BD-101

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.

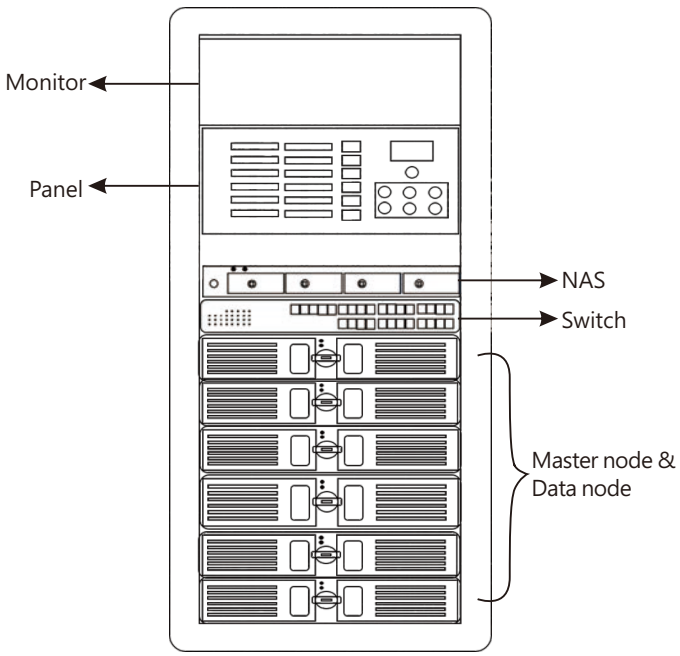
Features

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 2500W
   (4) Storage : 6TB
   (5) Node : Master node x 1, data node x 5
2. BD-Rack Cabinet (BD-10091)
   (1) Rack : ANSI/EIA standard, more than 30U
   (2) Cabinet door : Removable tempered glass door with lock x 1
   Removable side x 2
   (3) Switch : 24 port with network management
   (4) Environmental Monitoring : Fan x 4
   (5) Power Monitoring : Current meter x 2
   (6) 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) Monitor : 19 inch
   (2) NAS : 4 Bay
   (3) 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 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

List of Modules