapReduce

Amazon Redshift

Amazon Data Pipeline

Amazon S3

Management Console and APIs

Amazon DynamoDB

Stores structured data in tables, indexed by a primary key

Tables are a collection of items and Items are made up of attributes (columns)

Primary key can be:

- Single-attribute hash key
- Composite hash-range key
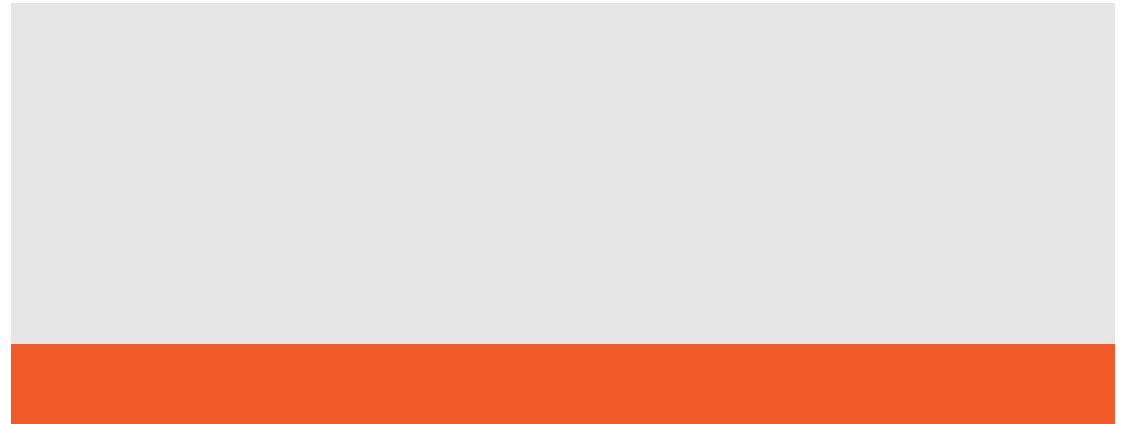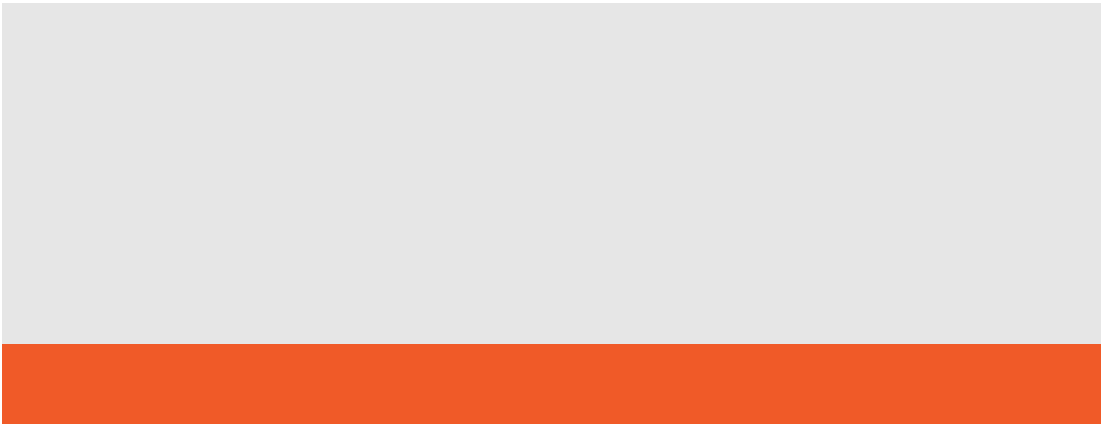
DynamoDB Features

Secondary indexes

Streams

Cross-region replication

Triggers

Schema-less

Two Ways to Search

ElastiCache

Open-source in-memory caching engines

- Memcached
  - Widely adopted memory object caching system
- Redis
  - Popular open-source in-memory key-value store
  - Supports data structures such as sorted sets and lists

Master / Slave replication and Multi-AZ

- Can be used to achieve cross AZ redundancy

Memcached vs. Redis

	Memcached	Redis
Cache to offload DB	√	√
Multithreaded performance	√	X
Horizontal scaling	√	X
Multi-AZ	X	√
Backup and restore	X	√
Pub/sub functionality	X	√
Sorting and ranking	X	√
Advanced data types	X	√
Persistence	X	√

Suggested Reading

Performance at Scale with Amazon ElastiCache:

https://d0.awsstatic.com/whitepapers/performance-at-scale-with-amazon-elasticache.pdf

Amazon Redshift

Fast and fully managed petabyte-scale relational data warehouse service

Analyze all your data using your existing business intelligence tools

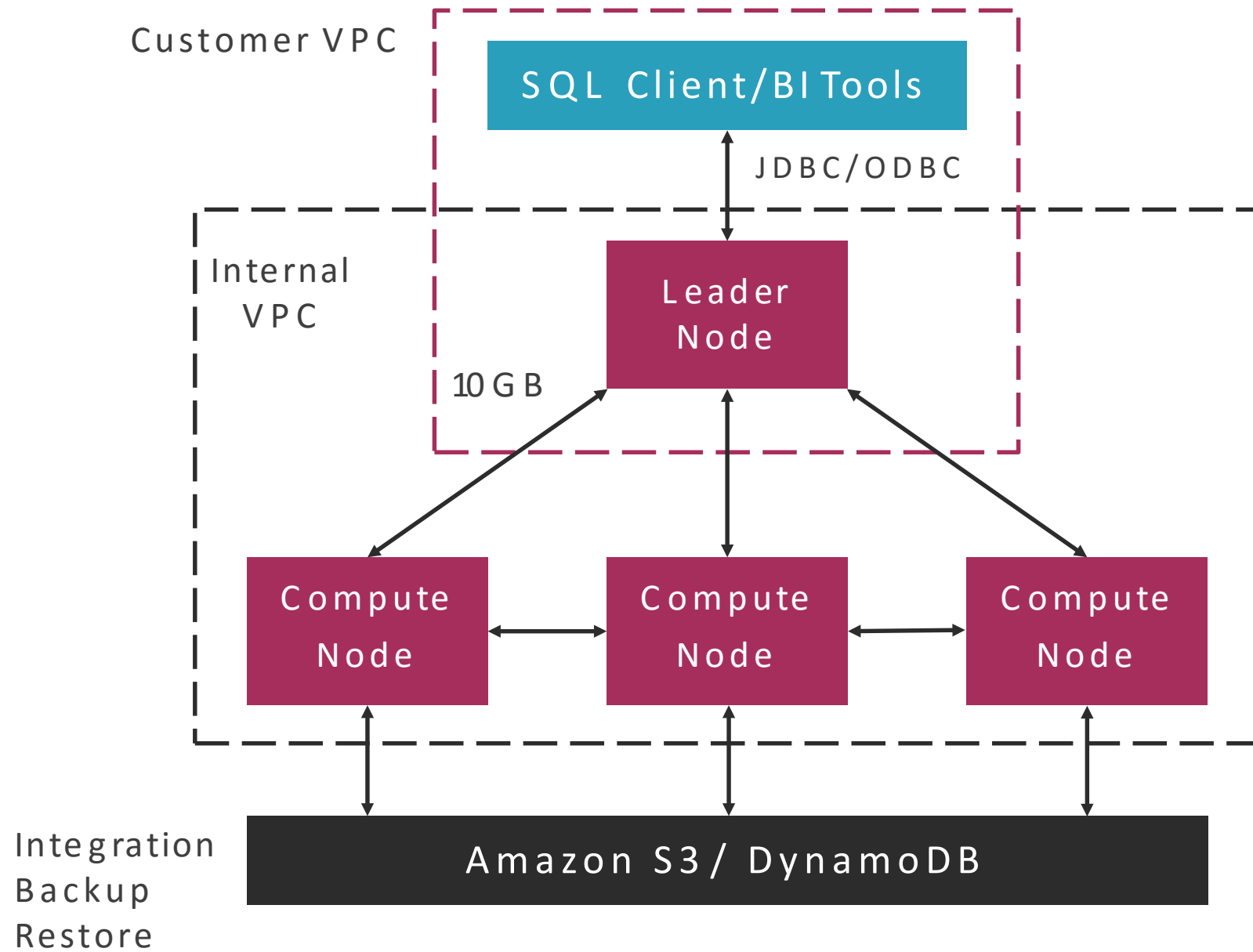HDD and SSD Platforms

Starts at $0.25/hour

Scale to $1,000/TB/Year

Amazon Redshift
Architecture

Leader Node

- Simple SQL end point
- Stores metadata
- Optimizes query plan
- Coordinates query execution

Compute Nodes

- Local columnar storage
- Parallel/distributed execution of all queries, loads, backups, restores, resizes

Backup and Fault Tolerance

Continuous/incremental backups

- Multiple copies within cluster
- Continuous and incremental backups to S3
- Continuous and incremental backups across regions
- Streaming restore

Backup and Fault Tolerance

Fault tolerance

- Disk failures
- Node failures
- Network failures
- Availability Zone/Region level disasters

Security

Load encrypted from S3

SSL to secure data in transit

Amazon VPC for network isolation

Encryption to secure data at rest

Audit logging and AWS CloudTrail integration

SOC 1/2/3, PCI-DSS, FedRAMP, BAA

Summary

Amazon DynamoDB

Amazon Redshift