Real Time Search and Analytics on Big Data

About the READMEs

Markdown (.md) is a human readable text format often used in version control software like GitHub. MarkDown easily converts to HTML and PDF so users can choose which format their prefer. Several text editors like Sublime Text 2 and TextMate have markdown bundles in which users can use to view a color formatted version of the files. If you do not have one of these text editors you can always use your OS default text editor to open the MarkDown files, use a browser to view the HTML files, or a PDF viewer to view the PDF files.

Introduction

Welcome to the Real Time Search and Analytics on Big Data tutorial. In this tutorial you will learn about Lucene, Solr, and how these search technologies can be used on big data in real time. Specifically we will cover:

* Lucene Indexing and Querying
* Running Solr Locally
* Using the Solr Administrative UI
* Understanding the Solr Schema
* Indexing Documents via SolrJ
* Translating SQL to Solr Queries
* Real Time Indexing

Tips

* If you lose track or fall behind with the tutorial, read the READMEs! You should be able to complete the exercises with the instructions provided.
* If the classroom is large, it will be tough for the instructor to provide any kind of 1 on 1 help. Please wait for a break to ask questions, or quietly ask a neighbor.
* If you are running Windows, I would suggest you set up an Ubuntu 12.04 LTS (http://www.ubuntu.com/download/desktop) virtual machine with VirtualBox (https://www.virtualbox.org/wiki/Downloads).

Prerequisites

Setup

I would recommend the user has the following

* Internet connectivity will be required to build the Maven projects as Maven will automatically download any dependencies specified in the POM file.
* Linux based or Mac OS X Operating System. Windows users can use Cygwin.
* Eclipse IDE for J2EE (not required, but helpful to browse code and write JUnit tests). I use Eclipse Juno but older versions should work.
* Maven2Eclipse Plugin for Eclipse (not required, but aids in importing/building Maven projects)
* Maven version 3 or higher (required to build shaded executable jars)

Knowledge

Students should have some unix knowledge (basic commands like cd, ls, mkdir, etc). It is helpful but not required to know Java and SQL.

To Download Eclipse

* Navigate your web browser to http://www.eclipse.org/downloads/
* Download the version that corresponds with your system. I use the Eclipse IDE for Java EE Developers
* Installation instructions may vary per your OS. For Linux and Mac OS X the download will include an executable called eclipse that you can run directly from the download folder (or if you would prefer you can move the download into an applications directory).

To install the M2E Plugin

* Eclipse -> Help -> Install New Software
* You will see the following page
* Select add next to the work with box
* Name: Whatever you would like to name the plugin
* Location: http://download.eclipse.org/technology/m2e/releases
* Select OK.
* Check the box that comes up, and press next.
* Continue to press next, accept agreement, etc until the plugin is installed.

To import the project in Eclipse with the M2E Plugin

* File -> Import
* Maven -> Existing Maven Project
* Browse
* Find github directory on your computer -> open
* Select all projects -> finish
* Pat yourself on the back, all of the projects are imported

To Install Maven 3.0

* Extract the distribution archive, i.e. apache-maven–3.0.4-bin.tar.gz to the directory you wish to install Maven 3.0.4. These instructions assume you chose /usr/local/apache-maven. The subdirectory apache-maven–3.0.4 will be created from the archive.
* In a command terminal, add the M2_HOME environment variable, e.g. export M2_HOME=/usr/local/apache-maven/apache-maven–3.0.4.
* Add the M2 environment variable, e.g. export M2=$M2_HOME/bin.
* Optional: Add the MAVEN_OPTS environment variable to specify JVM properties, e.g. export MAVEN_OPTS="-Xms256m -Xmx512m". This environment variable can be used to supply extra options to Maven.
* Add M2 environment variable to your path, e.g. export PATH=$M2:$PATH.
* Make sure that JAVA_HOME is set to the location of your JDK, e.g. export JAVA_HOME=/usr/java/jdk1.5.0_02 and that $JAVA_HOME/bin is in your PATH environment variable.
* Run mvn –version to verify that it is correctly installed.

Sample Data

Sample data is provided in the /data directory.