SE ZG522: Big Data Systems

13 Amazon's Storage and DynamoDB

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

May 07, 2022 ONLINE (WILP @ BITS-Pilani Jan-May 2022)

http://ktiwari.in/bds

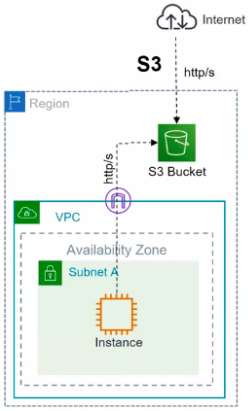
Amazon's Storage Services

- S3 (Simple Storage Service):
  Object storage. Independent access (API). Use link to get the data. read→change→write. Flat hirarchy. Amazon S3 Glacier.
- EBS (Elastic Block Storage):
  Connected to a Elastic Compute Cloud (EC2) machine. Not sharable. Data is in **block** (use own file format FAT, EXT3, ...). Multiple could be attached to one EC2 (like C: D: drive)
- EFS (Elastic File Service):
  Shared over network, like SAN. Mount anywhere. Hirarchical. File system is installed. Accessed through EC2.

Comparision

IOPS (Input Output operations Per Sec), Throughput, and Latency

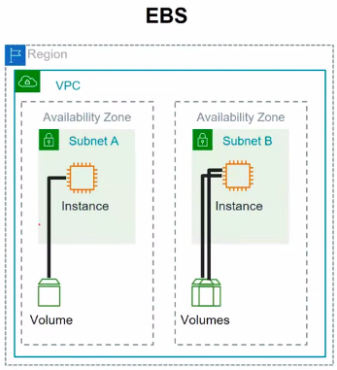
Architecture S3



S3 (Simple Storage Service):

- **Object** storage.
- Independent access (API).
- Use link to get the data.
- read→change→write.
- Flat hirarchy.
- Amazon S3 Glacier.
- Cheap

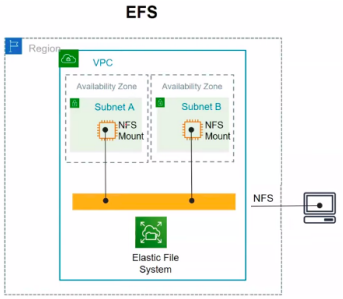
Architecture EBS



EBS (Elastic Block Storage):

- Connected to a Elastic Compute Cloud (EC2) machine.
- Not sharable.
- Data is in **block** (use own file format FAT, EXT3, ...).
- Multiple ones could be attached to single EC2 (like C: D: drive)
- Low latency
- High Availability

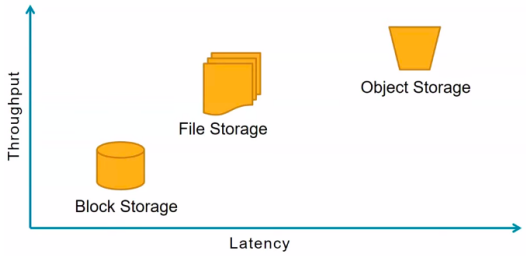
Architecture EFS



EFS (Elastic File Service):

- Shared over network, like SAN.
- Mount anywhere.
- Hirarchical.
- File system is installed.
- Accessed through EC2.
- Container Storage

Comparision



Cost: $S3 < EBS < EFS$

Amazon SimpleDB Database [1]

- Cloud based database.
- No installation or management
- 3-Node availability,
- low cost,
- scalable

[1] See: https://www.youtube.com/watch?v=4wMB5qQvA8c

Amazon DynamoDB [2]

- Fast and flexible/scalable NoSQL database (key-value pairs)
- Replicated at three distinct locations. Choose evential/strong consistency. So write is costly.
- No schema, only primary key. Atrribute + Item + Table

Users: Netflix, Nasdaq, duoLingo, etc.

[2] https://www.youtube.com/watch?v=k0fcbRj-pZE

Comparision

Features	DYNAMODB	RDS	RedShift	Aurora
Scaling	Scales Seamlessly	Relatively Easy to Scale	More Complexity in Scaling	Relatively Easy to Scale
Storage	Unlimited Storage	64TB	2PB	64TB with RDS
Pricing	Pay-per-Use Model	Generally Cheaper than Others	On-Demand model with added costs	Pricing depends upon RDS
Maintenance	Maintained by AWS	Maintained by AWS	Requires more Maintenance	No Maintenance Required

- Want RDBMS? RDS
- Complex and analytical queries? RedShift
- Non analatycal big database? Aurora
- Lot of transactions? DynamoDB

Amazon Relational Database Service (RDS)

DynamoDB Advantages [3]

- On demand capacity
- ACID support built in
- On demond backup
- Point in time recovery
- Encryption at rest
- Events (insertion/deletion) can be treated as streams

Can be accessed using

1) Console
2) Command Level Interface,
3) API

[3] see: https://www.youtube.com/watch?v=0Z_nvCVRms8

Thank You!

Thank you very much for your attention!