

Amazon EMR



Amazon EMR

- **Amazon Elastic MapReduce (Amazon EMR) is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data.**
- **Big data platform for processing vast amounts of data using open source tools such as**
 - Apache Spark,
 - Apache Hive,
 - Apache HBase,
 - Apache Flink,
 - Apache Hudi, and
 - Presto.



Amazon EMR

- Amazon EMR makes it easy to set up, operate, and scale your big data environments **by automating** time-consuming tasks like provisioning capacity and tuning clusters.
- With EMR you can run petabyte-scale analysis at less than half of the cost of traditional on-premises solutions and over 3x faster than standard Apache Spark.
- You can run workloads on Amazon EC2 instances, on Amazon Elastic Kubernetes Service (EKS) clusters, or on-premises using EMR on AWS Outposts.



Benefits

- **Easy to use**
- **Low cost**
- **Elastic**
- **Reliable**
- **Secure**
- **Flexible**



Use Cases

- **Machine learning**
- **Extract, transform, load (ETL)**
- **Clickstream analysis**
- **Real-time streaming**
- **Interactive analytics**



Demo



Thank you!!!



References

- [**https://docs.aws.amazon.com/index.html**](https://docs.aws.amazon.com/index.html)

