Data science is the exploration of data via the scientific method to discover meaning or insight and the construction of software systems that utilize such meaning and insight in a business or scientific context. (note: nothing about clusters, Hadoop, Spark, etc.)
The Big “Loop”

Ask A Question → Evaluate Results → Implement The Analysis → Form a Hypothesis
The Actual DS Cycle

1. Ask A Question
2. Form a Hypothesis
3. Acquire Data
4. Clean Data
5. Explore Data and Design Features
6. Build Model
7. Evaluate/Visualize Results
8. Deploy and Implement The Analysis
Data Science with Hadoop and Spark

- Background on Data Science and Hadoop/Spark
- Data Ingest
- Data Visualization
- Data Science Methods
- Source code

http://www.clustermonkey.net/Practical-Data-Science-with-Hadoop-and-Spark
Areas of Data Science

Covered in Book

- Data Munging (preparation)
- Machine Learning with Hadoop
- Predictive Modeling
- Clustering
- Anomaly Detection with Hadoop
- Natural Language Processing