|  |
| --- |
| Michigan State University |
| Installing KFS SolrCloud with KFS Solr Rest |
|  |

|  |
| --- |
| Rob Tucker  6/29/2016 |

Contents

[Overview 2](#_Toc454978814)

[Server Topology 2](#_Toc454978815)

[Install Zookeeper 3](#_Toc454978816)

[Install Solr Instances 3](#_Toc454978817)

[KFS Solr Configuration Files 3](#_Toc454978818)

[Upload Zookeeper/SolrCloud Configuration Files 4](#_Toc454978819)

[Create the KFS SolrCloud Collections 4](#_Toc454978820)

[Update Properties with Desired Database Connection Parameters 6](#_Toc454978821)

[Importing KFS Data into the SolrCloud 7](#_Toc454978822)

# Overview

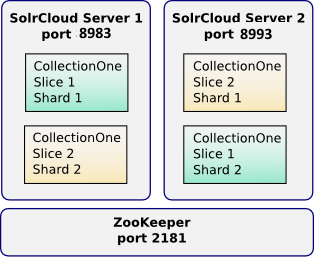
This document describes how to setup a simple 2 node Zookeeper/SolrCloud that will support a solrrest spring-boot application installation running against KFS data. The required application versions are listed below:

1. Java 8
2. Zookeeper 3.4.8
3. Solr 6.1.0
4. Banana 1.6.0
5. kfssolrrest 1.0-SNAPSHOT

Additional installation files noted in the document should be included with the communication containing this document.

# Server Topology

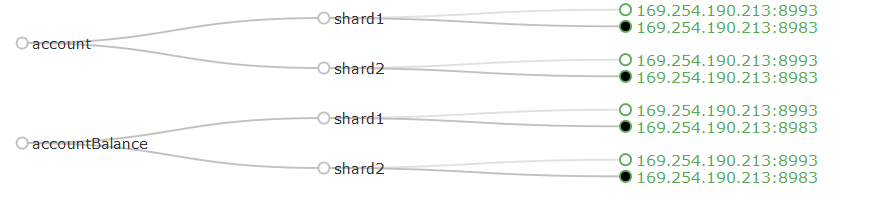
This document describes a simple 2 solr node configuration which contains multiple KFS-related collections. Collections are mapped to kfs tables. Each solr node will contain 2 shards for each collection. Data distribution across the nodes/shards is controlled by Zookeeper based on unique key definitions defined in the SolrCloud configuration. The diagram below shows a simple 2 node installation with both Solr instances and Zookeeper running on the same server using different ports.



Web API (REST)

Client Applications

An example screenshot from the SolrAdmin app:



# Install Zookeeper

Save the attached file zookeeper-3.4.8.tar.gz or you can download from http://mirrors.ibiblio.org/apache/zookeeper/zookeeper-3.4.8/

* unzip zookeeper into desired location on zookeeper server
* create zookeeper configuration file: < zookeeper-home>/conf/zoo.cfg with entries

tickTime=2000

dataDir=<desired-zookeeper-data-directory>

clientPort=2181

* start zookeeper: <zookeeper-home>/bin/zkServer.sh

# Install Solr Instances

Save the attached file solr-6.1.0.zip or you can download from http://apache.claz.org/lucene/solr/6.1.0

* unzip solr6 into desired location on the selected solr servers
* copy JDBC driver jar to <solr-home>/server/lib on each solr server
* create a webapps directory under <solr-home>/server and copy the included banana.war file into this directory

# KFS Solr Configuration Files

The attached file kfs-solr-install.zip contains Zookeeper/SolrCloud configuration files and batch scripts to support the KFS REST services. Unzip the contents of this file to an easily accessible work folder. The contents are described below:

1. In the root there are batch scripts to aid in the installation
2. In the scripts directory are module-specific batch scripts to upload configuration, create solr collections and other configuration work’
3. Schemas directory contains collection-specific solr configuration files

Both windows .bat files and Linux .sh files have been included. The testing has been done on my local machine running windows .bat files so the .sh files may need a little work.

# Upload Zookeeper/SolrCloud Configuration Files

Edit the OS appropriate batch script upload-config and configure the environment variables as shown below:

**SCRIPTS\_DIR** – full path to the configuration scripts directory described in the previous section

**PRIMARY\_SOLR\_INSTANCE\_ROOT\_DIR** – root directory of first solr installation

**SOLR\_SCHEMAS\_ROOT\_DIR** – full path to the schemas directory described in previous section

**ZOOKEEPER\_SERVER** – zookeeper server

**ZOOKEEPER\_PORT** – zookeeper port (probably 2181)

Ensure that Zookeeper is running and execute the upload-config batch script from the command line. This will upload all the SolrCloud collection-specific configuration to Zookeeper. This may take a bit of time. For the current KFS installation there are 354 collections defined.

# Create the KFS SolrCloud Collections

With configuration files uploaded to Zookeeper start all the Solr instances in cloud mode by executing a command similar to what is shown below:

C:\programs\solr-6.0.0-instance1\bin\solr.cmd start -cloud -s C:\programs\solr-6.0.0-instance1\server\solr -m 8g -p 8983 -z localhost:2181

The parameters are as follows:

* -cloud – starts solr in cloud mode
* -s solr server directory
* -m – max memory
* -p – solr port
* -z – Zookeeper host and port

When starting multiple Solr instances on the same server make sure you use different ports.

Once a solr instance is successfully started you should see some like:

Backing up C:\programs\solr-6.0.0-instance2\server\logs\solr.log

1 file(s) moved.

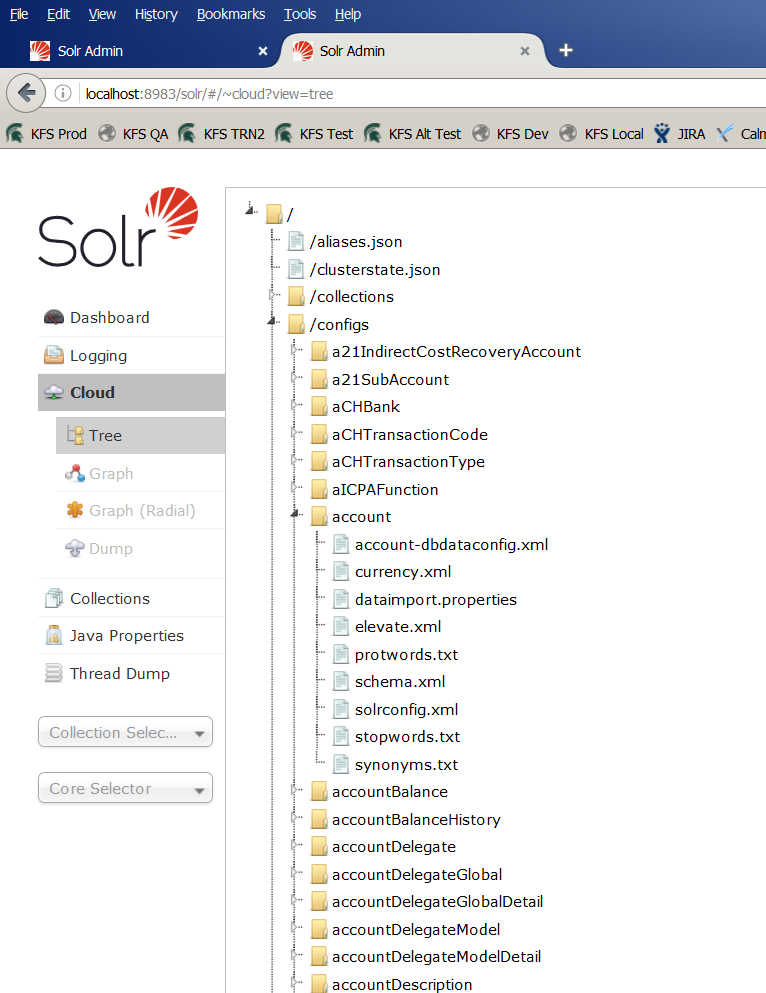
Backing up C:\programs\solr-6.0.0-instance2\server\logs\solr\_gc.log

1 file(s) moved.

Waiting up to 30 to see Solr running on port 8993

Started Solr server on port 8993. Happy searching!

You can now pull up the SolrAdmin app by hitting the primary Solr instance - <http://localhost:8983/solr/>. Click on Cloud->Tree and open up the configs folder and you should see your solr configurations that were loaded in the previous steps:



Edit the OS specific batch file create-collections extracted in an earlier step and set the environment variable as shown below:

**SCRIPTS\_DIR** – full path to the configuration scripts directory described in the previous section

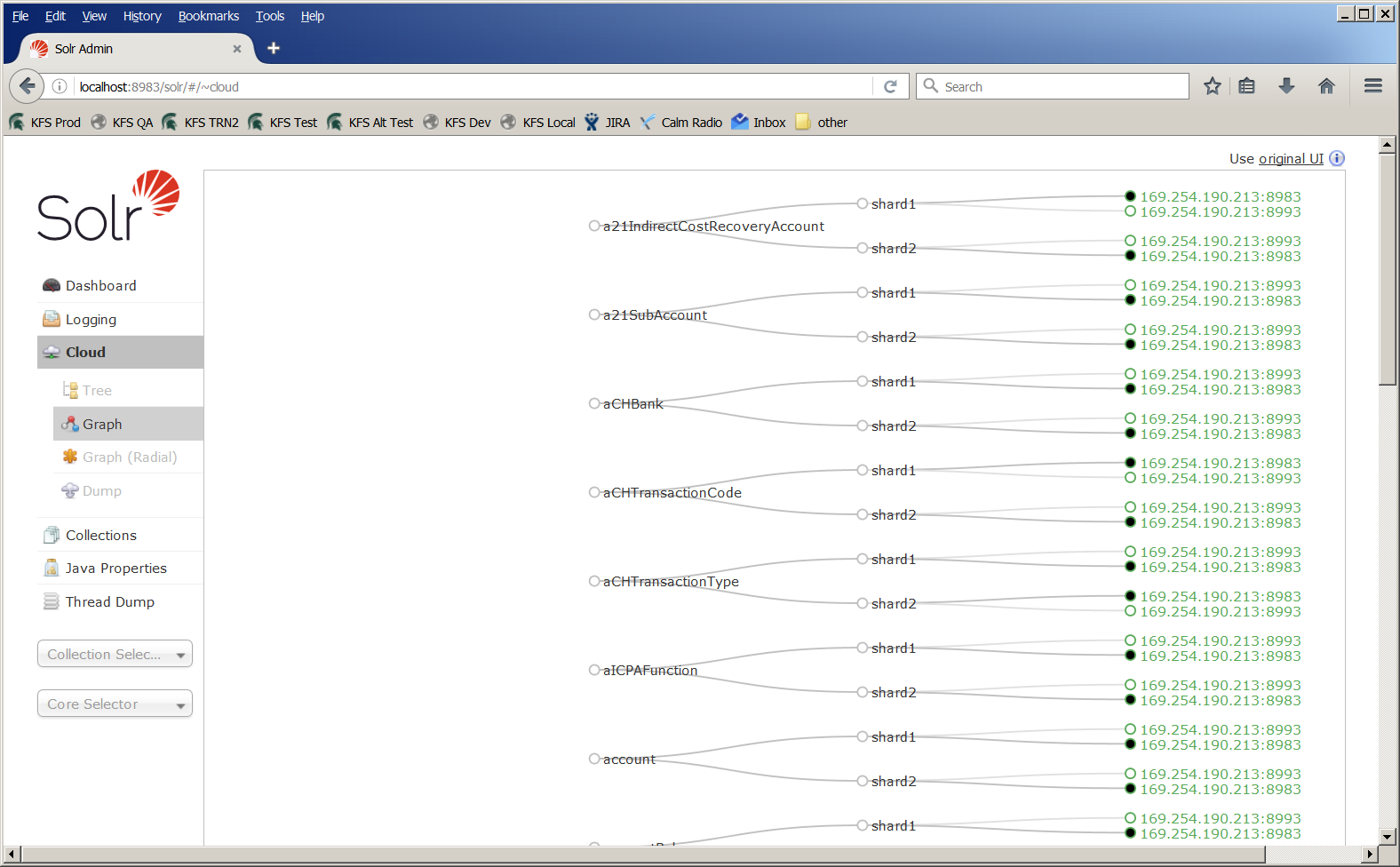
**PRIMARY\_SOLR\_SERVER** – host name of primary (first installed) Solr server

**PRIMARY\_SOLR\_PORT** - port of primary (first installed) Solr server

**CURL\_PATH** – full path to curl application (used for http rest calls to Zookeeper)

Ensure the Zookeeper and all the Solr instances are up and running then execute the create-collections batch script from the command line. The will create all the defined collections on the Solr instances. This will take some time.

After the collections are created pull up SolrAdmin and you should see something similar to what is shown below:



# Update Properties with Desired Database Connection Parameters

In order to pull the KFS data from the database and index into the SolrCloud we need database connection information. The individual collections have a predefine data import configuration defined with placeholders for the database-specific connection information as shown below:

<dataConfig>

<dataSource driver="${solr.database.driver}" url="${solr.database.url}" user="${solr.database.user}" password="${solr.database.password}" />

<document>

<entity name="Account" query="SELECT…

<field column="FIN\_COA\_CD" name="chartOfAccountsCode"/>

<field column="ACCOUNT\_NBR" name="accountNumber"/>

The database-specific values must be appended to the core.properties file in each solr shard for this to work. To do this edit the OS-specific batch script(s) update-db-connection-properties.bat and update-core-properties.bat for windows and the file ) update-db-connection-properties.ash for Linux. This script must be run on each installed Solr instance. Modify the environment variables as shown below:

**SOLR\_INSTANCE\_ROOT\_DIR** – full path to root directory of the Solr instance

**SCRIPTS\_RUN\_DIRECTORY** – full path to location of top level batch scripts

**JDBC\_DB\_DRIVER** – JDBC drive class name (oracle.jdbc.driver.OracleDriver)

**JDBC\_DB\_URL** – database JDBC connection URL

**DB\_USER** – database user name

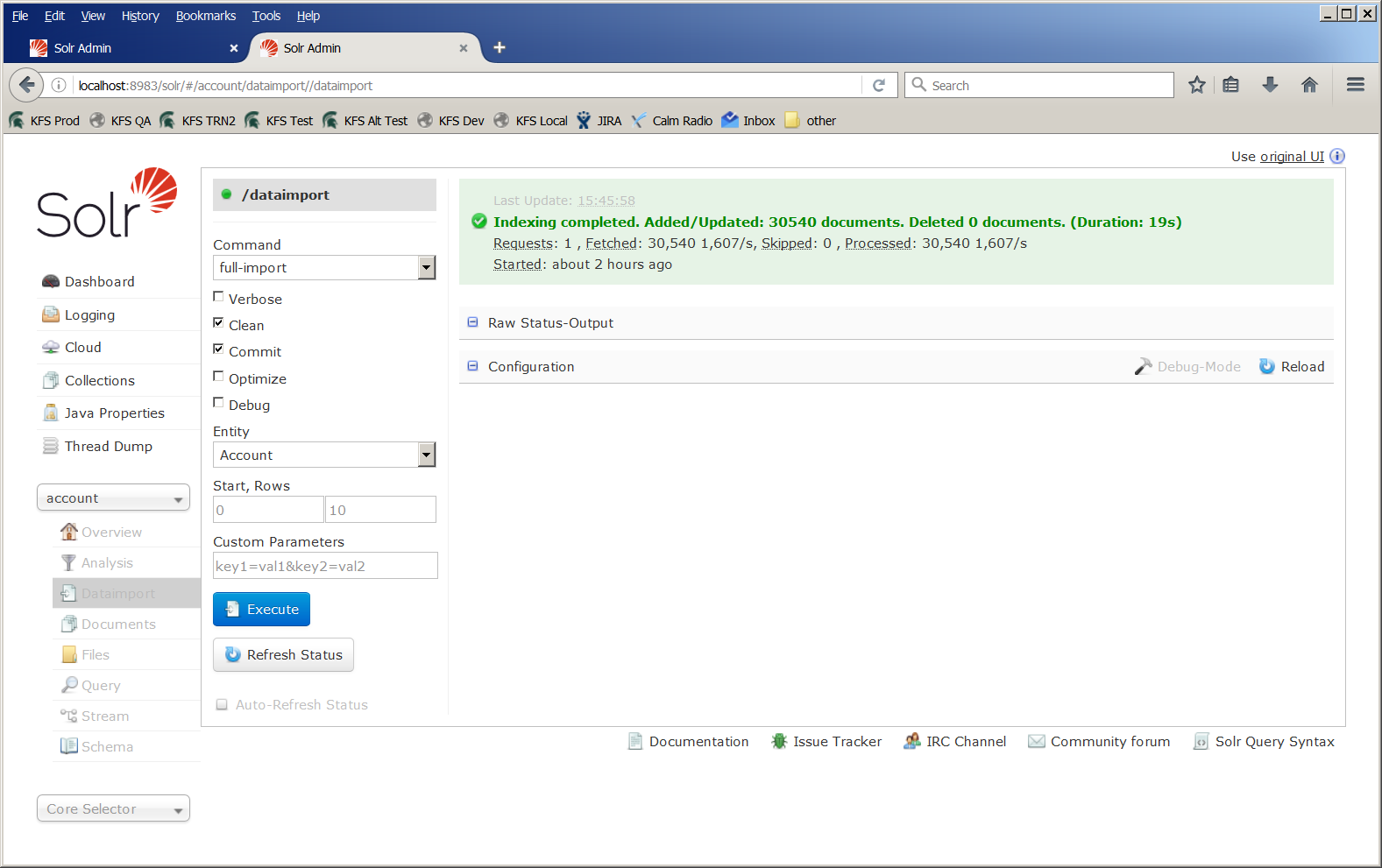
**DB\_PASS** – database password

After running the scripts above shutdown and restart all Zookeeper and all Solr instances

# Importing KFS Data into the SolrCloud

After successfully restarting Zookeeper and all Solr instances you can now import KFS data. There are 2 ways to do this.

1. Import individual collection data via the SolrAdmin interface as shown below:



1. Edit OS-specific batch script full-data-import and set the environment variables as shown below:

**SCRIPTS\_DIR** – full path to the configuration scripts directory described in the previous section

**PRIMARY\_SOLR\_SERVER** – host name of primary (first installed) Solr server

**PRIMARY\_SOLR\_PORT** - port of primary (first installed) Solr server

**CURL\_PATH** – full path to curl application (used for http rest calls to Zookeeper)

Execute script full-data-import from the command line and this should do a full data import on all configured collections. I have not tried this one yet and it would probably take quite a while. You can modify the module-specific full-data-import scripts to import specific collections as desired.