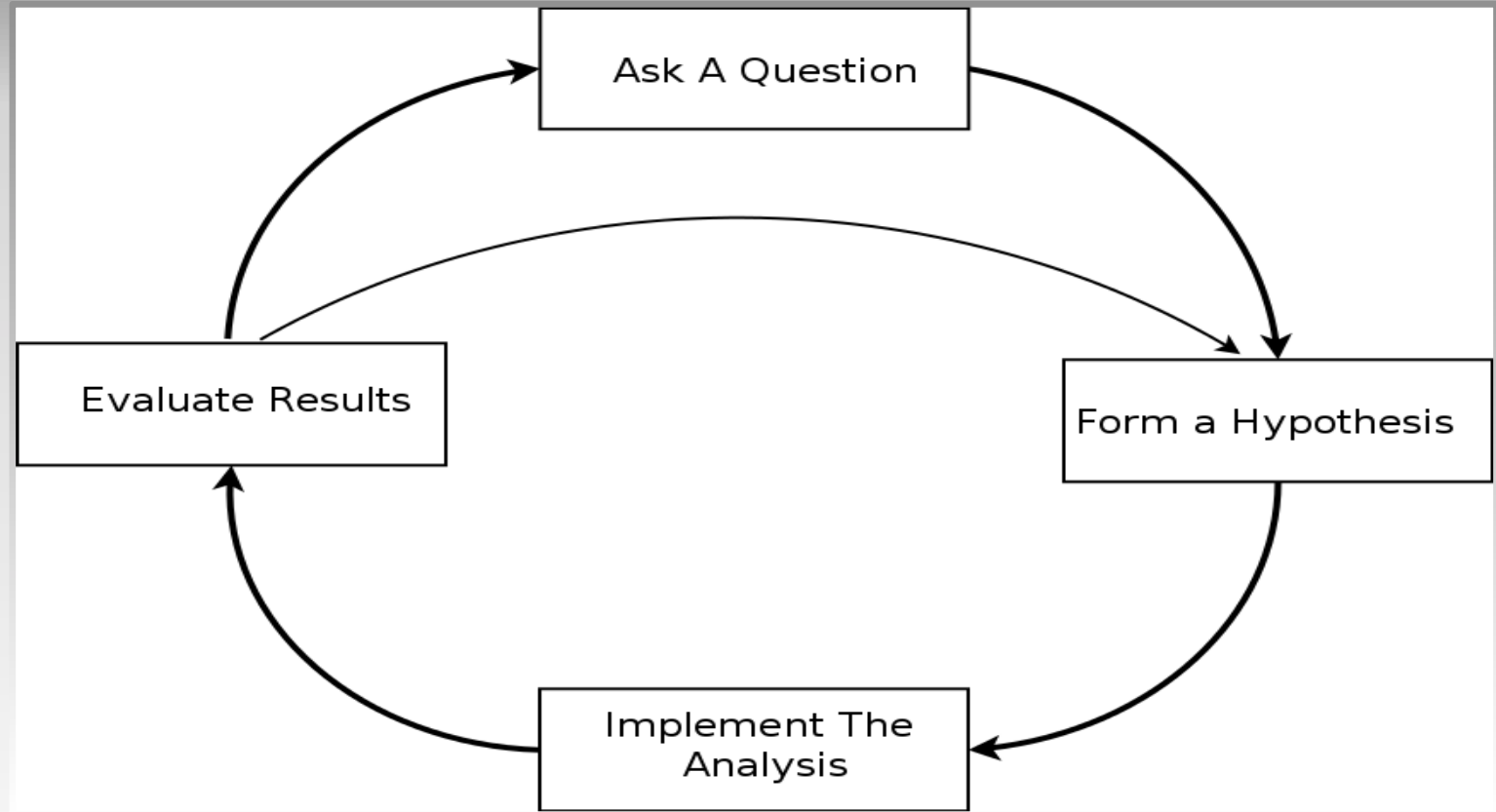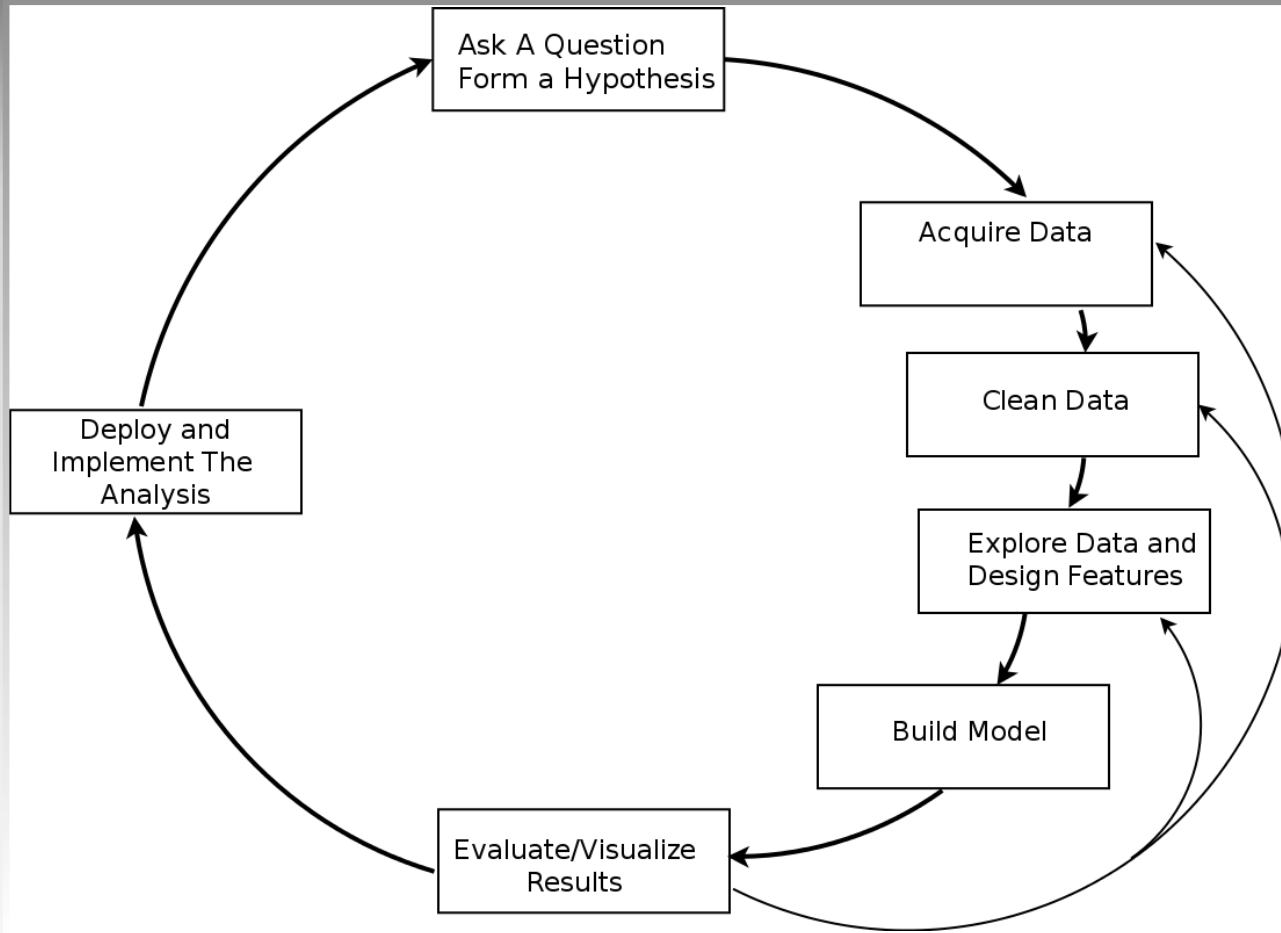# Data Science

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.)
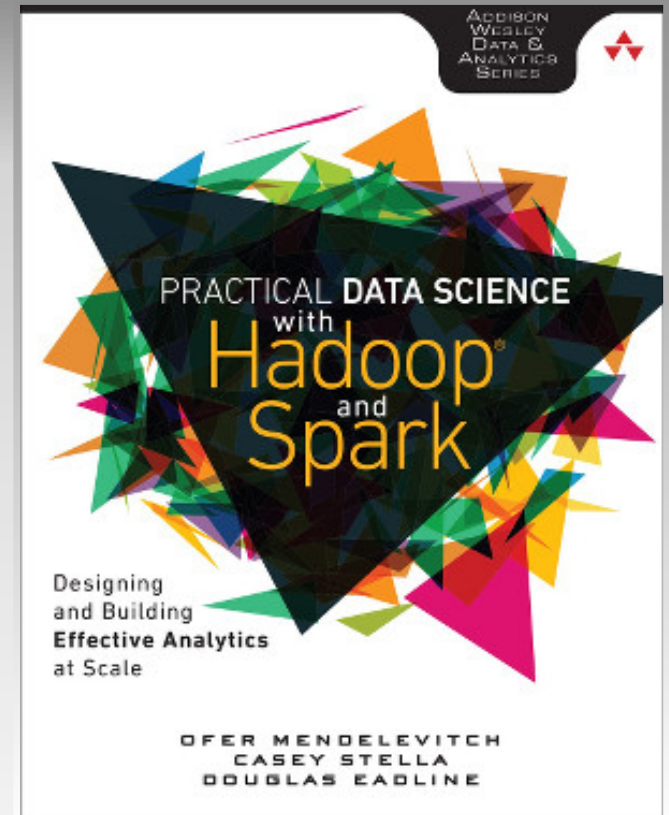
The Big "Loop"

The Actual DS Cycle

Data Science with Hadoop and Spark

- **Backgound on Data Science and Hadoop/Spark**

- **Data Injest**

- **Data Vizualization**

- **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**