Big Data Hands-on Workshop

Nov 8-9 (Sat-Sun) 9am-5pm
@ WHITE SPACE Houston
2914 White Oak Drive, Houston, TX  77007

Advance registration fee is $245
Pay @ bit.ly/bigdutd by Nov 3

Massive amounts of structured and unstructured data is being generated from online user activities on e-commerce and social-media websites. Big Data, as this massive amount of data is called, leads to many challenges:

- How do companies manage and process this massive amount of data?
- How do companies automatically learn hidden trends and patterns in this data?
- How do companies gather actionable intelligence to improve their bottom line?

This workshop will cover tools and techniques that directly address these challenges. This workshop is meant for technical managers and software developers. No advanced knowledge is needed. The workshop will be focused on hand-on practice labs with minimal lectures along the way. After attending this workshop you will be able to manage, analyze and learn from Big Data using commonly available tools.

**Featured Topics:**

- Storing and processing Big Data: Hadoop, Mapreduce, NoSQL
- Tools and techniques for analyzing Big Data
- Fundamentals of learning from Big Data
- Practical applications of Big Data
- Processing and Analyzing data with Hive and Advanced Hive

(See reverse side for the detailed agenda)
# Detailed Agenda

## Day 1: Introduction to BigData

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
</table>
| 9:00 – 10:30 | **Workshop Introduction**  
**Importance of BigData in today’s industry**  
**Big Data from Technology prospective**     |
| 10:30 – 10:45 | 15 minute Break                                                            |
| 10:45 – 12:00 | **Introduction to Hadoop**  
**Understanding Hadoop Ecosystem**  
**Introduction to HDFS**  
**Traditional File System VS HDFS**     |
| 1:00 – 2:15 | **Architecture and working of HDFS Federation**  
**Google Research Paper (GFS)**     |
| 2:15 – 2:30 | 15 Minute Break                                                            |
| 2:30 – 3:45 | **NoSQL**  
**Introduction to Hortonworks Hadoop Platform, set up computers**  
**Using HDP and interfaces**  
**Lab 1 (Hands on Practice)**     |
| 3:45 – 4:00 | 15 Minute Break                                                            |
| 4:00 – 5:00 | **Practice essential Hadoop Commands**  
**Demo Pig/Hive**  
**Lab 2 (Practice basic Hadoop Commands)**  
**Lab 3 (Practice advanced Hadoop Commands)**     |

## Lunch (12:00 – 1:00)

## Day 2: Diving in deep with Hive

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 – 10:00</td>
<td>Review Session and Q&amp;A from Day 1</td>
</tr>
<tr>
<td>10:00 – 10:15</td>
<td>15 Minute Break</td>
</tr>
</tbody>
</table>
| 10:15 – 11:15 | **Introduction to Hive**  
**Demo - Basic Hive syntax**  
**Lab 1 – Practice Hive basics**     |
| 11:15 – 11:30 | 15 minute Break                                                            |
| 11:30 – 12:00 | **Hive Data Types**  
**Demo - Loading and Storing Data with Hive**     |
| 1:00 – 2:15 | **Lab 2 – Exploring Data using Hive**  
**Advanced Hive**  
**Defining relations and joins, performing analysis using Hive Query Lang.**     |
| 2:15 – 2:30 | 15 Minute Break                                                            |
| 2:30 – 3:30 | **Lab 3 – Practice Hive Joins**  
**Demo - Hive User Defined Functions**  
**HCatalog**     |
| 3:30 – 3:45 | 15 Minute Break                                                            |
| 3:45 – 4:40 | **Lab 4 – Process, analyze and export data using Hive**  
One –on –one help and Q&A session     |
| 4:40 – 5:00 | Feedback and review                                                       |