



TRENDNOLOGIES

Upskill & Upgrade Your IT Career

Bigdata/Hadoop, Spark full stack specialization program

100% Placement Guarantee

Supporting Job seekers in career change and upskill program

We work with job seekers to meet them exactly where they are by offering personalized, customized support to meet their needs while assisting them in navigating their next career opportunity. Faced with the current realities of an ever-changing workforce, we believe our placement services must meet individuals exactly where they are. Our customized and personalized Training and Placement programs offers better career in Bigdata.

More details visit WWW.TRENDNOLOGES.COM

Placement Process



MODE OF TRAINING

- Classroom Training
- Online Instructor-Led Live Training
- Online self-paced training sessions

Big Data Certification - Become A BigData Hadoop Certified Professional

Get Certified By

- Cloudera CCP Data Engineer
- CCA SPARK AND HADOOP DEVELOPER
- Databricks Certified Associate Developer for Apache Spark 3.0
- Databricks Certified Professional Data Scientist
- HDP Certified Developer (HDP CD)
- CCA Administrator
- CCA HDP Administrator

Pre-requisites

There are **no pre-requisites** for **Big Data Hadoop and spark course**. However, a basic understanding of computer science technicalities and Basic knowledge of Linux commands and SQL will be helpful but is not mandatory. Don't worry, we will cover Linux commands/scripting and SQL in detail in our course.

Syllabus

Topic 1 :Overview of Big Data & HDFS Concepts along with Linux Commands

- Bigdata Intro
- HDFS Architecture
- Hadoop Setup
- Linux commands
- HDFS commands
- Quiz
- Assessment

Topic 2: Linux Scripting in detail

- History
- Architecture
- Development Commands
- Env Variables
- File Management
- Directories Management
- Admin Commands
- Advance Commands

Syllabus

Topic 2: Linux Scripting in detail- Cont..

- Shell Scripting
- Groups and User managements
- Permissions
- Important directory structure
- Disk utilities
- Compression Techniques
- Misc Commands
- Kernel, Shell
- Terminal , SSH, GUI
- Automation & Scripting in Linux
- Hands On Exercises
- Quiz
- Assessment

Topic 3: SQL

- SQL Introduction
- SQL Syntax
- SQL Select, Distinct, Where, AND, Or, Not, Order By
- Insert Into, Null Values, Update, Delete, Select Top, Min & Max, Count, Avg, Sum
- SQL Like, Wildcards, In, Between, Aliases
- SQL Joins (Inner, Left, Right, Full, Self, UNION)
- SQL Group By, Having, Exists, Any, All
- SQL Select Into, Insert Into Select, Case, Null Functions, Stored Procedures, Comments, Operators
- SQL Database
- SQL Create DB , Drop DB, Backup DB
- SQL Create Table, Drop Table, Alter Table

Syllabus

Topic 3: SQL- Cont..

- SQL Constraints, Not Null, Unique, Primary Key, Foreign Key, Check, Default
- SQL Index
- SQL Auto Increment
- SQL Dates
- Views
- SQL Subqueries
- SQL Window Functions
- Performance Tuning SQL Queries
- SQL Hosting
- SQL Data Types
- SQL Cheat Sheet
- Quiz
- Assessment

Topic 4: Deep Dive in Hadoop

- What is Hadoop?
- Evolution of Hadoop
- Features of Hadoop
- Characteristics of Hadoop
- Hadoop compared with Traditional Dist. Systems
- When to use Hadoop
- Limitations of Hadoop
- Components of Hadoop (HDFS, MapReduce, YARN)
- Hadoop Architecture
- Daemons in Hadoop Version 1 & 2
- How Data is stored in Hadoop
- Cluster, Datacenter, Spilt, Block, Rack Awareness, Replication, Heartbeat)
- Hadoop 1.0 Limitation
- NameNode High Availability
- Quiz/Assessment

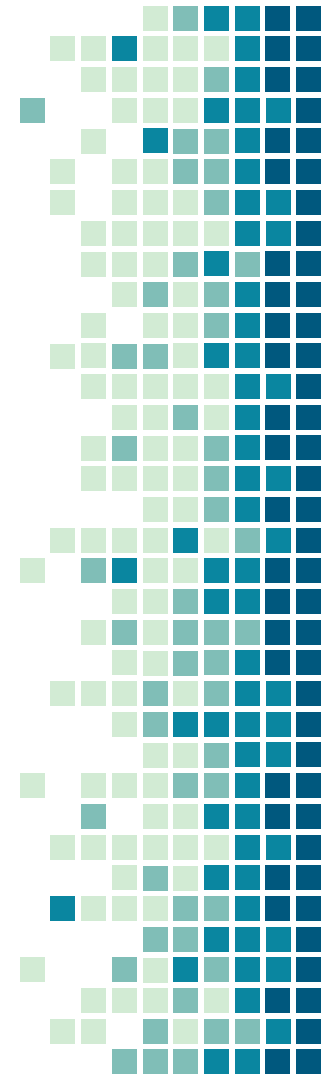
Syllabus

Topic 5: MapReduce - Distributed Computing Framework

- MapReduce Intro
- MapReduce – Theory / Depth
- MapReduce Programming concept
- MapReduce Practical
- Different types of files supported (Text, Sequence, and Avro)
- MapReduce Job submission in YARN Cluster in details
- Tweaking mappers and reducers
- Mapreduce package and deployment
- Quiz
- Assessment

Topic 6: Apache Sqoop - Moving Data into Hadoop (Vice- Versa)

- Sqoop Fundamentals
- Sqoop Exercise's
- Sqoop Import
- Sqoop Incremental
- Sqoop Export
- Sqoop Job
- Sqoop Merge
- Best practices & performance tuning
- Sqoop Use cases
- Quiz
- Assessment



Syllabus

Topic 7: Apache Hive Basics

- Introduction to Apache Hive
- Understanding Apache Hive
- Hive Practical
- Hive to know more
- Quiz

Topic 8: Advance Hive Advance & Components - Part 1

- Hive Definition Level Optimizations : Theory
- Hive Definition Level Optimizations : Practical
- Hive Query Level Optimizations : Theory
- Hive Query Level Optimizations : Practical
- Hive Windowing Functions
- Hive Ranking
- Hive Sorting
- Quiz/Assessment

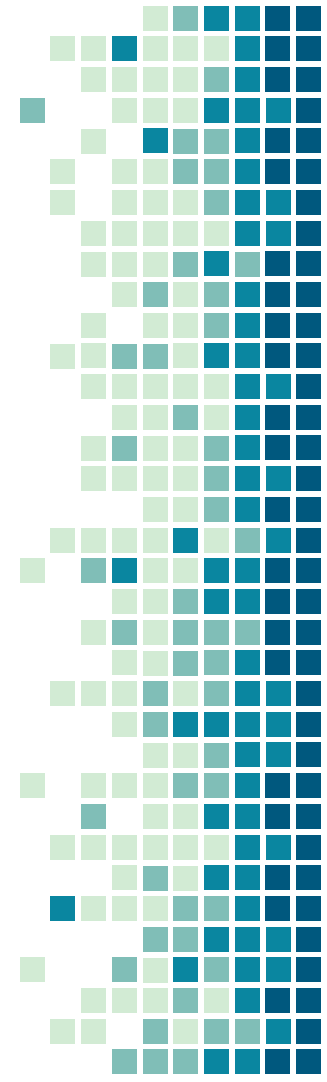
Topic 9: Advance Hive Advance & Components - Part 2

- Hive File Format
- Hive File Format – Practical
- Hive Compression Techniques
- Hive Vectorization & Changing the Hive Engine
- Hive Thrift Server
- Hive MSCK Repair
- Hive Miscellaneous
- Hive Optimization Techniques Revisit
- Hive SCD
- Hive / HBase Integration
- Hive Schema evolution (AVSC) use cases using AVRO dataset
- Quiz
- Assessment

Syllabus

Topic 10: Hue, Ambari, Cloudera Manager

- Introduction
- Cluster formation guide and implementation
- Deployment in Cloud
- Full Visibility of Cluster Health
- Metrics & Dashboards
- Heat Maps
- Configurations
- Services, Alerts, Admin activities
- Provisioning, Managing and Monitoring Hadoop Clusters
- Hue Introduction
- Access Hive
- Query executor
- Data browser
- Access Hive, HCatalog, Oozie, File Browser
- Hortonworks/Cloudera
- Cluster Design
- Different nodes (Gateway, Ingestion, Edge)
- System consideration
- Commands (fsck, job, dfsadmin, distcp, balancer)
- Schedulers in RM (Capacity, Fair, FIFO)
- View all services in Ambari & Cloudera Manager



Syllabus

Topic 11: Intro to NOSQL - HBase

- HBase Basics
- CAP Theorem
- HBase Architecture
- HBase Practical
- Storage Hierarchy – Characteristics
- Table Design
- HMaster & Regions
- Region Server & Zookeeper
- Inside Region Server (Memstore, Blockcache, HFile, WAL)
- Minor/Major Compactions
- Role of Zookeeper
- HBase Shell
- Introduction to Filters
- Row Key Design
- Performance Tuning

- Cassandra Overview
- Cassandra Architecture
- Integration with Hive
- Integration with Hadoop (Mini Project)
- Quiz
- Assessment

Topic 12: Phoenix

- Overview of Phoenix
- Introduction
- Architecture
- History
- Phoenix Hbase Integration
- Hbase table, view creation
- SQL & UDFs
- SQL Line & PSQL Line of Phoenix
- Phoenix Load & Query engine

Syllabus

Topic 12: Phoenix - Cont..

- Understanding co processor Configurations
- Hive -> HBase -> Phoenix integration
- Creation of views in phoenix
- Load bulk data using PSQL
- Serverlog Aggregation usecases

Topic 13: Oozie (Workflow Orchestration)

- Introduction
- History – Why Oozie
- Components
- Architecture
- Workflow Engine
- Nodes
- Workflow
- Coordinator
- Action (MapReduce, Hive, Spark, Shell & Sqoop)

- Introduction to Bundle
- Email Notification
- Error Handling
- Installation
- Workouts
- Orchestration of end to end tools
- Scheduling of data pipeline
- Invoking shell script, Sqoop, Hive & Spark

Topic 14: Python

- Python Introduction
- Evolution
- Application
- Features
- Installation & Configuration
- Objectives
- Flow Control
- Variables

Syllabus

Topic 14: Python - Cont..

- Data types
- Functions
- Modules
- OOPS
- Python for Spark
- Structures
- Collection types
- Looping Constructs
- Dictionary, Tuples & File I/O

Topic 15: Learning Scala - A Guide to Functional Programming

- Scala and Spark Setup
- Scala Basics
- Scala Functional Programming
- Scala Object Oriented Sessions
- Quiz/Assessment

Topic 16: Apache Spark - General Purpose Cluster Computing Framework

- Scala Interview Prep Kit
- Spark Fundamental Theory
- Spark Fundamental Practical
- Quiz
- Assessment

Topic 17: YARN

- Introduction to YARN
- YARN Architecture
- YARN Components
- YARN Long-lived & Short-lived Daemons
- YARN Schedulers
- Job Submission under YARN
- Multi tenancy support of YARN
- YARN High Availability
- YARN Fault tolerance handling

Syllabus

Topic 17: YARN - Cont..

- MapReduce job submission using YARN
- YARN UI
- History Server
- YARN Dynamic allocation
- Containerization of YARN
- Quiz
- Assessment

Topic 18: Apache Spark Use Cases in Depth

- Spark Real-Time Examples
- Spark Shared Variables
- YARN Revisit
- Spark on YARN Architecture
- Spark in depth
- Quiz
- Assessment

Topic 19: Spark StructuredAPI - Part 1

- Spark in depth continued
- Spark DataFrames, DataSets
- Quiz
- Assessment

Topic 20: Spark StructuredAPI - Part 2

- Spark in depth continued
- Quiz
- Assessment

Topic 21: Spark Performance Tuning - Part 1

- Spark Performance Tuning
- Quiz
- Assessment

Syllabus

Topic 22: Spark Performance Tuning - Part 2

- Spark Broadcast Join With Low level API (RDD)
- Spark Broadcast Join With Structured API (DataFrames)
- Spark Submit: Client Mode vs. Cluster Mode
- Spark Join Optimizations
- Spark Advance Optimization: Sort Aggregate vs Hash Aggregate
- Spark Catalyst, Tungsten, AST Optimizer
- Spark Connecting to External Source
- Quiz/Assessment

Topic 23: Spark Streaming

- Spark Real Time Processing
- Understanding Discretized Stream (DStream) in Spark Streaming
- Stream Processing in Spark – Word Count Example

- Understanding Stateless and Stateful Transformations in Spark Streaming
- Stateless Transformation – Word Count Example Using Eclipse IDE
- Stateful Transformation – Word Count Example
- Working with Sliding Windows
- Quiz/Assessment

Topic 24: Spark Advance Streaming - Structured

- Spark Structured Streaming – Part1
- Spark Structured Streaming – Part2
- Spark Structured Streaming – Part3
- Quiz
- Assessment

Syllabus

Topic 25: Apache Kafka

- Kafka Introduction
- Applications, Cluster Setup
- Broker fault tolerance
- Architecture
- Components
- Partitions & Replication
- Distribution of messages
- Producer & Consumer workload distribution
- Topics management
- Brokers
- Installation
- Workouts
- Console Publishing / Console Consuming
- Topic options
- Offset Management
- Cluster deployment in cloud

Topic 26: NIFI

- Nifi Introduction
- Core Components
- Architecture
- Nifi Installation & Configuration
- Fault tolerance
- Data Provenance Routing, mediation, transformation & routing
- Nifi -> Kafka -> Spark integration
- Workouts
- Scheduling
- Real time streaming
- Kafka producer & consumer
- File streaming with HDFS integration
- Data provenance / Packaging NIFI templates
- Rest API Integration / Twitter data capture
- Quiz/Assessment

Syllabus

Topic 27: Big Data on Cloud Part 1 (AWS S3, EMR, Athena + Glue)

- Introduction to Cloud Computing And Running Spark Code on AWS EMR
- Fundamentals of AWS for Bigdata Developer
- AWS Storage, Networking & CLI
- AWS EMR: Launch a EMR Cluster Using Advanced Options
- AWS Athena
- AWS Athena with Glue
- Quiz/Assessment

Topic 28: Big Data on Cloud Part 2 (Redshift, Glue, Airflow)

- Database vs. Datawarehouse vs. Data lake
- AWS Redshift Sessions
- AWS Glue
- Apache Airflow

- Apache Airflow – Workflow Management Platform
- Airflow Fundamentals Sessions
- Airflow Practical Pipeline Sessions
- Quiz/Assessment

Topic 29: CI/CD Pipeline (GitHub, Maven, & Jenkins)

- DevOps Basics
- Versioning
- Create and use a repository
- Start and manage a new branch
- Make changes to a file and push them to GitHub as commits
- Open and merge a pull request
- Create Story boards
- Desktop integration
- Maven integration with Git
- Create project in Maven

Syllabus

Topic 29: CI/CD Pipeline (GitHub, Maven, & Jenkins)- Cont..

- Add scala nature
- Maven operations
- Adding and updating POM
- Managing dependencies with the maven repository
- Building and installing maven
- Maven fat & lean jar build with submit

Topic 30: Real Time Project classes

- Real Time Projects on Bigdata which are diverse in nature covering various data sets from multiple domains such as Banking, Healthcare, Telecommunication, Social media, Insurance, and E-commerce.

Topic 31: Real Time Projects on Bigdata

- Project Statement
- Dataset
- Architectural Diagram and Solution
- Task segregation
- Project Demo Session
- Code and snippets
- Documentation

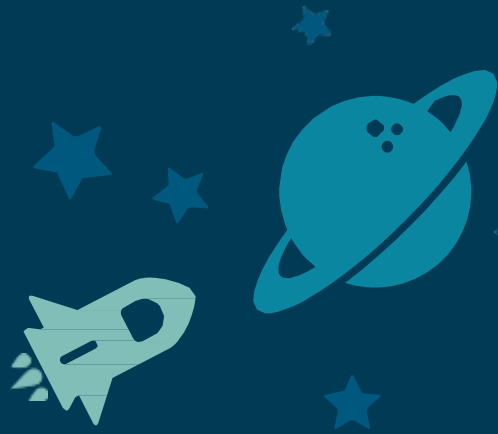
Topic 32: Interview Questions

Topic 33: Mock Interview and Mock Interview Answers

Topic 34: Resume Prep session and sample resumes

Know about your Trainers

- *Administrated and mastered by highly skilled industry experts Mr. Jai and Mr. Sundar with 27+ combined years of strong IT experience in leading MNC companies as Bigdata Architect and Senior Technical consultant.*
- *7+ Years of Hands-on expertise in Big Data & Analytics experience.*
- *Both Trainers are certified by Cloudera, Amazon Web Services and Azure.*
- *7+ years of delivering exceptional service to candidates in Bigdata Training and placements. Trained and Placed over 3000+ professionals in Big Data Hadoop and Spark.*
- *Experience professional and Placement oriented instructional technical trainers.*
- *Proven ability to plan and execute programs on time.*
- *Knowledge of "current best practices" and the latest research in the use of technology to enhance teaching and learning.*



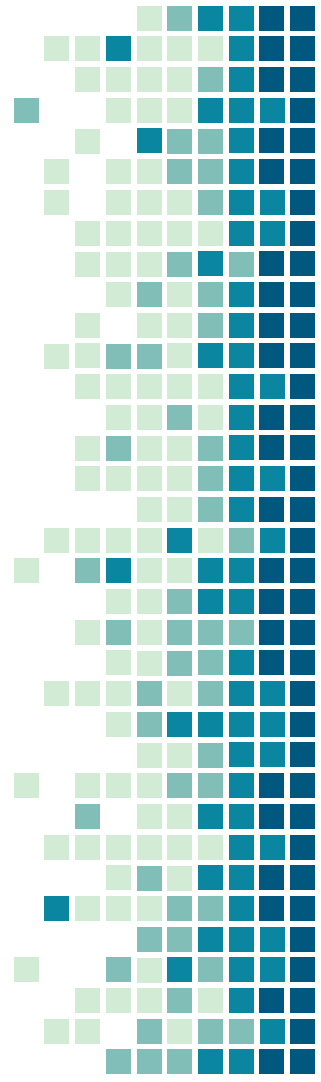
We're here to help

- Customized support addressing your needs
- Personalized approach consulting, planning and support
- Personalized career counselling
- Goal setting and establishing a career roadmap
- Focused on Emerging Technologies
- Technical Skills and personality screening
- In-depth review of a jobseeker's existing resume and creation of a customized new powerful resume.
- Establish an online presence in your job search with the latest channels.
- Networking, salary and negotiation strategies.
- Interviewing techniques and support
- Domestic and Global Placements
- Interview and Job Support
- Technical support after placement
- On-site assistance
- Re-deployment support



Why Trendnologies?

- 10+ years of delivering exceptional service to candidates in Training and placements.
- Fast Track Placement in 30 days.
- 100% Placement assurance.
- Managed by IT Experts as Trainers
- Customized support addressing your needs.
- Personalized approach / dedicated attention.
- Personalized career counselling.
- Personalized consulting, planning and support.
- Job oriented theoretical & Practical Classes.
- More focus on hands-on and use cases.
- Focused on case studies and Real-time projects.
- Complete end to end training soft material.
- Guidance for Resume preparation and interview preparation.
- Bigdata training along with Career guidance and Certification guidance.
- Comprehensive career solutions across India, USA, UK and UAE.



Key Features



About Bigdata – Data is Everywhere

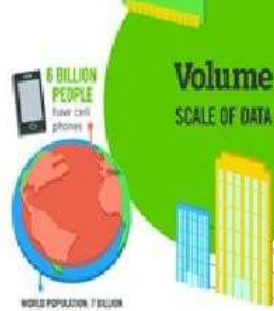
BIG DATA CLASSIFICATION



Big Data Statistics

40 ZETTABYTES

(40 trillion gigabytes)
of data will be created by
2020, an increase of 300
times from 2005



It's estimated that
2.5 QUINTILLION BYTES
(2.5 TRILLION GIGABYTES)
of data are created each day



Most companies in the
U.S. have at least
100 TERABYTES
(100,000 GIGABYTES)
of data stored

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources, such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data,
with 1.9 million in the United States



As of 2011, the global size of
data in healthcare was
estimated to be

150 EXABYTES
(100 million gigabytes)



**30 BILLION
PIECES OF CONTENT**
are shared on Facebook
every month



Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated
there will be
**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**

**4 BILLION+
HOURS OF VIDEO**
are watched on
YouTube each month



400 MILLION TWEETS
are sent per day by about 200
million monthly active users

**1 IN 3 BUSINESS
LEADERS**

don't trust the information
they use to make decisions



In one survey, more than
half of respondents were unsure of
how much of their data was
accurate

Veracity UNCERTAINTY OF DATA

Poor data quality costs the US
economy around
\$0.1 TRILLION A YEAR



The New York Stock Exchange
captures
**1 TB OF TRADE
INFORMATION**
during each trading session



Velocity ANALYSIS OF STREAMING DATA

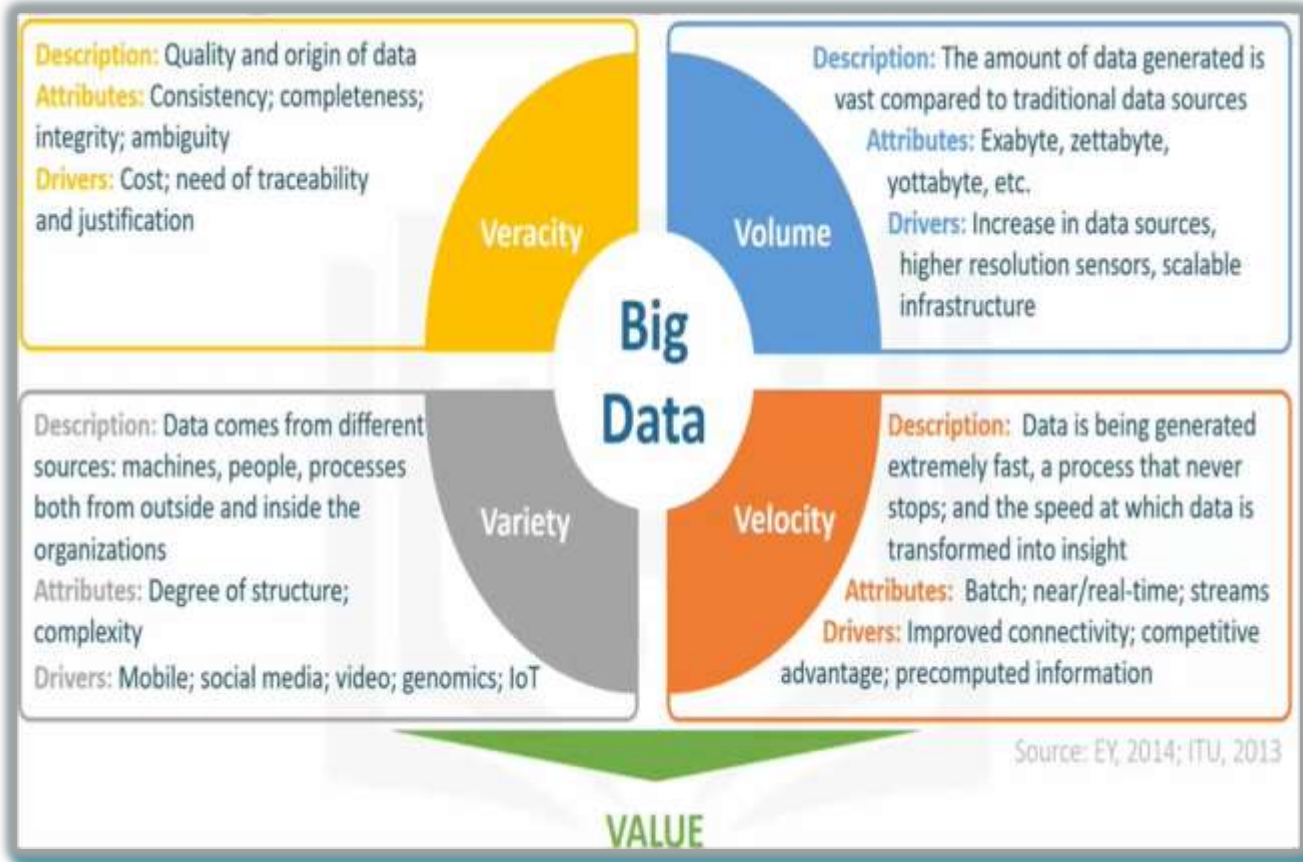
Modern cars have close to
100 SENSORS
that monitor items such as
fuel level and tire pressure



By 2016, it is projected
there will be
**18.9 BILLION
NETWORK
CONNECTIONS**
+ almost 2.5 connections



Characteristics of Big Data



Big data and Hadoop in Different Domains

Banking and Finance Sector

The banking and Finance industries face some of the challenges like card frauds, tick analytics, archival of audit trail, enterprise credit risk reporting, etc. They use Hadoop to get an early warning for security fraud and trade visibility. They use Hadoop to transform and analyze customer data for better insights, pre-trade decision-support analytics, etc.

Communication, Media and Entertainment

The communication, media, and entertainment industries face some challenges like collecting and analyzing consumer data for insights, finding patterns in real-time media usage, using social media, and mobile content. Using Hadoop, these companies analyze customers' data for better insights, create content for different target audiences.

Healthcare Providers

The Healthcare sectors by using Hadoop analyzes the unstructured format of data that includes patient history, disease case histories. This helps them to effectively treat the patients effectively based on previous case histories. Also, by identifying the disease that is common in a particular area, precautions can be taken, and medicines can be made available to those areas.

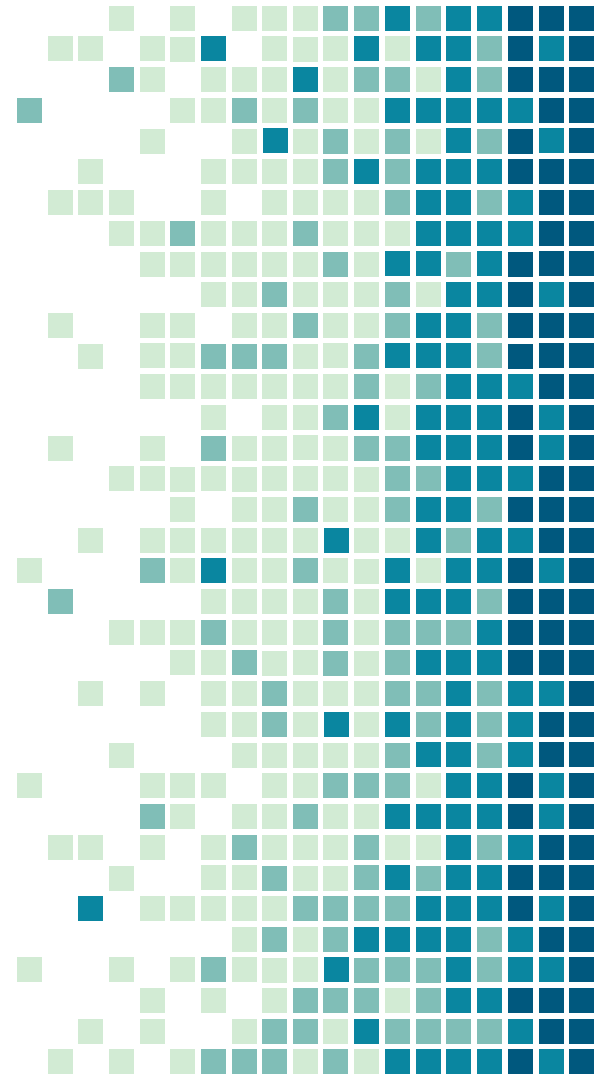
The University of Florida uses free public health data and Google Maps to visualize data that allows for faster identifying the spread of chronic disease.

Education

The education sector uses big data significantly. A University of Tasmania having 26000 students has deployed LMS(Learning Management System) that tracks the log time, how much time students spend on different pages and overall progress of the student over time.

Government

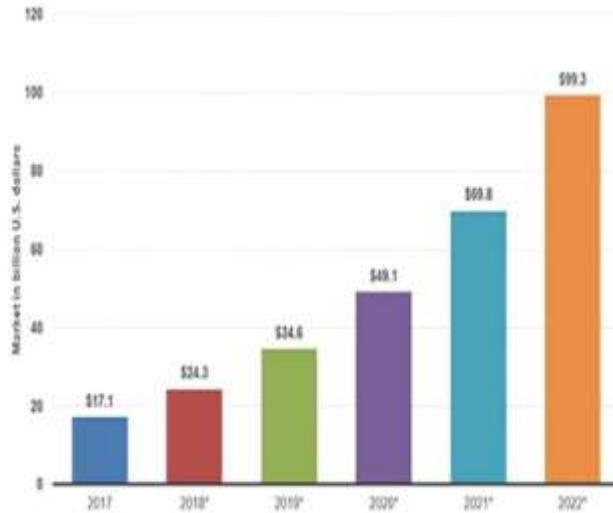
There are various government schemes that are in execution and are generating data tremendously. The Food and Drug Administration(FDA) is using Big Data to detect and study the patterns of food-related diseases, allowing for faster treatment responses.



Future of Hadoop

Big Data and Hadoop Market Size Forecast Worldwide 2017-2022

Size of Hadoop and Big Data Market Worldwide From 2017 To 2022
(in billion U.S. dollars)



As per the **Forbes** report, the **Hadoop and the Big Data** market will reach **\$99.31B** in 2022 attaining a **28.5% CAGR**.

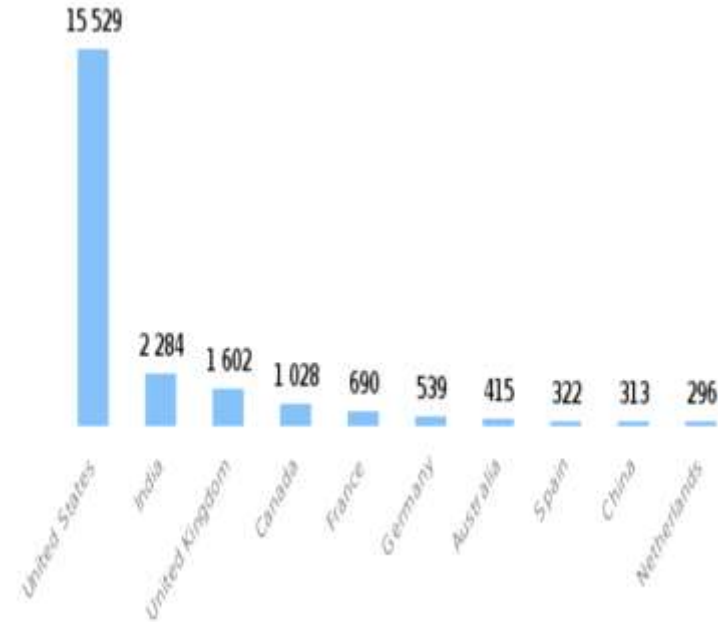
Thus learning Hadoop is the milestone to upskill your career in IT sectors as well as in many other domains.

COMPANIES USING HADOOP

The research indicates that Hadoop has good market prospects in many industries. With the arrival of the digital universe, we are dealing with the data explosion. As time passes, new technologies are continuously emerging, contributing a pool of data.

From Computing IT sector industries to industries like hospitals and healthcare, education, finance, telecommunication, retail, etc. all have Hadoop applications running on them. With the realization of the advantages of big data analysis, the adoption of Hadoop is increasingly exponentially

Distribution of companies using Apache Hadoop by Country



Job Predictions in Big data Analysis

HIGHEST PAYING ANALYTICAL SKILLS

Skill Name	Average Salary
MongoDB	\$115,256
Data Science	\$111,200
Big Data	\$112,568
NoSQL	\$110,596
Apache Hive	\$115,256
Machine Learning	\$116,235
Apache Hadoop	\$113,258
Predictive Analytics	\$118,334
MapReduce	\$119,895

Big Data Analytics Job Titles & Salaries

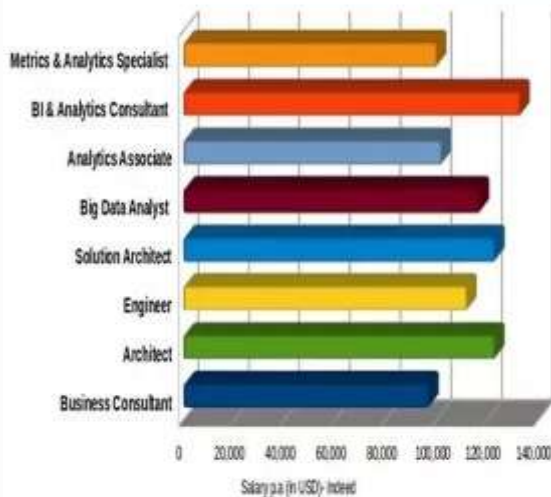


Image Source: Indeed

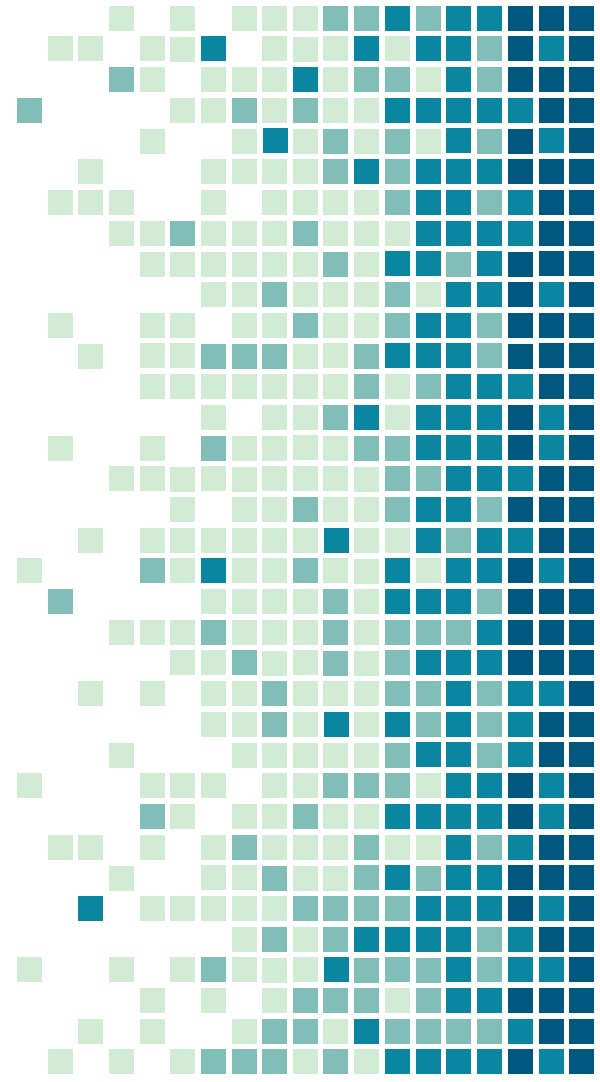
By 2023, the big data analytics market will reach \$103 billion as per the predictions. IBM predicts that the demand for the data scientist will soar 28%.

Over 97.2% of organizations are investing in big data and AI.

As per the PwC report, by 2022, in the US only there will be around 2.7 million job vacancies for data analytics and data science. The largest companies like Cisco, Dell, EY, IBM, Google, Siemens, Twitter, OCBC bank, are looking for Hadoop professionals to process and profit from the sea of available data.

The average annual salary in the UK is £66,250-£66,750 for Hadoop jobs and \$92,512 to \$102,679 for Hadoop developers, as per Indeed.

TARGET AUDIENCE



Our Students are Placed - Globally



BRANCHES - CHENNAI

Annanagar Branch I

*F22, 2C & 2D, Second floor,
Second Main Road,
Annanagar East,
Chennai - 600102
86819 62962*

Medavakkam Branch

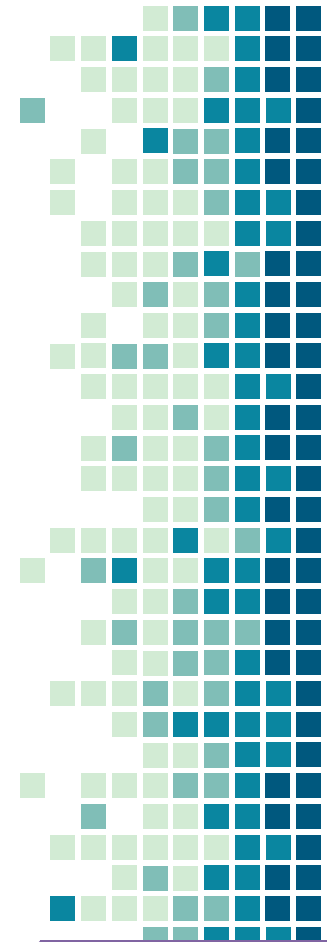
*#29, M.S Complex,
Second Floor,
Anna Street,
Medavakkam.
Chennai - 600100
84288 88962*

Annanagar Branch II

*New no. 83, Old no. M-9-B, M
Block
Annanagar East,
Near K4 Roundtana,
Chennai - 600102
86809 62962*

Placement Division

*F22, 3D third floor,
Second Main Road,
Annanagar East,
Chennai - 600102.
91763 62962*



BRANCHES - COIMBATORE

Coimbatore Branch I

*#778, Mettupalyam Road,
R S Puram
Coimbatore - 641002
93600 93938*

Coimbatore Placement Division

*KB complex 105 ,
Lokamanya Street East,
R S Puram ,
Coimbatore - 641002
8428444962*

BANGALORE Branch

*#421, 4th floor
South block Manipal
centre,
MG Road
Bengaluru - 560042
84286 66962*



Thanks
Contact with Bigdata
Experts @8270962962

Visit our website for more
details :

WWW.TRENDNOLOGIES.COM

