

BIG DATA

ANALYTICS FOR BEGINNERS



Dr.A.L.Sreenivasulu | Dr.Vempati Krishna
Dr.V.L.Pavani | C.Nagesh

INDEX

- CHAPTER 1 BIG DATA INTRODUCTION
 - 1.1 INTRODUCTION
 - 1.2 BIG DATA CHARACTERISTICS
 - 1.3 A BIG DATA ARCHITECTURE
 - 1.3.1 CHALLENGES IN DESIGNING BIG DATA ARCHITECTURE
 - 1.3.2 BENEFITS OF BIG DATA ARCHITECTURE
 - 1.4 TYPES OF BIG DATA
 - 1.5 FEATURES OF BIG DATA ANALYTICS AND REQUIREMENTS
 - 1.6 THE BUILDING BLOCKS OF HADOOP
 - 1.7 BIG DATA TECHNOLOGY
 - 1.8 BIG DATA VS. BUSINESS INTELLIGENCE
 - 1.9 BENEFITS OF USING BIG DATA
 - 1.10 DISADVANTAGES OF BIG DATA**

- CHAPTER 2 **WORKING WITH BIG DATA**
 - 2.1 HADOOP - INTRODUCTION
 - 2.2 HADOOP FRAMEWORK
 - 2.3 HADOOP ARCHITECTURE
 - 2.3.1 WORK WITH HADOOP
 - 2.3.2 ADVANTAGES OF HADOOP
 - 2.4 HDFS
 - 2.4.1 ARCHITECTURE OF HDFS
 - 2.4.2 HDFS OBJECTIVES
 - 2.5 HADOOP FILE SYSTEMS
 - 2.5.1 GOOGLE FILE SYSTEM
 - 2.5.2 HADOOP DISTRIBUTED FILE SYSTEM
 - 2.5.3 HDFS-MASTERNODE NODES
 - 2.5.4 SCALABILITY COMPARISON BETWEEN GFS AND HDFS
 - 2.5.5 DATA FLOW INPUT AND OUTPUT
 - 2.6 HDFS OPERATIONS IN HADOOP
 - 2.6.1 ACTIVATING HDFS
 - 2.6.2 LISTING HDFS FILES
 - 2.6.3 DATA ENTRY INTO HDFS
 - 2.6.4 DATA RETRIEVAL FROM HDFS
 - 2.6.5 THE HDFS IS BEING SHUT DOWN.
 - 2.6.6 HDFS ADVANTAGES
 - 2.7 HADOOP CLUSTER

- 2.7.1 HADOOP CLUSTER BASIC ARCHITECTURE
- 2.7.2 BENEFITS OF HADOOP CLUSTERS

CHAPTER 3 MAP REDUCE :

- 3.1 INTRODUCTION
 - 3.1.1 FEATURES OF MAPREDUCE
 - 3.1.2 HADOOP MAPREDUCE ALGORITHM WORK
- 3.2 A WEATHER DATASET
- 3.3 MAPREDUCE IN JAVA
 - 3.3.1 THE NEW JAVA MAPREDUCE API IS AS FOLLOWS:
- 3.4 BASIC PROGRAMS OF HADOOP MAPREDUCE: COMBINER
- 3.5 BASIC PROGRAMS OF HADOOP MAPREDUCE PARTITIONER
- 3.6 BASIC PROGRAMS OF HADOOP MAPREDUCE PHASE OF REDUCTION
- 3.7 BASIC PROGRAMS OF HADOOP MAPREDUCE PHASE OF REDUCTION RECORD WRITER

CHAPTER 4 HADOOP I/O:

- 4.1 INTRODUCTION
- 4.2 WRITABLE INTERFACE
 - 4.2.1 WRITABLECOMPARABLE INTERFACE
 - 4.2.2 THE IMPORTANCE OF WRITABLE IN HADOOP
 - 4.2.3 WRITABLE INTRODUCED IN HADOOP
- 4.3 WRITABLE BE USED IN HADOOP
- 4.4 WRITABLE CLASSES – HADOOP DATA TYPES
- 4.5 WRITABLE WRAPPERS FOR JAVA PRIMITIVES
- 4.6 RAWCOMPARATOR WILL SPEED UP YOUR HADOOP MAP/REDUCE (MR)
- 4.7 COMPARATORS THAT ARE MADE TO ORDER.

CHAPTER 5 PIG:

- 5.1 INTRODUCTION
 - 5.1.1 APACHE PIG APPLICATIONS

- 5.1.2 History of the Apache Pig
- 5.2 IMPORTANCE OF THE APACHE PIG
- 5.3 PIG CHARACTERISTICS
- 5.4 Apache Pig Vs MapReduce
- 5.5 APACHE PIG – ARCHITECTURE
- 5.6 PIG LATIN DATA MODEL
- 5.7 THE PIG LATIN APPLICATION FLOW IN HADOOP
- 5.8 WORKING THROUGH THE ABCS OF PIG LATIN
- 5.9 PIG LATIN – BASICS
- 5.10 EVALUATING LOCAL AND DISTRIBUTED MODES OF RUNNING PIG SCRIPTS:
- 5.11 PIG SCRIPT INTERFACES IN HADOOP
- 5.12 PIG SCRIPT EXECUTION | APACHE PIG RUNNING SCRIPTS AND COMMENTS

CHAPTER 6 HIVE:

- 6.1 INTRODUCTION TO THE HIVE
 - 6.1.1 MAPREDUCE OPERATIONS CAN BE CARRIED OUT IN A VARIETY OF WAYS:
 - 6.1.2 IMPORTANT HIVE CHARACTERISTICS
 - 6.1.3 HIVE CAN BE INTERACTED WITH USING METHODS
 - 6.1.4 THE FOLLOWING ARE SOME OF THE MOST CRITICAL ELEMENTS OF HIVE:
 - 6.1.5 RELATIONAL DATABASES VS. HIVE:
 - 6.1.6 FEATURES OF HIVE
- 6.2 HIVE ARCHITECTURE
- 6.3 HIVE JOB EXECUTION FLOW:
- 6.4 DIFFERENT MODES OF HIVE
- 6.5 WORKFLOW OF HIVE
- 6.6 HIVE - DATA TYPES
 - 6.6.1 TYPES OF STRINGS
- 6.7 HIVE - CREATE DATABASE

- 6.8 HIVE - DROP DATABASE
 - 6.9 HIVE - CREATE TABLE
 - 6.10 LOAD DATA STATEMENT IN TABLE
 - 6.11 HIVE - ALTER TABLE
 - 6.12 REPLACE STATEMENT
 - 6.13 HIVE - DROP TABLE
 - 6.14 PARTITIONING IN HIVE
 - 6.15 HIVE - BUILT-IN OPERATORS
 - 6.15.1 RELATIONAL OPERATORS
 - 6.15.2 ARITHMETIC OPERATORS
 - 6.15.3 LOGICAL OPERATORS
 - 6.15.4 COMPLEX OPERATORS
 - 6.16 HIVE - BUILT-IN FUNCTIONS
 - 6.16.1 AGGREGATE FUNCTIONS
 - 6.17 HIVE - VIEW AND INDEXES
 - 6.18 HIVEQL - SELECT-WHERE
 - 6.18.1 HIVEQL - SELECT-GROUP BY
 - 6.19 JOINS
 - 6.19.1 LEFT OUTER JOIN
 - 6.19.2 RIGHT OUTER JOIN
 - 6.19.3 FULL OUTER JOIN
 - 6.20 LIMITATIONS OF HIVE
 - 6.21 APACHE HIVE'S ADVANTAGES
 - 6.22 QUERY LANGUAGE FOR HIVE (HQL)
- CHAPTER 7 BDA APPLICATIONS**
- 7.1 INTRODUCTION
 - 7.2 BIG DATA APPLICATIONS IN HEALTHCARE

- 7.3 BIG DATA ANALYTICS IN EDUCATION
- 7.4 BIG DATA ANALYTICS IN INDUSTRIES
- 7.5 BIG DATA ANALYTICS IN SOCIAL MEDIA
- 7.6 BIG DATA ANALYTICS IN CYBER SECURITY
- 7.7 BIG DATA ANALYTICS IN E-COMMERCE

ABOUT THE AUTHORS



Dr.A.L.Sreenivasulu did his B-Tech(CSE), M-Tech(CSE) & Ph.D(CSE) from JNTUA University Ananthapuramu , He got 20 years of teaching experience at various engineering colleges and published various National & International journals. He is currently working as professor, CSE Dept at GATES Institute of Technology, Gooty. His areas of interest include Wireless sensor Networks, Big Data Analytics, Computer Networks and Network Security. He is a member of Institution of Engineers India. He guided more than 60 UG projects and 20 PG Projects.



Dr.Vempati Krishna completed Ph.D. from Rayalaseema University, Kurnool, Andhra Pradesh, India, having 21 years of Teaching and Research experience. He is currently working as Professor, CSE department at TKR College of Engineering and Technology, Hyderabad. The author area of interests are Image Processing, Data Analytics , Data Science and IoT. He published various papers in reputed national and international journals and Conferences. He guided more than 80 UG projects and PG Projects.



Dr.V.L.Pavani Currently working as an Associate Professor and HOD in the Department of Computer Applications , Madanapalle Institute of Technology and Science, Madanapalle. Secured PhD from Sri Krishnadevaraya University in Computer Science and Technology. Overall 17+ years of experience in Computer Science and Applications education and a certified IGIP International Engineering Educator. Good track record of research success with multiple published articles. Guided hundreds of academic projects with a minimum of 6 projects a year which includes most of the real world problems.



C.Nagesh Currently working as an Associate Professor and HOD in the Department of Computer Science and Engineering at Tadipatri Engineering College, Anantapuramu. He did his B-Tech at SVUA, M-Tech from JNTUH. He had 15 years of teaching experience and 3 years of research experience. His research areas are Machine Learning, Big Data, Data Mining and Networks. He published more than 15 national and international journals. He guided 12 PG and 20 UG projects.



SA SOUTH ASIAN
ACADEMIC PUBLICATIONS

<https://www.saap.org.in/>

E-mail id :saapbooks@gmail.com