



**STEVENS**  
INSTITUTE *of* TECHNOLOGY  
THE INNOVATION UNIVERSITY®

***Big Data Analytics***

***Vinay Rajesh Gor***

## Table of Contents

1. Introduction to Data Analytics .....	3
2. Introduction to Big Data Analytics.....	3
3. Big Data Analytics tools.....	4
3.1 Amazon Kinesis.....	4
3.1.1 Kinesis Video Streams.....	5
3.1.2 Kinesis Data Streams.....	5
3.1.3 Kinesis Data Firehouse.....	6
3.1.4 Kinesis Data Analytics.....	6
3.2 Apache Kafka.....	7
3.3 Microsoft Azure & Power BI.....	8
3.4 Oracle Advanced Analytics.....	9
4. Gartner Magic Quadrant for Analytics and BI Platforms.....	10
5. Conclusion.....	11

## List of Figures

Fig 1. How Kinesis Video Stream Works.....	5
Fig 2. How Kinesis Data Stream Works.....	5
Fig 3. How Kinesis Data Firehose.....	6
Fig 4. How Kinesis Data Analytics works.....	6
Fig 5. How Apache Kafka Works.....	7
Fig 6. How Azure and Power BI work together.....	8
Fig 7. How Oracle Advanced Analytics works.....	9
Fig 8. Magic Quadrant for Analytics and Business Intelligence Platforms.....	10

## **1. Introduction to Data Analytics**

With the increase in the amount of generated data across the world, companies have started using data to improve their business by understanding, analyzing, predicting and implementing business solutions generated by trends and patterns obtained from this data. This study of data is called “Data Analytics”. According to the statistics, increasing data accessibility by 10% can increase the revenue of fortune 1000 companies by \$65 million which makes it important for companies to harness data and implement data in their business.

Data analytics is a broad field and can be used for variety of businesses purposes but there are 4 primary types of data analytics,

- a. Descriptive Analytics - Deals with Data Exploration
- b. Diagnostics Analytics - Deals with findings of initial data exploration
- c. Predictive Analytics - Deals with prediction and likely occurrence using past and current data
- d. Prescriptive Analytics - Deals with making improved decisions based on patterns obtained from predictive analytics

While traditionally many tools including Microsoft Excel and other tools & Softwares are used for Data analytics huge amounts of data now called as “Big Data” makes it difficult and have several disadvantages,

- Cannot handle and store huge amount of data (>1 GB)
- Slow in processing
- Low efficiency
- Cannot be used for Real time analytics and monitoring
- Cannot be configured with other sources, tools and softwares

## **2. Introduction to Big Data Analytics**

Big data is a field of data that deals with too complex or large data sets both structured and unstructured. Several big data tools are available in the market with their own capabilities, performances, and uses. These tools include AWS Kinesis, Apache Kafka, IBM Cognos Analytics, Tableau, Power Bi, AWS Quicksight, Einstein Analytics, Looker, etc. Companies use different tools for different uses and purposes. This depends on the architecture they started with, cost effectiveness and usefulness to drive business. This paper will discuss some of the tools used for Big data analytics.

Big Data tools can play several significant roles for data analytics and have several advantages over traditional data analytics tools.

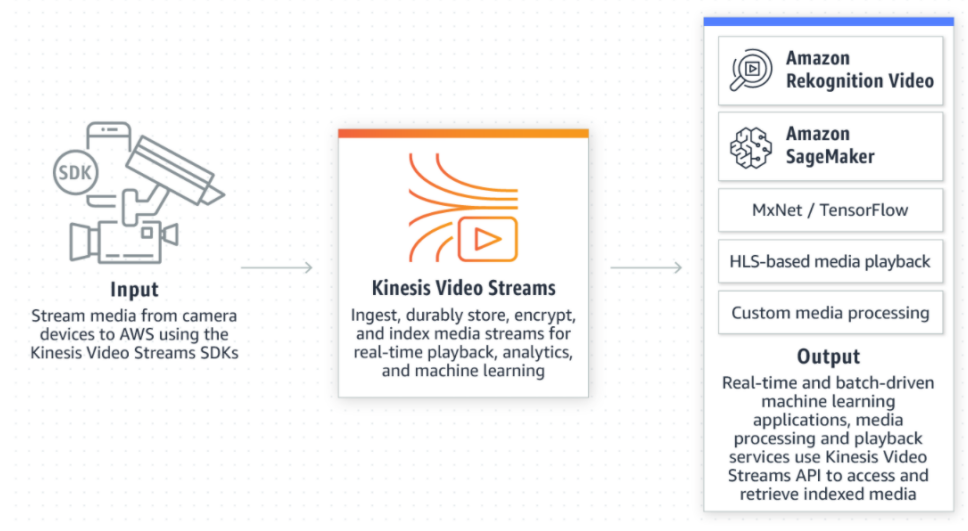
### 3. Big data analytics tools

#### 3.1 Amazon Kinesis

Amazon Web Services launched Kinesis in 2013 which processes the large data in real time. Kinesis used Amazon Kinesis Data Streams to collect large streams of data records, these large streams of data records are in real-time. Data processing is done through Kinesis Data Streams applications which reads the data streams as data records and then real time processed data records are sent to dashboards. This real time processed data can be sent to dashboards for data analytics applications.

As per requirement there are 4 Kinesis platforms available,

- **Amazon Kinesis Video Streams** - Amazon Kinesis Video Streams captures, stores and processes media streams including video streams for analytics.



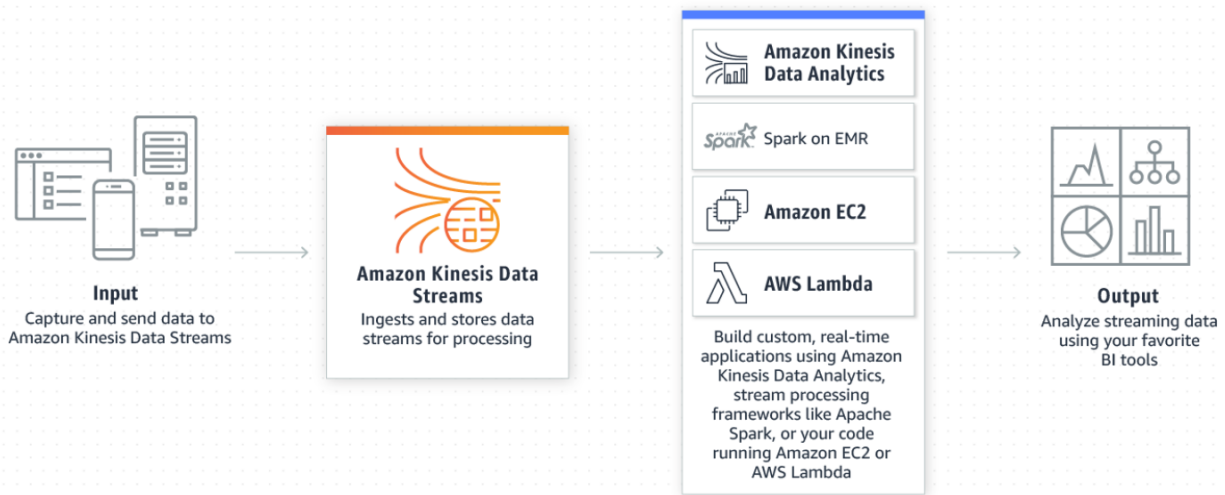
*Fig 1. How Kinesis Video Stream Works*

*Note: From AWS Amazon Kinesis Video Streams*

Uses in Big Data Analytics:-

1. Security Monitoring
2. Emergency Monitoring
3. Prevent Crime
4. Industrial Automation
5. Smart Home & Devices
6. Sentiment Detection

- **Amazon Kinesis Data Streams** - Amazon Kinesis Data Streams can capture massive (gigabytes per second) real time data streams. These data streams can be social media feeds, financial transactions, IT logs, location tracking, databases which can be fed in data streams to obtain a real time dashboard.



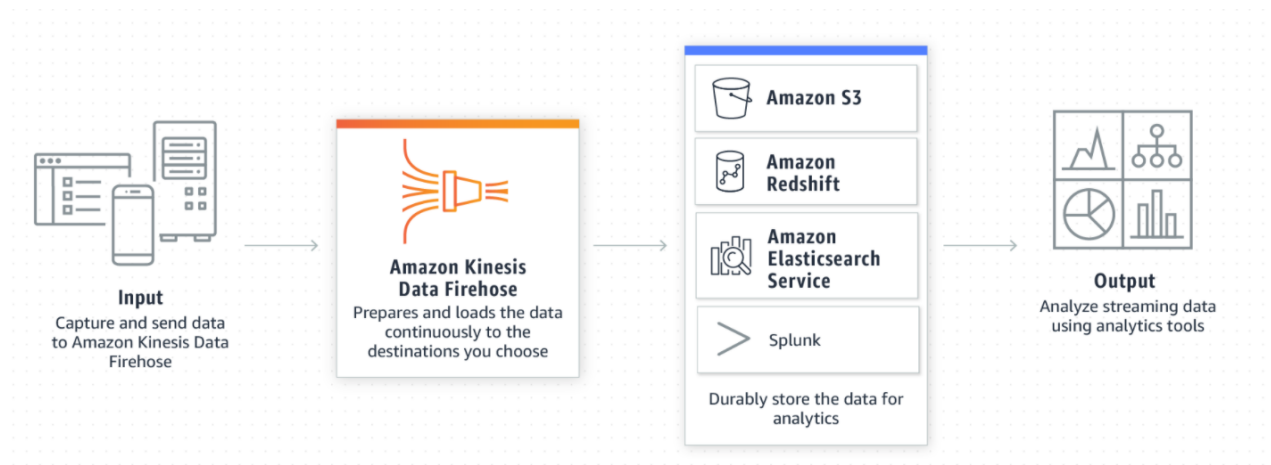
*Fig 2. How Kinesis Data Stream Works*

*Note: From AWS Amazon Kinesis Data Streams*

Uses in Big Data Analytics:-

1. Real time price estimation
2. Real time detecting and fixing issues
3. Transmit and process website logs
4. Provide insights for offers, pricing and transactions

- **Amazon Kinesis Data Firehose** - Kinesis Data Firehose can load the data into data lakes, data stores and analytical services which can transform data to other Amazon tools to deliver real time analytics. To durably store the data other AWS services can be used according to their integration including Amazon S3, Amazon Redshift, Amazon Elastic Service and etc. Kinesis stores the data continuously in real-time which can help users to obtain and analyze data in real-time.



*Fig 3. How Kinesis Data Firehose*

*Note: From AWS Amazon Kinesis Data Firehose*

Uses in Big Data Analytics:-

1. Real time analytics of metrics, insights and dashboards
2. Analyzing log
3. Clickstream analytics
4. Security & Monitoring
5. Sport analytics

- **Amazon Kinesis Data Analytics** - Amazon Kinesis Data Analytics collects, stores, transforms and analyzes data in real time with Apache Flink. Kinesis reduces complexity to build and integrate with other aws tools.



*Fig 4. How Kinesis Data Analytics works*

*Note: From AWS Amazon Kinesis Data Analytics*

Uses in Big Data Analytics:-

1. ETL (Extract, Transform and Load) applications
2. Real time analytics for query streaming data
3. Analyze patterns and anomalies

### 3.2 Apache Kafka

Apache Kafka is a large ecosystem open-source platform interface which can integrate with almost all tools (AWS, Elasticsearch, Postgres etc.) and all languages (Python, R, Java, Scala). Apache Kafka has streams of events which can be processed such as transformation, joins, filters, groups. Zookeeper provides support to manage clusters. With high scalability, low latency, permanent storage and high availability Kafka has made its way to become used by most of the fortune 100 companies.

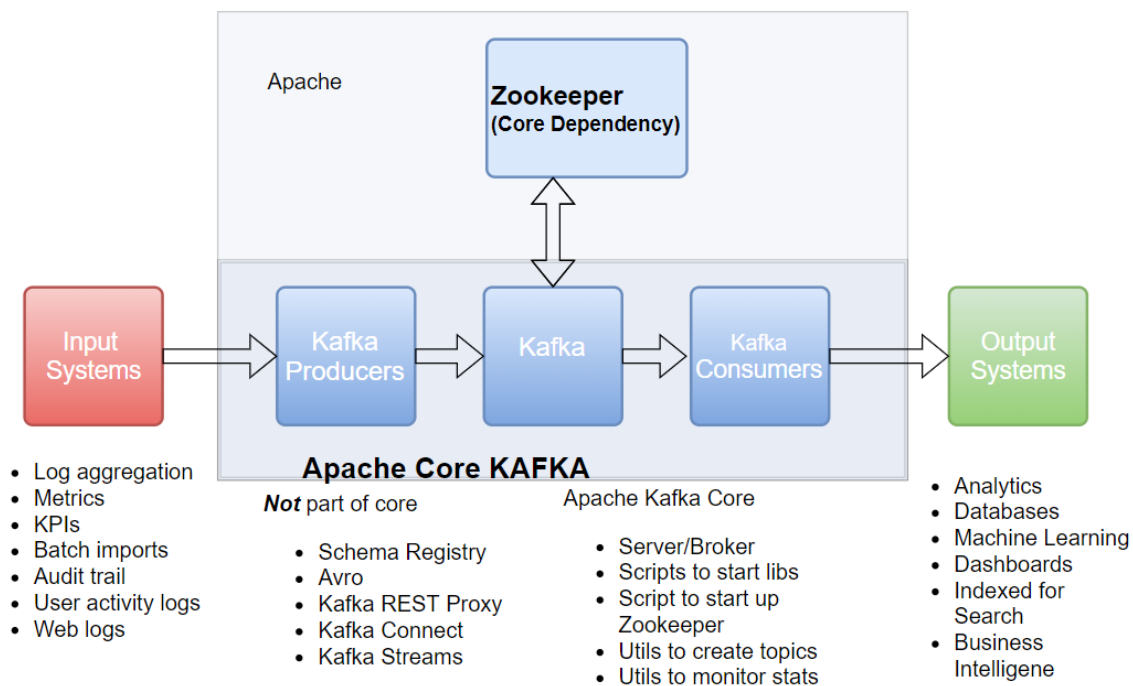


Fig 5. How Apache Kafka Works

[Note: From cloudurable.com](http://cloudurable.com)

Uses in Big Data Analytics:-

1. Real time processing of website activities such as views on page, type of searches, time of search
2. Real time analyzing metrics, statistics and monitoring
3. Kafka is used for aggregation of log as a solution
4. Sourcing of events which are time-ordered records

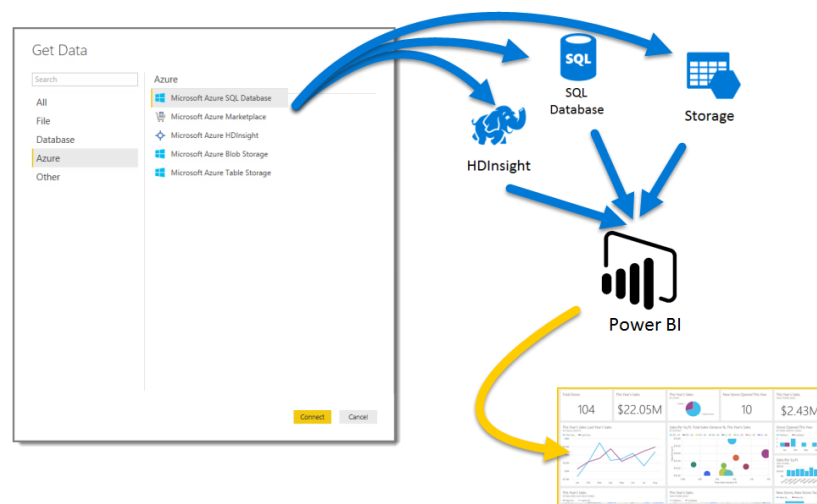
### 3.3 Microsoft

Microsoft has been the top player in the field of Data starting from its Inception. Microsoft has many products including Microsoft Azure which handles cloud computing and Power BI an analytical tool which in combination are used by many companies.

**Microsoft Power BI** - Microsoft Power BI is an analytical platform used by individuals and companies for visualizations, exploration and business intelligence. Power BI provides data security, data sensitivity, end-to-end encryption and monitoring analytics in real-time.

**Microsoft Azure** - Microsoft Azure is a cloud computing platform that can be used for computing, storage, analytics etc. Azure Synapse Analytics is an analytics service which brings data services including analytics and business intelligence from preparing, managing and exploring data from a single unified platform. The choice of programming language in Azure can be limitless and can connect up to 95 native connectors.

Azure having handling capacity of petabytes of data and Power BI analyzing data can deal with big data analytics in an efficient way thus using microsoft apps Azure and Power BI can be integrated for Big data analytics providing best of both worlds. This combination can provide visualizations, analytics of petabytes of data in seconds.



*Fig 6. How Azure and Power BI work together*

[Note: From Microsoft.com](#)

Uses in Big Data Analytics:-

1. Automation and analysis in real-time
2. Data Intelligence & processing for large scale data



### 3.4 Oracle

Oracle is one of the visionaries in Big data analytics. Oracle Advanced analytics (OAA) with In-database data mining algorithms and open source R algorithms along with compatibility of other languages such as SQL & NOSQL and workflow of GUI and IDEs Oracle enables enterprise analytical applications. OAA can be easily worked with SQL queries and with open source R algorithms and using action verbs (Predict, Classify, Associate, Regress) which makes coding minimal and hence can be used by analysts having less or no experience in programming but still want to used Analytics as a part of their business.

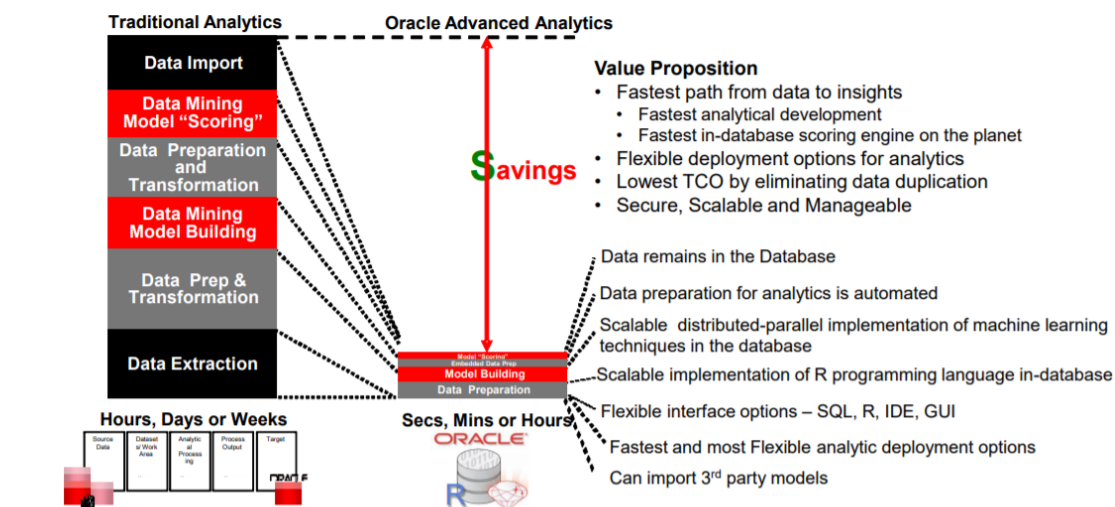


Fig 7. How Oracle Advanced Analytics works

[Note: Source Oracle.com](http://Source Oracle.com)

With OAA traditional analytics preparation such as Data Import, Mining, Extraction, Transformation, Loading is reduced which makes it easy for analysts to start working on projects with flexible automation, interface options, and fast analytical deployments.

Uses in Big Data Analytics:-

1. Fraud detection in real-time
2. Anomaly detection for industrial applications
3. Predictive analytics for financial/accounting
4. Security and suspicious detection
5. Sentimental analysis
6. Social Networking analysis

#### 4. Gartner Magic Quadrant for Analytics and BI Platforms

Gartner, a major research and advisory firm providing consultation, information and advice in various domains, launches various research articles every year which also includes articles for Analytics and BI Platforms. This magic quadrant reflects major tools and technologies used for Analytics and BI across companies. These companies are placed as Challengers, Leaders, Niche Players and Visionaries. Microsoft has been named top leader for the 14th consecutive year, along with Microsoft major leaders in markets include Tableau and Qlik which are becoming favourite analytical tools. Major challengers include Google(Looker), Domo and Microstrategy which are more or less similar to each other and other tools found in leaders with some micro differences. AWS has been a top Niche player having a wide variety of data and cloud services which makes utilization of AWS in every specialized field. Visionaries tools include Oracle, SAS, SAP which are relatively new to the market but are making a quick way in the analytics and BI field.



Fig 8. Magic Quadrant for Analytics and Business Intelligence Platforms

Note: Source Gartner (February 2021)

## **5. Conclusion**

With different types of tools and technologies available in the market, there are a variety of options available for companies including type of services required, cloud computing capabilities, server computing, type of file/files companies are using, companies can choose as per their configuration and integration. While there are no right or wrong tools for Analytics and BI but integration of the company to obtain that analytics is what matters.

## **References:-**

<https://www.mastersindatascience.org/learning/what-is-data-analytics/>

<https://aws.amazon.com/kinesis/video-streams/?amazon-kinesis-video-streams-resources-blog.sort-by=item.additionalFields.createdDate&amazon-kinesis-video-streams-resources-blog.sort-order=desc>

<https://aws.amazon.com/kinesis/data-streams/>

<https://aws.amazon.com/kinesis/data-firehose/>

<https://aws.amazon.com/kinesis/data-analytics/>

<https://kafka.apache.org/>

<https://kafka.apache.org/uses>

<https://azure.microsoft.com/en-us/services/synapse-analytics/#customer-stories>

<https://powerbi.microsoft.com/en-us/power-bi-and-azure/>

<https://powerbi.microsoft.com/en-us/>

<https://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataoracleadvanalytics11gr2-1930894.pdf>

<https://info.microsoft.com/ww-Landing-2021-Gartner-MQ-for-Analytics-and-Business-Intelligence-Power-BI.html?LCID=EN-US>