

Project Related Links

Connecting HBASE and querying using Spark

1. <https://databricks.com/session/apache-spark-apache-hbase-connector-feature-rich-and-efficient-access-to-hbase-through-spark-sql>
2. <https://www.youtube.com/watch?v=ILKTzPWaB1I>
3. <https://www.youtube.com/watch?v=jJnc7vCQP9c>

Practical demo of real time analytics

1. <https://www.youtube.com/watch?v=Db1t6pG7a94>

Connecting Cassandra and Spark

1. <https://www.youtube.com/watch?v=pzbN5WdsyLs>
2. <https://www.youtube.com/watch?v=ex3cygoepLA>

HIVE and Spark

1. <https://www.youtube.com/watch?v=PFLIC9EWVKc>

Querying your data interactively- Options:

1. You can set up drill and can query data from multiple database
2. Phoenix can be used to integrate pig and query with HBase
3. Presto can be used to query both from Cassandra and Hive

Linear Regression with Map Reduce:

1. <https://dzone.com/articles/linear-regression-with-map-reduce>

Analyzing Streams of Data:

1. You can use data published either from Flume and Kafka and do spark streaming
2. Storm, Flink are other options to do stream analysis

Best Books for further Studies: Almost all O reilly publications are good

1. Advanced Analytics with Spark
2. Learning Spark
3. Hadoop Application Architectures
4. Elasticsearch: The Definitive guide
5. Using Flume
6. Hadoop: The Definitive guide
7. **Apache Zookeeper**
8. **Building Applications on Mesos**
9. **MangoDB-The Definitive guide**
10. **Programming Hive**
11. **Cassandra: The Definitive guide**
12. **HBase: The Definitive guide**
13. **Spark Documentation** apache.spark.org is enough for spark

14.