|	
Name	: Ritik Vinod Nandanwar
Roll No	: 59
Division	: A
Registration No	: 2018BCS111
Subject	: Distributed Computing
Date	: 20 Aug 2021
Title	: Case Study of Amazon Web Services

Information:

As we all know the technology is spreading in every part of our globe. Day by day new technologies are coming in market which blow our mind. One of the example is Amazon which started very long back nearly 27 years ago. "The Jeff Bezos", who is not The Jeff Bezos at that time launched a company called Cadabra, later known as Amazon, out of his house in Seattle as book selling company. The company grows exponentially in last 15 years, there were 2200 daily visits at the starting year and now in June 2021 visitors(desktop and mobile) count rise to 2.7 billion. And now amazon sells nearly about 12 million its own product and if we include the marketplaces it sells nearly 343 million products. To keep an account of this it created Amazon Web Service (AWS) which is made up of many different cloud computing products and services. The highly profitable division of Amazon provides servers, storage, networking, remote computing, email, mobile development, and security. AWS can be broken into three main products: EC2, Amazon's virtual machine service, Glacier, a low-cost cloud storage service, and S3, Amazon's storage system. As of April 2021, Canalys reports that the worldwide cloud market grew 35% this quarter to $41.8 billion. Aws is leading this market with 32% of the market. Let us look at the How it is done.

Infrastructure:
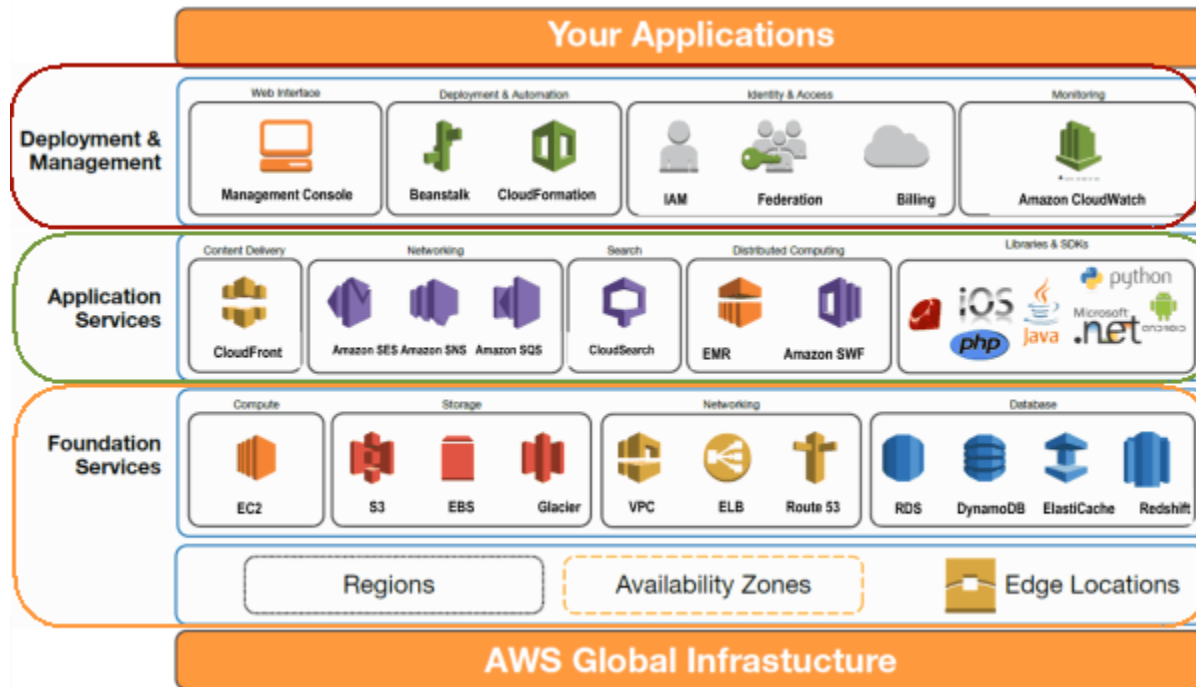
The AWS Global Cloud infrastructure is the most secure, extensive, and reliable cloud platform. With millions of active customers and tens of thousands of partners globally, AWS has the largest and most dynamic ecosystem. Customers across virtually every industry and of every size, including start-ups, enterprises, and public sector organizations, are running every imaginable use case on AWS. It spans 81 availability Zones with 25 geographic regions around the world.
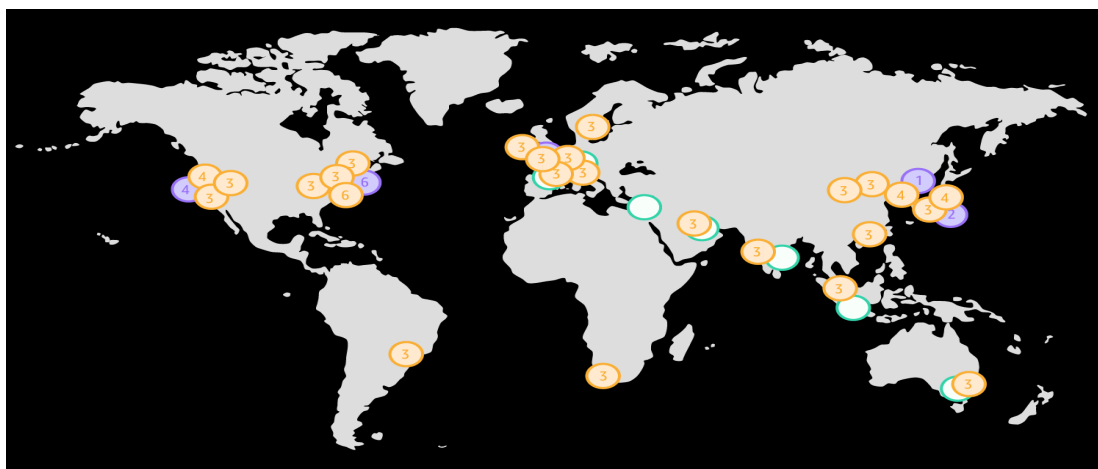
Scalability :

It allows the companies to be extremely flexible and take advantage of the conceptually infinite scalability of the cloud. Customers uset to over provision to ensure they had enough capacity to handle their business operations at the peak level of activity. Now, they can provision the amount of resources that they actually need, knowing they can instantly scale up or down along with the needs of their business, which also reduces cost and improves the customer's ability to meet their user's demands. Companies can quickly spin up resources as they need them, deploying hundreds or even thousands of servers in minutes.

Reliability :

A reliable workload starts with upfront design decisions for both software and infrastructure. Our architecture choices will impact our workload behavior across all five Well Architected Pillars. For reliability, there are specific patterns that must be followed. With AWS, workload developers have their choice of languages and technologies to use. AWS SDKs take the complexity out of coding by providing language-specific APIs for AWS services..

Performance :

Inorder to provide performance AWS consists :
! Democratize Advanced Technologies
! As AWS has big spanning area it provide lower latency at minimum cost
! Use serverless architecture
! Quick testing using multiple instances ,storage or configurations
! using Mechanical Sympathy

Flexibility :

The AWS Global Infrastructure gives us the flexibility of choosing how and where we want to run our workloads, and when we are using the same network, control plane, API's, and AWS services. If we would like to run our applications globally we can choose from any of the AWS Regions and AZs.

Security :

With AWS, we can control where data is stored, who can access it, and what we allow our organization to consume at any given moment. Automating security tasks on AWS enables us to be more secure by reducing human configuration errors and giving my team time to focus on other work . All data flowing across the AWS global network that interconnects our data centers and regions is automatically encrypted at the physical layer before it leaves AWS secured facilities.

Amazon EMR and SWF (as Distributed Computing):

Amazon SWF helps developers build, run, and scale background jobs that have parallel or sequential steps. We can think of Amazon SWF as a fully-managed state tracker and task coordinator in the cloud. If our apps take more than 500 milliseconds to complete, we need to track the state of processing, and you need to recover or retry if a task fails, Amazon SWF can help you. It promotes a separation between the control flow of our background jobs stepwise logic and tkh actual units of work that contain our unique business logic. It runs within Amazon's high availability data centres, so the state tracking and task processing is available whenever applications need them. Video encoding using Amazon S3 and Amazon EC2. In this use case , large videos are uploaded to Amazon S3 in chunks. The upload of chunks has to be monitored. After a chunk is stored and uploaded , it is encoded by downloading it to an Amazon S3. Failure could occur during this process due to one or more chunks encountering encoding errors. Such failures need to be detected and handled through Amazon SWFs cloud workflow Management.
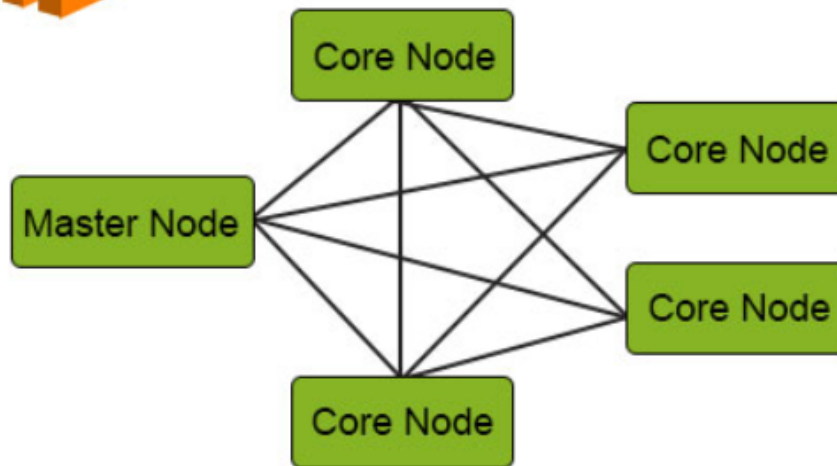

How Amazon SWF works

Example use case: Video transcoding application

Start > Encode video > Add DRM protection > Finish
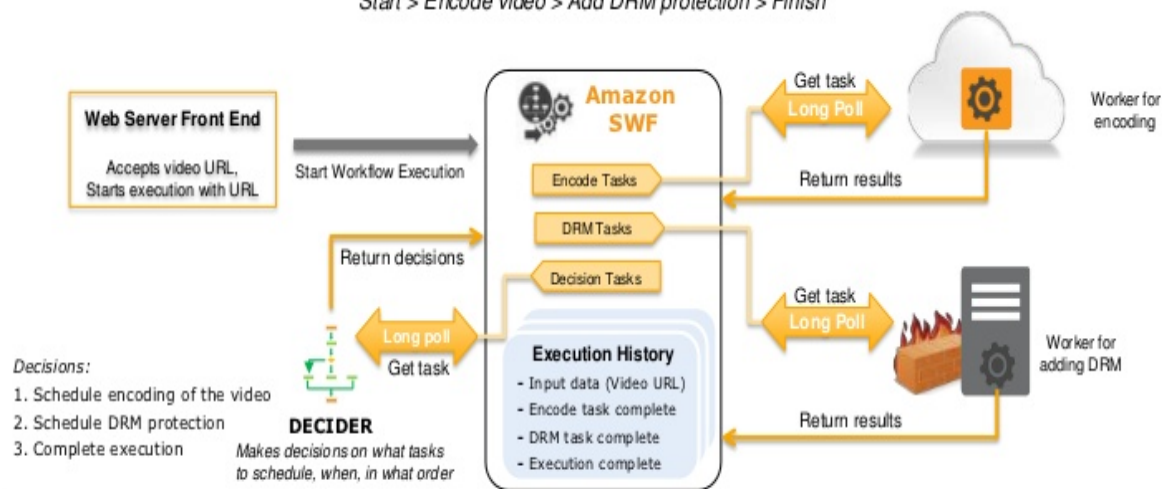
Amazon EMR Cluster

Amazon Elastic MapReduce (Amazon EMR) is a web service that makes it easy to quickly and cost-effectively process vast amounts of data.Amazon EMR uses Hadoop, an open source framework, to distribute our data and processing across a resizable cluster of Amazon EC2 instances. Amazon EMR is used in a variety of applications, including log analysis, web indexing, data warehousing, machine learning, financial analysis, scientific simulation, and bioinformatics. Customers launch millions of Amazon EMR clusters every year.

Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase , Apache Flink, Apache Hudi, and Presto. Amazon EMR makes it easy to set up, operate, and scale our big data environments by automating time- consuming tasks like provisioning capacity and turning clusters. With EMR we can run petabyte-scale analysis at less than half of the cost of traditional on-premises solutions and over 3x faster than standard Apache Spark .EMR spends less time monitoring and turning our cluster. EMR is turned for the cloud and constantly monitors our cluster - retrying failed tasks and automatically replacing poorly performing instances. EMR automatically configures EC2 firewall settings, controlling network access to instances and launches clusters in an Amazon Virtual Private Cloud(VPC). Machine learning uses EMR built in machine learning tools, including Apache Spark MLlib, Tensorflow, and Apache MXNet for scalable machine learning algorithms, and uses custom AMIs and bootstrap actions to easily add our preferred libraries and tools to create our own predictive analytics toolset. EMR can be used to process vast amounts of genomic data and other large scientific datasets quickly and efficiently. Researchers can access genomic data hosted for free on AWS.

Summary:

During the course of the case study it is understandable what impact AWS is making on our industries with the help of Distributed System. Amazon Web Services has always been focused on improving efficiency of their infrastructure in every aspect. The results of a study by 451 Research show that AWSs infrastructure is 3.6 times more energy efficient than the median of U.S. enterprise data centres surveyed. It is now the leader in the market of cloud computing with steadily growing their market value.

References:

1) Amazon AWS Well-Architected Framework

   https://wa.aws.amazon.com/wellarchitected/2020-07-02T19-33-23/index.en.html

2) Amazon EMR

   https://aws.amazon.com/emr/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc

3) Amazon Global Infrastructure

   https://aws.amazon.com/about-aws/global-infrastructure/?nc2=h_ql_le_int_gi

4) Amazon Simple Workflow Services

   https://aws.amazon.com/swf/

5) AWS vs Azure vs Google Cloud Market Share 2021: What the Latest Data Shows

   https://www.parkmycloud.com/blog/aws-vs-azure-vs-google-cloud-market-share/