

All about Data Engineering

# Data Engineering using AWS - 101

by Sachin Chandrashekhar

Data Engineering Hub

<https://masterclass.sachin.cloud>



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon S3

Amazon Simple Storage Service (S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. It's ideal for data storage, backup, and analytics.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon S3

Amazon S3 provides features like scalable storage, versioning, data encryption, lifecycle management, and integration with various AWS services. It supports diverse use cases such as data lakes, backups, and archival storage.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## What is Amazon Redshift?

Amazon Redshift is a fully managed petabyte-scale data warehouse service in the cloud, designed for large-scale data storage and real-time data analysis.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon Redshift

- Redshift offers fast query performance, scalable architecture, columnar storage, and data compression. It integrates with various data sources and supports advanced analytics and machine learning integrations.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS Glue

AWS Glue is a fully managed ETL service that makes it easy to prepare and transform data for analytics, using a serverless architecture to automate data integration tasks.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS Glue

AWS Glue includes a data catalog for metadata management, job scheduling, data transformation with Spark, and support for various data sources and targets. It simplifies ETL processes with automated schema discovery and code generation.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Amazon EMR Overview

Amazon Elastic MapReduce (EMR) is a cloud-native big data platform that allows processing vast amounts of data quickly and cost-effectively using open-source tools like Hadoop, Spark, and Hive.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon EMR

- EMR provides managed clusters with automatic scaling, integration with S3 for storage, and support for a wide range of big data processing frameworks. It offers flexibility and cost-effectiveness for processing large datasets.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

# Understanding Amazon Kinesis

Amazon Kinesis is a platform for real-time data streaming and analytics, providing powerful tools for processing and analyzing large streams of data in real-time.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon Kinesis

- Kinesis offers services like Kinesis Data Streams for real-time data ingestion, Kinesis Data Firehose for data delivery to storage and analytics services, and Kinesis Data Analytics for real-time data processing.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS Lambda

AWS Lambda is a serverless compute service that lets you run code in response to events without provisioning or managing servers, enabling you to build scalable and cost-effective data processing workflows.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS Lambda

- Lambda offers automatic scaling, built-in fault tolerance, and integration with various AWS services for event-driven execution. It supports multiple programming languages and provides fine-grained access control with IAM.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon RDS

Amazon Relational Database Service (RDS) is a managed relational database service that supports various database engines, offering scalability, high availability, and security for data-driven applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon RDS

- RDS provides automated backups, patching, and failover support. It allows you to choose from multiple database engines like MySQL, PostgreSQL, and Oracle, and integrates with other AWS services for enhanced functionality.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon DynamoDB

Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability, ideal for key-value and document data models:



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon DynamoDB

DynamoDB offers single-digit millisecond response times, automatic scaling, built-in security features, and global replication. It supports flexible data models and is optimized for high-traffic applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS Fargate

AWS Fargate is a serverless compute engine for containers that works with Amazon ECS. It allows you to run containers without managing the underlying infrastructure, simplifying containerized application deployment.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS Fargate

- Fargate handles container orchestration, scaling, and provisioning, enabling you to focus on application development. It integrates with ECS for task management and supports both Linux and Windows containers.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon MSK

Amazon Managed Streaming for Apache Kafka (MSK) is a fully managed service that makes it easy to build and run applications that use Apache Kafka for processing streaming data.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon MSK

MSK provides a fully managed Kafka service with automatic scaling, data replication, and monitoring. It integrates with other AWS services and offers seamless migration from on-premises Kafka clusters.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon AppFlow

Amazon AppFlow is a fully managed integration service that enables you to securely transfer data between AWS services and SaaS applications like Salesforce, ServiceNow, and SAP, without writing any code.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon AppFlow

- AppFlow provides bidirectional data transfer, data transformation capabilities, and pre-built connectors for popular SaaS applications. It also supports encryption and data filtering to ensure secure and relevant data transfers.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

# Introduction to AWS Systems Manager Parameter Store

AWS Systems Manager Parameter Store is a secure, scalable, and highly available service to centrally manage configuration data and secrets as text parameters. It provides a simple interface to store and retrieve configuration information.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS Systems Manager Parameter Store

Parameter Store enables you to manage configuration data for applications and data engineering workflows, such as database connection strings, API keys, and environment variables, ensuring secure and consistent access.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS Secrets Manager

AWS Secrets Manager is a fully managed service that makes it easy to rotate, manage, and retrieve database credentials, API keys, and other secrets throughout their lifecycle, enhancing security and compliance.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS Secrets Manager

Secrets Manager allows you to securely store and manage sensitive information, automate secret rotation with built-in integrations for RDS databases, and access secrets using the AWS SDK or CLI.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon MWAA

Amazon Managed Workflows for Apache Airflow (MWAA) is a managed orchestration service that makes it easier to run data pipelines using Apache Airflow in the cloud. MWAA handles scaling, patching, and securing Airflow environments.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon MWAA

MWAA offers automatic scaling, high availability, and seamless integration with AWS services, reducing the operational overhead of managing Apache Airflow environments and allowing data engineers to focus on building workflows.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon EventBridge

Amazon EventBridge is a serverless event bus service that allows you to connect application data from your own applications, SaaS applications, and AWS services to create event-driven applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon EventBridge

- EventBridge enables real-time data processing by routing events from different sources to target AWS services like Lambda, Step Functions, or SQS, making it an essential tool for building data-driven applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon Athena

Amazon Athena is an interactive query service that enables you to analyze data directly in Amazon S3 using standard SQL. It's serverless, so there's no infrastructure to manage, and it scales automatically.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon Athena

- Athena supports a variety of data formats, including CSV, JSON, ORC, and Parquet. It integrates with AWS Glue Data Catalog for metadata management and provides pay-per-query pricing, making it cost-effective for on-demand queries.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon QuickSight

Amazon QuickSight is a scalable, serverless, embeddable business intelligence (BI) service that provides interactive data visualizations and insights. It helps you create and share rich dashboards and reports.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon QuickSight

QuickSight offers machine learning insights, customizable dashboards, and support for various data sources. It provides fast and interactive visualizations and scales automatically with user needs.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon EC2

Amazon Elastic Compute Cloud (EC2) provides scalable compute capacity in the cloud. It allows you to run virtual servers, manage storage, and scale resources based on your application's needs.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon EC2

- EC2 offers a variety of instance types optimized for different workloads, flexible storage options, and auto-scaling. It integrates with other AWS services for enhanced performance and management.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon CloudWatch

Amazon CloudWatch provides monitoring and observability for AWS resources and applications. It collects and tracks metrics, collects and monitors log files, and sets alarms.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon CloudWatch

CloudWatch offers detailed insights into system performance, automated alarms, and event-driven actions. It integrates with AWS services to provide comprehensive monitoring and logging capabilities.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS IAM

AWS Identity and Access Management (IAM) enables you to manage access to AWS services and resources securely. You can create and manage users, groups, and permissions to control access.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS IAM

- IAM supports multi-factor authentication, role-based access control, and detailed access policies. It integrates with AWS services to enforce security and compliance requirements.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon SNS

Amazon Simple Notification Service (SNS) is a fully managed messaging service that enables you to send messages to multiple subscribers, including SMS, email, and application endpoints.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon SNS

- SNS supports pub/sub messaging, message filtering, and integration with other AWS services. It provides scalable and reliable messaging for distributed applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to Amazon SQS

Amazon Simple Queue Service (SQS) is a fully managed message queuing service that allows you to decouple and scale microservices, distributed systems, and serverless applications.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of Amazon SQS

- SQS offers standard and FIFO queues, ensuring high throughput and message order. It provides reliable message delivery and automatic scaling to handle varying workloads.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Introduction to AWS KMS

AWS Key Management Service (KMS) is a managed service that makes it easy to create and control encryption keys used to encrypt your data, providing centralized key management.



# Data Engineering Hub

- Sachin Chandrashekhar  
<https://masterclass.sachin.cloud>

## Key Features of AWS KMS

- KMS supports automatic key rotation, integrates with other AWS services for encryption, and provides detailed logging of key usage. It ensures data protection and compliance.



All about Data Engineering



I teach Real-world AWS Data Engineering.  
Go to the link below

by Sachin Chandrashekhar

<https://masterclass.sachin.cloud>

Save