Big data is one of the most popular terms in recent years. More and more industries begin to use the findings of big data analysis to formulate future strategic decisions. Therefore, the demand for big data related talents is increasing accordingly. The Big Data Experiment Lab is designed for big data analysis skills training. The rich and comprehensive training courseware enables students to learn big data analytical skills effectively and efficiently.

● What is big data ?

Big data refers to a collection of large data sets with diversified structures. In the past, big data was usually used for data analysis or statistical applications in enterprises, and it was limited to analyzing historical data. Today, data analysis advanced from analyzing historical data to prediction of future trends with improved accuracy. This is due to the development of clouds and the Internet of Things, lowered costs for storage equipment, rapid growth in data volumes, and the improvement in software technologies. Big data is no longer just data processing, but an unprecedented tool for business intelligence.

The Characteristics of Big Data

Volume

In various fields such as financial services, energy management, biomedicine, and multimedia communities, a large number of data sets is being generated every second.

Velocity

Whenever data is sent to servers, it will be analyzed immediately and bring real-time modifications on the previous results so as to obtain latest findings with maximum data values.

Variety

Diverse data includes structured and unstructured data, such as text, location, sound, videos, and pictures, all of which can be interactively analyzed to reveal correlations between data sets.

Veracity

Is the data source correct? Is the data accurately recorded even if it's true? Are there any anomalies in data sets? Wrong data sources may result in deviations in the analysis results and affect the accuracy of the predictions. Therefore, ensuring authenticity of data sources is also one of the key points of big data analysis.

Value

The greatest values of big data analysis lie in digging out the data that is valuable for future trends from massive data, and performing in-depth analysis through artificial intelligence and machine learning to improve accuracy.

Hadoop - The System Foundation of Big Data

Hadoop is an open source software framework that can successfully solve various problems, such as file storage, file backup and data processing. Therefore, it is widely used and has become the mainstream technology of big data analysis.

Spark - Data Processing for Big Data

Speed is very important in data processing when it comes to big data. An important feature of Spark is that it can be operated in memory, which makes Spark more efficient in data analysis and calculation than Map Reduce.

Python - The Extraction of Big Data

Python is a programming language commonly used in various fields. It can crawl large amounts of effective data from the network in a low-cost and automated manner. The powerful data processing capability is the main reason why Python becomes an important programming language when analyzing big data.

The System Features

1. System independence: it can be operated without any internet connection nor any additional hardware/software installation. The cabinet design makes it easy to move.
2. Convenience: the troubleshooting function provides system restoring function, so users can troubleshoot quickly. Through 6 different models of random data generation, users can generate datasets suitable for different algorithms easily.
3. Expansion: it can be applied to different researches and experiments for big data analysis, and it can also be combined with IoT-enabled devices to store and analyze various data sets from sensors to achieve cross-domain applications.
4. Richness : it provides comprehensive courseware for big data analysis
   (1) 9 different algorithms & 20+ classic big data analysis examples
   (2) Tools, such as Hadoop, Yarn, Spark, Hive, and HBase, are introduced and applied.

Learning Objectives

1. Data cleaning, regularization and standardization
2. Architecture and configuration of the big data ecosystem
3. Comparison of assorted databases
4. Use various algorithms to extract, store, retrieve and analyze big data
5. Integration of big data analysis and artificial intelligence

BD-101B

Big Data Experiment Lab

Big data is an area that involves extraction, storage, retrieval, and analysis of data sets, especially those that are too voluminous and complex for traditional data processing application software to process. With the advent of technological advances as well as the accessibility and transparency of information, big data analysis is becoming increasingly prominent and prevalent in modern research, and more and more industries have adopted big data analysis to generate strategic and operational value.

Several data statistical analysis methods have been developed to analyze data sets of massive volume, variety, and velocity. As “big data” proliferates, data statistical analysis methods become one of the most determinant factors in determining whether or not a data set is valuable in terms of operability, timeliness, and predictivity. As a result, the demand for data scientists in today's job market is growing and will continue to grow.

Unlike traditional big data equipment, this Big Data Educational Equipment is a light weight big data equipment designed with flexibility. Simply power it on, immediately a classroom is converted into a big data lab. With our fast troubleshooting function, user can easily restore the system with a few steps if disorder occurs. Students can learn the entire big data ecosystem and the operation of our equipment with the comprehensive big data textbook provided.

Features

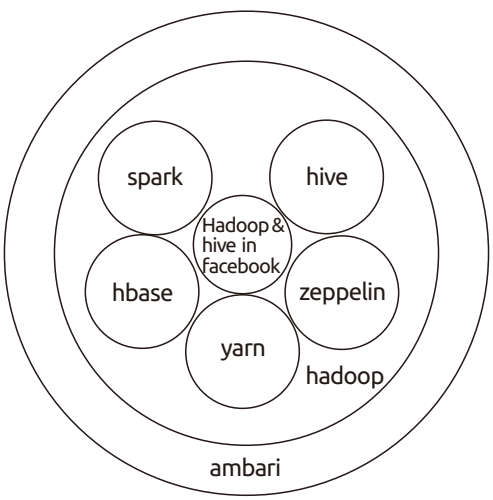
1. Standalone System
   (1) Standalone operation system without the requirement of extra hardware / software installations.
   (2) Easily-moved cabinet architecture that convenient for teaching in any classroom without special setting.
2. 24-7 Monitor System
   (1) Check and update the CPU, network, hardware real-time situation through 24-7 monitor system.
   (2) Refresh button which can renew real-time data on monitor panel.
3. Comprehensive Big Data Ecosystem Training
   (1) Standalone big data server which provides 20+ case studies with step-by-step experiment instructions.
   (2) Including necessary big data introduction and instructions, i.e. Hadoop, Yarn, Spark, Hive, HBase, etc.
4. User Friendly
   Simply proceed experiments through web-IDE, without any complex setting on parameters and operating system.
5. Fast Troubleshooting
   Easily restore the disordered nodes to defaults by system reset function, needless to spend plenty of time sorting problems out.
6. Random Generator
   6 different model random generators. One click generates available data for corresponding algorithm.
7. Applications of Big Data & Extensions to AI
   (1) The system can be extended as the core for learning in big data, AI, etc.
   (2) The system can be used as a tool for users to design an algorithm to analyze data collected from other sources.

Specifications

1. BD-Cluster (BD-10001)
   (1) CPU : 6 core 12 threads
   (2) RAM : More than 190G
   (3) POWER : More than 1000W
   (4) Storage : 3TB
   (5) Node : Master node x 1, data node x 2
2. BD-Rack Cabinet (BD-10091)
   (1) Rack : ANSI/EIA standard, more than 19U
   (2) Cabinet door : Removable tempered glass door with lock x 1
   Removable side x 2
   (3) Environmental Monitoring : Fan x 2
   (4) Power Monitoring : Current meter x 1
   (5) Wireless LAN : IEEE 802.11
3. BD-System Monitor (BD-10002)
   (1) CPU usage display
   (2) Hard disk usage display
   (3) Network usage display
   (4) Ram usage display
   (5) Refresh button x 1
   (6) Random generate dataset button x 6
4. BD-Data Server (BD-10003)
   (1) NAS : 2 Bay
   (2) DataSet : Up to 20 case studies, each case includes 100+ thousand data

5. Tooling
   (1) Framework : Ambari
   (2) File system : Hadoop
   (3) Source manage : Yarn
   (4) Computer : Spark, Mapreduce
   (5) Database : Hive, HBase
   (6) IDE : Zeppelin
   (7) Interactive application
   (8) Restore : Clonezilla
6. Accessories(BD-19001)
   (1) Mouse/keyboard : wireless x 1, wired x 1
   (2) Cable : RJ45 x 7
   (3) Experiment manual x 1 set

List of Modules



List of Experiments

1. Python Experiments
   (1) Python introduction
   (2) Web-scraping by python
   (3) Data process
2. Yarn Experiments
   (1) Yarn configuration
   (2) End process
   (3) Compare configuration experiment
3. Hive / HBase Experiments
   (1) Hive / MySQL
   (2) HBase / NoSQL
   (3) Compare databases experiment
4. Spark & Zeppelin
   (1) Alternating Least Squares matrix factorization
   (2) Decision tree
   (3) Pipeline
   (4) Resilient distributed dataset
   (5) Support Vector Machine
   (6) Binary classification
   (7) Naive Bayes binary
   (8) Mean-shift
   (9) Decision tree regression
5. AI Related
   (1) Neural networks
   (2) Facial recognition

For more details contact:
Quality Quest Engineers
Chennai : 600006
Tamilnadu, India
Ph: +91 98400092362