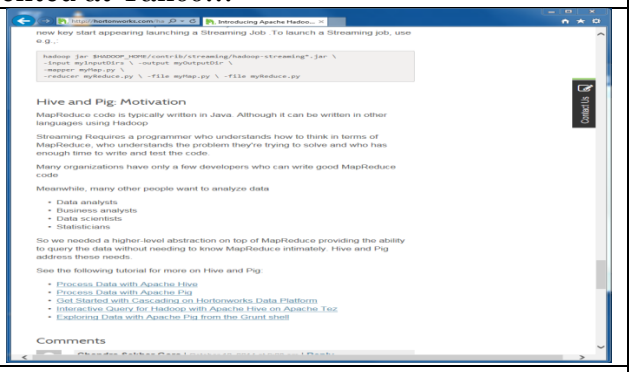
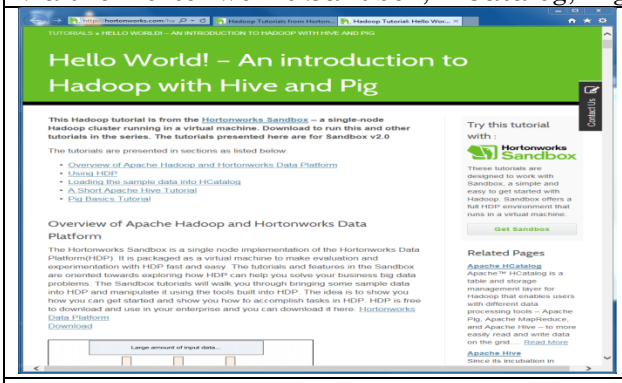
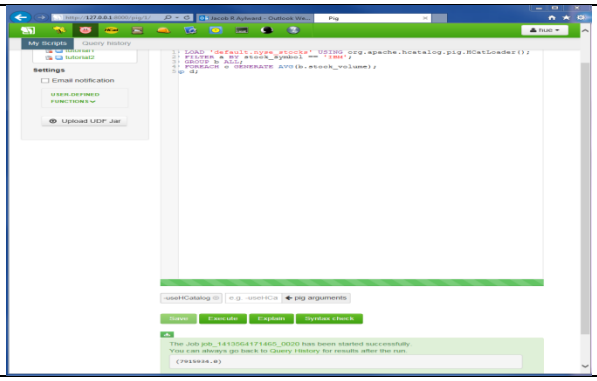
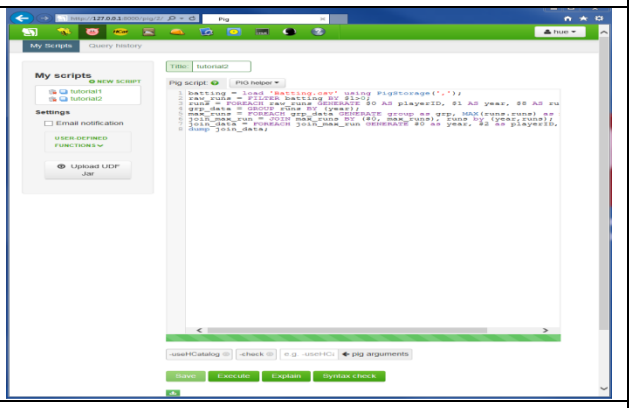
OCTOBER 27, 2014

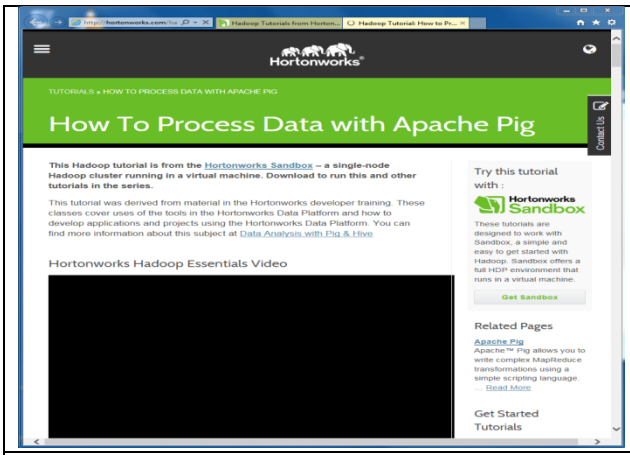
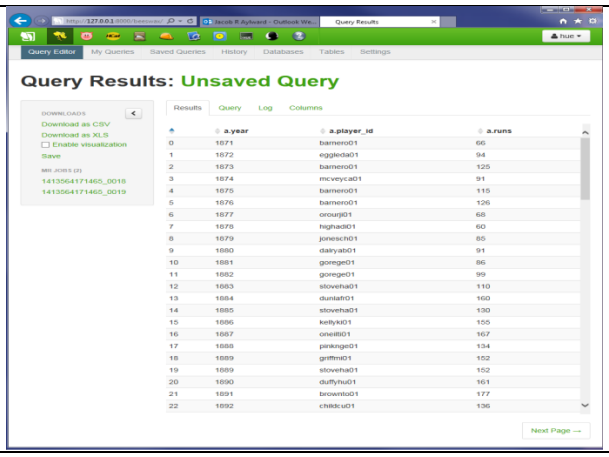
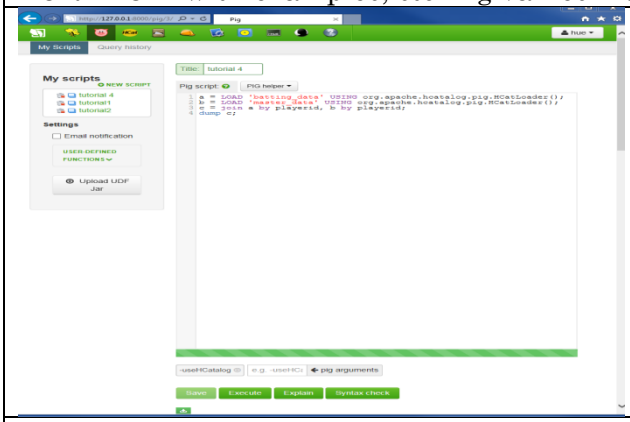
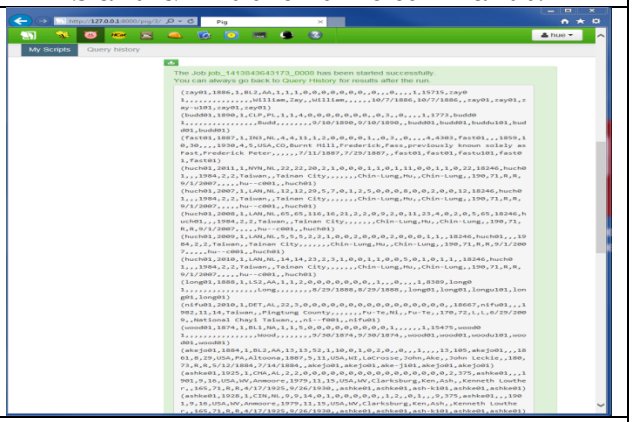
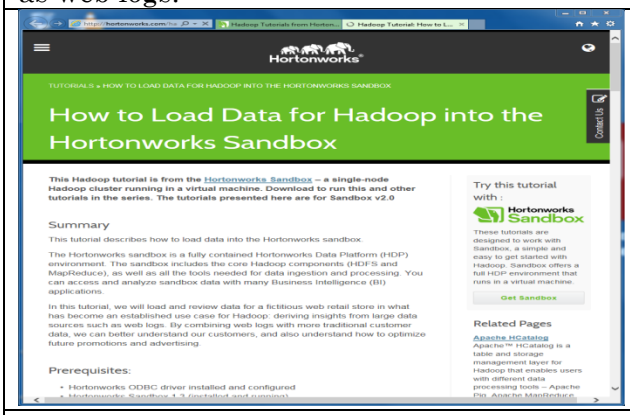
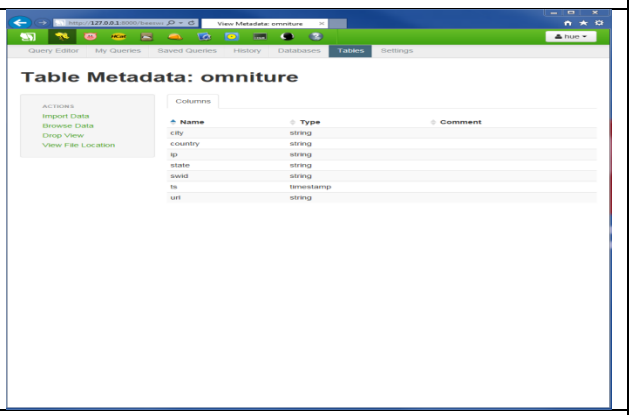
JACOB AYLWARD

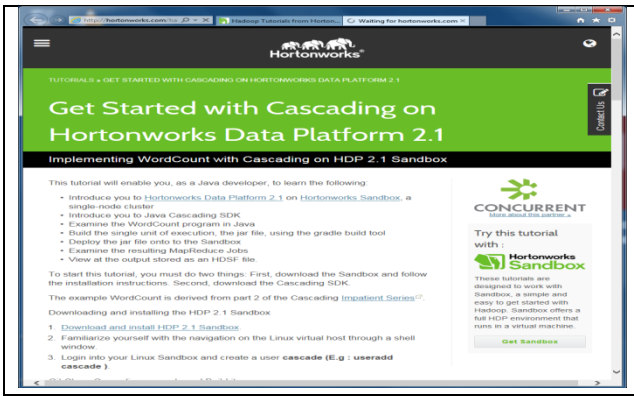
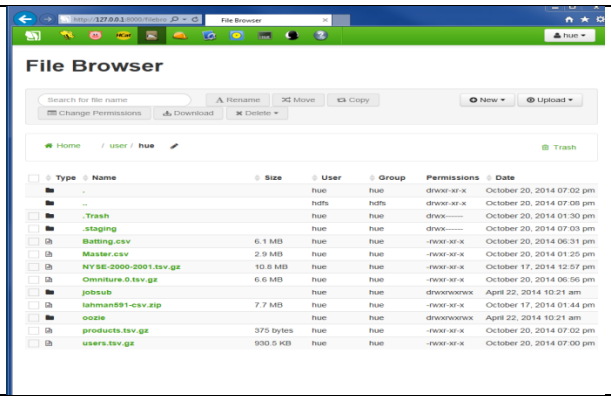
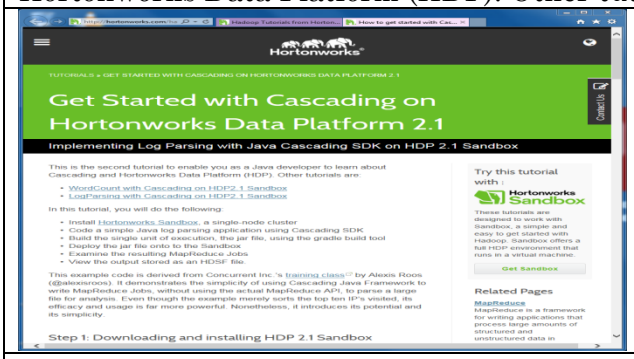
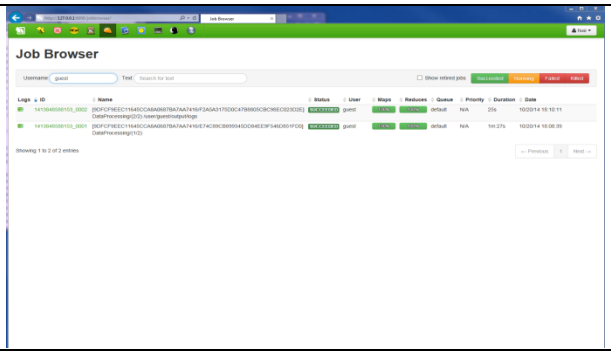
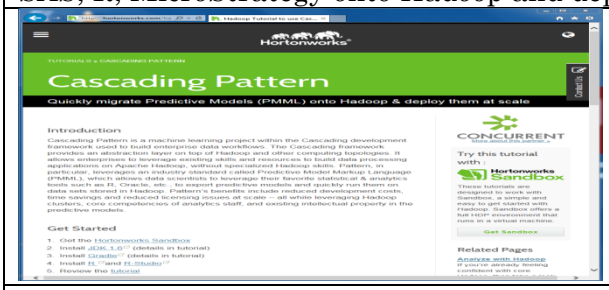
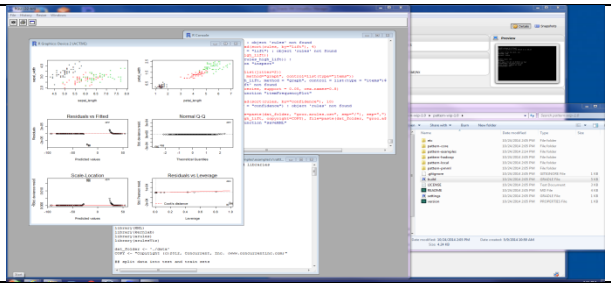
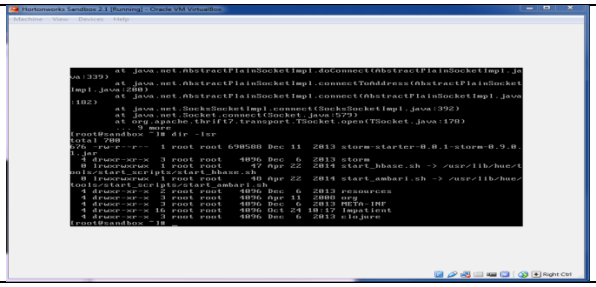
HORTONWORKS TUTORIAL

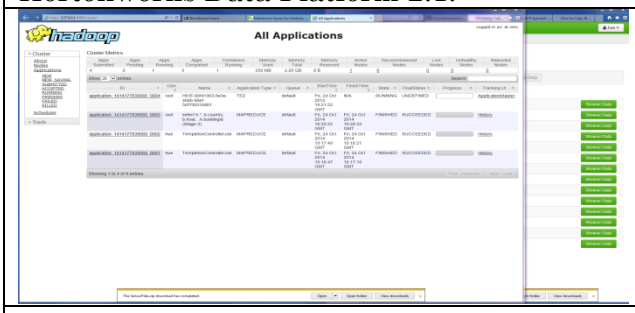
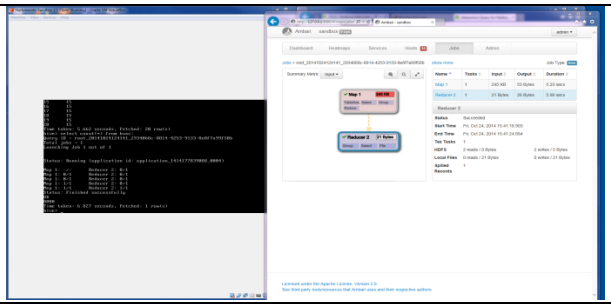
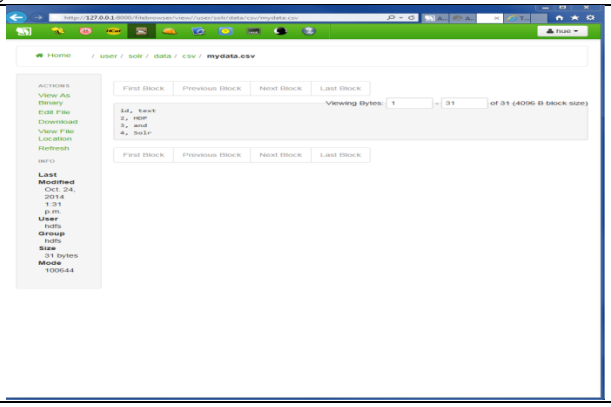
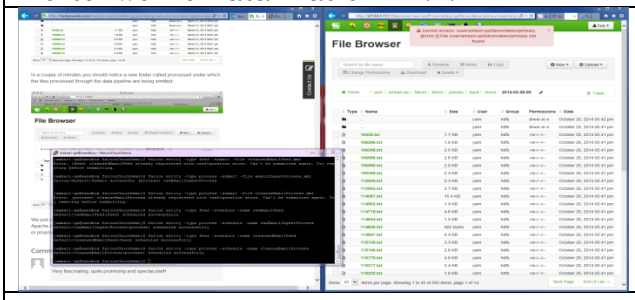
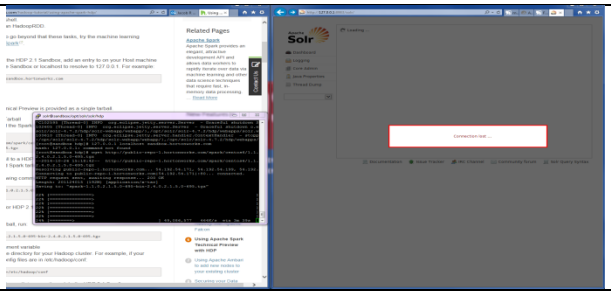
JACOB R. AYLWARD
FROSTBURG STATE UNIVERSITY
101 Braddock Rd, Frostburg, MD 21532

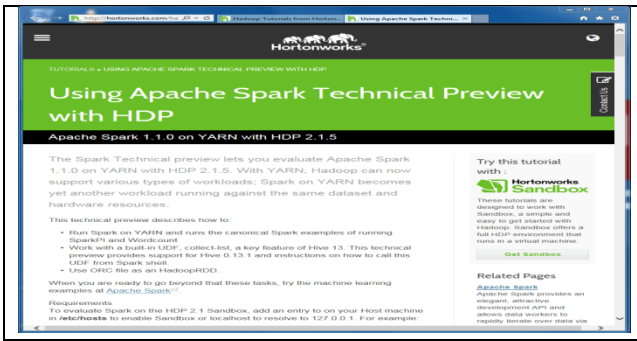
Jacob Aylward

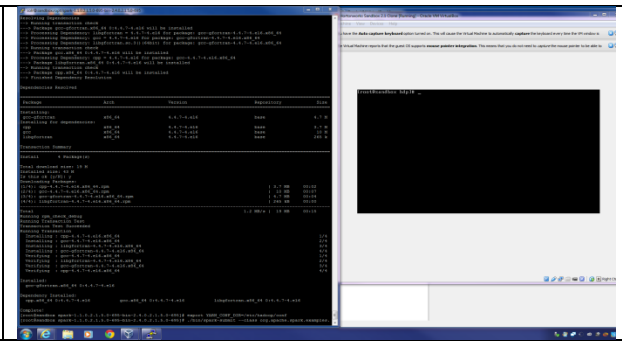
Hello World Tutorials:	
Tutorial 1	Introducing Apache Hadoop to Developers
Apache Hadoop is a community driven open-source project governed by the Apache Software Foundation. It was originally implemented at Yahoo…	
	
Tutorial 2	Hello World! – An introduction to Hadoop with Hive and Pig
This Hadoop tutorial provides a short introduction into working with big data in Hadoop via the Hortonworks Sandbox, HCatalog, Pig and Hive.	
	
Tutorial 3	How to Process Data with Apache Hive
This Hadoop tutorial shows how to process Data with Hive using a set of Baseball statistics on American players from 1871-2011.	
	
Tutorial 4	How to Process Data with Apache Pig
This Hadoop tutorial shows how to process Data with Apache Pig using a set of Baseball statistics on American Players from 1871-2011.	

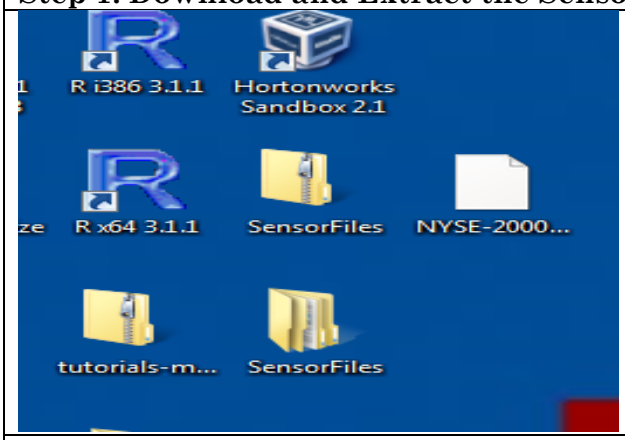
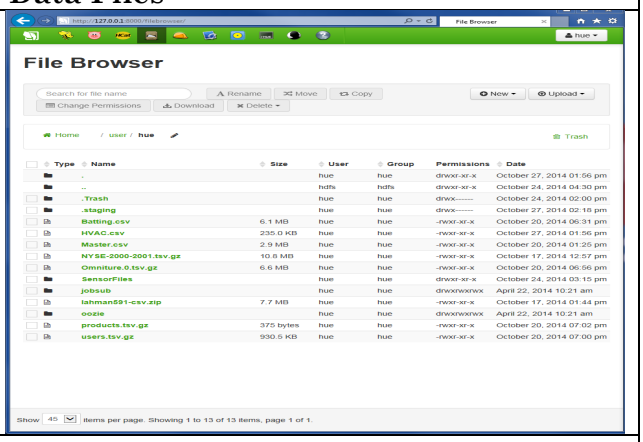
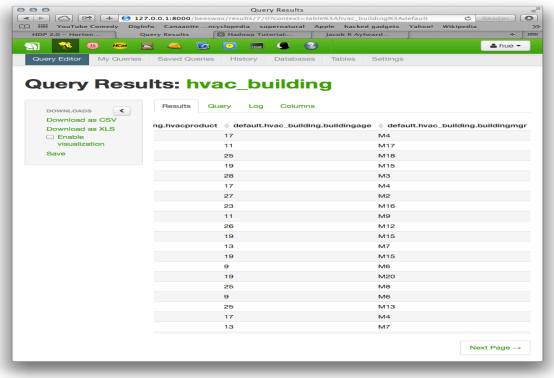
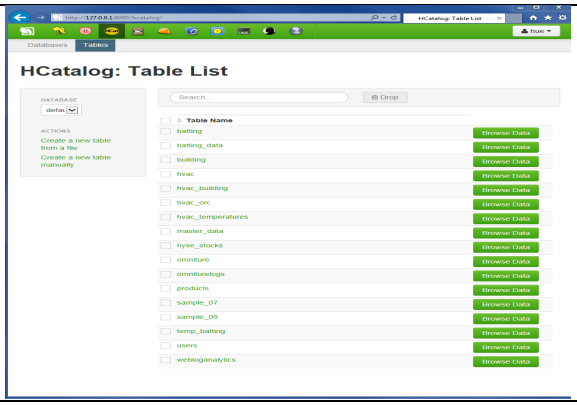
<table>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Tutorial 5</td><td>Exploring Data with Apache Pig from the Grunt shell</td></tr>
<tr><td colspan="2">In thus tutorial, you will learn how to load a data file into HDFS; Learn about ‘FILTER, FOREACH’ with examples; storing values into HDFS and Grunt shell’s file commands.</td></tr>
<tr><td> 
</td><td>  </td></tr>
<tr><td>Tutorial 6</td><td>How to Load Data for Hadoop into the Hortonworks Sandbox</td></tr>
<tr><td colspan="2">In this tutorial, we will load and review data for a fictitious web retail store in what has become an established use cause for Hadoop: deriving insight from large data sources such as web logs.</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Tutorial 7</td><td>Get Started with Cascading on Hortonworks Data Platform 2.1</td></tr>
<tr><td colspan="2">How to get started with Cascading and Hortonworks Data Platform using the Word Count Example.</td></tr>
</table>

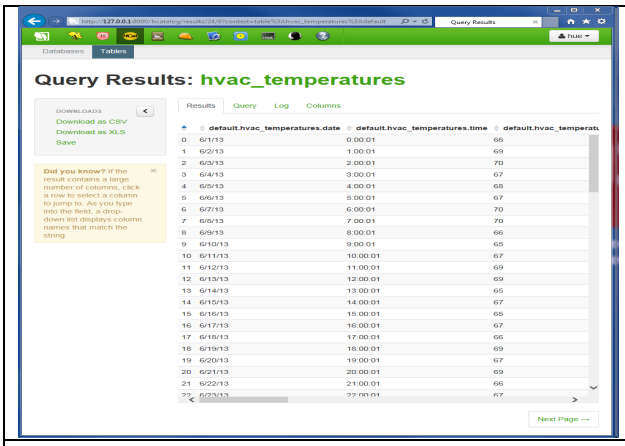
	
Tutorial 8	Get Started with Cascading on Hortonworks Data Platform 2.1
This is the second tutorial to enable you as a Java developer to learn about Cascading and Hortonworks Data Platform (HDP). Other tutorials are:…	
	
Tutorial 9	Cascading Pattern
Learn how to use Cascading Pattern to quickly migrate Predictive Models (PMML) from SAS, R, MicroStrategy onto Hadoop and deploy them at scale.	
	
Tutorial 10	Processing streaming data in Hadoop with Apache Storm
How to use Apache Storm to process real-time streaming data in Hadoop with Hortonworks Data Platform.	
	
Tutorial 11	Interactive Query for Hadoop with Apache

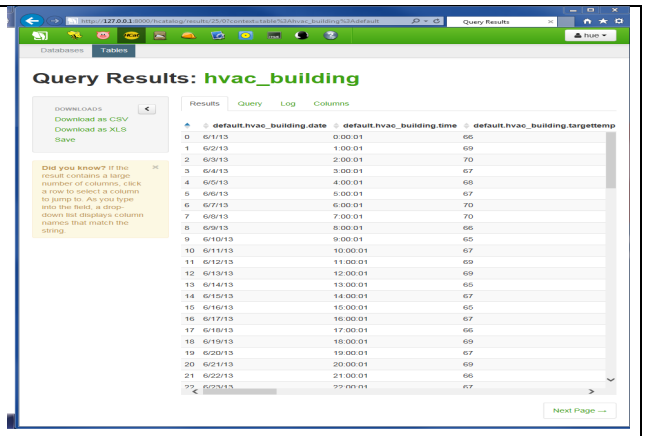
	Hive on Apache Tez
How to use Apache Tez and Apache Hive for Interactive Query with Hadoop and Hortonworks Data Platform 2.1.	
	
Tutorial 12	Searching Data with Apache Solr
In this tutorial we will walk through how to run Solr in Hadoop with the index (solr data files) stored on HDFS and using a map reduce jobs to index files.	
	
Tutorial 13	Define and Process Data Pipelines in Hadoop with Apache Falcon
Use Apache Falcon to define an end-to-end data pipeline and policy for Hadoop and Hortonworks Data Platform 2.1.	
	
Tutorial 14	Using Apache Spark Technical Preview with HDP
This guide describes how to run Spark on YARN. It also provides the canonical examples of running SparkPL and Wordcount with Spark shell.	



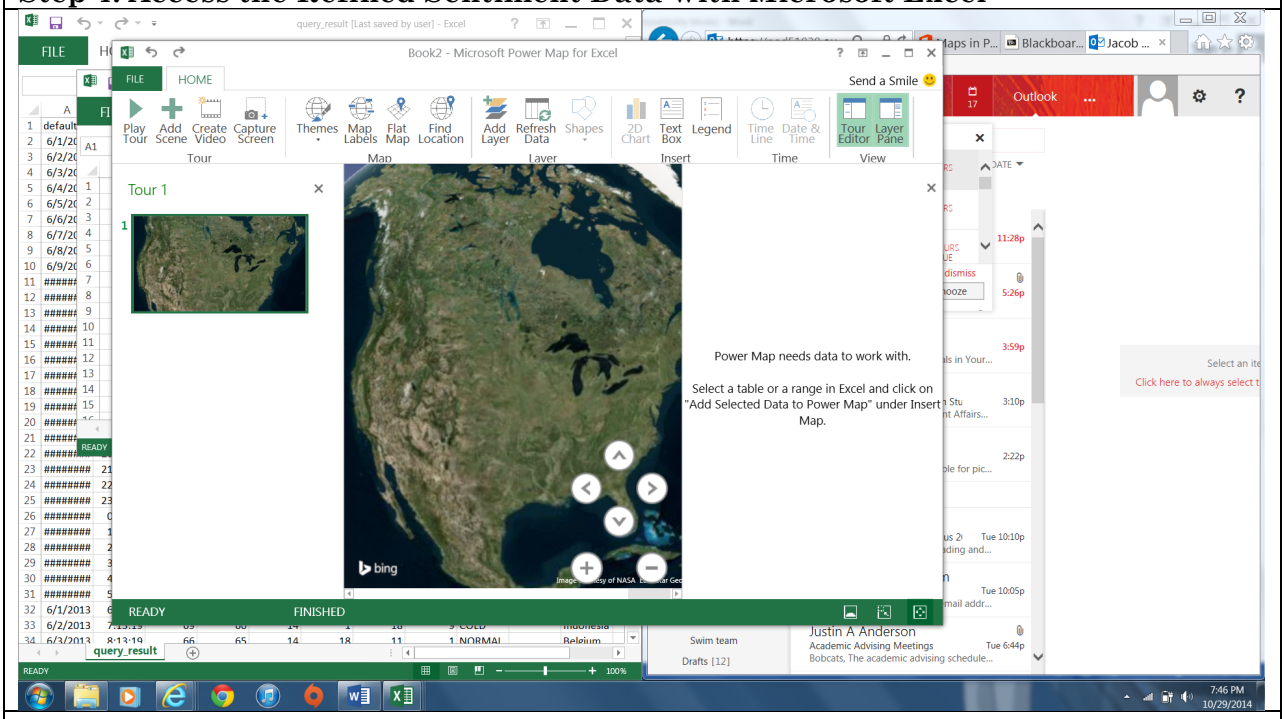


Real World Examples:	
Example 7	How to Analyze Machine and Sensor Data
The tutorial describes how to refine data from heating, ventilation, and air conditioning (HVAC) systems using the Hortonworks Data Platform, and how to analyze the refined sensor data to maintain optimal building temperatures.	
Step 1: Download and Extract the Sensor Data Files	
	
Step 2: Load the Sensor Data into the Hortonworks Sandbox	
	
Step 3: Run Two Hive Scripts to Refine the Sensor Data	





Step 4: Access the Refined Sentiment Data with Microsoft Excel



Step 5: Visualize the Sensor Data Using Excel Power View