

# AWS Storage & Databases & Analytics

Thomas Le Moullec. AWS Solutions Architect  
September 29th 2020

# Agenda

- Storage types and characteristics
- Block Storage - EBS
- File Storage - EFS
- Object Storage – S3
- Data Transfer
- Databases on AWS
- Data Analytics on AWS

# Storage Types



Block Storage

**Access:** Select \* from table;

**Ex:** Hard disks, SAN Storage Arrays

**Use Cases:** Databases ,file on disks

**Benefits:** Low latency, Read Efficiency,  
Address blocks, Consistent I/O



File Storage

**Access:** C:\folderpath\music.mp4

**Ex:** NAS, Windows File Servers

**Use Cases:** File System: Access entire file

**Benefits:** File Access, File Locking



Object Storage

**Access:** GET or PUT /prefix/Pofbwy56

**Ex:** OpenStack Swift, Ceph

**Use Cases:** Data Lake, Storage Pool

**Benefits:** Rich Metadata, File access,  
Unstructured and Scalability

# Storage Characteristics

## Durability

Measure of expected data loss

2 copies on 1 site : 99.99 % Durability

2 copies on 2 sites : 99.999 % Durability

3 copies on 3 sites : 99.999999999 % Durability

9 9 of durability: if you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years

# Storage Characteristics

## Availability

Measure of expected downtime

AWS Services offer a Service Level Agreement (SLA) to provide benefits to customer if SLA are not met

# Storage Characteristics

## Security

Security measures for at-rest and in-transit data

## Cost

Amount per storage unit, e.g. \$ / GB. Or per data transfer out/in

## Scalability

Upward flexibility, storage size, number of users

## Performance

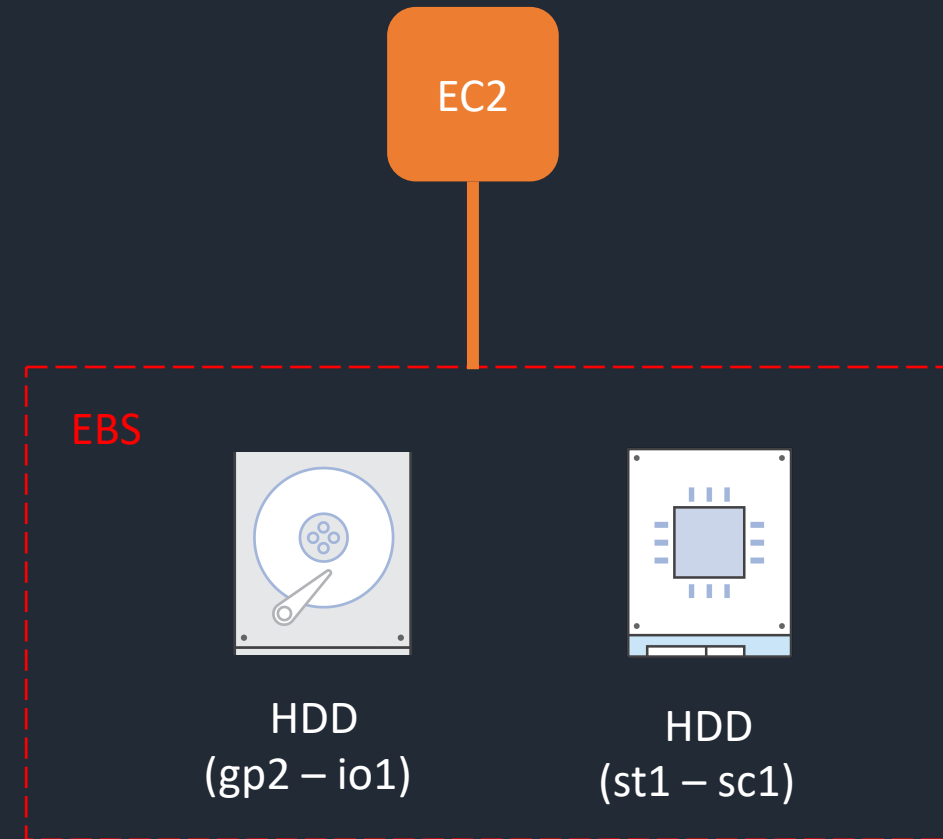
Performance metrics (bandwidth)

## Integration

Ability to interact via API or with other services

# Block Storage - EBS

- Block Storage as a Service
- Create, attach, detach, re-attach volume(s) to an EC2 **over network**
- **Redundancy** within AZ ; 99,999 Reliability
- **Scalability**: Scale up or down
- **Backup**: Point in Time Snapshots in S3 (Cross Region)
- **Performance**: Different type of Disks (Low latency, I/O)



# File Storage - EFS

- Fully Managed scalable file system for EC2 instances
- Can be shared across thousands of instances
- Scale Storage and Performance **with Elasticity** (Petabyte Scale) with number of files
- **Seamless Integration: NFS** based, standard OS APIs
- Accessible from onPremise
- Pay for the storage used
- Highly Available and Durable with Multi AZ



Amazon **Elastic File Storage**





# Object Storage – S3

- API and HTTPS requests
- Highly Durable - 99,99999999% and Infinite Scale
- Multiple Tiers with different pricing (Standard, Infrequent Access, one AZ)
- Lifecycle rules, Event, Versioning, Cross Region
- Advanced Security (Encryption, ACL, Bucket Policies) and Compliance (Audit)
- Data Lake, “Database”, Website Hosting, Media
- Used by many services (Logs, Backup, Snapshots)



Amazon Simple Storage Service



# Data Transfer



S3 Transfer Acceleration



Amazon Kinesis Data Streams



AWS Database Migration Service



AWS Snowball Edge



AWS DataSync

**+11 Options** for Transfer

# Traditional Databases

RDBMS is used for **all workloads**:

- Key-Value
- Complex Query
- Warehousing
- OLTP / OLAP Transactions
- Analytics












Punitive Licensing Model: **Oracle**

Web Tier

App Tier

RDBMS : One DB for ALL

# Databases on AWS - 15+ Purpose Built

Database type	Use cases	AWS service
Relational	Traditional applications, ERP, CRM, e-commerce	 <a href="#">Amazon Aurora</a>  <a href="#">Amazon RDS</a>  <a href="#">Amazon Redshift</a>
Key-value	High-traffic web apps, e-commerce systems, gaming applications	 <a href="#">Amazon DynamoDB</a>
In-memory	Caching, session management, gaming leaderboards, geospatial applications	 <a href="#">Amazon ElastiCache for Memcached</a>  <a href="#">Amazon ElastiCache for Redis</a>
Document	Content management, catalogs, user profiles	 <a href="#">Amazon DocumentDB (with MongoDB compatibility)</a>
Wide column	High scale industrial apps for equipment maintenance, fleet management, and route optimization	 <a href="#">Amazon Keyspaces (for Apache Cassandra)</a>
Graph	Fraud detection, social networking, recommendation engines	 <a href="#">Amazon Neptune</a>
Time series	IoT applications, DevOps, industrial telemetry	 <a href="#">Amazon Timestream</a>
Ledger	Systems of record, supply chain, registrations, banking transactions	 <a href="#">Amazon QLDB</a>

# How to choose a Database?

**Understand Data Characteristics:** Transactional, performance

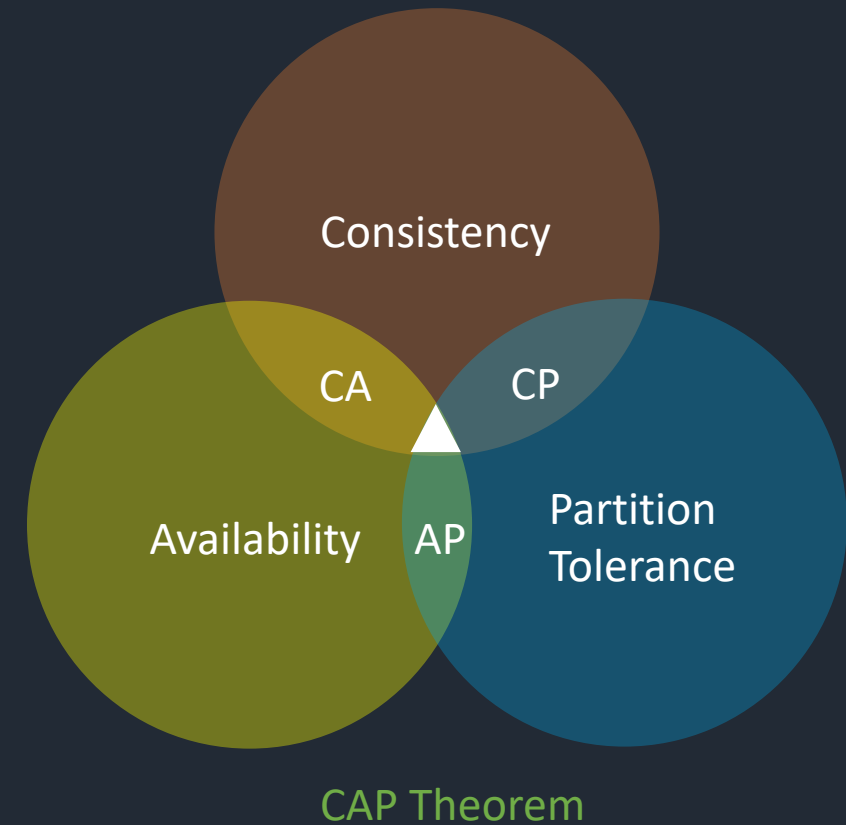
**Evaluate Available Options:** Databases and Storage services

**Collect and Record Database performance metrics:** TPS, Queries queue, latency

**Understand Access Patterns**

**Optimization:** Indexing, Key Distribution, Nodes cluster design

**Skills in House**



# Key features of a Managed DB

Managed by  
Customer

- App optimization
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net

Database on-premises

- App optimization
- Scaling
- High availability
- Database backups
- DB software patches
- DB software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, net

Database on EC2

Managed  
by AWS

- App optimization
  - Scaling
  - High availability
  - Database backups
  - DB software patches
  - DB software installs
  - OS patches
  - OS installation
  - Server maintenance
  - Rack and stack
  - Power, HVAC, net
- Amazon RDS

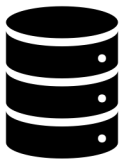
# Amazon RDS – Managed Relational DB



Highly Secure: Encryption, Connectivity Access, Network



Availability: Multi AZ, Failover



Performance: Easy Scale Compute and Storage, Read Replica, Disks types (Provisioned IOPS)



Managed: Easy Deploy, Patching, Backup, Snapshots

Popular Engines supported

ORACLE®



Microsoft SQL Server

# Amazon RDS - Aurora

Fully Managed MySQL and PostgreSQL compatible, cloud native – from AWS



Performance & Scalability

5x throughput of MySQL, 3x PostgreSQL, up to 15 replicas



Availability & Durability

Self-healing, 6 copies of data across three Availability Zones



Cost

Commercial Grade Database at 1/10th of the cost

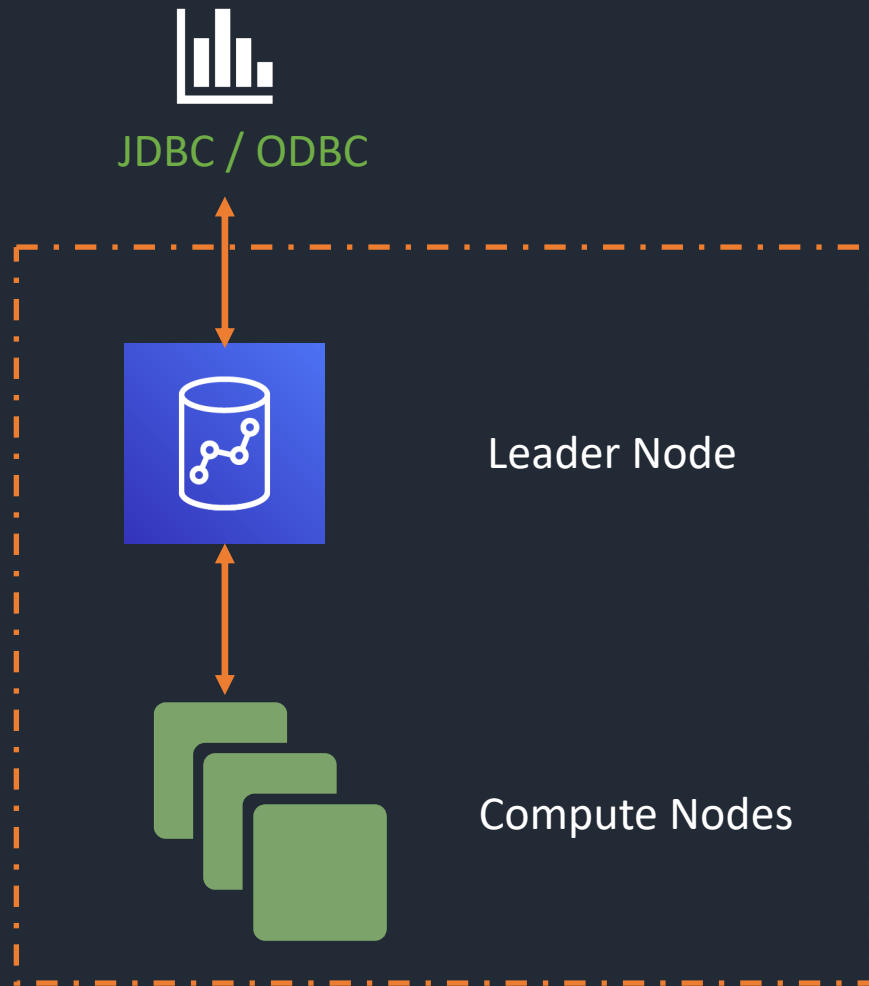


Serverless Feature

Perfect for Testing environment, no idle servers.  
A lot of other advanced features (Multi Master, Global Tables)



# Data Warehousing - Redshift



**50% LESS EXPENSIVE** THAN ALL OTHER CLOUD DATA WAREHOUSES

**Fully Managed:** No complex DBA tasks, No licenses, Spectrum, Federated Queries

**Relational Data Warehouse:** Complex relational queries

**Massively parallel:** Queue Management, Nodes Design (Leader, Compute)

**Scale:** Petabyte Scale, elastic scaling on demand – Up & Down

**Optimized Storage:** Loading, Columnar Storage, Compression,

**Pay for compute and storage independently:** RA3 new storage type.  
Others: DC DS

**Security:** Integrates with VPC, IAM and keys KMS

# NoSQL - DynamoDB



Fully Managed: Zero Admin, Backups,  
no capacity planning



Secure



Consistency



Advanced Features: Global Tables, Stream,  
Lambda integration



Highly Available (99.99%), Scalable,  
Durable (Data replicated in 3 AZ)

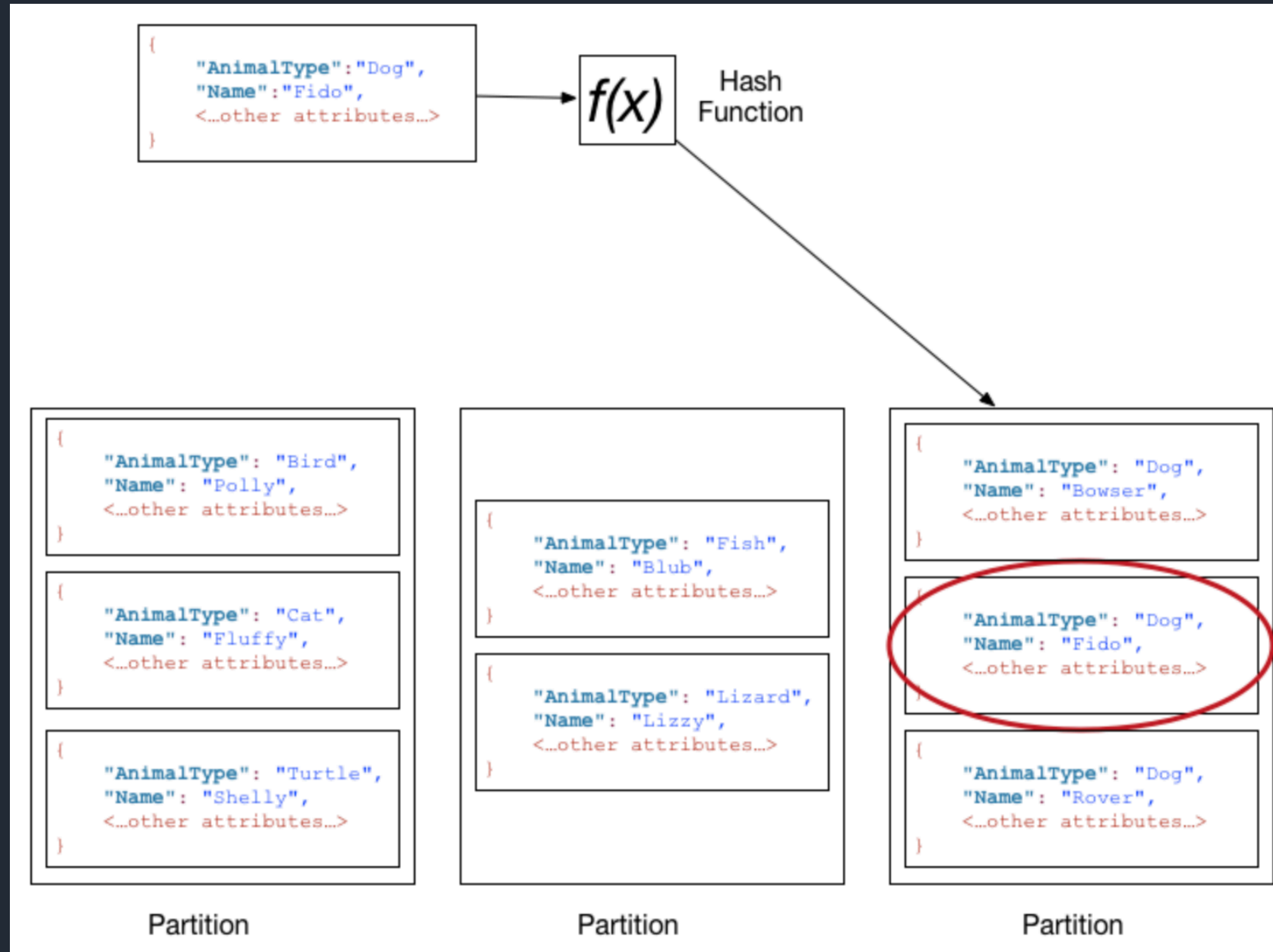


Single-Digit millisecond latency + DAX

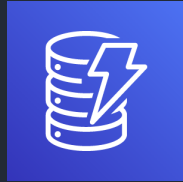


Cost Optimized: Serverless with pay for what you use  
- Pay for Writes, Reads and Storage in GB

# DynamoDB – Data Distribution

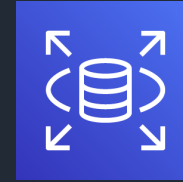


# NoSQL vs SQL



Amazon DynamoDB




















- Non Relational Data
- Dynamic Schema and unstructured Data
- Easy Management
- Good for unstructured such as doc or json
- No Servers & API driven



Amazon RDS

- Need Joins / Relational data and Transactional Operations
- Structured & Predefined Schema
- Good for multiple rows of transactions
- SQL Skills in house

# Data Analytics on AWS

Category	Use cases	AWS service
Analytics	Interactive analytics	 <b>Amazon Athena</b>
	Big data processing	 <b>Amazon EMR</b>
	Data warehousing	 <b>Amazon Redshift</b>
	Real-time analytics	 <b>Amazon Kinesis</b>
	Operational analytics	 <b>Amazon Elasticsearch Service</b>
	Dashboards and visualizations	 <b>Amazon QuickSight</b>
Data movement	Real-time data movement	 <b>Amazon Managed Streaming for Apache Kafka (MSK)</b>
		 <b>Amazon Kinesis Data Streams</b>
		 <b>Amazon Kinesis Data Firehose</b>
		 <b>Amazon Kinesis Data Analytics</b>
		 <b>Amazon Kinesis Video Streams</b>  <b>AWS Glue</b>
Data lake	Object storage	 <b>Amazon S3</b>  <b>AWS Lake Formation</b>
	Backup and archive	 <b>Amazon S3 Glacier</b>  <b>AWS Backup</b>
	Data catalog	 <b>AWS Glue</b>  <b>AWS Lake Formation</b>
	Third-party data	 <b>AWS Data Exchange</b>

# Thank you !

Find me on LinkedIn:  
Thomas LE MOULLEC

