

Amazon Web Services

Name: Shantanu Shukla

Enrolment No.: 01211805424

Course: M. Tech. (AI & DS)

Components

AWS offers various services according to a user's workload and tasks.

Some of the most popular components are listed below:

- **Data Management & Data Transfer**
- **Compute & Networking**
- **Storage**
- **Automation and Orchestration**
- **Operations & Management**
- **Visualization Tools**
- **Security and Compliance**

Data Management and Data Transfer

Running HPC applications in the cloud starts with moving the required data into the cloud. AWS Snowball is a data transport solution that securely transfers large amounts of data into and out of the AWS Cloud. Using Snowball addresses common challenges with large-scale data transfers including high network costs, long transfer times, and security concerns. AWS DataSync is a data transfer service that makes it easy for you to automate moving data between on-premises storage and Amazon S3 or Amazon Elastic File System (Amazon EFS). DataSync automatically handles many of the tasks related to data transfers that can slow down migrations or burden your IT operations, including running your own instances, handling encryption, managing scripts, network optimization, and data integrity validation. AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your datacenter, office, or colocation environment, which in many cases can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.

Compute & Networking

The AWS HPC solution lets you choose from a variety of compute instance types that can be configured to suit your needs, including the 3rd generation Intel® Xeon® processors, 3rd generation AMD EPYC processors, Arm-based AWS Graviton2 processors, the latest NVIDIA GPU-based instances, and field programmable gate array (FPGA) powered instances.

Compute intensive: Most customers will find that either the Hpc6a or C5n instances will meet the core requirements of most compute-intensive workloads. These instances are designed to address many common workloads from computational fluid dynamics (CFD), computer aided engineering (CAE), materials science, to reservoir simulation. Hpc6a instances feature 96 3rd generation AMD EPYC processors with up to 3.6 GHz all-core turbo frequency, and 384 GB RAM. Hpc6a instances offer up to 65% better price performance over comparable compute-optimized, x86 based instances. Customers with applications that run best on the Intel compiler or need to maximize performance per core for application licensing costs, should consider the C5n instances. C5n instances feature the Intel Xeon Scalable Platinum 8000 series (Skylake) processor with a sustained all core Turbo CPU clock speed of up to 3.5 GHz.

Storage

Storage options and storage costs are critical factors when considering an HPC solution. AWS offers flexible object, block, or file storage for your transient and permanent storage requirements. Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2. Provisioned IOPS allows you to allocate storage volumes of the size you need and to attach these virtual volumes to your EC2 instances. Amazon Simple Storage Service (S3) is designed to store and access any type of data over the Internet and can be used to store the HPC input and output data long term and without ever having to do a data migration project again. Amazon FSx for Lustre is a high performance file storage service designed for demanding HPC workloads and can be used on Amazon EC2 in the AWS cloud. Amazon FSx for Lustre works natively with Amazon S3, making it easy for you to process cloud data sets with high performance file systems. When linked to an S3 bucket, an FSx for Lustre file system transparently presents S3 objects as files and allows you to write results back to S3. You can also use FSx for Lustre as a standalone high-performance file system to burst your workloads from on-premises to the cloud. By copying on-premises data to an FSx for Lustre file system, you can make that data available for fast processing by compute instances running on AWS. Amazon Elastic File System (Amazon EFS) provides simple, scalable file storage for use with Amazon EC2 instances in the AWS Cloud.

Automation and Orchestration

Automating the job submission process and scheduling submitted jobs according to predetermined policies and priorities are essential for efficient use of the underlying HPC infrastructure. AWS Batch lets you run hundreds to thousands of batch computing jobs by dynamically provisioning the right type and quantity of compute resources based on the job requirements. AWS Parallel Computing Service is a managed service for building and operating managed Slurm clusters. AWS Parallel Cluster is an open-source distributed tool used to assemble and operate HPC clusters. Amazon EnginFrame is a web portal designed to provide efficient access to HPC-enabled infrastructure using a standard browser. EnginFrame provides you a user-friendly HPC job submission, job control, and job monitoring environment.

Operations & Management

Monitoring the infrastructure and avoiding cost overruns are two of the most important capabilities that can help an HPC system administrators efficiently manage your organization's HPC needs. Amazon CloudWatch is a monitoring and management service built for developers, system operators, site reliability engineers (SRE), and IT managers. CloudWatch provides you with data and actionable insights to monitor your applications, understand and respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health. AWS Budgets gives you the ability to set custom budgets that alert you when your costs or usage exceed (or are forecasted to exceed) your budgeted amount.

Visualization Tools

The ability to visualize results of engineering simulations without having to move massive amounts of data to/from the cloud is an important aspect of the HPC stack. Remote visualization helps accelerate the turnaround times for engineering design significantly. Amazon DCV enables you to remotely access 2D/3D interactive applications over a standard network. In addition, Amazon AppStream 2.0 is another fully managed application streaming service that can securely deliver application sessions to a browser on any computer or workstation.

Security and Compliance

Security management and regulatory compliance are other important aspects of running HPC in the cloud. AWS offers multiple security related services and quick-launch templates to simplify the process of creating a HPC cluster and implementing best practices in data security and regulatory compliance. The AWS infrastructure puts strong safeguards in place to help protect customer privacy. All data is stored in highly secure AWS data centers. AWS Identity and Access Management (IAM) provides a robust solution for managing users, roles, and groups that have rights to access specific data sources. Organizations can issue users and systems individual identities and credentials, or provision them with temporary access credentials using the Amazon Security Token Service (Amazon STS). AWS manages dozens of compliance programs in its infrastructure. This means that segments of your compliance have already been completed. AWS infrastructure is compliant with many relevant industry regulations such as HIPAA, FISMA, FedRAMP, PCI, ISO 27001, SOC 1, and others.