Hadoop: Introduction

page 1

Meet the expert: Barry Solomon has over 23 years of experience as a consultant. He has developed with Fortran, C, C , Visual Basic, Java, and Visual C#. His extensive database experience includes working with Microsoft Access, Microsoft SQL Server, MySQL, and Oracle. His expertise now includes working with big data, Hadoop in particular, and all of its attending ecosystems as the limitations have been exceeded in most modern database systems.

Prerequisites: This course assumes that the users have an understanding of working with databases and database systems. The user should also be familiar with syntax commands for Linux.

Runtime: 02:01:01

Course description: In this course we are going to look at the necessity of big data in today's world and how it fits into your organizations future. Then we will look at one big data framework in particular, Hadoop, as it is fully open source and driven by the community. We will examine some of the pieces that comprise Hadoop and demonstrate some of its functionality. There are so many use cases where big data can enhance your organizations competitive edge - analyzing social media, sensor data, click stream data, geographic analysis, emails, the list goes on. Hopefully you have a better understanding, not only of what big data and Hadoop are, but, more importantly, where they fit into your organizations structure and what they bring to the table.

Course outline:

Purpose of Big Data
- Introduction
- End of the Line
- OLTP and OLAP
- Storage
- Big Data as Supercomputer
- Scalability
- Hard Drives
- Parallelism
- Whose Data is it?
- Being Competitive and Relevant
- What is Big Data
- Variety, velocity and volume
- Leveraging and ROI
- Data Data Everywhere
- Throw it in the Lake of Data
- Summary

Use Cases
- Introduction
- Use Cases
- Real Time vs Batch Processing
- What About Databases
- OLTP and OLAP
- Appliances
- Mix and Match
- Schema on Write, on Read
- NoSQL
- Summary

Hadoop
- Introduction
- What do I get
- Hadoop
- File System
- MapReduce
- YARN
- Ecosystem
- Pig
- Hive
- Mahout and Oozie
- NoSQL
- Sqoop
- Ambari
- ZooKeeper
- The other pieces
- Tez
- Summary

Hadoop Demo
- Introduction
- Where do we go?
- Demo: Download
- Demo: Putty
- Demo: Web Interface
- Demo: Back to Putty
- Demo: PIG
- Demo: HIVE Table
- Demo: Ambari
- Demo: Query
- Summary