Overview of The Hadoop Ecosystem and Big Data

Introduction

This document presents an overview of the Hadoop Ecosystem and Big Data. An attempt is made to cover all elements of the Hadoop Ecosystem and related technologies that apply to the Big Data space. There are some elements presented that, while, technically, are not part of the Hadoop Ecosystem, they relate to it and are utilized in conjunction to Hadoop and Big Data.

Note: At this time, this document is evolving and as such should be viewed as incomplete and "under construction. Every effort will be made to add to the document in a very frequent manner until it reaches a complete state (which may be never, as the Hadoop Ecosystem and Big Data Technologies are evolving and appear to be continually in a state of flux.

Hadoop

What is Hadoop?

A Little History of Hadoop

Motivation for Hadoop

The HDFS

Map Reduce 2 (Java Api)

Map Reduce With Other Languages

Spark

Batch Processing

Stream Processing

Spark Shell Languages

Python

Scala

Spark SQL

Integration With MongoDb

Pig

Hive

HBase

Sqoop

Kafka

Storm

Flume

Spark