Module 1:
Big Data Analytics & IoT

Introduction

Faculty Introduction

Ekpe Okorafor, PhD

AFFILIATIONS:

- Accenture - Applied Artificial Intelligence
- African University of Science & Technology
- Nigerian British University

RESEARCH INTERESTS:

- Generative AI
- Machine Learning
- Big Data & Advanced Analytics
- Knowledge Graph
- Information Assurance and Cybersecurity

FUN & INTERESTS

- Drawing & Painting
- Hiking
- Soccer
- DJ & Dancing

IoT and Big Data Analytics

Introductions
(30 Seconds)

1. Name

2. Nationality

3. Affiliated Institution

4. Field of Study

5. Fun Fact

IoT and Big Data Analytics

Today's Program

09:00 - 18:30   IoT and Big Data Analytics Workshop

- Location: Adriatico Guest House - Denardo Lecture Hall
- 09:00   Introduction to Big Data & IoT Analytics Problem Scope. Analysis of Large Scale Real-Time and Streaming Data 1h30'
- 10:30   Coffee break 30'
- 11:00   Introduction to Kafka 2h0'
- 13:00   Lunch Break 1h0'
- 14:00   Lab – Install & Verify Docker Environment for Kafka 2h0'
- 16:00   Coffee break 30'
- 16:30   Lab – Creating Topics & Passing Messages. Group Discussion on Problem 2h0'

Module 1

Big Data & Analytics

Learning Objectives

Upon successful completion of this module, you will be able to describe Big Data, examples and types. You will also be ale to recognize the characteristics of Big Data and, describe an approach for handling and processing Big Data.

IoT and Big Data Analytics

What is Data?

What is Data?

- The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media

What is Big Data?

What is Big Data?

- Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

IoT and Big Data Analytics

What is Big Data?

IoT and Big Data Analytics

Examples of Big Data

The **New York Stock Exchange** is an example of Big Data that generates about one terabyte of new trade data per day.

Social Media: The statistic shows that 500+terabytes of new data get ingested into the databases of social media site Facebook, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

A single Jet engine can generate 10+terabytes of data in 30 minutes of flight time. With many thousand flights per day, generation of data reaches up to many Petabytes.

IoT and Big Data Analytics

Types of Big Data

There are three types of Big Data:

- Structured

- Unstructured

- Semi-structured

Structured

- Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured'

- Data stored in a relational database management system is one example of a 'structured' data.

Example of Structured Data

Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

IoT and Big Data Analytics

Unstructured

- Any data with unknown form or structure is classified as unstructured data.

- A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.

Examples of Un-structured Data

The output returned by 'Google Search'

IoT and Big Data Analytics
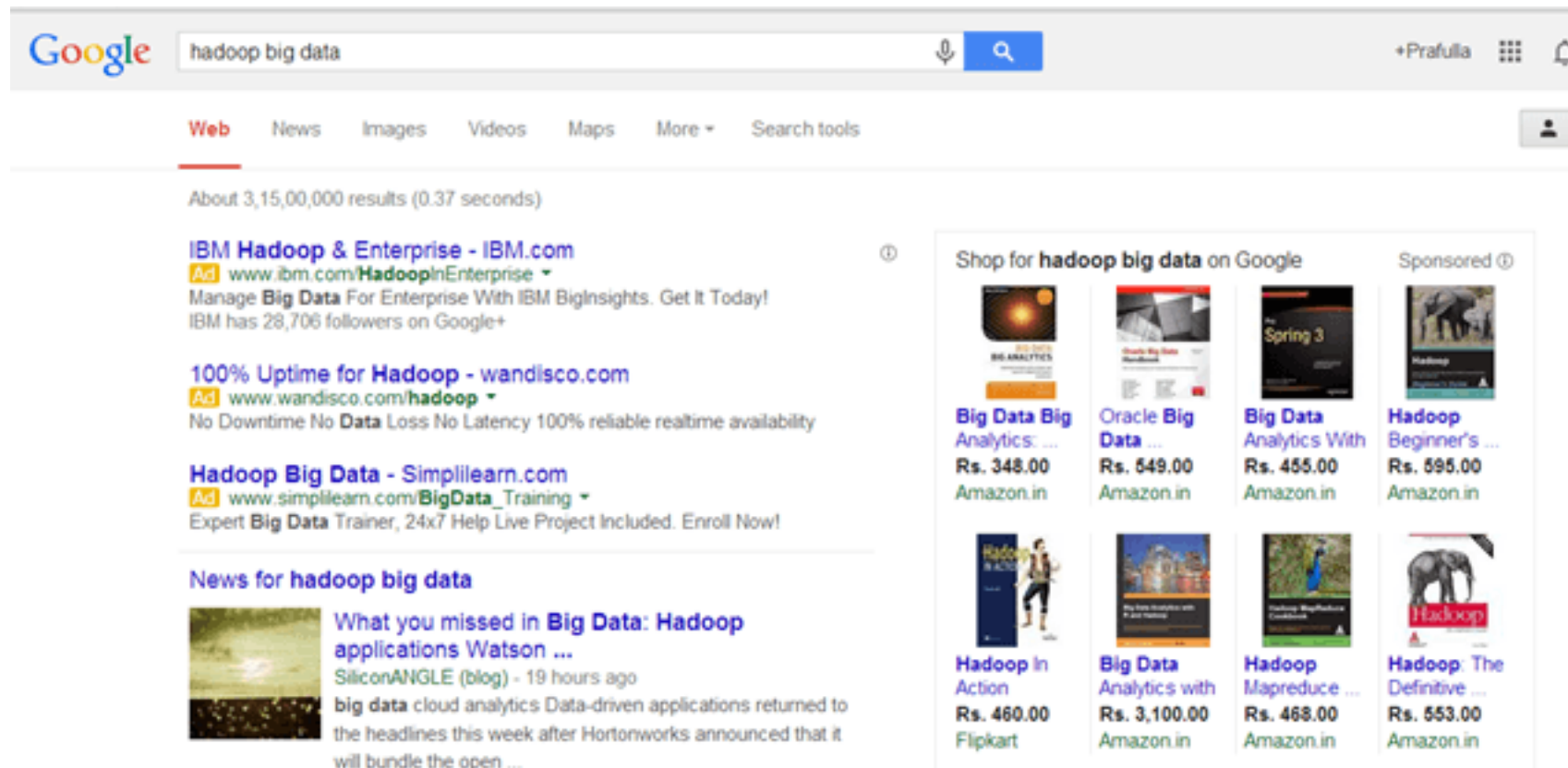
Semi-structured

- Semi-structured data is data that can't be organized in relational databases or doesn't have a strict structural framework yet does have some structural properties or loose organizational framework.

- Semi-structured data includes text that is organized by subject or topic or fit into a hierarchical programming language, yet the text within is open-ended, having no structure itself.

- Example of semi-structured data is a data represented in an XML file.

Examples of Semi-structured Data

Personal data stored in an XML file

```
<rec><name>Prashant
Rao</name><sex>Male</sex><age>35</age></rec>
<rec><name>Seema
R.</name><sex>Female</sex><age>41</age></rec>
<rec><name>Satish
Mane</name><sex>Male</sex><age>29</age></rec>
<rec><name>Subrato
Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah
J.</name><sex>Male</sex><age>35</age></rec>
```

IoT and Big Data Analytics

Characteristics of Big Data

Big data can be described by the following characteristics:

- Volume

- Variety

- Velocity

- Variability

Characteristics of Big Data

- **Volume** - The name Big Data itself is related to a size which is enormous.

- **Variety** - Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.

- **Velocity** - The term 'velocity' refers to the speed of generation of data.

- **Variability** - This refers to the inconsistency which can be shown by the data at times

IoT and Big Data Analytics

Advantages of Big Data Processing

- Ability to process Big Data in DBMS brings in multiple benefits, such as-
  - Businesses can utilize outside intelligence while taking decisions
- Access to social data from search engines and sites like Facebook, twitter are enabling organizations to fine tune their business strategies.
  - Improved customer service
- Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.
  - Early identification of risk to the product/services, if any
  - Better operational efficiency
- Big Data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.

Summary

- Big Data definition : Big Data meaning a data that is huge in size. Bigdata is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.

- Big Data analytics examples includes stock exchanges, social media sites, jet engines, etc.

- Big Data could be 1) Structured, 2) Unstructured, 3) Semi-structured

- Volume, Variety, Velocity, and Variability are few Big Data characteristics

- Improved customer service, better operational efficiency, Better Decision Making are few advantages of Bigdata

IoT and Big Data Analytics

IoT and Big Data Analytics