



# Practical - 10

2CS702 – Big Data Analytics

Harshit Gajipara

19BCE059

## Aim:

Case Study: Use following platforms for solving any big data analytic Problem of your choice.

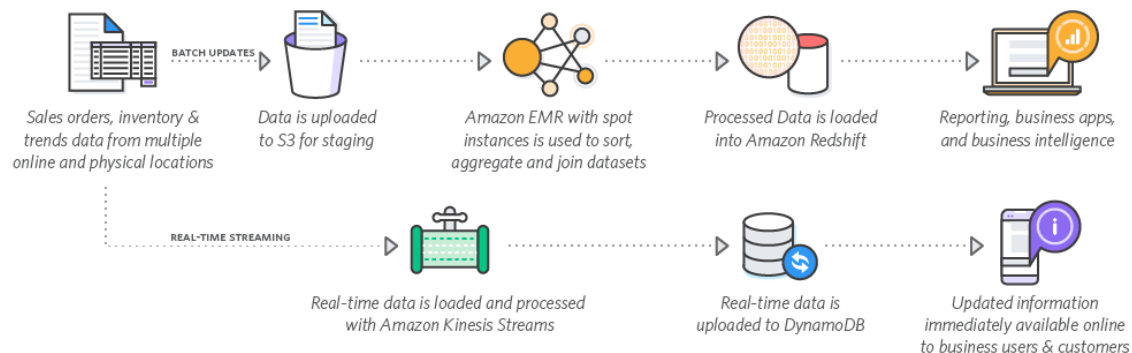
1. Amazon web services 2. Microsoft Azure 3. Google App engine

### 1. Amazon Web Services

With AWS you can build virtually any big data application. These are just a few examples of how organizations are driving value from big data with AWS.

- On Demand Big Data Analytics:

With AWS you can build an entire analytics application to power your business. Scale a Hadoop cluster from zero to thousands of serves within just a few minutes and then turn it off again when you are done. This means you can process big data workloads in less time and at a lower cost.

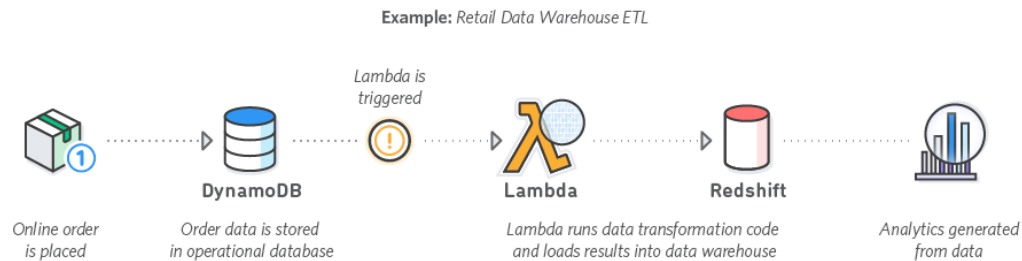


- Clickstream Analysis:

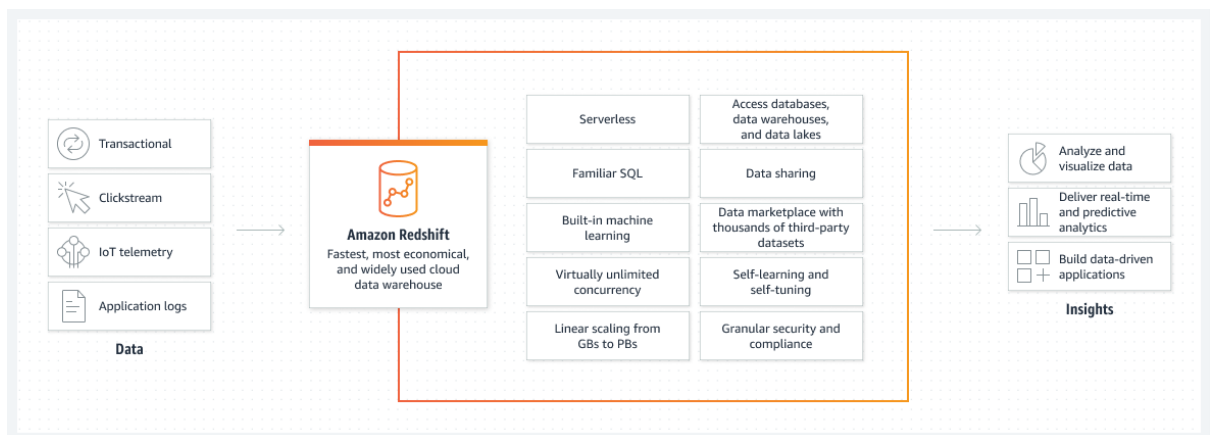
Improve your customer's digital experience and get a better understanding of your website. Collect, process, analyze, and visualize clickstream insights in real-time with AWS.

- Event-Driven Extract, Transform and Load (ETL):

Use AWS Lambda to perform data transformations - filter, sort, join, aggregate, and more - on new data, and load the transformed datasets into Amazon Redshift for interactive query and analysis.



## Amazon Redshift



## Amazon Lake Formation



## 2. Microsoft Azure

Microsoft Azure Big Data Analytics solutions provide a comprehensive solution which turn your data into actionable insights. Comprehensive solution which turn your data into actionable insights.

- Azure Synapse Analytics: Limitless analytics service with unmatched time to insight.
- Azure Databricks: A fully managed, fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure.
- Azure HDInsight: A fully managed cloud Hadoop and Spark service backed by 99.9% SLA for your enterprise.
- Azure Data Factory: A data integration service to orchestrate and automate data movement and transformation.
- Azure Stream Analytics: Real-time data stream processing from millions of IoT devices.
- Azure Analysis Service: A fully managed on-demand pay-per-job analytics service with enterprise-grade security, auditing and support.

### 3. Google App Engine:

The features of Google App Engine include:

- Empower everyone to get insights by removing traditional constraints of scale, performance and cost to solve business challenges.
- Optimize business outcomes and customer experiences with real-time intelligence and predictive insights using industry-leading AI.
- Maximize value from data with a flexible, open and multi-cloud platform. The ways in which Google App Engine in Big Data Analytics applications includes:
- Data Warehouse Modernization: Get more data from your warehouse.
- Data Lake Modernization: Make the most of your data lake,
- Streaming Analytics: Extract business value from real-time data.
- Business Intelligence: Drive better outcomes through data-driven experiences.
- Geospatial Analytics and AI: Find value in location-based workloads.
- Datasets: Enhance analytics and AI with pre-built datasets.

	<b>Amazon Web Services</b>	<b>Microsoft Azure</b>	<b>Google App Engine</b>
Cloud-based enterprise data warehouse (EDW) that uses massively parallel processing (MPP) to quickly run complex queries across petabytes of data.	Redshift	Synapse Analytics	Used for hosting web applications
A simple and safe service for sharing big data.	Lake Formation	Data Share	Google Cloud Datastore
Fully managed, low latency, and distributed big data analytics platform that runs complex queries across petabytes of data. Highly optimized for log and time series data.	Amazon Timestream	Azure Time Series Insights	API for data query (supports frameworks like django, flask etc)
Apache Spark-based analytics platform.	EMR	Databricks	---
Fully managed, low latency, distributed big data analytics platform to run complex queries across petabytes of data.	EMR	Azure Data Explorer	Low level API
Managed Hadoop service. Deploy and manage Hadoop clusters in Azure.	EMR	HDInsight	---
Massively scalable, secure data lake functionality built on Azure Blob Storage.	EMR	Data Lake Storage	GAE Filestore (virtual)
NoSQL key-value store for rapid development using massive semi-structured datasets.	Dynamo DB	Azure Cosmos DB	Big Table, Google Firestore
Serverless computing	AWS Lambda	Azure Functions	Google cloud
Relational Database	RDS	SQL, MySQL, PostgreSQL	GQL

**Conclusion:**

In this practical, we analyse how AWS, Microsoft Azure and Google App Engine provides solutions for big data analytics.