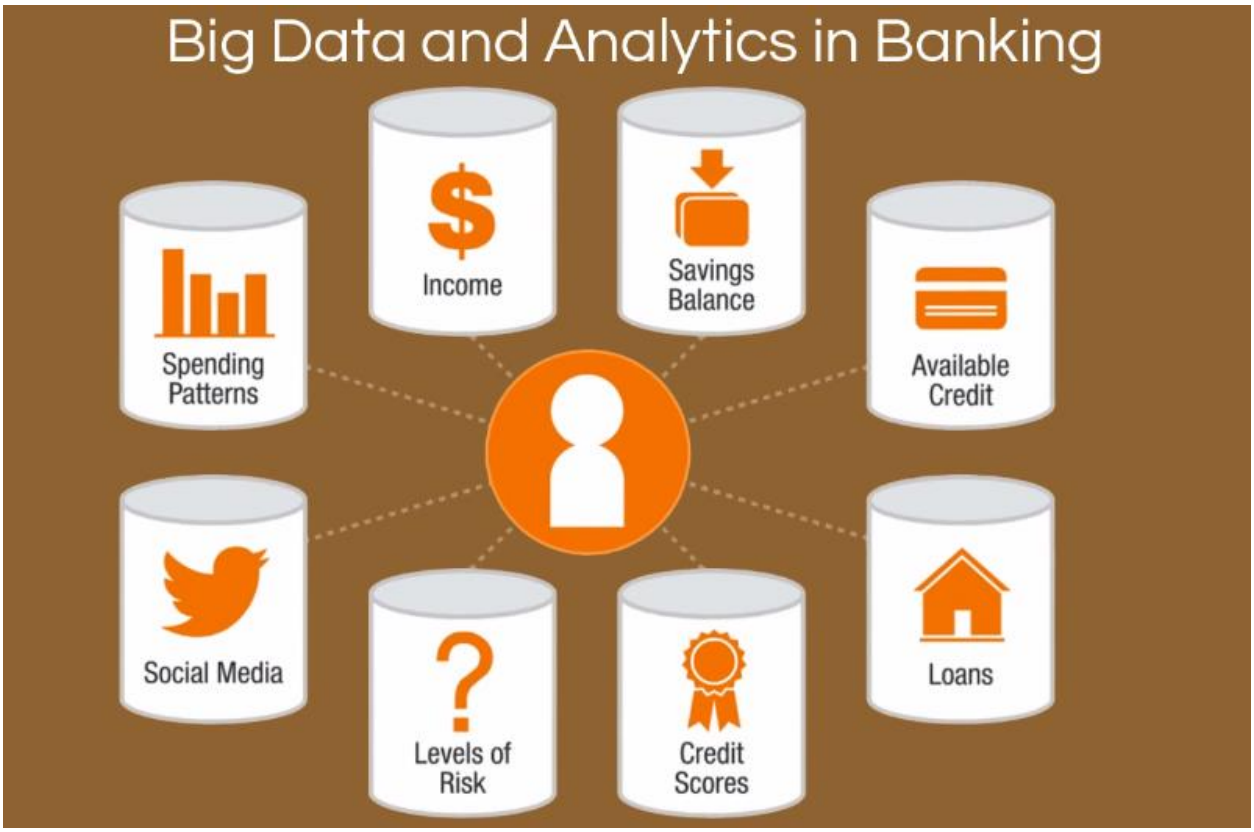
DR.MANIKANTAN'S COLLECTION LINKS FOR BIGDATA COMMUNITY FROM DATA BANK – HARD DISK 7

BIGATA PROJECT



Project Video : BANKING Data Analytics

https://youtu.be/iQrao1C7juk

Project Video : E-Commerce Data Analytics

https://youtu.be/ishBF1Xu8g8

Project Video : Music Data Analystics

https://youtu.be/8ewDcIeVPn8

Link 1: Aviation Data Analysis Using Apache Hive

Link 2: Aviation data analysis using Apache Pig

Learn to implement the scheduling of Hadoop Jobs using using Jenkins and Rundeck.

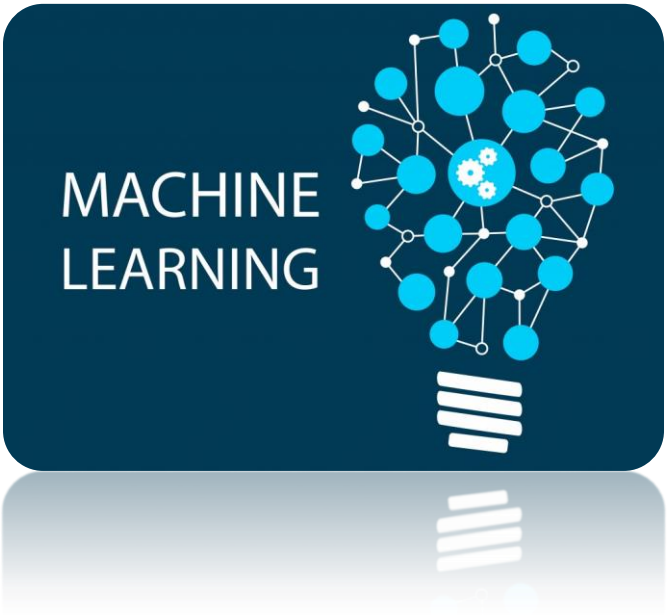
Link 3: Scheduling Hadoop Jobs using Jenkins.

Link 4: Scheduling Hadoop Jobs using Rundeck.

Since Machine Learning is heavily implemented along with Big Data technologies like Hadoop and spark,I am giving some use cases on machine learning with Spark.

http://bigdata.ices.utexas.edu/project/gene-disease/

http://bdgenomics.org/



Link 5: Machine Learning with Spark - Part 1

Link 6: Machine Learning with Spark - Part 2

Link 7: Machine Learning with Spark on Bank Use Case - Part 3

Link 8: Machine Learning with Spark - Part 4

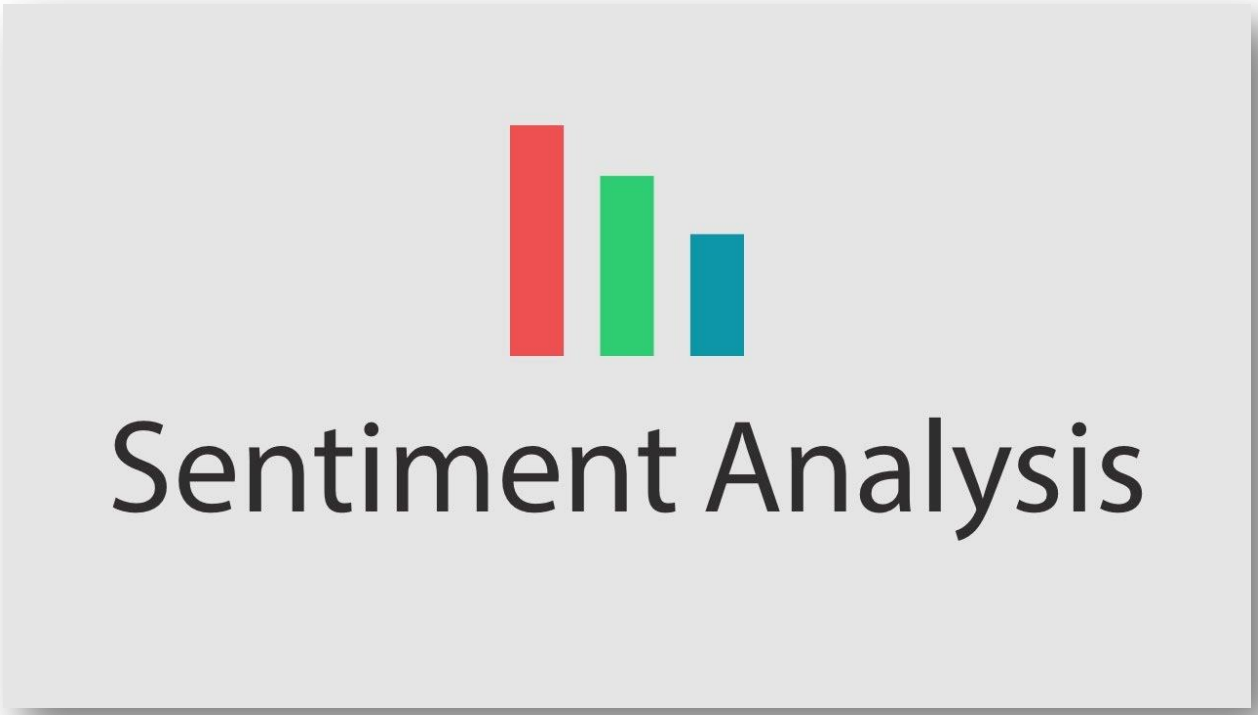
Link 9: Machine Learning with Spark - Part 5

Link 10: Distributed SQL engine for Big Data

Hadoop use cases in the below part of this answer.

Link 1: Spark Streaming and Kafka Integration

You can also perform sentiment analysis on demonetization using Apache Pig using below link:



Link 2: Sentiment Analysis on Demonetization – Pig Use Case

Also perform analysis on Aviation data using Apache Pig , Hive and Tableau.

Find the below links for small use cases on MapReduce in hadoop.

The below two links will help you to get hold on Map Reduce concepts:

Link 13: Map Reduce Use Case - Uber Data Analysis

Link 14: MapReduce Use Case-Youtube Data Analysis

Link 15: Map reduce Use case – Titanic Data Analysis

Below links are related to sentiment analysis using Hadoop's various components like Pig and Hive.

Link 16: Pig Use Case - Weblog Analysis

Link 17: Pig Use Case – The Daily Show Data Analysis Part – I

Link 18: Pig Use Case – The Daily Show Data Analysis Part – II

Link 19:Determining Popular Hashtags in Twitter Using Pig

Link 20: Sentiment Analysis on Twitter – TimeZone wise analysis

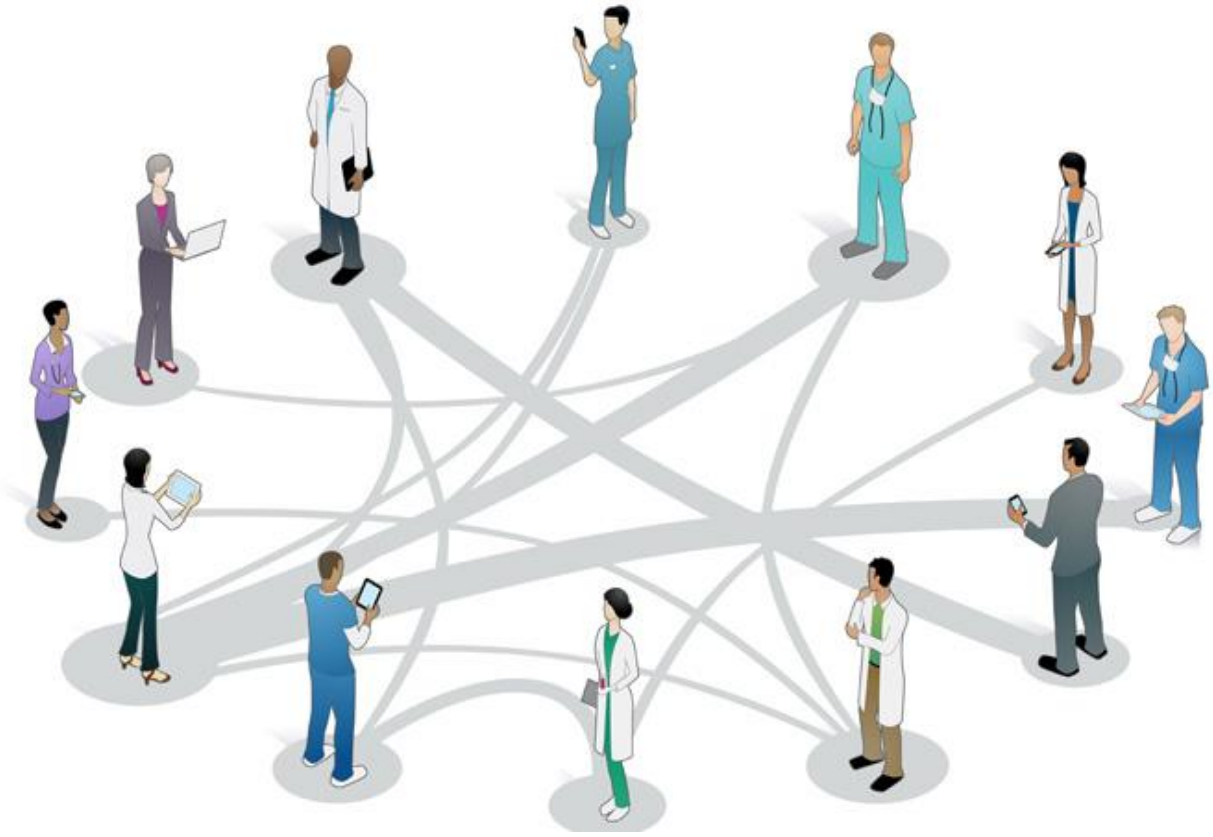
Link 21: Hive Use case – Counting Hashtags Using Hive

Link 22: Sentiment Analysis on Tweets with Apache Pig Using AFINN Dictionary

Link 23: Sentiment Analysis on Tweets with Apache Hive Using AFINN Dictionary

Link 24: Pokemon Data Analysis using Apache Hive

For beginner's level use cases in Spark , refer the below links:



Link 25: HealthCare Use Case With Apache Spark

Link 26: Introduction to Spark RDD and Basic Operations in RDD

Link 27: Analyzing New York Crime Data Using SparkSQL

Link 28: Spark Use Case – Travel Data Analysis

Link 29: Spark Use Case – Uber Data Analysis

Link 30: Spark Use Case – Analyzing MovieLens Dataset

Link 31: Spark Use Case – Social Media Analysis

Link 32: Spark SQL Use Case – 911 -Emergency Helpline Number Data Analysis

Link 33: Building Spam filter Engine using Spark.

Apart from these free links there are few paid programs which will help to get real exposure in Big Data Projects.

Link: Mastering Hadoop With Industrial Projects.

Learning Programs

MapReduce VS Spark:

- Aadhaar dataset analysis
- Inverted Index Example
- Secondary Sort Example
- Wordcount Example

Sentimental Analysis

- Spark Streaming part 1: Real time twitter sentiment analysis
- Spark streaming part 2: Real time twitter sentiment analysis using Flume
- Spark streaming part 3: Real time twitter sentiment analysis using kafka
- Data guarantees in Spark Streaming with Kafka integration

STROM and KAFKA

Using Storm to analyze twitter data and integration with kafka:

- Realtime stream processing using Apache Storm - Part 1
- Realtime stream processing using Apache Storm and Kafka - Part 2

SOURCE CODE

Also checkout my github repositories for the code:

Flight Project Source Code :
https://github.com/benedekh/bigdata-projects

- aadhaar-dataset-analysis
- inverted-index
- secondarysort
- wordcount
- stdatalabs/SparkTwitterStreamAnalysis
- stdatalabs/StormTwitterStreamAnalysis

Meet-Up Learning Collections for big data chapter

- Sep 12, 2017
  Introduction to Spark Structured Streaming - Part 11 : Event Time
- Sep 7, 2017
  Introduction to Spark Structured Streaming - Part 10 : Ingestion Time
- Sep 1, 2017
  Introduction to Spark Structured Streaming - Part 9 : Processing Time Window
- Aug 31, 2017
  Introduction to Spark Structured Streaming - Part 8 : Time Abstraction
- Aug 18, 2017
  Introduction to Spark Structured Streaming - Part 7 : Checkpointing State
- Aug 16, 2017
  Introduction to Spark Structured Streaming - Part 6 : Stream Enrichment using Static Data Join
- Aug 11, 2017
  Introduction to Spark Structured Streaming - Part 5 : File Streams
- Aug 8, 2017
  Introduction to Spark Structured Streaming - Part 4 : Stateless Aggregations
- Aug 6, 2017
  Introduction to Spark Structured Streaming - Part 3 : Stateful WordCount
- Aug 1, 2017
  Introduction to Spark Structured Streaming - Part 2 : Source and Sinks
- Jul 25, 2017
  Introduction to Spark Structured Streaming - Part 1 : DataFrame Abstraction to Stream
- Jul 11, 2017
  Migrating to Spark 2.0 - Part 10 : Second Meetup Talk
- Jun 23, 2017
  Migrating to Spark 2.0 - Part 9 : Hive Integration
- Jun 20, 2017
  Migrating to Spark 2.0 - Part 8 : Catalog API

- Jun 14, 2017
  Migrating to Spark 2.0 - Part 7 : SubQueries
- Jun 9, 2017
  Migrating to Spark 2.0 - Part 6 : Spark ML Transformer API
- Jun 7, 2017
  Migrating to Spark 2.0 - Part 5 : Meetup Talk
- May 10, 2017
  Migrating to Spark 2.0 - Part 4 : Cross Joins
- May 8, 2017
  Migrating to Spark 2.0 - Part 3 : DataFrame to Dataset
- May 3, 2017
  Scalable Spark Deployment using Kubernetes - Part 9 : Service Update and Rollback
- May 2, 2017
  Scalable Spark Deployment using Kubernetes - Part 8 : Meetup Talk
- Apr 15, 2017
  Migrating to Spark 2.0 - Part 2 : Built-in CSV Connector
- Apr 13, 2017
  Migrating to Spark 2.0 - Part 1 : Scala Version and Dependencies
- Mar 6, 2017
  Scalable Spark Deployment using Kubernetes - Part 7 : Dynamic Scaling and Namespaces
- Feb 26, 2017
  Scalable Spark Deployment using Kubernetes - Part 6 : Building Spark 2.0 Two Node Cluster
- Feb 26, 2017
  Scalable Spark Deployment using Kubernetes - Part 5 : Building Spark 2.0 Docker Image
- Feb 23, 2017
  Scalable Spark Deployment using Kubernetes - Part 4 : Service Abstractions
- Feb 17, 2017
  Scalable Spark Deployment using Kubernetes - Part 3 : Kubernetes Abstractions
- Feb 15, 2017
  Scalable Spark Deployment using Kubernetes - Part 2 : Installing Kubernetes Locally using Minikube
- Feb 13, 2017

Scalable Spark Deployment using Kubernetes - Part 1 : Introduction to Kubernetes

- Nov 22, 2016
  Statistical Data Exploration using Spark 2.0 - Part 3 : Outlier Detection using Quantiles
- Oct 22, 2016
  Statistical Data Exploration using Spark 2.0 - Part 2 : Shape of Data with Histograms
- Oct 21, 2016
  Statistical Data Exploration using Spark 2.0 - Part 1 : Five Number Summary
- Sep 25, 2016
  Interactive Workflow Management using Azkaban : API Driven Workflow Management for Spark
- Aug 30, 2016
  Anatomy of Spark Catalyst - Part 2 : Meetup Talk
- Aug 25, 2016
  Functional Programming in Rust - Part 2 : Functional Combinators
- Aug 24, 2016
  Functional Programming in Rust - Part 1 : Function Abstraction
- Aug 5, 2016
  Anatomy of Spark Catalyst - Part 1 : Meetup Talk
- Jul 3, 2016
  Introduction to Spark 2.0 - Part 7 : Meetup Talk on Spark 2.0 API
- Jul 3, 2016
  Evolution of Apache Spark : Journey of Spark in 1.x Series
- May 20, 2016
  Introduction to Spark 2.0 - Part 6 : Custom Optimizers in Spark SQL
- May 19, 2016
  Introduction to Spark 2.0 - Part 5 : Time Window in Spark SQL
- May 17, 2016
  Introduction to Spark 2.0 - Part 4 : Introduction to Catalog API
- May 11, 2016
  Introduction to Spark 2.0 - Part 3 : Porting Code from RDD API to Dataset API
- May 11, 2016
  Introduction to Spark 2.0 - Part 2 : Wordcount in Dataset API
- May 11, 2016
  Introduction to Spark 2.0 - Part 1 : Spark Session API

- May 5, 2016
  Introduction to Flink Streaming - Part 10 : Meetup Talk
- Apr 28, 2016
  Introduction to Flink Streaming - Part 9 : Event Time in Flink
- Apr 27, 2016
  Introduction to Flink Streaming - Part 8 : Understanding Time in Flink Streaming
- Apr 17, 2016
  Apache Beam : Next Step in Big Data Unification
- Apr 6, 2016
  Introduction to Flink Streaming - Part 7 : Implementing Session Windows using Custom Trigger
- Apr 5, 2016
  Introduction to Flink Streaming - Part 6 : Anatomy of Window API
- Mar 14, 2016
  Introduction to Flink Streaming - Part 5 : Window API in Flink
- Mar 12, 2016
  Introduction to Flink Streaming - Part 4 : Understanding Flink's Advanced Stream Processing using Google Cloud Dataflow
- Mar 10, 2016
  Introduction to Flink Streaming - Part 3 : Running Streaming Applications in Flink Local Mode
- Mar 8, 2016
  Introduction to Flink Streaming - Part 2 : Discretization of Stream using Window API
- Mar 7, 2016
  Introduction to Flink Streaming - Part 1 : WordCount
- Mar 7, 2016
  Interactive Scheduling using Azkaban - Part 2 : Challenges in scheduling interactive workloads
- Mar 5, 2016
  What's New in Spark : Tales from Spark Summit East - Framework Improvements
- Mar 4, 2016
  Introduction to Spark 2.0 : A Sneak Peek At Next Generation Spark
- Mar 3, 2016
  Interactive Scheduling using Azkaban - Part 1 : Setting up Solo Server
- Feb 17, 2016

Building Distributed Systems from Scratch - Part 2 : Handling third party libraries

- Feb 7, 2016

Introduction to Hadoop (HDFS & Map/Reduce) for Spark developers

- Jan 11, 2016

Introduction to Apache Flink - Meetup talk

- Dec 6, 2015

Introduction to Apache Flink for Spark Developers : Flink vs Spark

- Dec 2, 2015

Building Distributed Systems from Scratch - Part 1

- Nov 20, 2015

Akka HTTP testing

- Nov 13, 2015

JSON in Akka HTTP

- Nov 13, 2015

Akka HTTP Hello world

- Sep 19, 2015

Introduction to Machine learning with Spark

- Aug 5, 2015

Improving Mobile payments with Real time Spark

- Aug 5, 2015

Anatomy of Data Frame API : Deep dive into Spark SQL Data Frame API

- Jun 30, 2015

Anatomy of Data Source API : Deep dive into Spark SQL Data Source API

- Jun 8, 2015

Structured data processing with Spark SQL - Meetup Video

- May 26, 2015

Analysing CSV data in Spark : Introduction to Spark Data Source API - Part 2

- May 25, 2015

Introduction to Spark Data Source API - Part 1

- May 1, 2015

An Introduction to Spark Streaming- Meetup Video

- Apr 6, 2015

Handling empty batches in Spark streaming

- Mar 31, 2015

Anatomy of RDD : Deep dive into spark RDD abstraction - Meetup video

- Mar 11, 2015

Extending Spark API

- Feb 26, 2015

Introduction to Apache Spark - Meetup video

- Feb 26, 2015

Apache Spark is not a one-trick pony : Going beyond in-memory processing

- Feb 7, 2015

Pipe in Spark

- Jan 6, 2015

A Simple Akka Remote example

- Jan 5, 2015

Building Read it later service on MEAN stack - Part 2

- Jan 4, 2015

Building Read it later service on MEAN stack - Part 1

- Jan 3, 2015

In Pursuit of the Unknown : 17 equations that changed the world - book review

- Jan 2, 2015

History of Apache Spark : Journey from Academia to Industry

- Dec 22, 2014

Running scala programs on YARN

- Dec 15, 2014

Implementing shuffle in Mesos

- Dec 15, 2014

Distributing third party libraries in Mesos

- Dec 3, 2014

sizeof operator for Java/Scala

- Nov 18, 2014

Kryo disk serialization in Spark

- Nov 16, 2014

Functional programming in C++

- Nov 11, 2014

Custom mesos executor in Scala

- Nov 4, 2014

Granular resource management in Mesos

- Oct 26, 2014

Mesos Hello world in Scala

- Oct 24, 2014

Mesos single node cluster on Ubuntu

- Sep 29, 2014

Evaluating Spark RDD's for side effects

- Sep 26, 2014

Boot custom recovery without flashing

- Aug 18, 2014

Sbt on ubuntu

- Aug 3, 2014

Google, it's time - We want Scala for Android

- Jul 30, 2014

Fold in spark

- Jul 29, 2014

Dog fight - How Apple and Google Went to War and Started a Revolution

- Jun 29, 2014

Converting Matlab file to Spark RDD

- Jun 27, 2014

Glom in spark

Before Starting Hadoop Project – Analyze first in internet what are the common project are available ?

1) https://www.javaworld.com/article/2972303/big-data/the-7-most-common-hadoop-and-spark-projects.html

Use CASE

Find the below links for small use cases on MapReduce in hadoop.

The below two links will help you to get hold on Map Reduce concepts:

Link 1: MapReduce Use Case-Youtube Data Analysis

Link 2: Map reduce Use case – Titanic Data Analysis

Below links are related to sentiment analysis using Hadoop's various components like Pig and Hive.

Link 3: Sentiment Analysis on Twitter – TimeZone wise analysis

Link 4: Hive Use case – Counting Hashtags Using Hive

Link 5: Sentiment Analysis on Tweets with Apache Pig Using AFINN Dictionary

Link 6: Sentiment Analysis on Tweets with Apache Hive Using AFINN Dictionary

For beginner level case study and basic concepts related to Spark RDD, refer the below links:

Link 7: HealthCare Use Case With Apache Spark

Link 8: Introduction to Spark RDD and Basic Operations in RDD

Useful links for AI :

http://www.jst.go.jp/erato/kawarabayashi/english/index.html

https://www.ngdata.com/big-data-analysis-resources/

http://www.datamation.com/applications/hadoop-and-big-data-60-top-open-source-tools-1.html

https://adtmag.com/blogs/dev-watch/2016/05/asf-big-data-projects.aspx