

Day 8/13

Module 8 - AI ML and Data Analytics

Date: _____
DELTA PUN

AI - A broad field focused on the development of intelligent ^{computer} systems capable of performing human like tasks.

ML → A type of AI for training machines to perform complex tasks without explicit instructions.

#1 AWS AI/ML Solutions

Tier 1 - Pre built AI Services.

(I) Language Services (Text/speech Related)

Amazon comprehend - Text analysis (sentiment, entities)

Amazon Translate - Language Translation

Amazon Transcribe → Speech to Text

Amazon Polly → Text to Speech

(II) Computer Vision & Search Services (Images/ Search)

Amazon Rekognition - Image/Video Analysis

Amazon Textract - Document text Extraction

Amazon Kendra - Intelligent enterprise search

(III) Conversational AI & Personalisation

Amazon Lex - Chatbot / Voicebot

Amazon Personalise → Recommendation engine

Tier 2 - ML Services

~~And~~ Train & deploy your own model without
infrastructure

Amazon SageMaker AI - Apha custom ML model
build, deploy karha ho → use this

Tier 3 - ML Framework & Infrastructure

ML Frameworks (Py torch, Tensorflow etc) - khud
coding karke model banana ho, use this

AWS ML Infrastructure (EC2, EMR, ECS) - High
performance compute Chahiye ML training ke
liye, use these

EMR? → Elastic MapReduce

Large-scale data processing, ML training on huge
datasets, log analysis / Big data processing service

2 AWS Generative AI Solutions

Deep Learning - Deep learning is a subset of
ML that uses multi-layer neural networks
to learn complex patterns from large data

Generative AI (Gen AI)

Generative AI uses large pre-trained foundation models (FMs) to create new content like text, images, audio or code.

AWS Gen AI Services

① Amazon SageMaker Jumpstart

Ready to deploy - pre-trained models and ML solutions with quick customization.

② Amazon Bedrock - Fully managed service to access and customize foundation models via a single API.

③ Amazon Q - AI assistant that integrates with company data to answer questions and boost productivity.

#3 Deeper into above Services

Amazon SageMaker Jumpstart

Pre-trained ML/FMs ka library inside SageMaker - quickly deploy, fine-tune and experiment with models in few clicks

Use when:

→ Need fast prototype

→ Pre-trained model customize karna hai

→ Compare multiple models before finalizing

Amazon Bedrock

Fully managed service to access & customize multiple foundation models via single API

- production-ready GenAI app banana hai
- Multiple FMs test karne hai (Claude, SDEK)
- Enterprise security + scalability chahiye.

Amazon Q Business

Enterprise AI assistant - company data se answer de, workflows automate kare.

Use when:

- Employees ko internal knowledge search chahiye
- Enterprise data se insights extract karna ho.

Amazon Q Developer

AI coding assistant - code suggest kare, generate kare, review kare

Use when:

Developers ko faster coding chahiye
 Aur code review + security suggestions

#4 Intro to Data Analytics

ETL Process

- Extract data from various sources
- Transform it into a consistent, usable format
- Load it into a destination system, like a data warehouse

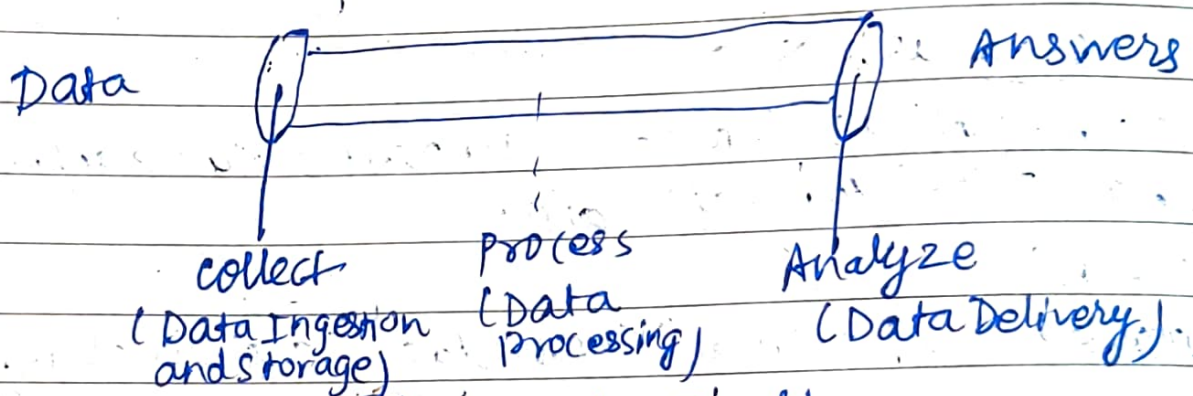


Fig - Data Pipeline

AWS Data Pipeline Services

- Amazon Kinesis and AWS Glue for data ingestion
- Amazon S3 and Amazon Redshift for storage
- Amazon EMR for processing
- Amazon QuickSight for visualisation.

Redshift: Large scale structured data ko store karke SQL se analyze karne ke liye
 S3: Files store karne ke liye, not an analytics engine

#5 Data pipelines on AWS.

① Data Ingestion - It involves moving data from source systems into your chosen storage solution.

Amazon Kinesis Data Streams

- Ingest data in real-time
- Deliver to multiple data consumers

Amazon Data Firehouse

- Ingest data in near real-time
- Batch compress and encrypt data before loading

② Data Storage Services

To gain insights, data is commonly consolidated into a single location.

Amazon S3 - It is a popular choice for data lakes. It is fully elastic and can store any amount of structured or unstructured data.

Amazon Redshift - Amazon Redshift is a fully managed service that can store petabytes of structured and semi-structured data.

Store + Analyze.

(III) Data Cataloging Services (AWS Glue Data Catalog)

→ Cataloging your data with metadata provides an inventory of your organization's data.

→ Fully managed metadata repository

→ Discover, catalog and manage data

→ Automate data discovery

→ Data governance and compliance support.

(IV) Data processing Service

→ Data processing services clean and transform your data so it's ready to be analyzed

AWS Glue

→ Fully managed ETL Service

→ Create ETL jobs visually with built-in Scheduling

→ Simple, code free data processing

Amazon EMR

→ Perform large-scale complex data processing using popular frameworks (such as Apache Spark)

- Ideal for organisations with big data exposure.
- More control and Flexibility.

① Data Analysis and Visualisation Services
Queries and Visualisation tools help you develop important insights about your data.

Amazon Athena

- Fully managed serverless services.
- Ideal for standard, infrequent SQL queries
- Analyze data across environments

Amazon Redshift

- Fully managed data warehouse service.
- Ideal for complex queries
- Fine-grained query control
- Advanced features

Amazon QuickSight

- Create interactive business intelligence dashboards and reports.
- Scales to tens and thousands of users
- Ideal for technical and non-technical users.
- Build, Analyze and Share.

Amazon OpenSearch Service

- perform real-time search, monitoring and analysis of business and operational data.
- Ideal for app monitoring, log analytics and website search.