Hewlett Packard Enterprise

Big Data: From a Technical Perspective with Hands-on Exercises BD120S

HPE course number	BD120S
Course length	2 Days
Delivery mode	ILT
View schedule, local pricing, and register	**View now**
View related courses	**View now**

Why HPE Education Services?

- Hands-on practice
- Customized on-site delivery
- More than 80 training locations worldwide
- Online instructor-led and self-paced training at www.hpe.com/education
- Comprehensive student materials
- Experienced and best-in-the-field HPE instructors
- Focus on job-specific skills State-of-the-art classroom facilitie

This 2-day course introduces the essential knowledge of Big Data from a technical hands-on perspective. The characteristics of Big Data and the challenges working with Big Data will be illustrated. An integrated information architecture that incorporates Big Data is introduced. Then, the components of Big Data technologies will be described with Big Data storage and Big Data processing as two fundamental pillars. NoSQL database and storage system for Big Data storage will be explained. Hadoop - with HDFS, MapReduce, Pig Latin and Hive – for Big Data processing will be covered. Useful analytic techniques with highlights on statistics, machine learning and pattern mining will be described to further strengthen the know-how about Big Data processing. This course integrates both lecture and practical hands-on exercises. Students can practice different Big Data technologies using well-prepared virtual environment and exercise scripts.

Audience

- IT Professionals
- Programmers
- Administrators and Data Analysts

Prerequisites

We recommend attendees to have basic information technology knowledge before attending this course.

Key topics

1.Big Data Overview	• What is Big Data? • Challenges in extracting useful information andinsights from Big Data • Data volume: growing amount of data	• Data variety: handling structured, semi-structured, unstructured data • Data velocity: derive insights quick forconsumption • Data veracity: create values from insights
2.Big Data: Integrated InformationArchitecture	• Incorporating data of multiple structures • Structured data • Semi-structured data • Unstructured data	• Data Analysis Lifecycle • Identify data sources for analysis • Data storage • Data processing – pre-processing, analysisand visualization
3.Components in Big Data solution	• Emergence of Big Data technologies • Components HDFS • MapReduce • NoSQL storage/database	• Hive/ Pig • Flume • Statistical tool and data mining tools • Other ecosystem components
4.Big Data Storage	• Emergence of NoSQL database/storage system • How NoSQL works • Data storage models • Key-value store	• Column-family store • Document store • Graph store • Choosing the data storage models based ondata characteristics
5.Big Data Processing: Using Hadoop	• What is Hadoop? • Hadoop • Market segments, their size and growth • Hadoop architecture • - Hadoop Distributed File System (HDFS) and itsconcepts	• - MapReduce and its concepts • Integration of disparate data storage system • Practical development tips and techniques • An interactive game to illustrate the working mechanismof Hadoop
6.Big Data Processing : Using scripts	• MapReduce development versus Script-based development • Pig Latin: How does it work? • Example: Abstract MapReduce jobs using Pig Latin	• Hive: How does it work? • Example: Perform ad-hoc Big Data query with Hive
7.Big Data Processing: Analytic techniques	• Using statistics to perform analytics (Finding co-occurrence pattern) • Using machine learning library to perform analytics(Finding recommendation items)	• Using data mining to perform analytics (Enhancing searchquality)
8.Hands-on Exercises	• Getting started with NoSQL for real-time data access • Integrate NoSQL in web application • Use Hadoop in pseudo-distributed mode • Monitor HDFS using convenient Web Console • Monitor MapReduce using convenient Web Console	• Use Hive/Pig scripts to discover insights on Hadoop • Calculate aggregated statistics of large-scale data on Hadoop • Perform intelligent recommendation on Hadoop • Perform pattern recognition on Hadoop

Learn more at
hpe.com/ww/bigdata

Follow us:

BD120S A.00, January 2019, a00064769enw

Hewlett Packard
Enterprise