

Aim: AWS Case Study: Amazon.com.



Theory: About AWS

- Launched in 2006, Amazon Web Services (AWS) began exposing key infrastructure services to businesses in the form of web services -- now widely known as cloud computing.
- The ultimate benefit of cloud computing, and AWS, is the ability to leverage a new business model and turn capital infrastructure expenses into variable costs.
- Businesses no longer need to plan and procure servers and other IT resources weeks or months in advance.
- Using AWS, businesses can take advantage of Amazon's expertise and economies of scale to access resources when their business needs them, delivering results faster and at a lower cost.
- Today, Amazon Web Services provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers hundreds of thousands of businesses in 190 countries around the world.
- Amazon.com is the world's largest online retailer. In 2011, Amazon.com switched from tape backup to using Amazon Simple Storage Service (Amazon S3) for backing up the majority of its

Oracle databases. This strategy reduces complexity and capital expenditures, provides faster backup and restore performance, eliminates tape capacity planning for backup and archive, and frees up administrative staff for higher value operations. The company was able to replace their backup tape infrastructure with cloud-based Amazon S3 storage, eliminate backup software, and experienced a 12X performance improvement, reducing restore time from around 15 hours to 2.5 hours in select scenarios.



With data center locations in the U.S., Europe, Singapore, and Japan, customers across all industries are taking advantage of the following benefits:

- **Low Cos**
- **Agility and Instant Elasticity**
- **Open and Flexible**
- **Secure**

The Challenge

As Amazon.com grows larger, the sizes of their Oracle databases continue to grow, and so does the sheer number of databases they maintain. This has caused growing pains related to backing up legacy Oracle databases to tape and led to the consideration of alternate strategies including the use of Cloud services of Amazon Web Services (AWS), a subsidiary of Amazon.com. Some of the business challenges Amazon.com faced included:

- Utilization and capacity planning is complex, and time and capital expense budget are at a premium. Significant capital expenditures were required over the years for tape hardware, data center space for this hardware, and enterprise licensing fees for tape software. During that time, managing tape infrastructure required highly skilled staff to spend time with setup, certification and engineering archive planning instead of on higher value projects. And at the end of every fiscal year, projecting future capacity requirements required time consuming audits, forecasting, and budgeting.
- The cost of backup software required to support multiple tape devices sneaks up on you. Tape robots provide basic read/write capability, but in order to fully utilize them, you must invest in proprietary tape backup software. For Amazon.com, the cost of the software had been high, and added significantly to overall backup costs. The cost of this software was an ongoing budgeting pain point, but one that was difficult to address as long as backups needed to be written to tape devices.

- Maintaining reliable backups and being fast and efficient when retrieving data requires a lot of time and effort with tape. When data needs to be durably stored on tape, multiple copies are required. When everything is working correctly, and there is minimal contention for tape resources, the tape robots and backup software can easily find the required data. However, if there is a hardware failure, human intervention is necessary to restore from tape. Contention for tape drives resulting from multiple users' tape requests slows down restore processes even more. This adds to the recovery time objective (RTO) and makes achieving it more challenging compared to backing up to Cloud storage.

Why Amazon Web Services?

Amazon.com initiated the evaluation of Amazon S3 for economic and performance improvements related to data backup. As part of that evaluation, they considered security, availability, and performance aspects of Amazon S3 backups. Amazon.com also executed a cost-benefit analysis to ensure that a migration to Amazon S3 would be financially worthwhile. That cost benefit analysis included the following elements:

- Performance advantage and cost competitiveness. It was important that the overall costs of the backups did not increase. At the same time, Amazon.com required faster backup and recovery performance. The time and effort required for backup and for recovery operations proved to be a significant improvement over tape, with restoring from Amazon S3 running from two to twelve times faster than a similar restore from tape. Amazon.com required any new backup medium to provide improved performance while maintaining or reducing overall costs. Backing up to on-premises disk based storage would have improved performance, but missed on cost competitiveness. Amazon S3 Cloud based storage met both criteria.
- Greater durability and availability. Amazon S3 is designed to provide 99.999999999% durability and 99.99% availability of objects over a given year. Amazon.com compared these figures with those observed from their tape infrastructure, and determined that Amazon S3 offered significant improvement.
- Less operational friction. Amazon.com DBAs had to evaluate whether Amazon S3 backups would be viable for their database backups. They determined that using Amazon S3 for backups was easy to implement because it worked seamlessly with Oracle RMAN.

- Strong data security. Amazon.com found that AWS met all of their requirements for physical security, security accreditations, and security processes, protecting data in flight, data at rest, and utilizing suitable encryption standards.

The Benefits

With the migration to Amazon S3 well along the way to completion, Amazon.com has realized several

benefits, including:

- Elimination of complex and time-consuming tape capacity planning. Amazon.com is growing larger

and more dynamic each year, both organically and as a result of acquisitions. AWS has enabled Amazon.com to keep pace with this rapid expansion, and to do so seamlessly. Historically, Amazon.com business groups have had to write annual backup plans, quantifying the amount of tape storage that they plan to use for the year and the frequency with which they will use the tape resources. These plans are then used to charge each organization for their tape usage, spreading the cost among many teams. With Amazon S3, teams simply pay for what they use, and are billed for their usage as they go. There are virtually no upper limits as to how much data can be stored in Amazon S3, and so there are no worries about running out of resources. For teams adopting Amazon S3 backups, the need for formal planning has been all but eliminated.

- Reduced capital expenditures. Amazon.com no longer needs to acquire tape robots, tape drives, tape inventory, data center space, networking gear, enterprise backup software, or predict future tape consumption. This eliminates the burden of budgeting for capital equipment well in advance as well as the capital expense.
- Immediate availability of data for restoring – no need to locate or retrieve physical tapes. Whenever a DBA needs to restore data from tape, they face delays. The tape backup software needs to read the tape catalog to find the correct files to restore, locate the correct tape, mount the tape, and read the data from it. In almost all cases the data is spread across multiple tapes, resulting in further delays. This, combined with contention for tape drives resulting from multiple users' tape requests, slows the process down even more. This is especially severe during critical events such as a data center outage, when many databases must be restored simultaneously and as soon as possible. None of these problems occur with Amazon S3. Data restores can begin immediately, with no waiting or tape queuing – and that means the database can be recovered much faster.

- Backing up a database to Amazon S3 can be two to twelve times faster than with tape drives. As one example, in a benchmark test a DBA was able to restore 3.8 terabytes in 2.5 hours over gigabit Ethernet. This amounts to 25 gigabytes per minute, or 422MB per second. In addition, since Amazon.com uses RMAN data compression, the effective restore rate was 3.37 gigabytes per second. This 2.5 hours compares to, conservatively, 10-15 hours that would be required to restore from tape.
- Easy implementation of Oracle RMAN backups to Amazon S3. The DBAs found it easy to start backing up their databases to Amazon S3. Directing Oracle RMAN backups to Amazon S3 requires

only a configuration of the Oracle Secure Backup Cloud (SBC) module. The effort required to configure the Oracle SBC module amounted to an hour or less per database. After this one-time setup, the database backups were transparently redirected to Amazon S3.

- Durable data storage provided by Amazon S3, which is designed for 11 nines durability. On occasion, Amazon.com has experienced hardware failures with tape infrastructure – tapes that break, tape drives that fail, and robotic components that fail. Sometimes this happens when a DBA is trying to restore a database, and dramatically increases the mean time to recover (MTTR). With the durability and availability of Amazon S3, these issues are no longer a concern.
- Freeing up valuable human resources. With tape infrastructure, Amazon.com had to seek out engineers who were experienced with very large tape backup installations – a specialized, vendor-specific skill set that is difficult to find. They also needed to hire data center technicians and dedicate them to problem-solving and troubleshooting hardware issues – replacing drives, shuffling tapes around, shipping and tracking tapes, and so on. Amazon S3 allowed them to free up these specialists from day-to-day operations so that they can work on more valuable, business-critical engineering tasks.
- Elimination of physical tape transport to off-site location. Any company that has been storing Oracle backup data offsite should take a hard look at the costs involved in transporting, securing and storing

their tapes offsite – these costs can be reduced or possibly eliminated by storing the data in Amazon S3.

As the world's largest online retailer, Amazon.com continuously innovates in order to provide improved customer experience and offer products at the lowest possible prices. One such innovation has been to replace tape with Amazon S3 storage for database backups. This

innovation is one that can be easily replicated by other organizations that back up their Oracle databases to tape.

Products & Services

- Compute
- Content Delivery
- Database
- Deployment & Management
- E-Commerce
- Messaging
- Monitoring
- Networking
- Payments & Billing

- Storage
- Support
- Web Traffic
- Workforce

Products & Services

Compute

- › **Amazon Elastic Compute Cloud (EC2)**
- Amazon Elastic Compute Cloud delivers scalable, pay-as-you-go compute capacity in the cloud.
- › **Amazon Elastic MapReduce**
- Amazon Elastic MapReduce is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data.
- › **Auto Scaling**
- Auto Scaling allows to automatically scale our Amazon EC2 capacity up or down according to conditions we define.

Content Delivery

- - › Amazon CloudFront
- Amazon CloudFront is a web service that makes it easy to distribute content with low latency via a global network of edge locations.

Database

- - › Amazon SimpleDB
- Amazon SimpleDB works in conjunction with Amazon S3 and AmazonEC2 to run queries on structured data in real time.
- - › Amazon Relational Database Service (RDS)
- Amazon Relational Database Service is a web service that makes it easy to set up, operate, and scale a relational database in the cloud.
- - › Amazon ElastiCache
- Amazon ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud.

E-Commerce

- - ›**Amazon Fulfillment Web Service (FWS)**
- Amazon Fulfillment Web Service allows merchants to deliver products using Amazon.com's worldwide fulfillment capabilities.

Deployment & Management

- ✓
 - ›**AWS Elastic Beanstalk**
- ✓

AWS Elastic Beanstalk is an even easier way to quickly deploy and manage applications in the AWS cloud. We simply upload our application, and Elastic Beanstalk automatically handles the deployment details of capacity provisioning, load balancing, auto-scaling, and application health monitoring.
- - ›**AWS CloudFormation**
- AWS CloudFormation is a service that gives developers and businesses an easy way to create a collection of related AWS resources and provision them in an orderly and predictable fashion.

Monitoring

- - ›**Amazon CloudWatch**
- Amazon CloudWatch is a web service that provides monitoring for AWS cloud resources, starting with Amazon EC2

Messaging

- - ›**Amazon Simple Queue Service (SQS)**

- Amazon Simple Queue Service provides a hosted queue for storing messages as they travel between computers, making it easy to build automated workflow between Web services.
- ›**Amazon Simple Notification Service (SNS)**
- Amazon Simple Notification Service is a web service that makes it easy to set up, operate, and send notifications from the cloud.
- ›**Amazon Simple Email Service (SES)**
- Amazon Simple Email Service is a highly scalable and cost-effective bulk and transactional email-sending service for the cloud.

Workforce

- ›**Amazon Mechanical Turk**
- Amazon Mechanical Turk enables companies to access thousands of global workers on demand and programmatically integrate their work into various business processes.

Networking

- ›**Amazon Route 53**

- Amazon Route 53 is a highly available and scalable Domain Name System (DNS) web service.
- ›**Amazon Virtual Private Cloud (VPC)**
- Amazon Virtual Private Cloud (Amazon VPC) lets you provision a private, isolated section of the Amazon Web Services (AWS) Cloud where we can launch AWS resources in a virtual network that you define. With Amazon VPC, we can define a virtual network topology that closely resembles a traditional network that you might operate in your own datacenter.
- ›**AWS Direct Connect**
- AWS Direct Connect makes it easy to establish a dedicated network connection from your premise to AWS, which in many cases can reduce our network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.
- ›**Elastic Load Balancing**
- Elastic Load Balancing automatically distributes incoming application traffic across multiple Amazon EC2 instances.

Payments & Billing

- ›**Amazon Flexible Payments Service (FPS)**
- Amazon Flexible Payments Service facilitates the digital transfer of money between any two entities, humans or computers.
- ›**Amazon DevPay**

- Amazon DevPay is a billing and account management service which enables developers to collect payment for their AWS applications.
- Storage
- ›**Amazon Simple Storage Service (S3)**
- Amazon Simple Storage Service provides a fully redundant data storage infrastructure for storing and retrieving any amount of data, at any time, from anywhere on the Web.
- ›**Amazon Elastic Block Store (EBS)**
- ✓ Amazon Elastic Block Store provides block level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes are off-instance storage that persists independently from the life of an instance.
- ›**AWS Import/Export**
- AWS Import/Export accelerates moving large amounts of data into and out of AWS using portable storage devices for transport.

Support

- ›**AWS Premium Support** AWS Premium Support is a one-on-one, fast-response support channel to help you build and run applications on AWS Infrastructure Services.

Web Traffic

- ›**Alexa Web Information Service**

- Alexa Web Information Service makes Alexa's huge repository of data about structure and traffic patterns on the Web available to developers.
- ›Alexa Top Sites
- Alexa Top Sites exposes global website traffic data as it is continuously collected and updated by Alexa Traffic Rank.

Amazon CloudFront

- → Amazon CloudFront is a web service for content delivery.
- → Amazon CloudFront is a web service for content delivery.
- → It integrates with other Amazon Web Services to give developers and businesses an easy way to distribute content to end users with low latency, high data transfer speeds, and no commitments.
- → Amazon CloudFront delivers our static and streaming content using a global network of edge locations.
- → Requests for our objects are automatically routed to the nearest edge location, so content is delivered with the best possible performance.

Amazon CloudFront

- → Amazon CloudFront is optimized to work with other Amazon Web Services, like Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2).
- → Amazon CloudFront also works seamlessly with any origin server, which stores the original, definitive versions of our files.

-
- → Like other Amazon Web Services, there are no contracts or monthly commitments for using Amazon CloudFront _ we pay only for as much or as little content as you actually deliver

through the service.

Amazon Simple Queue Service (Amazon SQS)

- → Amazon Simple Queue Service (Amazon SQS) offers a reliable, highly scalable, hosted queue for storing messages as they travel between computers.
- → By using Amazon SQS, developers can simply move data between distributed components of their applications that perform different tasks, without losing messages or requiring each

component to be always available.

- → Amazon SQS makes it easy to build an automated workflow, working in close conjunction with the Amazon Elastic Compute Cloud (Amazon EC2) and the other AWS infrastructure web

services.



Amazon Simple Queue Service (Amazon SQS)

- → Amazon SQS works by exposing Amazon's web-scale messaging infrastructure as a web service.
- → Any computer on the Internet can add or read messages without any installed software or special firewall configurations.
- → Components of applications using Amazon SQS can run independently, and do not need to be on the same network, developed with the same technologies, or running at the same time

BigTable

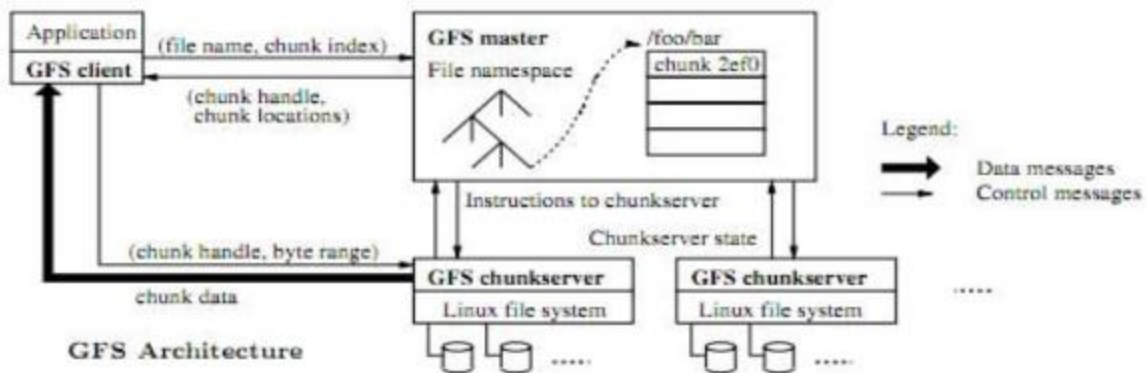
- → Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers.
- → Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance.
- → These applications place very different demands on Bigtable, both in terms of data size (from URLs to web pages to satellite imagery) and latency requirements (from backend bulk processing to real-time data serving).

- → Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products.

The Google File System(GFS)

- → The Google File System (GFS) is designed to meet the rapidly growing demands of Google's data processing needs.
- → GFS shares many of the same goals as previous distributed file systems such as performance, scalability, reliability, and availability.
- → It provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of clients.
- → While sharing many of the same goals as previous distributed file systems, file system has successfully met our storage needs.
- → It is widely deployed within Google as the storage platform for the generation and processing of data used by our service as well as research and development efforts that require large data sets.

- → The largest cluster to date provides hundreds of terabytes of storage across thousands of disks on over a thousand machines, and it is concurrently accessed by hundreds of clients.



Conclusion:

Thus we have studied a case study on amazon web services.