

Big Data Analysis

(MA60306)

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 1
January 4, 2023

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data
- ▷ Velocity - Speed of the generation of new data

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data
- ▷ Velocity - Speed of the generation of new data
- ▷ Veracity - Accuracy of the data

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data
- ▷ Velocity - Speed of the generation of new data
- ▷ Veracity - Accuracy of the data

Challenges

- ▷ data storage
- ▷ data visualization
- ▷ data processing

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data
- ▷ Velocity - Speed of the generation of new data
- ▷ Veracity - Accuracy of the data

Challenges

- ▷ data storage
- ▷ data visualization
- ▷ data processing

How big is big?

Big Data Analysis

Dealing with Large and complex data sets that are difficult to process using traditional **DBMS** techniques such as RDBMS

The characteristics of big data

- ▷ Volume - amount of data
- ▷ Variety - diversity of the data
- ▷ Velocity - Speed of the generation of new data
- ▷ Veracity - Accuracy of the data

Challenges

- ▷ data storage
- ▷ data visualization
- ▷ data processing

How big is big? What are the machinery?

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics
- ▷ machine-generated data - without human intervention, IoT

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics
- ▷ machine-generated data - without human intervention, IoT
- ▷ *graph*-based or network data - social networks, SPARQL

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics
- ▷ machine-generated data - without human intervention, IoT
- ▷ *graph*-based or network data - social networks, SPARQL
- ▷ audio, image, and video - video games

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics
- ▷ machine-generated data - without human intervention, IoT
- ▷ *graph*-based or network data - social networks, SPARQL
- ▷ audio, image, and video - video games
- ▷ streaming data

Big Data Analysis

Types of data

- ▷ structured - depends on a data model, exp. excel table, SQL
- ▷ unstructured - for example, emails
- ▷ natural language - unstructured, requires knowledge of specific data science techniques and linguistics
- ▷ machine-generated data - without human intervention, IoT
- ▷ *graph*-based or network data - social networks, SPARQL
- ▷ audio, image, and video - video games
- ▷ streaming data