What Is Big Data?
- **Industry Needs and Solutions**
    - What's So *Big* About Big Data?
    - A Brief History of Hadoop
        - Google
        - Nutch
    - What Is Hadoop?
        - Derivative Works and Distributions
        - Hadoop Distributions
        - Core Hadoop Ecosystem
        - Important Apache Projects for Hadoop
            - Hive
            - Pig
            - SQOOP
            - HCatalog
            - HBase
            - Flume
            - Mahout
            - Ambari
            - Oozie
            - Zookeeper
        - The Future for Hadoop
- **Big Data in the Real World**
    - Common Industry Analytics
        - Telco
        - Energy
        - Retail

Setting Up for Big Data
- **Confi guring Your First Big Data Environment**
    - Getting Started
    - Getting the Install
    - Running the Installation
        - On-Premise Installation: Single-Node Installation
    - Validating Your New Cluster
        - Logging into HDInsight Service
        - Verify HDP Functionality in the Logs
    - Common Post-Setup Tasks
        - Loading Your First Files
        - Verifying Hive and Pig

Storing and Managing Big Data

- **HDFS, Hive, HBase, and HCatalog**
    - Exploring the Hadoop Distributed File System
    - Explaining the HDFS Architecture
        - Interacting with HDFS
    - Exploring Hive: The Hadoop Data Warehouse Platform
        - Designing, Building, and Loading Tables
        - Querying Data
        - Confi guring the Hive ODBC Driver
    - Exploring HCatalog: HDFS Table and Metadata Management
    - Exploring HBase: An HDFS Column-Oriented Database
        - Columnar Databases
        - Defi ning and Populating an HBase Table
        - Using Query Operations
- **Storing and Managing Data in HDFS**
    - Understanding the Fundamentals of HDFS

- HDFS Architecture
- NameNodes and DataNodes
- Data Replication
  - Using Common Commands to Interact with HDFS
    - Interfaces for Working with HDFS
    - File Manipulation Commands
    - Administrative Functions in HDFS
  - Moving and Organizing Data in HDFS
    - Moving Data in HDFS
    - Implementing Data Structures for Easier Management
    - Rebalancing Data

- **Adding Structure with Hive**
  - Understanding Hive's Purpose and Role
    - Providing Structure for Unstructured Data
    - Enabling Data Access and Transformation
    - Differentiating Hive from Traditional RDBMS Systems
    - Working with Hive
  - Creating and Querying Basic Tables
    - Creating Databases
    - Creating Tables
    - Adding and Deleting Data
    - Querying a Table
  - Using Advanced Data Structures with Hive
    - Setting Up Partitioned Tables
    - Loading Partitioned Tables
    - Using Views
    - Creating Indexes for Tables

- **Expanding Your Capability with HBase and HCatalog**
  - Using HBase
    - Creating HBase Tables
    - Loading Data into an HBase Table
    - Performing a Fast Lookup
    - Loading and Querying HBase
  - Managing Data with HCatalog
    - Working with HCatalog and Hive
    - Defi ning Data Structures
    - Creating Indexes
  - Creating Partitions
  - Integrating HCatalog with Pig and Hive
  - Using HBase or Hive as a Data Warehouse

Working with Your Big Data

- **Effective Big Data ETL with SSIS, Pig, and Sqoop**
  - Combining Big Data and SQL Server Tools for Better Solutions
    - Why Move the Data?
    - Transferring Data Between Hadoop and SQL Server
  - Working with SSIS and Hive
    - Connecting to Hive
  - Configuring Your Packages
    - Loading Data into Hadoop
    - Getting the Best Performance from SSIS
  - Transferring Data with Sqoop
    - Copying Data from SQL Server
    - Copying Data to SQL Server
  - Using Pig for Data Movement

- Transforming Data with Pig
- Using Pig and SSIS Together
  - o Choosing the Right Tool
    - Use Cases for SSIS
    - Use Cases for Pig
    - Use Cases for Sqoop

- **Data Research and Advanced Data Cleansing with Pig and Hive**
  - o Getting to Know Pig
    - When to Use Pig
    - Taking Advantage of Built-in Functions
    - Executing User-defined Functions
    - Using UDFs
  - o Using Hive
    - Data Analysis with Hive
    - Types of Hive Functions
    - Extending Hive with Map-reduce Scripts
    - Creating a Custom Map-reduce Script
- **Big Data Analytics**

  - o Data Science, Data Mining, and Predictive Analytics
    - Data Mining
    - Predictive Analytics
  - o Mahout
    - What is Apache Mahout?
    - What Mahout Does
      - Collaborative filtering
      - Clustering
      - Classification
      - Frequent itemset mining

- **Real Time Case Study**