CHAPTER 11

Analytics and Machine Learning

**Introduction**

Cloud computing, through AWS, facilitates the integration of analytics and machine learning, allowing businesses to gain insights from large datasets and implement innovative solutions. This chapter focuses on AWS services specifically designed for analytics and machine learning, providing essential tools for data-driven decision-making and advanced artificial intelligence applications.

* Analytics
  + Amazon Athena
  + Amazon CloudSearch
  + Amazon DataZone (Preview)
  + Amazon EMR
  + Hosted Hadoop framework
  + Amazon FinSpace
  + Amazon Kinesis
  + Amazon Managed Streaming for Apache Kafka (MSK)
  + Amazon OpenSearch Service
  + Amazon QuickSight
  + Amazon Redshift
  + AWS Clean Rooms (Preview)
  + AWS Data Exchange
  + AWS Data Pipeline
  + AWS Glue
  + AWS Lake Formation
* Machine Learning
  + Amazon Augmented AI
  + Amazon Bedrock
  + Amazon CodeGuru
  + Amazon Comprehend
  + Amazon DevOps Guru
  + Amazon Elastic Inference
  + Deep learning inference acceleration
  + Amazon Forecast
  + Amazon Fraud Detector
  + Amazon HealthLake
  + Amazon Kendra
  + Amazon Lex
  + Amazon Lookout for Equipment
  + Amazon Lookout for Metrics
  + Amazon Monitron
  + Amazon Omics
  + Amazon Personalize
  + Amazon Polly
  + Amazon Rekognition
  + Amazon SageMaker
  + Amazon SageMaker Ground Truth
  + Amazon Textract
  + Amazon Transcribe
  + Amazon Translate
  + Apache MXNet on AWS
  + AWS Deep Learning AMIs
  + AWS Deep Learning Containers
  + AWS DeepComposer
  + AWS DeepLens
  + AWS DeepRacer
  + AWS Inferentia
  + AWS Panorama
  + PyTorch on AWS
  + TensorFlow on AWS
  + Amazon CodeWhisperer

This chapter explains the features and applications of various AWS services, including analytics and machine learning. It highlights how these tools support data-driven decision-making and artificial intelligence for businesses.

Part 1: Analytics

Analytics is essential in cloud computing for deriving insights from large datasets. This section highlights key AWS services for efficient data processing, analysis, and visualization.

A diagram of a software company

Description automatically generated

Figure 1 Architecture: Data is collected from multiple data sources across the enterprise, including software-as-a-service (SaaS) applications, edge devices, logs, streaming media, and social networks (AWS Documentation).

Amazon Athena:

Amazon Athena is a serverless query service that lets users analyze data in Amazon S3 with SQL queries, enabling immediate analysis without complex data transformations or infrastructure (AWS, n.d.a).

Amazon CloudSearch:

Amazon Athena is a serverless query service (AWS, Amazon CloudSearch, n.d.).

Amazon DataZone:

Amazon DataZone (Preview) offers a secure data-sharing environment and collaboration in the life sciences. It supports genomics and biomedical research by facilitating safe collaboration on sensitive data (AWS, Amazon DataZone, n.d.).

Amazon EMR (Elastic MapReduce):

Amazon EMR is a big cloud-based data platform that processes large datasets using popular frameworks such as Apache Spark and Apache Hadoop. EMR enables scalable and cost-effective data processing, making it a fundamental tool for big data analytics (AWS, Amazon EMR, n.d.).

\*\*Hosted Hadoop Framework:

\*\* AWS provides a hosted Hadoop framework that allows users to quickly deploy and manage Hadoop clusters, enabling scalable and reliable distributed data processing (AWS, AWS Analytics, n.d.)

\*\*Amazon FinSpace:

\*\* Designed for the financial industry, Amazon FinSpace enhances data management, analytics, and collaboration, addressing the specific needs of financial data workflows (AWS, Amazon FinSpace, n.d.).

\*\*Amazon Kinesis:

\*\* Amazon Kinesis offers services for real-time processing of streaming data at scale, allowing applications to ingest, buffer easily, and process data for effective real-time analytics (AWS, Amazon Kinesis, n.d.).

Amazon Managed Streaming for Apache Kafka (MSK):

\*\*Amazon MSK\*\*: A fully managed Kafka service that streamlines the deployment and management of Apache Kafka clusters for reliable streaming data and analytics (AWS, Amazon Managed Streaming for Apache Kafka (MSK), n.d.).

\*\*Amazon OpenSearch Service

\*\*: A managed Elasticsearch service that simplifies the deployment and operation of Elasticsearch clusters, enabling scalable search solutions (AWS, Amazon OpenSearch Service, n.d.).

\*\*Amazon QuickSight\*\*: A cloud-based business analytics service that allows users to create interactive dashboards and visualizations for better data insights (AWS, Amazon QuickSight, n.d.).

\*\*Amazon Redshift

\*\*: A fully managed data warehouse service optimized for high-performance analysis, enabling complex queries on large datasets (AWS, Amazon Redshift, n.d.).

\*\*AWS Clean Rooms

\*\*: A secure environment for analyzing sensitive data, ensuring compliance with regulatory requirements (AWS, AWS Clean Rooms (Preview, n.d.).

\*\*AWS Data Exchange\*\*: A marketplace for discovering and subscribing to third-party datasets, facilitating data collaboration and accessibility (AWS, AWS Data Exchange, n.d.).

\*\*AWS Data Pipeline

\*\*: A service for automating data movement and transformation between AWS services, making data pipeline management (AWS, AWS Data Pipeline, n.d.)easier.

\*\*AWS Glue

\*\*: A fully managed ETL service that automates data preparation, providing a serverless environment for efficient data integration (AWS, AWS Glue, n.d.)

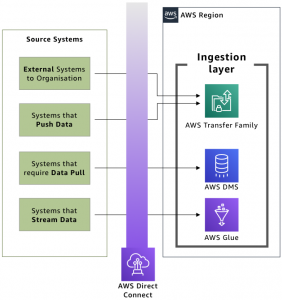
.

Figure 2 Ingestion layer against source systems (AWS Blog).

AWS Lake Formation:

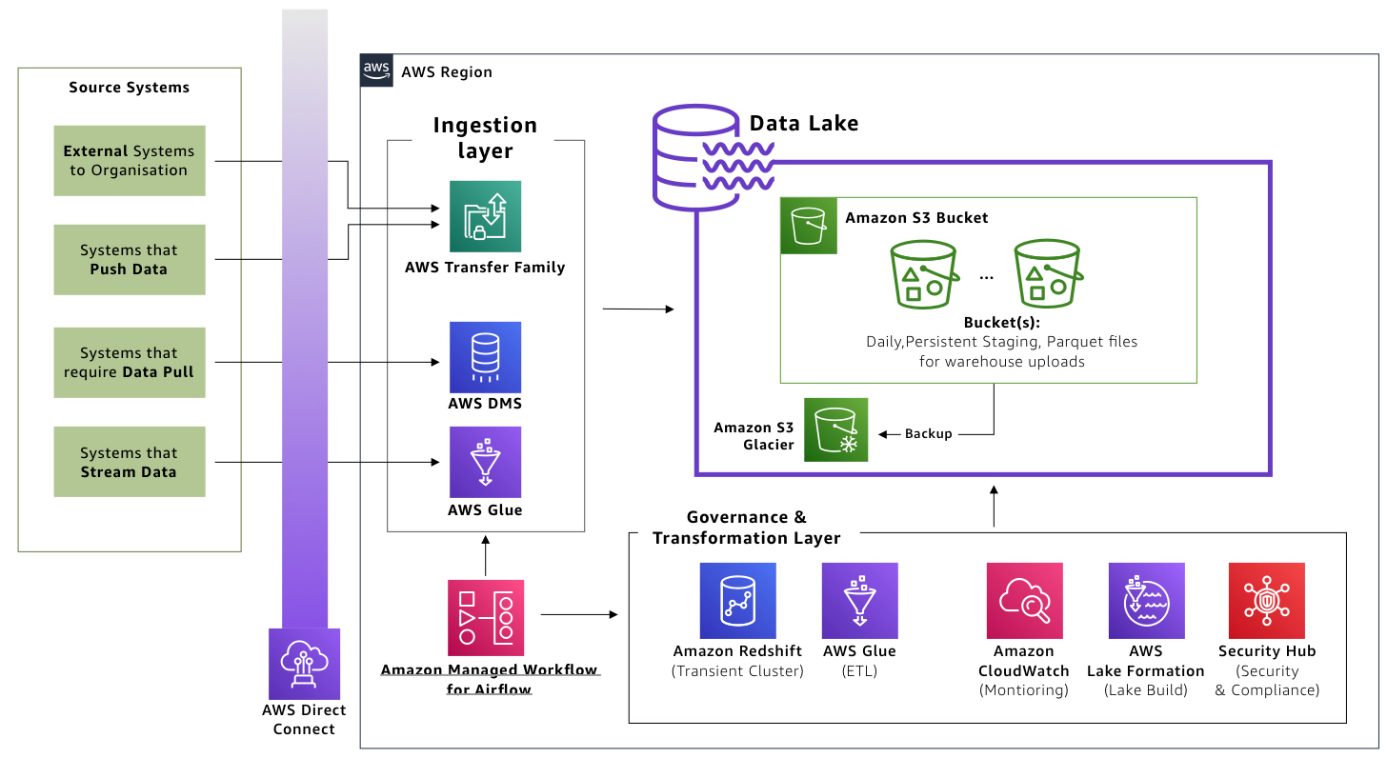


Figure 3 The Governance and transformation layer prepares data in the lake (AWS Blog).

AWS Lake Formation simplifies the creation, security, and management of data lakes by providing tools for data ingestion, security, and access control (AWS, AWS Lake Formation, n.d.)

. This overview highlights the wide range of AWS analytics services available for processing, analyzing, and visualizing data. As we explore each service, you'll learn how to apply these tools to various analytics

tasks. In machine learning, AWS offers a robust suite of services for building custom models and integrating pre-trained solutions.

Part 2: Machine Learning

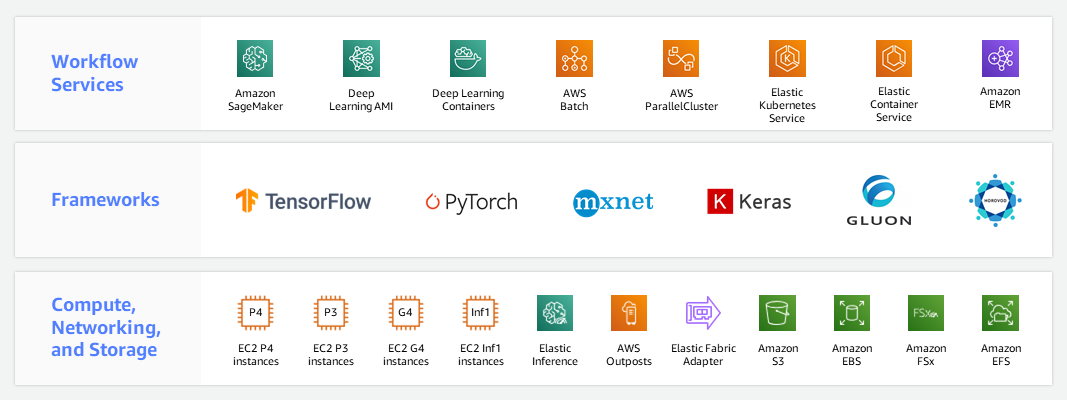


Figure 4 AWS Machine Learning Infrastructure (AWS Documentation).

\*\*Amazon Augmented AI (A2I):

\*\* This service allows developers to create custom machine learning workflows that incorporate human review, enhancing the reliability and accuracy of predictions (AWS, Amazon Augmented AI, n.d.).

\*\*Amazon Bedrock:

\*\* Bedrock simplifies the entire machine learning process, from data preparation and model training to deployment, making it easier to develop scalable machine learning applications (AWS, Amazon Bedrock, n.d.).

\*\*Amazon CodeGuru:

\*\* CodeGuru automatically reviews code, provides improvement recommendations, and enhances code performance and reliability (AWS, Amazon CodeGuru, n.d.)

using machine learning.

\*\*Amazon Comprehend:\*\* This service performs natural language processing tasks, extracting insights and relationships from text while supporting multiple languages to help developers build advanced language applications (AWS, Amazon Comprehend, n.d.).

\*\*Amazon DevOps Guru

\*\*: Utilizes machine learning to identify operational issues and anomalies, automate problem detection, and provide actionable insights to improve application reliability (AWS, Amazon DevOps Guru, n.d.).

\*\*Amazon Elastic Inference

\*\*: Enhances machine learning instances with cost-effective GPU-powered inference acceleration to improve deep learning inference (AWS, Amazon Elastic Inference, n.d.).

\*\*Deep Learning Inference Acceleration

\*\*: AWS delivers purpose-built hardware, like AWS Inferentia, for high throughput and low-latency inference (AWS, AWS Deep Learning Containers, n.d.)

to boost machine learning model performance.

\*\*Amazon Forecast\*\*: A fully managed service that automates forecasting using machine learning for applications like demand forecasting and financial planning (AWS, Amazon Forecast, n.d.)

.

\*\*Amazon Fraud Detector\*\*: Leverages machine learning to analyze historical data and create custom models, helping to detect and prevent online fraud.

\*\*Amazon HealthLake\*\*: A HIPAA-eligible service that allows healthcare providers to securely store, transform, and analyze health data using machine learning for structured data extraction from unstructured medical information.

\*\*Amazon Augmented AI (A2I):\*\* This service allows developers to create custom machine learning workflows that incorporate human review, enhancing the reliability and accuracy of predictions (AWS, Amazon Fraud Detector, n.d.).

\*\*Amazon

Bedrock:\*\* Bedrock simplifies the entire machine learning process, from data preparation and model training to deployment, making it easier to develop scalable machine learning applications.

\*\*Amazon CodeGuru:\*\* CodeGuru automatically reviews code, provides improvement recommendations, and enhances code performance and reliability using machine learning.

\*\*Amazon Comprehend:\*\* This service performs natural language processing tasks, extracting insights and relationships from (AWS, Amazon HealthLake, n.d.)text while supporting multiple languages to help developers build advanced language applications.

\*\*Amazon Kendra:

\*\* An intelligent search service that leverages machine learning to help organizations integrate robust search capabilities into their applications, facilitating easy access to relevant information (AWS, Amazon Kendra, n.d.).

\*\*Amazon Lex:

\*\* A service designed to streamline the creation of conversational interfaces with natural language understanding, powering chatbots and interactive voice response systems to enhance user interactions through machine learning (AWS, Amazon Lex, n.d.).

\*\*Amazon Lookout for Equipment:

\*\* This service employs machine learning to detect abnormal behaviors in equipment by analyzing sensor data, enabling proactive maintenance and reducing downtime (AWS, Amazon Lookout for Equipment, n.d.).

\*\*Amazon Lookout for Metrics:

\*\* A machine learning service that identifies anomalies in metrics, automates monitoring key performance indicators, and provides timely alerts for unusual patterns (AWS, Amazon Lookout for Metrics, n.d.).

\*\*Amazon Monitron:

\*\* An end-to-end solution for equipment monitoring that integrates sensors, a gateway, and machine learning to anticipate equipment failures before they occur (AWS, Amazon Monitron, n.d.).

\*\*Amazon Omics:

\*\* A comprehensive service for analyzing genomic data at scale, utilizing machine learning to help researchers extract meaningful insights and drive advancements in scientific discoveries within the life sciences (AWS, Amazon Omics, n.d.).

\*\*Amazon Personalize:\*\*

Amazon Personalize is a machine learning service that facilitates the creation of personalized recommendations for users. It customizes recommendations for products, content, and more (AWS, Amazon Personalize, n.d.)by analyzing user behavior.

\*\*Amazon Polly:\*\*

Amazon Polly converts text into lifelike speech using advanced machine-learning techniques. With support for multiple languages and various voices, Polly enables developers to integrate natural-sounding speech into their applications (AWS, Amazon Polly, n.d.).

\*\*Amazon Rekognition:\*\*

Amazon Rekognition is a powerful image and video analysis service that leverages machine learning. It can identify objects, people, text, scenes, and activities, making it an invaluable tool for content analysis and security applications (AWS, Amazon Rekognition, n.d.).

\*\*Amazon SageMaker:\*\*

Amazon SageMaker is a fully managed machine learning service encompassing the entire ML workflow. It simplifies the processes of model building, training, and deployment, allowing developers to concentrate on creating robust machine-learning applications (AWS, Amazon SageMaker, n.d.).

\*\*Amazon SageMaker Ground Truth:\*\*

Amazon SageMaker Ground Truth is a data labeling service that utilizes machine learning to lower labeling costs and enhance annotation accuracy. It streamlines the creation of high-quality training datasets for machine learning (AWS, Amazon SageMaker, n.d.)purposes.

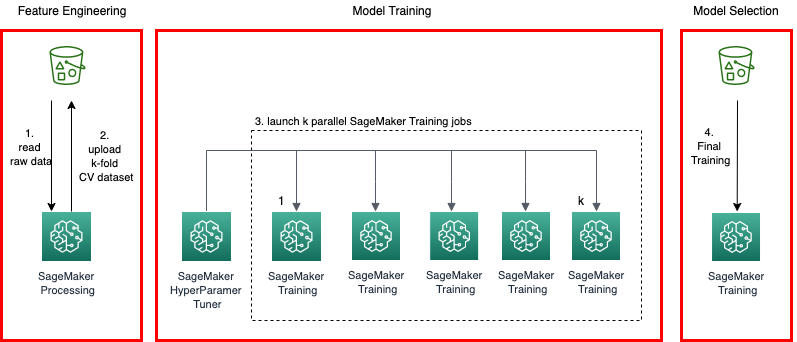


Figure 5 Cross-Validation Machine Learning Model Pipeline at Scale with Amazon SageMaker (AWS Documentation).

\*\*Amazon Textract:\*\*

Amazon Textract is a fully managed Optical Character Recognition (OCR) service powered by machine learning. It automates the extraction of text, forms, and tables from scanned documents, streamlining the data extraction (AWS, Amazon Textract, n.d.)process.

\*\*Amazon Transcribe:\*\*

Amazon Transcribe delivers automatic speech recognition (ASR) services through machine learning. This service accurately converts spoken language into written text, allowing applications to transcribe audio content effectively (AWS, Amazon Transcribe, n.d.).

\*\*Amazon Translate:\*\*

Amazon Translate is a neural machine translation service that facilitates translation between multiple languages. It utilizes machine learning to provide accurate, natural-sounding translations suitable for various applications (AWS, Amazon Translate, n.d.).

\*\*Apache MXNet on AWS:\*\*

AWS supports Apache MXNet, an open-source deep learning framework. With AWS infrastructure, developers can use MXNet's scalability and flexibility to efficiently build and deploy machine learning models (AWS, Apache MXNet on AWS, n.d.).

\*\*AWS Deep Learning AMIs:\*\*

AWS provides Deep Learning Amazon Machine Images (AMIs), which offer a collection of deep learning frameworks. These AMIs simplify the setup process for creating a deep learning environment on EC2 instances.

\*\*AWS Deep Learning Containers:\*\*

AWS Deep Learning Containers supply pre-configured Docker images tailored for deep learning applications. These containers create a consistent and reproducible environment for executing machine learning workloads (AWS, AWS Deep Learning Containers, n.d.).

\*\*AWS DeepComposer:\*\*

AWS DeepComposer is an innovative machine learning-enabled keyboard that empowers developers to create music using generative AI models. It showcases the creative potential of integrating machine learning with music composition (AWS, AWS DeepComposer, n.d.).

\*\*AWS DeepLens:\*\*

AWS DeepLens is a deep learning-enabled video camera designed to aid in developing computer vision applications. It provides a hands-on approach to learning and implementing deep learning models in practical scenarios (AWS, AWS DeepLens, n.d.).

\*\*AWS DeepRacer:\*\*

AWS DeepRacer is an autonomous 1/18th scale race car that facilitates reinforcement learning. It allows developers to deepen their understanding of machine learning concepts within an interactive and competitive racing environment (AWS, AWS DeepRacer, n.d.).

\*\*AWS Inferentia:\*\*

AWS Inferentia is a custom-built chip designed to accelerate deep learning inference workloads. With its high throughput and low latency, Inferentia significantly enhances the performance of machine-learning models (AWS, AWS Inferentia, n.d.).

\*\*AWS Panorama:\*\*

AWS Panorama is a machine learning appliance that empowers on-premises cameras with computer vision capabilities. It enables the local analysis of video feeds, creating new possibilities for applications in industrial automation and (AWS, AWS Panorama, n.d.)other fields.

\*\*PyTorch on AWS:\*\*

AWS offers support for PyTorch, an open-source deep learning framework. Leveraging AWS infrastructure, developers can enjoy the flexibility and efficiency of PyTorch in building and deploying machine learning models (AWS, PyTorch on AWS, n.d.)effectively.

\*\*TensorFlow on AWS:\*\*

AWS provides comprehensive support for TensorFlow, another open-source machine learning framework. Developers can use AWS's scalability and power to build and train machine learning models using TensorFlow (AWS, TensorFlow on AWS, n.d.).

\*\*Amazon CodeWhisperer:\*\*

Amazon CodeWhisperer is a service that utilizes machine learning to assist developers in writing code more efficiently. Offering context-aware suggestions enhances the coding experience and speeds up development workflows (AWS, Amazon CodeWhisperer, n.d.)

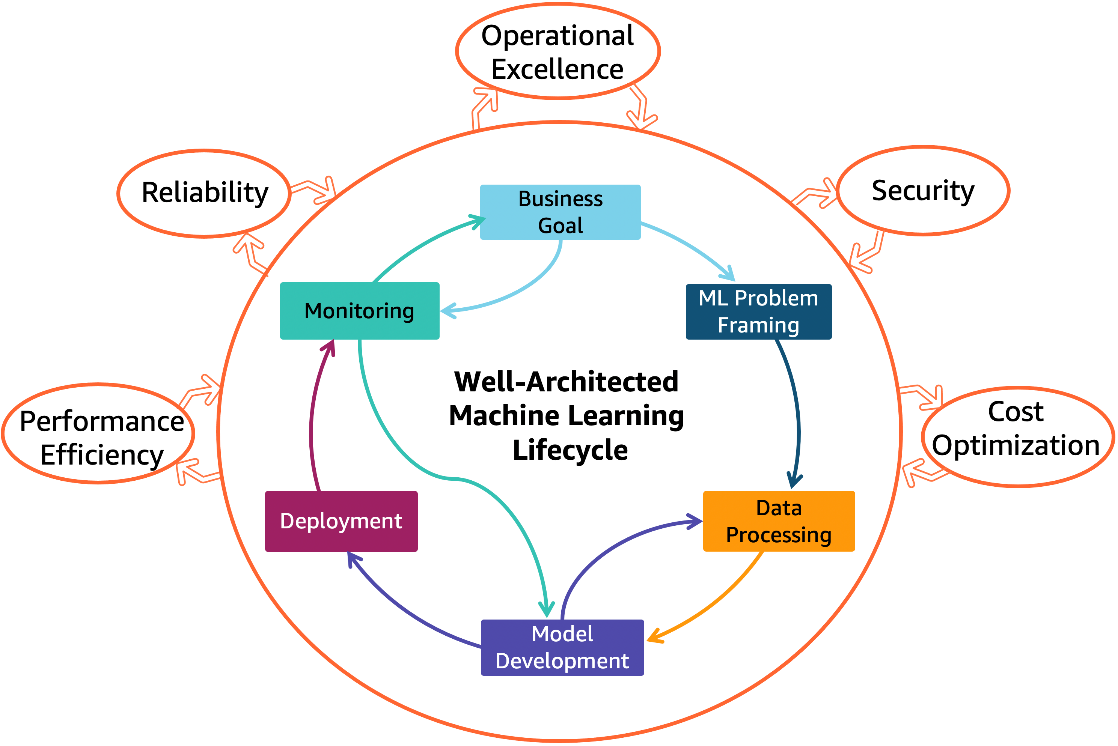
****.

Figure 6 AWS Well-Architected Machine Learning Lens (AWS Architecture).

**\*\*Chapter 11: Conclusion\*\***

Chapter 11 provides an overview of Analytics and Machine Learning within AWS, showcasing various services that enhance data-driven decision-making and automation.

\*\*Part 1: Analytics\*\*

AWS offers several analytics services for managing large datasets. Amazon Athena enables real-time data analysis through a serverless query service for Amazon S3 (AWS, n.d.a), while Amazon QuickSight provides data visualization (AWS, Amazon QuickSight, n.d.)

tools. Amazon EMR processes large datasets using frameworks like Apache Spark and Hadoop (AWS, Amazon Managed Streaming for Apache Kafka (MSK), n.d.), and Amazon Redshift is a powerful data warehousing solution (AWS, Amazon Redshift, n.d.)

. AWS Glue and AWS Lake Formation assist with data integration and transformation, (AWS, AWS Glue, n.d.)supporting effective analytics frameworks (AWS, AWS Lake Formation, n.d.).

\*\*Part 2: Machine Learning\*\*

In machine learning, AWS provides services for developers and data scientists, with Amazon SageMaker offering end-to-end capabilities (AWS, Amazon SageMaker, n.d.)

for model development. Other services like Amazon Comprehend (AWS, Amazon Comprehend, n.d.)and Amazon Rekognition deliver natural language processing and computer vision functionalities (AWS, Amazon Rekognition, n.d.). Amazon Personalize helps create tailored user experiences, and Amazon Forecast generates accurate predictions for demand for

ecasting.

Overall, Chapter 11 highlights AWS's commitment to making analytics and machine learning accessible. It provid (AWS, Amazon Personalize, n.d.)es a range of services that address the diverse needs of businesses and position data as a strategic asset. Chapter 11 provides an overview of Analytics and Machine Learning within AWS, showcasing various services that enhance data-driven decision-making and (AWS, Amazon Forecast, n.d.)automation.

\*\*Part 1: Analytics\*\*

\*\*Part 2: Machine Learning\*\*

In machine learning, AWS provides services for developers and data scientists, with Amazon SageMaker

offering end-to-end capabilities for model development. Other services like Amazon Comprehend and Amazon Rekognition deliver natural language processing and computer vision functionalities. Amazon Personalize helps create tailored user experiences, and Amazon Forecast generates accurate predictions for demand forecasting.

Chapter 11 highlights AWS's commitment to making analytics and machine learning accessible, providing services that address businesses' diverse needs, and positioning data as a strategic asset.

**References**

AWS. (n.d.). Amazon Augmented AI. Retrieved from https://aws.amazon.com/augmented-ai/

AWS. (n.d.). Amazon Bedrock. Retrieved from https://aws.amazon.com/bedrock/

AWS. (n.d.). Amazon CloudSearch. Retrieved from https://aws.amazon.com/cloudsearch/

AWS. (n.d.). Amazon CodeGuru. Retrieved from https://aws.amazon.com/codeguru/

AWS. (n.d.). Amazon CodeWhisperer. Retrieved from https://aws.amazon.com/codewhisperer/

AWS. (n.d.). Amazon Comprehend. Retrieved from https://aws.amazon.com/comprehend/

AWS. (n.d.). Amazon DataZone. Retrieved from https://aws.amazon.com/datazone/

AWS. (n.d.). Amazon DevOps Guru. Retrieved from https://aws.amazon.com/devops-guru/

AWS. (n.d.). Amazon Elastic Inference. Retrieved from https://aws.amazon.com/elastic-inference/

AWS. (n.d.). Amazon EMR. Retrieved from https://aws.amazon.com/emr/

AWS. (n.d.). Amazon FinSpace. Retrieved from https://aws.amazon.com/finspace/

AWS. (n.d.). Amazon Forecast. Retrieved from https://aws.amazon.com/forecast/

AWS. (n.d.). Amazon Fraud Detector. Retrieved from https://aws.amazon.com/fraud-detector/

AWS. (n.d.). Amazon HealthLake. Retrieved from https://aws.amazon.com/healthlake/

AWS. (n.d.). Amazon Kendra. Retrieved from https://aws.amazon.com/kendra/

AWS. (n.d.). Amazon Kinesis. Retrieved from https://aws.amazon.com/kinesis/

AWS. (n.d.). Amazon Lex. Retrieved from https://aws.amazon.com/lex/

AWS. (n.d.). Amazon Lookout for Equipment. Retrieved from https://aws.amazon.com/lookout-for-equipment/

AWS. (n.d.). Amazon Lookout for Metrics. Retrieved from https://aws.amazon.com/lookout-for-metrics/

AWS. (n.d.). Amazon Managed Streaming for Apache Kafka (MSK). Retrieved January 28, 2025, from https://aws.amazon.com/msk/

AWS. (n.d.). Amazon Monitron. Retrieved from https://aws.amazon.com/monitron/

AWS. (n.d.). Amazon Omics. Retrieved from https://aws.amazon.com/omics/

AWS. (n.d.). Amazon OpenSearch Service. Retrieved from https://aws.amazon.com/opensearch-service/

AWS. (n.d.). Amazon Personalize. Retrieved from https://aws.amazon.com/personalize/

AWS. (n.d.). Amazon Polly. Retrieved from https://aws.amazon.com/polly/

AWS. (n.d.). Amazon QuickSight. Retrieved from https://aws.amazon.com/quicksight/

AWS. (n.d.). Amazon Redshift. Retrieved from https://aws.amazon.com/redshift/

AWS. (n.d.). Amazon Rekognition. Retrieved from https://aws.amazon.com/rekognition/

AWS. (n.d.). Amazon SageMaker. Retrieved from https://aws.amazon.com/sagemaker/

AWS. (n.d.). Amazon Textract. Retrieved from https://aws.amazon.com/textract/

AWS. (n.d.). Amazon Transcribe. Retrieved from https://aws.amazon.com/transcribe/

AWS. (n.d.). Amazon Translate. Retrieved from https://aws.amazon.com/translate/

AWS. (n.d.). Apache MXNet on AWS. Retrieved from https://aws.amazon.com/mxnet/

AWS. (n.d.). AWS Analytics. Retrieved Januay 28, 2025, from https://aws.amazon.com/analytics/

AWS. (n.d.). AWS Clean Rooms (Preview. Retrieved from https://aws.amazon.com/clean-rooms/

AWS. (n.d.). AWS Data Exchange. AWS Data Exchange. Retrieved from https://aws.amazon.com/data-exchange/

AWS. (n.d.). AWS Data Pipeline. Retrieved from https://aws.amazon.com/datapipeline/

AWS. (n.d.). AWS Deep Learning Containers. Retrieved from https://aws.amazon.com/machine-learning/containers/

AWS. (n.d.). AWS DeepComposer. Retrieved from https://aws.amazon.com/deepcomposer/

AWS. (n.d.). AWS DeepLens. Retrieved from https://aws.amazon.com/deeplens/

AWS. (n.d.). AWS DeepRacer. Retrieved from https://aws.amazon.com/deepracer/

AWS. (n.d.). AWS Glue. Retrieved from https://aws.amazon.com/glue/

AWS. (n.d.). AWS Inferentia. Retrieved from https://aws.amazon.com/inferentia/

AWS. (n.d.). AWS Lake Formation. Retrieved from https://aws.amazon.com/lake-formation/

AWS. (n.d.). AWS Panorama. Retrieved January 28, 2025, from https://aws.amazon.com/panorama/

AWS. (n.d.). Deep Learning Inference Acceleration. Retrieved from https://aws.amazon.com/deep-learning/

AWS. (n.d.). PyTorch on AWS. Retrieved from https://aws.amazon.com/pytorch/

AWS. (n.d.). TensorFlow on AWS. Retrieved from https://aws.amazon.com/tensorflow/

AWS. (n.d.a). *Amazon Athena.* Retrieved Jamuary 26, 2025, from AWS Documentation: https://aws.amazon.com/athena/