

Introduction to AWS - Genomics Beginners



Clinical Cancer Genomics

powered by aws



Document	Introduction to AWS - Genomics Beginner
Creator	Venkatesh Chellappa
KI Course Code	5534
Date	24/04/2023 - 28/04/2023

Background

AWS is a cloud computing platform that provides a wide range of services for various industries including genomics.

We will explore what AWS is, the services available, and how it can be useful for genomics.

Introduction

AWS is a cloud computing platform that provides a range of services, including computing, storage, databases, analytics, and more. It was launched by Amazon in 2006 and has since become one of the most popular cloud computing platforms worldwide.

History, Overview and Insights

AWS started as an internal project within Amazon to provide its e-commerce infrastructure as a service to other companies. It was later opened to the public, and today, it's a leading cloud computing provider with a vast ecosystem of services and partners.

Overview of the AWS ecosystem: AWS has a vast ecosystem of services that can be used for various purposes, including hosting websites, processing data, building applications, and more. Its services are organized into categories such as compute, storage, databases, networking, security, and more.

How AWS has transformed the IT industry: AWS has transformed the IT industry by making it easier and more affordable to use computing resources. It has democratized access to cloud computing, allowing businesses and organizations of all sizes to leverage the power of the cloud.

Overview of Services

Brief descriptions of primary services offered by AWS: AWS offers hundreds of services, but some of the primary ones include:

- **Compute:** services for running applications and workloads, such as EC2 (Elastic Compute Cloud) and Lambda.
- **Storage:** services for storing and managing data, such as S3 (Simple Storage Service), EFS (Elastic File System), and Glacier.
- **Databases:** services for managing databases, such as Aurora, RDS (Relational Database Service), and DynamoDB.
- **Networking:** services for connecting resources, such as VPC (Virtual Private Cloud) and Direct Connect.
- **Security:** services for securing resources and data, such as IAM (Identity and Access Management) and WAF (Web Application Firewall).

AWS Services for Genomics

Overview of AWS services essential for genomics: In addition to the services listed in Slide 3, AWS offers several other services that are essential for genomics, including:

- Aurora: a highly scalable and available relational database service that can be used to store genomic data and metadata.
- RDS: a managed relational database service that can be used to run popular database engines such as MySQL, PostgreSQL, and SQL Server that can be used to store and manage genomic data.
- DynamoDB: a NoSQL database service that can be used to store and manage genomic data and metadata.

AWS compute services including EC2, Lambda, EMR, and Batch: These services can be used to run bioinformatics tools, workflows, and pipelines for genomic data analysis.

AWS Storage services including S3, EFS, and Glacier: These services can be used to store and manage genomic data in the cloud.

AWS Database services including Aurora, RDS, and DynamoDB: These services can be used to store and manage genomic data and metadata.

Benefits of using AWS for Genomics

Increased scalability and flexibility: AWS offers on-demand computing resources that can be easily scaled up or down based on the needs of the workload. This makes it easy to process large volumes of genomic data quickly and efficiently.

Cost savings and improved efficiency: AWS offers a pay-as-you-go pricing model, which means that you only pay for the resources you use. This can result in significant cost savings compared to traditional on-premises infrastructure. Additionally, AWS's managed services reduce the need for manual maintenance, allowing researchers to focus on their work.

Enhanced security and data privacy: AWS offers a wide range of security features, including network security, identity and access management, encryption, and compliance certifications. This makes it easy to secure genomic data and maintain data privacy.

Integration with widely used bioinformatic tools: AWS integrates with a wide range of bioinformatic tools and workflows, making it easy to move existing workloads to the cloud.

Case Study/Use Case

Case Study 1:

Example of a leading genomics company using AWS: One example of a leading genomics company using AWS is Illumina. Illumina is a global leader in genomic sequencing and array-based technologies.

Description of their use case: Illumina uses AWS to run its BaseSpace Sequence Hub, a cloud-based genomics analysis platform. BaseSpace Sequence Hub allows researchers to upload, process, and

analyze genomic data using a variety of tools and workflows.

Results obtained after transitioning to AWS: By transitioning to AWS, Illumina was able to increase the scalability and flexibility of its genomics analysis platform. This has enabled researchers to process and analyze large volumes of genomic data quickly and efficiently, leading to new insights and discoveries.

Case Study 2:

Another example of a leading genomics company using AWS is DNAexus. DNAexus is a cloud-based genomics platform that provides end-to-end solutions for sequencing data analysis and management.

Description of their use case: DNAexus uses AWS to power its platform, which allows researchers to analyze, store, and share genomic data. The platform uses AWS's scalable and flexible infrastructure to process large volumes of data quickly and efficiently.

Results obtained after transitioning to AWS: By transitioning to AWS, DNAexus was able to improve the scalability and flexibility of its platform. This has enabled researchers to process and analyze genomic data more efficiently, leading to new insights and discoveries. Additionally, using AWS has helped DNAexus reduce costs and increase the speed of its data processing.

Conclusion

We've discussed the primary AWS services relevant to genomics, including compute, storage, and database services. We've also highlighted the benefits of using AWS for genomics, such as increased scalability and flexibility, cost savings and improved efficiency, enhanced security and data privacy, and integration with widely used bioinformatic tools. Finally, we've given an example of leading genomics companies, Illumina and DNAexus, that use AWS to improve their genomics analysis.