BIG DATA USING HADOOP PROGRAM

An eleven-week in-depth program covering the Apache Hadoop™ framework and how it fits with Big Data

DePaul University's Big Data Using Hadoop Program is designed for IT professionals whose companies are new to the Big Data environment. Growing interest in Apache Hadoop™ and Big Data is driving the demand for professionals with skills to derive value from all data. Big Data refers to the ever increasing amounts of unstructured data being generated by websites, sensors, and other sources as well as transactional data from enterprise applications. Hadoop is an open-source implementation of frameworks for reliable, scalable, distributed computing and data storage, which can be used for managing Big Data. This program will walk through installation, provisioning, and ongoing resource management within a Hadoop cluster. The program also covers the full lifecycle of a Hadoop deployment using realistic hands-on lab experiments. Students will obtain a deeper understanding of Hadoop using real-world use cases to help provide an understanding of the power of Hadoop.

DePaul University has partnered with Cloudera® through the Cloudera Academic Partnership (CAP) to foster the next generation of Big Data professionals. DePaul uses Cloudera® academic program materials and its industry-leading software products and tools to deliver high-quality training programs with state-of-the-art Big Data processing technology. Students will learn how to design, build, launch, deploy, and scale an application using Hadoop in different platforms. Classroom lectures and demonstrations will be complemented by hands-on labs, reading assignments, case studies, and projects. In order to maximize learning, students will be required to bring their own laptop computer to every class session. While access to most cloud services explored in the program will be provided to students in class, there may be some cloud services that are only accessible via the use of a student’s own credit card. Students should expect to spend a small fee to access these services.

YOU WILL LEARN:

- Big Data characteristics, the Hadoop ecosystem, and core technologies of Hadoop
- How Hadoop Distributed File System (HDFS) and MapReduce work
- How to develop and test MapReduce applications
- How to use MapReduce combiners, partitioners, and the distributed cache
- How to implement common algorithms in Hadoop
- What hardware configurations are optimal for Hadoop clusters
- How to configure Hadoop's options for best cluster performance
- How to use ETL-like tools such as Sqoop and Flume to bring data into Hadoop
- How to install/run Hadoop in the Cloud
- How to integrate other tools with Hadoop such as Apache Pig™, Apache Hive™, etc.
- Use cases and best practices to process Big Data with Hadoop

DePaul University
College of Computing and Digital Media
Institute for Professional Development
243 S. Wabash Avenue, Room 301
Chicago, IL 60604-2300
(312)362-6282
ipd.cdm.depaul.edu
**BIG DATA USING HADOOP PROGRAM**

The Big Data Using Hadoop Program provides IT professionals with the technical skills required to successfully implement and maintain Hadoop clusters. The program is designed to provide a rapid immersion into Big Data with Hadoop. Students will be introduced to the Hadoop cluster, the ways to interact with the Hadoop file system (HDFS), and the job process using MapReduce. An introduction to different higher-level interfaces and tools to manage data process in the Hadoop cluster such as Apache Hive, Apache Pig, etc. is also presented. The program consists of an effective mix of interactive lecture and extensive use of hands-on lab exercises. The program will focus on the most popular Hadoop distribution—Cloudera®. Other Hadoop distributions will also be covered.

Students will gain knowledge of the skills needed to build and deploy applications using a variety of common services. Each student will have a flexible environment to access Hadoop along with sample code and scripts to learn best practices and real-world scenarios. Classroom lectures and demonstrations will be complemented by reading assignments, hands-on exercises, case studies, and projects. Students in the program are also expected to bring their own laptop computers to class.

**CURRICULUM**

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BIG DATA OVERVIEW</strong></td>
<td>Review the history of data models and usage, current state, and future trends. Cover Big Data industry and vendor landscape. Discuss Big Data adoption, maturity level, market, and job opportunities.</td>
</tr>
<tr>
<td><strong>HADOOP ECOSYSTEM</strong></td>
<td>Introduction to the Apache Software Foundation, the Hadoop ecosystem, products, tools, user community, and contributors. Review core Hadoop components. Cover fundamental concepts of parallel and distributed systems.</td>
</tr>
<tr>
<td><strong>PLANNING YOUR HADOOP CLUSTER</strong></td>
<td>General planning considerations including selecting the right hardware, software, storage, and network.</td>
</tr>
<tr>
<td><strong>INSTALLING HADOOP</strong></td>
<td>Hadoop installation and cluster configuration. Review advanced configuration parameters and configuration management tools.</td>
</tr>
<tr>
<td><strong>THE HADOOP DISTRIBUTED FILE SYSTEM (HDFS)</strong></td>
<td>Overview of HDFS architecture design, features, and security. Understand writing and reading files.</td>
</tr>
<tr>
<td><strong>MAPREDUCE</strong></td>
<td>Overview of basic MapReduce concepts and architecture. Major features and key components of MapReduce. Discussion of node failure and recovery.</td>
</tr>
<tr>
<td><strong>CLUSTER ADMINISTRATION AND MANAGEMENT</strong></td>
<td>An overview of cluster upgrading, cluster rebalancing - adding and removing cluster nodes. How to back up NameNode Metadata. How to monitor cluster, check HDFS status, and manage Hadoop’s log files.</td>
</tr>
</tbody>
</table>
Installing and managing other Hadoop projects

Installation of other tools such as Apache Hive, Apache Pig, Apache Impala, Apache Sqoop, Apache Flume, Apache Hue, Apache HBase, etc. to integrate with Hadoop.

Case studies and Industry trends

Case studies and best practices. Discussion of technology trends and job market.

General information

Admission

Applicants should have basic programming experience and experience using Windows and Linux commands. In addition, students are required to bring their own laptop computers to class. Existing knowledge of Hadoop is not required.

A substantial commitment of time is required for this intensive course of study. Final admission will be determined by the admissions committee on the basis of an applicant's overall qualifications, including work history and educational background.

Facilities

To promote the learning process, the Institute maintains special-purpose laboratories as well as dedicated classrooms equipped with state-of-the-art audio/visual equipment.

In addition, the college’s unique Course OnLine (COL) technology allows students to replay classes over the Internet. COL captures and replays five components of the classroom experience—audio, video, PC screen, whiteboard, and document camera input—and incorporates them into one interface to provide an innovative rebroadcast system.

Schedule

The Institute offers one section of the program each quarter. Classes meet one day per week. The option to take the program entirely online is also available.

Faculty

The faculty consists of a team of instructors from the College of Computing and Digital Media and experts in the industry. Faculty will be available throughout the program both in person and through electronic mail.
The college, through its School of Computing and its School of Cinema and Interactive Media, offers a variety of programs at the undergraduate and graduate levels. Over 1,000 students are enrolled in the college’s bachelor’s programs and over 2,000 students are enrolled in the master’s and Ph.D. programs making the college’s graduate program one of the largest in the country. The college offers more than 200 courses each quarter, many in the evening, in several locations: the Loop Campus, the Lincoln Park Campus, the O'Hare Campus, the Naperville Campus, and the Rolling Meadows Campus. Most of the degree programs are also available exclusively online.

Current offerings at the undergraduate level include:

**Offerings at the undergraduate level include:**
- Animation B.S. / Animation B.A.
- Computer Games Development B.S.
- Computer Graphics and Motion Technology B.S.
- Computer Science B.S.
- Computing B.A.
- Digital Cinema B.S. / Digital Cinema B.A.
- Graphic Design B.F.A.
- Information Assurance and Security Engineering B.S.
- Information Systems B.S.
- Information Technology B.S. / Information Technology B.A.
- Interactive Media B.S.
- Math and Computer Science B.S.
- Network Technologies B.S.

Offerings at the graduate level include:
- Animation M.A.
- Applied Technology M.S.
- Business Information Technology M.S.
- Cinema Production M.S.
- Computer Game Development M.S.
- Computer Graphics and Motion Technology M.S.
- Computational Finance M.S.
- Computer Science M.S.
- Computer, Information and Network Security M.S.
- E-Commerce Technology M.S.
- Human-Computer Interaction M.S.
- Information Systems M.S.
- Information Technology M.A.
- IT Project Management M.S.
- Network Engineering and Management M.S.
- Predictive Analytics M.S.
- Software Engineering M.S.
- J.D./M.A. in Computer Science Technology
- J.D./M.S. in Computer Science Technology
- Master’s of Fine Arts
  - Animation
  - Cinema
  - Screenwriting
- Ph.D. in Computer and Information Sciences

The Institute for Professional Development was formed by the college in 1984 to assist both individuals and businesses in keeping pace with the rapid development of computer technologies. The Institute currently offers a variety of intensive certificate programs in these areas:

- Advanced SQL
- Big Data and NoSQL
- Big Data Using Hadoop
- Cloud Computing Technologies
- Data Science for Business
- IPv6
- Java™ Developer
- Java™ Web Development
- .NET Developer
- Ruby on Rails®
- SQL Server® Business Intelligence
- SQL Server® Database Administration
- Web Development with JavaScript and HTML5
- Web Development with Python®

**APPLICATION PROCEDURE:**

Complete the enclosed application and return it with a non-refundable $40.00 application fee (check or money order made payable to DEPAUL UNIVERSITY) to:

DePaul University
Big Data Using Hadoop Program
Institute for Professional Development
243 S. Wabash Avenue, Room 301
Chicago, IL 60604-2300

The words "Apache Hadoop", “Apache Hive”, “Apache Pig”, “Amazon Web Services”, “Cloudera”, “SharePoint”, “SQL Server”, “Python”, “Java”, and “Ruby on Rails” are registered or unregistered trademarks in the United States of America and/or other countries. The Big Data Using Hadoop Program at DePaul University is an independent program of study and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by the Apache Software Foundation or any other external entities.