**AWS Analytics**

* [**Amazon Athena**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-athena)
* [**Amazon EMR**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-emr)
* [**Amazon CloudSearch**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-cloudsearch)
* [**Amazon Elasticsearch Service**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-elasticsearch-service)
* [**Amazon Kinesis**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-kinesis)
* [**Amazon Kinesis Data Firehose**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-kinesis-firehose)
* [**Amazon Kinesis Data Analytics**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-kinesis-analytics)
* [**Amazon Kinesis Data Streams**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-kinesis-streams)
* [**Amazon Kinesis Video Streams**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-kinesis-video)
* [**Amazon Redshift**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-redshift)
* [**Amazon QuickSight**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-quicksight)
* [**AWS Data Pipeline**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#aws-data-pipeline)
* [**AWS Glue**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#aws-glue)
* [**AWS Lake Formation**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#aws-lake-formation)
* [**Amazon Managed Streaming for Kafka (MSK)**](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-managed-streaming)

1. **Amazon Athena**

* [Amazon Athena](https://aws.amazon.com/athena) is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL.
* Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run.
* Athena is easy to use. Simply point to your data in Amazon S3, deﬁne the schema, and start querying using standard SQL.
* Most results are delivered within seconds. With Athena, there’s no need for complex **extract, transform, and load (ETL)** jobs to prepare your data for analysis. This makes it easy for anyone with SQL skills to quickly analyze large-scale datasets.
* Athena is out-of-the-box integrated with **AWS Glue Data Catalog**, allowing you to create a unified metadata repository across various services, crawl data sources to discover schemas and populate your Catalog with new and modified table and partition definitions, and maintain schema versioning.
* You can also use Glue’s fully-managed ETL capabilities to transform data or convert it into columnar formats to optimize cost and improve performance.

1. **Amazon EMR**

* [Amazon EMR](https://aws.amazon.com/emr) provides a managed Hadoop framework that makes it easy, fast, and cost-eﬀective to process vast amounts of data across dynamically scalable Amazon EC2 instances.
* You can also run other popular distributed frameworks such as Apache Spark, HBase, Presto, and Flink in Amazon EMR, and interact with data in other AWS data stores such as Amazon S3 and Amazon DynamoDB.
* EMR Notebooks, based on the popular Jupyter Notebook, provide a development and collaboration environment for ad hoc querying and exploratory analysis.
* Amazon EMR securely and reliably handles a broad set of big data use cases, including log analysis, web indexing, data transformations (ETL), machine learning, ﬁnancial analysis, scientiﬁc simulation, and bioinformatics.

1. **Amazon CloudSearch**

* [Amazon CloudSearch](https://aws.amazon.com/cloudsearch/) is a managed service in the AWS Cloud that makes it simple and cost-effective to set up, manage, and scale a search solution for your website or application.
* Amazon CloudSearch supports 34 languages and popular search features such as **highlighting, autocomplete, and geospatial search**.

**Amazon Elasticsearch Service**

* [Amazon Elasticsearch Service](https://aws.amazon.com/elasticsearch-service/) makes it easy to deploy, secure, operate, and scale Elasticsearch to search, analyze, and visualize data in real-time.
* With Amazon Elasticsearch Service, you get easy-to-use APIs and real-time analytics capabilities to power use-cases such as ***log analytics, full-text search, application monitoring, and clickstream analytics, with enterprise-grade availability, scalability, and security***.
* The service oﬀers integrations with open-source tools like Kibana and Logstash for data ingestion and visualization.
* It also integrates seamlessly with other AWS services such as [**Amazon Virtual Private Cloud (Amazon VPC)**](https://aws.amazon.com/vpc/)**,**[**AWS Key Management Service (AWS KMS)**](https://aws.amazon.com/kms/)**,**[**Amazon Kinesis Data Firehose**](https://aws.amazon.com/kinesis/data-firehose)**,**[**AWS Lambda**](https://aws.amazon.com/lambda)**, AWS Identity and Access Management (IAM),**[**Amazon Cognito**](https://aws.amazon.com/cognito/)**, and**[**Amazon CloudWatch**](https://aws.amazon.com/cloudwatch), so that you can go from raw data to actionable insights quickly.

1. **Amazon Kinesis**

* [Amazon Kinesis](https://aws.amazon.com/kinesis/) makes it easy to **collect, process, and analyze** **real-time, streaming data** so you can get timely insights and react quickly to new information.
* Amazon Kinesis offers key capabilities to cost-effectively process streaming data at any scale, along with the flexibility to choose the tools that best suit the requirements of your application.
* With Amazon Kinesis, you can ingest real-time data such as video, audio, application logs, website clickstreams, and IoT telemetry data for machine learning, analytics, and other applications.
* Amazon Kinesis enables you to process and analyze data as it arrives and respond instantly instead of having to wait until all your data is collected before the processing can begin.
* Amazon Kinesis currently oﬀers four services:
* **Kinesis Data Firehose,**
* **Kinesis Data Analytics,**
* **Kinesis Data Streams, and**
* **Kinesis Video Streams.**
  1. **Amazon Kinesis Data Firehose**
* [Amazon Kinesis Firehose](https://aws.amazon.com/kinesis/firehose/) is the easiest way to reliably load streaming data into data stores and analytics tools.
* It can **capture, transform, and load streaming data** into Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk, enabling **near real-time analytics** with existing business intelligence tools and dashboards you’re already using today.
* It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration.
* It can also batch, compress, transform, and encrypt the data before loading it, **minimizing the amount of storage used at the destination and increasing security.**
* You can easily create a **Firehose delivery stream** from the AWS Management Console, conﬁgure it with a few clicks, and start sending data to the stream from hundreds of thousands of data sources to be loaded continuously to AWS—all in just a few minutes.
* You can also configure your delivery stream to automatically convert the incoming data to columnar formats like Apache Parquet and Apache ORC, before the data is delivered to Amazon S3, for cost-effective storage and analytics.
  1. **Amazon Kinesis Data Analytics**
* Amazon Kinesis Data Analytics is the easiest way to **analyze streaming data, gain actionable insights, and respond to your business and customer needs in real time**.
* Amazon Kinesis Data Analytics reduces the complexity of building, managing, and integrating streaming applications with other AWS services.
* SQL users can easily query streaming data or build entire streaming applications using templates and an interactive SQL editor.
* Java developers can quickly build sophisticated streaming applications using open source Java libraries and AWS integrations to transform and analyze data in real-time.
* Amazon Kinesis Data Analytics takes care of everything required to run your queries continuously and scales automatically to match the volume and throughput rate of your incoming data.
  1. **Amazon Kinesis Data Streams**
* [Amazon Kinesis Data Streams (KDS)](https://aws.amazon.com/kinesis/streams/) is a massively scalable and durable real-time data streaming service.
* KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources such as **website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events**.
* The data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.
  1. **Amazon Kinesis Video Streams**
* [Amazon Kinesis Video Streams](https://aws.amazon.com/kinesis/streams/) makes it easy to securely stream video from connected devices to AWS for analytics, machine learning (ML), playback, and other processing.
* Kinesis Video Streams automatically provisions and elastically scales all the infrastructure needed to ingest streaming video data from millions of devices.
* It also durably stores, encrypts, and indexes video data in your streams, and allows you to access your data through easy-to-use APIs.
* Kinesis Video Streams enables you to playback video for live and on-demand viewing, and quickly build applications that take advantage of computer vision and video analytics through integration with Amazon Recognition Video, and libraries for ML frameworks such as Apache MxNet, TensorFlow, and OpenCV.

1. **Amazon Redshift**

* [Amazon Redshift](https://aws.amazon.com/redshift/) is a fast, scalable data warehouse that makes it simple and cost-effective to analyze all your data across your data warehouse and data lake.
* Redshift delivers ten times faster performance than other data warehouses by using machine learning, massively parallel query execution, and columnar storage on high-performance disk.
* You can setup and deploy a new data warehouse in minutes, and run queries across petabytes of data in your Redshift data warehouse, and exabytes of data in your data lake built on Amazon S3.
* You can start small for just $0.25 per hour and scale to $250 per terabyte per year, less than one-tenth the cost of other solutions.

1. **Amazon QuickSight**

* [Amazon QuickSight](https://quicksight.aws/) is a fast, cloud-powered business intelligence (BI) service that makes it easy for you to deliver insights to everyone in your organization.
* QuickSight lets you create and publish interactive dashboards that can be accessed from browsers or mobile devices.
* You can embed dashboards into your applications, providing your customers with powerful self-service analytics.
* QuickSight easily scales to tens of thousands of users without any software to install, servers to deploy, or infrastructure to manage.

1. **AWS Data Pipeline**

* [AWS Data Pipeline](https://aws.amazon.com/datapipeline) is a web service that helps you reliably process and move data between diﬀerent AWS compute and storage services, as well as on-premises data sources, at speciﬁed intervals.
* With AWS Data Pipeline, you can regularly access your data where it’s stored, transform and process it at scale, and eﬃciently transfer the results to AWS services such as [Amazon S3](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/storage-services.html#amazon-s3), [Amazon RDS](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/database.html#amazon-rds), [Amazon DynamoDB](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/database.html#amazon-dynamodb), and [Amazon EMR](https://docs.aws.amazon.com/whitepapers/latest/aws-overview/analytics.html#amazon-emr).
* AWS Data Pipeline helps you easily create complex data processing workloads that are fault tolerant, repeatable, and highly available.
* You don’t have to worry about ensuring resource availability, managing inter-task dependencies, retrying transient failures or timeouts in individual tasks, or creating a failure notification system.
* AWS Data Pipeline also allows you to move and process data that was previously locked up in on-premises data silos.

1. **AWS Glue**

* [AWS Glue](https://aws.amazon.com/glue) is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics.
* You can create and run an ETL job with a few clicks in the AWS Management Console.
* You simply point AWS Glue to your data stored on AWS, and AWS Glue discovers your data and stores the associated metadata (e.g. table definition and schema) in the AWS Glue Data Catalog. Once cataloged, your data is immediately searchable, queryable, and available for ETL.

1. **AWS Lake Formation**

* [AWS Lake Formation](https://aws.amazon.com/lake-formation) is a service that makes it easy to set up a secure data lake in days.
* A data lake is a centralized, curated, and secured repository that stores all your data, both in its original form and prepared for analysis.
* A data lake enables you to break down data silos and combine different types of analytics to gain insights and guide better business decisions.
* However, setting up and managing data lakes today involves a lot of manual, complicated, and time-consuming tasks.
* This work includes loading data from diverse sources, monitoring those data flows, setting up partitions, turning on encryption and managing keys, defining transformation jobs and monitoring their operation, re-organizing data into a columnar format, configuring access control settings, deduplicating redundant data, matching linked records, granting access to data sets, and auditing access over time.
* Creating a data lake with Lake Formation is as simple as defining where your data resides and what data access and security policies you want to apply.
* Lake Formation then collects and catalogs data from databases and object storage, moves the data into your new Amazon S3 data lake, cleans and classifies data using machine learning algorithms, and secures access to your sensitive data.
* Your users can then access a centralized catalog of data which describes available data sets and their appropriate usage.
* Your users then leverage these data sets with their choice of analytics and machine learning services, like Amazon EMR for Apache Spark, Amazon Redshift, Amazon Athena, Amazon SageMaker, and Amazon QuickSight.

1. **Amazon Managed Streaming for Kafka (MSK)**

* [Amazon Managed Streaming for Kafka (MSK)](https://aws.amazon.com/msk) is a fully managed service that makes it easy for you to build and run applications that use [Apache Kafka](https://kafka.apache.org/) to process streaming data.
* Apache Kafka is an open-source platform for building real-time streaming data pipelines and applications.
* With Amazon MSK, you can use Apache Kafka APIs to populate data lakes, stream changes to and from databases, and power machine learning and analytics applications.
* Apache Kafka clusters are challenging to setup, scale, and manage in production.
* When you run Apache Kafka on your own, you need to provision servers, configure Apache Kafka manually, replace servers when they fail, orchestrate server patches and upgrades, architect the cluster for high availability, ensure data is durably stored and secured, setup monitoring and alarms, and carefully plan scaling events to support load changes.
* Amazon Managed Streaming for Kafka makes it easy for you to build and run production applications on Apache Kafka without needing Apache Kafka infrastructure management expertise.
* That means you spend less time managing infrastructure and more time building applications.
* With a few clicks in the [Amazon MSK console](https://console.aws.amazon.com/msk/home?region=us-east-1) you can create highly available Apache Kafka clusters with settings and configuration based on Apache Kafka’s deployment best practices. Amazon MSK automatically provisions and runs your Apache Kafka clusters.
* Amazon MSK continuously monitors cluster health and automatically replaces unhealthy nodes with no downtime to your application.
* In addition, Amazon MSK secures your Apache Kafka cluster by encrypting data at rest.