Solr Configuration
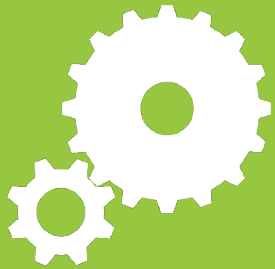
Xavier Morera
@xmorera
www.searchtechnologies.com

pluralsight
hardcore dev and IT training

Solr Configuration

Installation

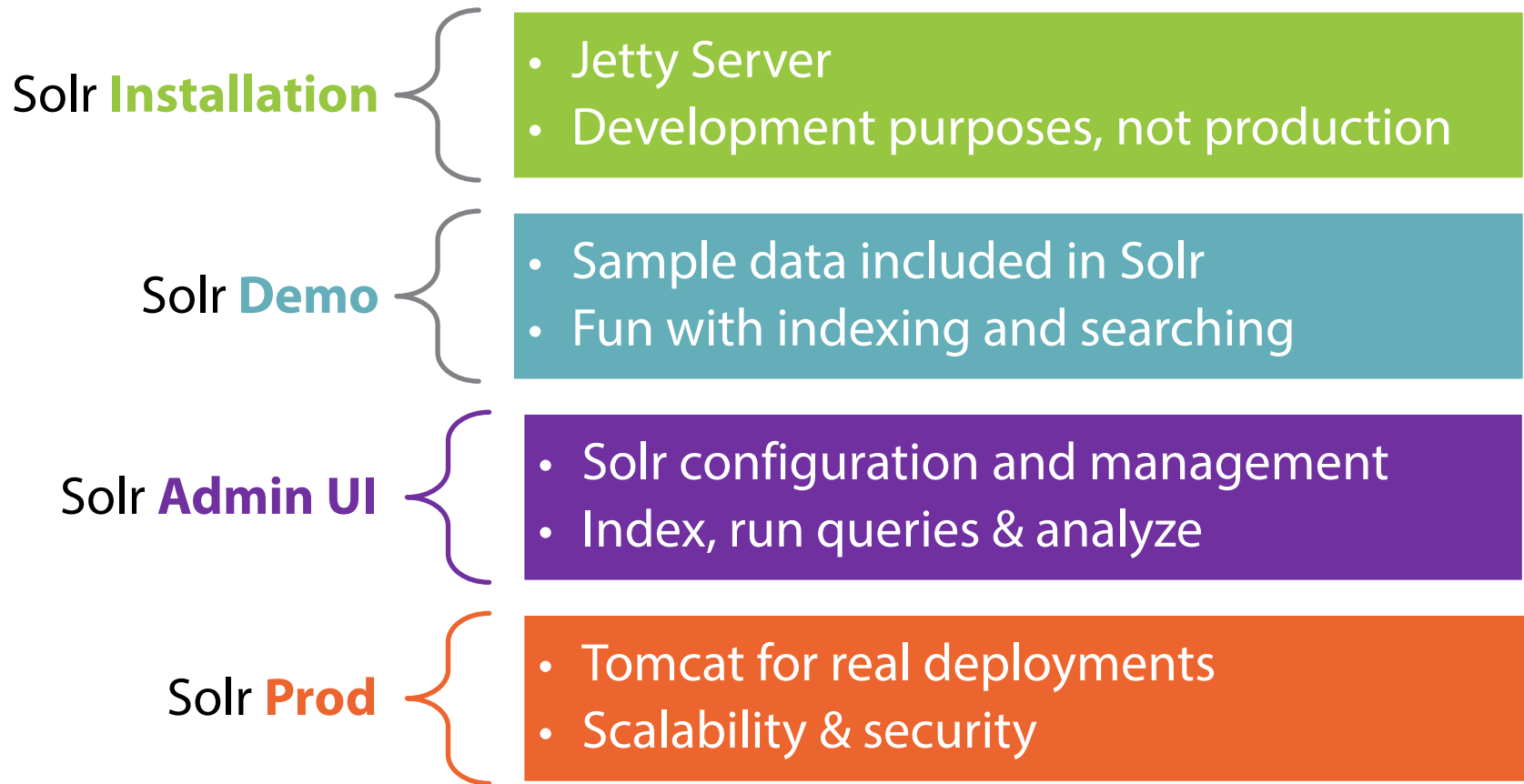D
Demo

Admin UI

Tomcat

Solr Configuration Agenda

Solr **Installation**
- Jetty Server
- Development purposes, not production

Solr **Demo**
- Sample data included in Solr
- Fun with indexing and searching

Solr **Admin UI**
- Solr configuration and management
- Index, run queries & analyze

Solr **Prod**
- Tomcat for real deployments
- Scalability & security

Solr Configuration Installation
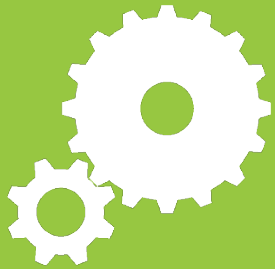
Installation	Demo	Admin UI	Tomcat

Getting Solr

- **Download from Apache.org**
  - https://lucene.apache.org/solr/
  - Latest version
  - What platform? (Windows here…)

- **Or just *search* for "download solr"**

solr-4.7.0.zip	3/28/2014 6:17 PM	153,950 KB	Compressed (zipped) Folder
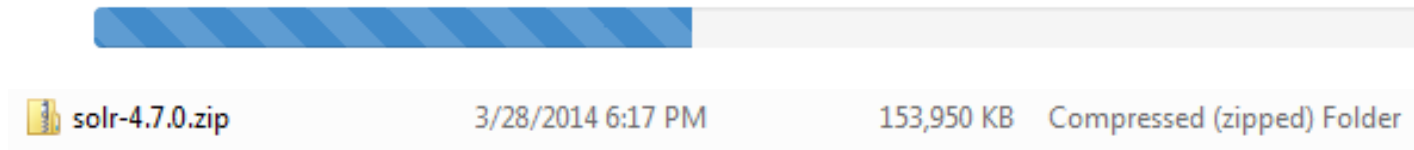
- **Do you need an older/specific version**
  - Apache archives
  - http://archive.apache.org/dist/lucene/solr/

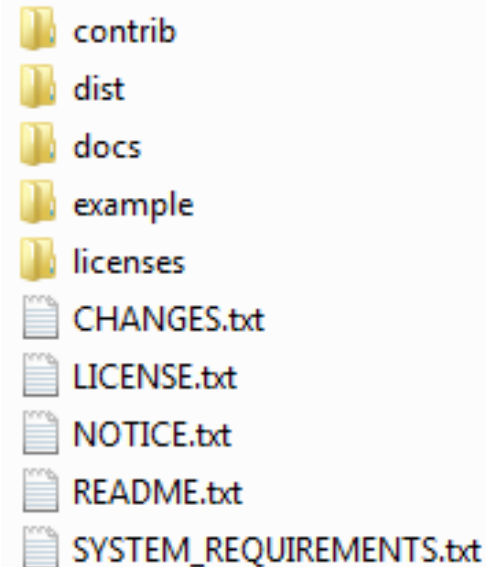# Unzip and Review Contents

- **System Requirements**
  - Prerequisite Java 1.6 or higher

- **contrib, dist, docs, licenses**

- **example!**
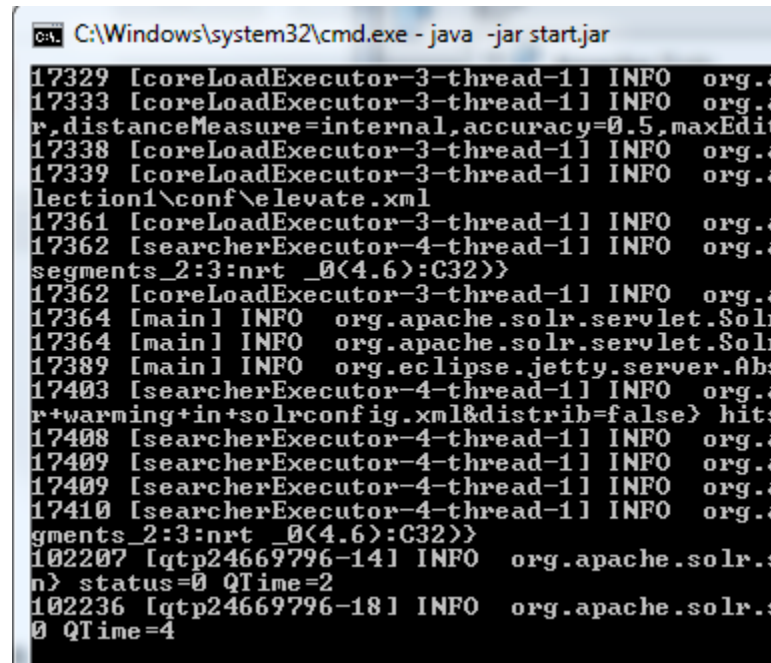 - Self contained Solr instance
  - Copy and rename ▷ 📁 pluralsight-solr

📁 contrib
📁 dist
📁 docs
📁 example
📁 licenses
📄 CHANGES.txt
📄 LICENSE.txt
📄 NOTICE.txt
📄 README.txt
📄 SYSTEM_REQUIREMENTS.txt

Please Start Your (Search) Engine!

- **Supported in Windows, Linux and OS X**
  - Windows in my case

- **Check java version (>= 1.6)**
  - Java –version

- **Use Jetty for development**

- **Start up Solr**
  - Java –jar start.jar

```
C:\Windows\system32\cmd.exe - java  -jar start.jar
17329 [coreLoadExecutor-3-thread-1] INFO   org.a
17333 [coreLoadExecutor-3-thread-1] INFO   org.a
r,distanceMeasure=internal,accuracy=0.5,maxEdit
17338 [coreLoadExecutor-3-thread-1] INFO   org.a
17339 [coreLoadExecutor-3-thread-1] INFO   org.a
lection1\conf\elevate.xml
17361 [coreLoadExecutor-3-thread-1] INFO   org.a
17362 [searcherExecutor-4-thread-1] INFO   org.a
segments_2:3:nrt _0(4.6):C32)}
17362 [coreLoadExecutor-3-thread-1] INFO   org.a
17364 [main] INFO   org.apache.solr.servlet.Sol
17364 [main] INFO   org.apache.solr.servlet.Sol
17389 [main] INFO   org.eclipse.jetty.server.Abs
17403 [searcherExecutor-4-thread-1] INFO   org.a
r+warming+in+solrconfig.xml&distrib=false> hits
17408 [searcherExecutor-4-thread-1] INFO   org.a
17409 [searcherExecutor-4-thread-1] INFO   org.a
17409 [searcherExecutor-4-thread-1] INFO   org.a
17410 [searcherExecutor-4-thread-1] INFO   org.a
gments_2:3:nrt _0(4.6):C32)}
102207 [qtp24669796-14] INFO   org.apache.solr.s
n> status=0 QTime=2
102236 [qtp24669796-18] INFO   org.apache.solr.s
0 QTime=4
```
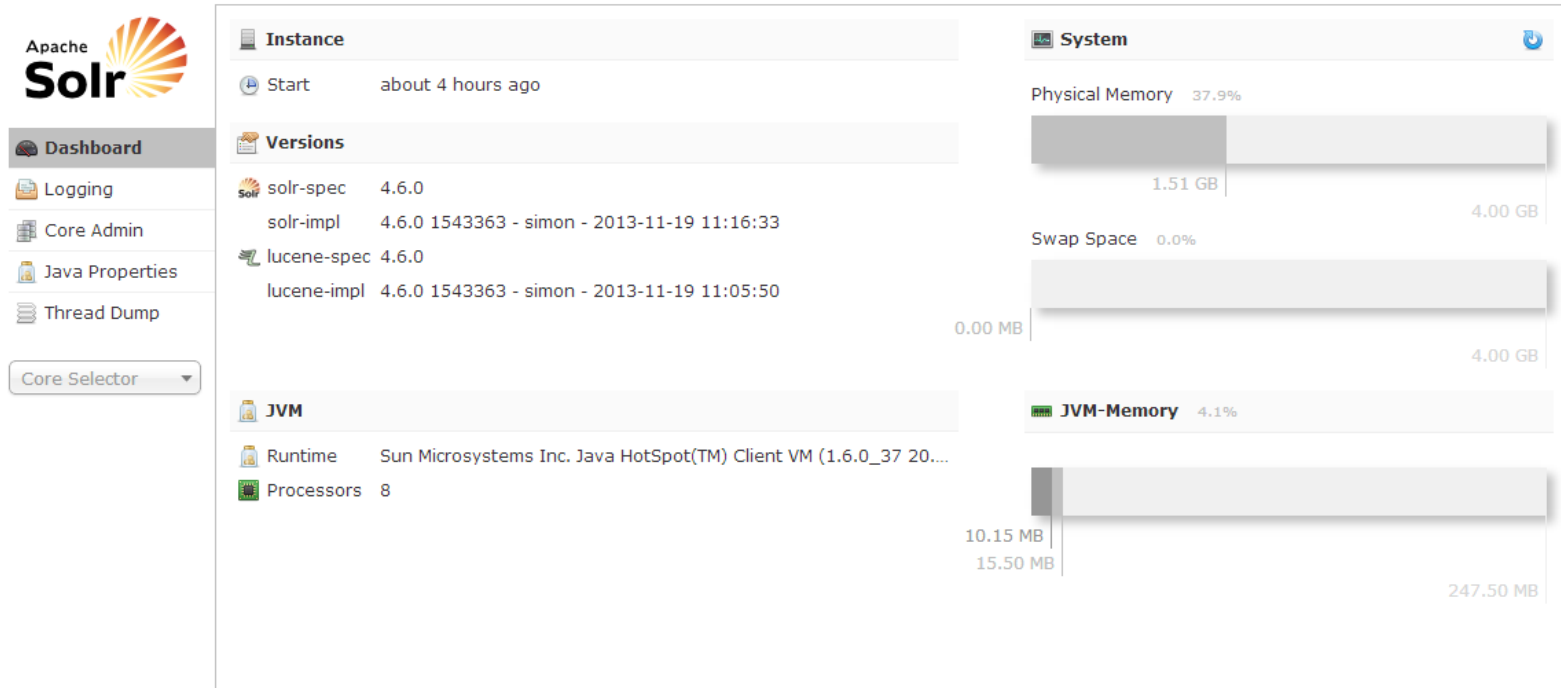
Solr Up & Running

- **Confirm operational**
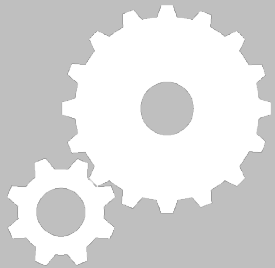
- **http://localhost:8983/solr/#/**

Solr Configuration Demo

Installation	**D** Demo	Admin UI	Tomcat

Let's Tour Around Solr

- **Install**

- **Get data into the index**

- **Run a query or two**

- **Admin UI afterwards**

Apache **Solr**

	Instance		System	
Start	4 months ago		Physical Memory	84.6%
Host	10.208.65.184			
CWD	C:\Program Files\Apache Software Foundation\To...			519.76 MB
Instance	c:\usadoscr\collection1			614.62 MB
Data	c:\usadoscr\collection1\data		Swap Space	12.6%
Index	C:\usadoscr\collection1\data\index			
Versions			1.08 GB	

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
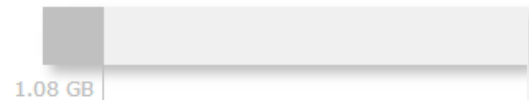
What Data Will I Use?

- **exampledocs → data for testing**
  - XML
  - CSV

- **post.jar**
  - Tool for posting documents
  - That follow a specific format

- **Java –jar post.jar -help**
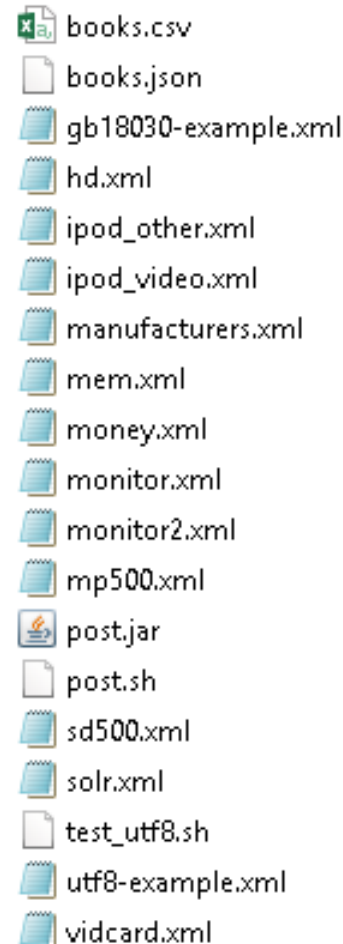
books.csv
books.json
gb18030-example.xml
hd.xml
ipod_other.xml
ipod_video.xml
manufacturers.xml
mem.xml
money.xml
monitor.xml
monitor2.xml
mp500.xml
post.jar
post.sh
sd500.xml
solr.xml
test_utf8.sh
utf8-example.xml
vidcard.xml

How to Upload Sample Data

- **cd \exampledocs**

- **Java –jar post.jar *.xml**

```
C:\pluralsight-solr\exampledocs>java -jar post.jar *xml
SimplePostTool version 1.5
Posting files to base url http://localhost:8983/solr/update using content-type a
pplication/xml..
POSTing file gb18030-example.xml
POSTing file hd.xml
POSTing file ipod_other.xml
POSTing file ipod_video.xml
POSTing file manufacturers.xml
POSTing file mem.xml
POSTing file money.xml
POSTing file monitor.xml
POSTing file monitor2.xml
POSTing file mp500.xml
POSTing file sd500.xml
POSTing file solr.xml
POSTing file utf8-example.xml
POSTing file vidcard.xml
14 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/update..
Time spent: 0:00:00.509
```

Other Upload Mechanisms

- **DataImportHandler**

- **Add XML, Json, CSV documents**

- **SolrNet or SolrJ**

- **Multiple tools that post documents to Solr's index**
  - ASPIRE
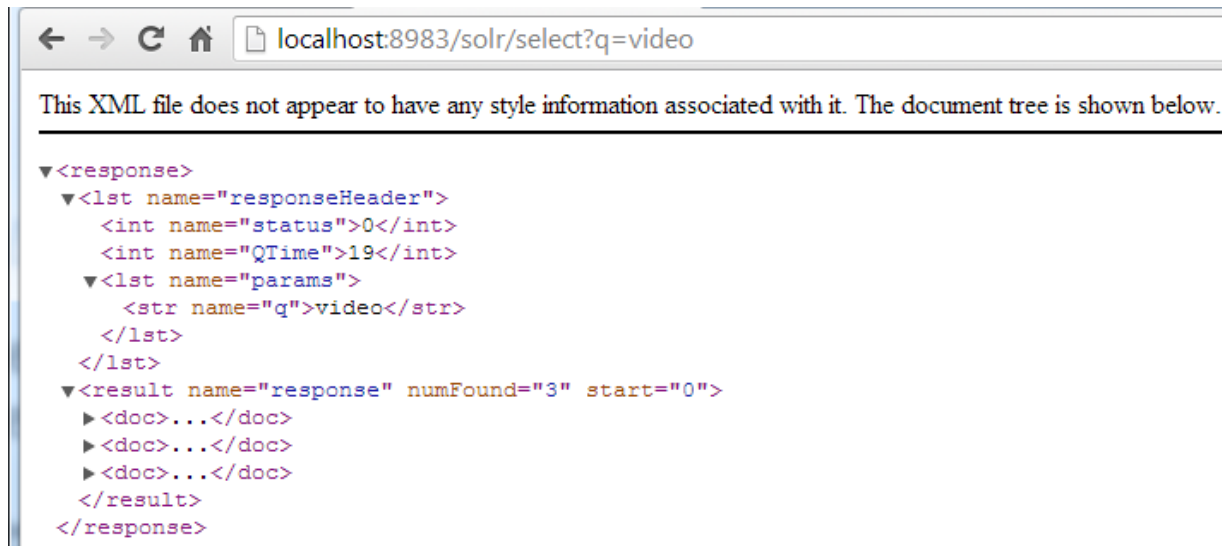
Ask a Question! (A.K.A. Query!)

- **http://localhost:8983/solr/select?q=video**

- **Url made up of**
  - Host name (localhost) & Port number (8983)
  - Application name (solr)
  - Request handler for queries (select)
  - Query itself (q=video).

```
← → C ⌂  🗋 localhost:8983/solr/select?q=video

This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<response>
  ▼<lst name="responseHeader">
      <int name="status">0</int>
      <int name="QTime">19</int>
    ▼<lst name="params">
        <str name="q">video</str>
      </lst>
    </lst>
  ▼<result name="response" numFound="3" start="0">
    ▶<doc>...</doc>
    ▶<doc>...</doc>
    ▶<doc>...</doc>
    </result>
</response>
```

Document

- **Let's review one doc**

```xml
▼<doc>
  <str name="id">EN7800GTX/2DHTV/256M</str>
  <str name="name">ASUS Extreme N7800GTX/2DHTV (256 MB)</str>
  <str name="manu">ASUS Computer Inc.</str>
  <str name="manu_id_s">asus</str>
  ▼<arr name="cat">
    <str>electronics</str>
    <str>graphics card</str>
  </arr>
  ▼<arr name="features">
    <str>NVIDIA GeForce 7800 GTX GPU/VPU clocked at 486MHz</str>
    <str>256MB GDDR3 Memory clocked at 1.35GHz</str>
    <str>PCI Express x16</str>
    <str>Dual DVI connectors, HDTV out, video input</str>
    <str>OpenGL 2.0, DirectX 9.0</str>
  </arr>
  <float name="weight">16.0</float>
  <float name="price">479.95</float>
  <str name="price_c">479.95,USD</str>
  <int name="popularity">7</int>
  <str name="store">40.7143,-74.006</str>
  <bool name="inStock">false</bool>
  <date name="manufacturedate_dt">2006-02-13T00:00:00Z</date>
  <long name="_version_">1458929687206035456</long>
</doc>
```

Changing Solr Port in Jetty

- **Change port in \etc\jetty.xml**

```
<Call name="addConnector">
  <Arg>
      <New class="org.eclipse.jetty.server.bio.SocketConnector">
        <Set name="host"><SystemProperty name="jetty.host" /></Set>
        <Set name="port"><SystemProperty name="jetty.port" default="8983"/></Set>
        <Set name="maxIdleTime">50000</Set>
        <Set name="lowResourceMaxIdleTime">1500</Set>
        <Set name="statsOn">false</Set>
      </New>
  </Arg>
</Call>
```

Solr Configuration Admin UI

Installation

Demo

Admin UI

Tomcat

Admin Section

- **Dashboard**

- **Logging**

- **Core Admin**

- **Java Properties**

- **Thread Dump**

- **Core Selector**

- **Footer**
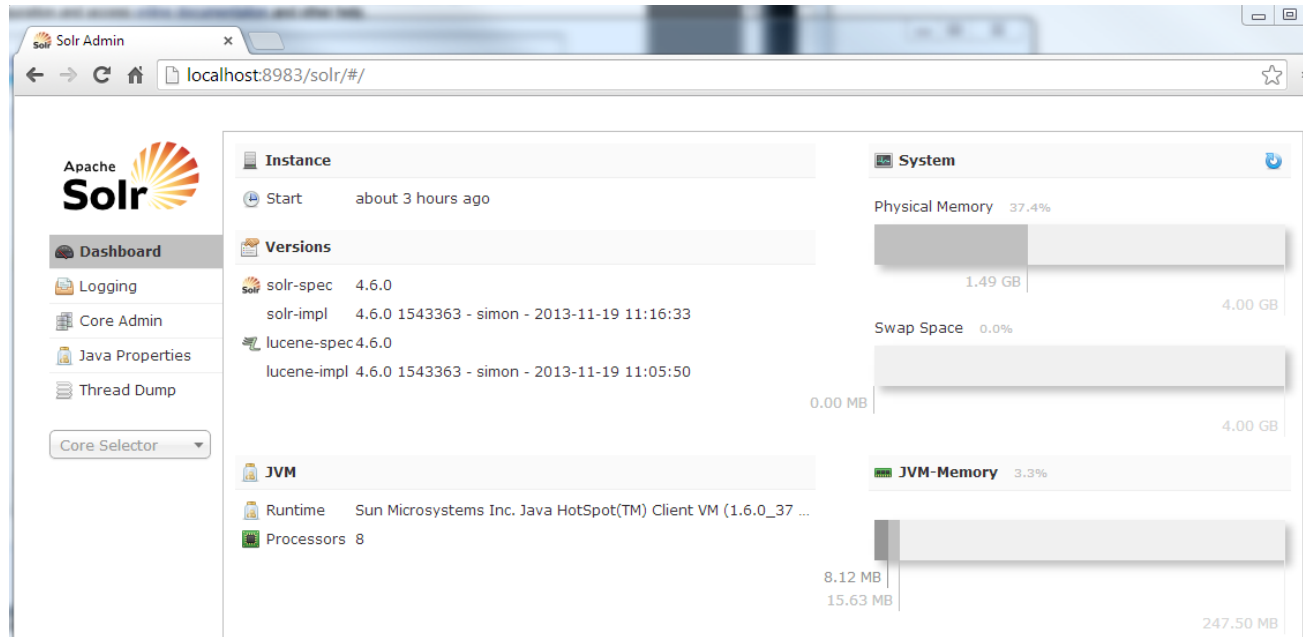
Footer → Getting Assistance

📄 Documentation   🐞 Issue Tracker   👥 IRC Channel   ✉ Community forum   🔲 Solr Query Syntax

- **Documentation → http://lucene.apache.org/solr/**

- **Issue Tracker → Official issue tracker**
  - http://issues.apache.org/jira/browse/SOLR

- **IRC Channel → irc.freenode.net, Port 7000, #solr**

- **Community Forum → Mailing lists and archives**

- **Solr Query Syntax → Wiki describing Solr's syntax**
  - http://wiki.apache.org/solr/SolrQuerySyntax

Dashboard

- **Left hand side navigation and center for selected option**

- **General Information**
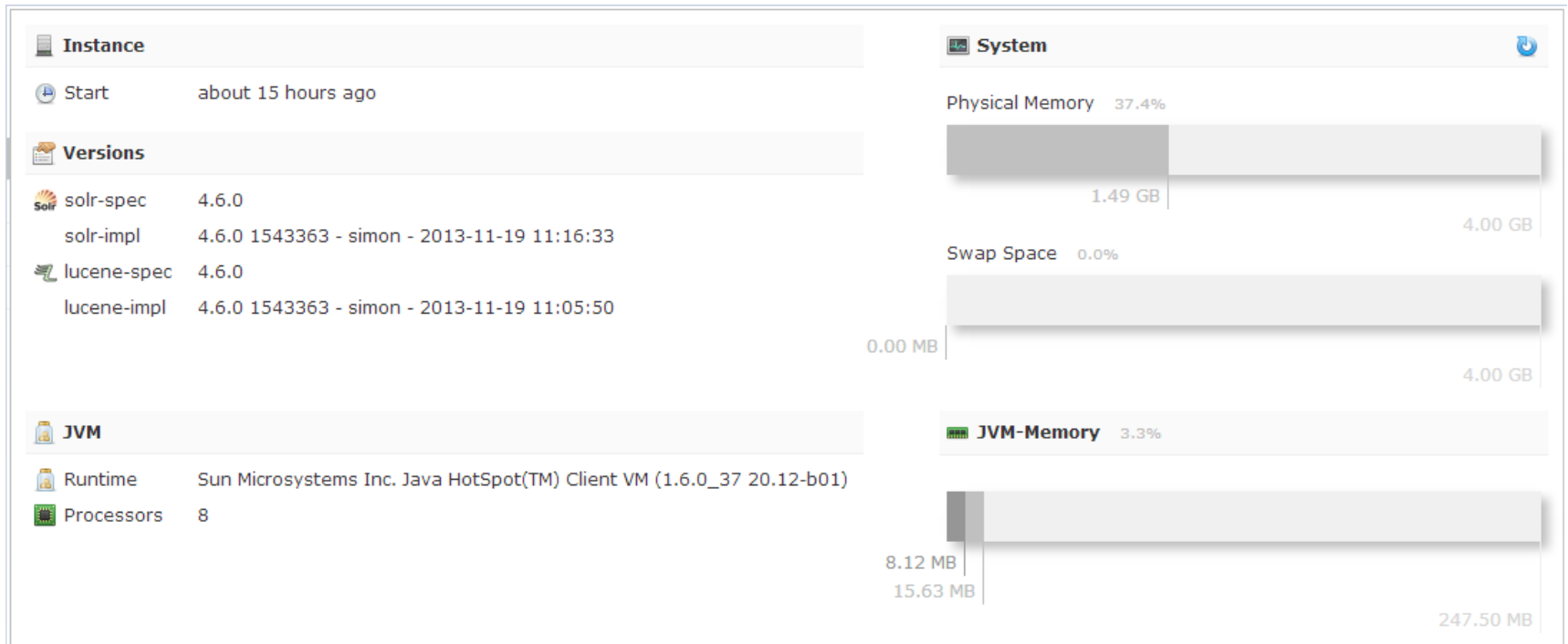
Instance			System	
Start	about 15 hours ago		Physical Memory 37.4%	
Versions				1.49 GB
solr-spec	4.6.0			4.00 GB
solr-impl	4.6.0 1543363 - simon - 2013-11-19 11:16:33		Swap Space 0.0%	
lucene-spec	4.6.0			
lucene-impl	4.6.0 1543363 - simon - 2013-11-19 11:05:50		0.00 MB	4.00 GB
JVM			JVM-Memory 3.3%	
Runtime	Sun Microsystems Inc. Java HotSpot(TM) Client VM (1.6.0_37 20.12-b01)		8.12 MB	
Processors	8		15.63 MB	247.50 MB

Logging

- **Messages from Solr's log files**

Logging

- **Select Log Level to change**

Logging

- **Multiple levels**

- **Can be configured via *logging.properties* files**

Level	Result
FINEST	Reports everything.
FINE	Reports everything but the least important messages.
CONFIG	Reports configuration errors.
INFO	Reports everything but normal status.
WARNING	Reports all warnings.
SEVERE	Reports only the most severe warnings.
OFF	Turns off logging.
UNSET	Removes the previous log setting.

Core Admin

- **Manage your cores (your what?)**

Add Core	Unload	Rename	Swap	Reload	Optimize

collection1

📦 Core

startTime:	19 minutes ago
instanceDir:	C:\pluralsight-solr\solr\collection1\
dataDir:	C:\pluralsight-solr\solr\collection1\data\

📊 Index

lastModified:	9 days ago
version:	3
numDocs:	32
maxDoc:	32
deletedDocs:	-
optimized:	✔
current:	✔
directory:	org.apache.lucene.store.NRTCachingDirectory: NRTCachingDirectory(org.apache.lucene.store.SimpleFSDirectory@C:\pluralsight-solr\solr\collection1\data\index lockFactory=org.apache.lucene.store.NativeFSLockFactory@300429; maxCacheMB=48.0 maxMergeSizeMB=4.0)
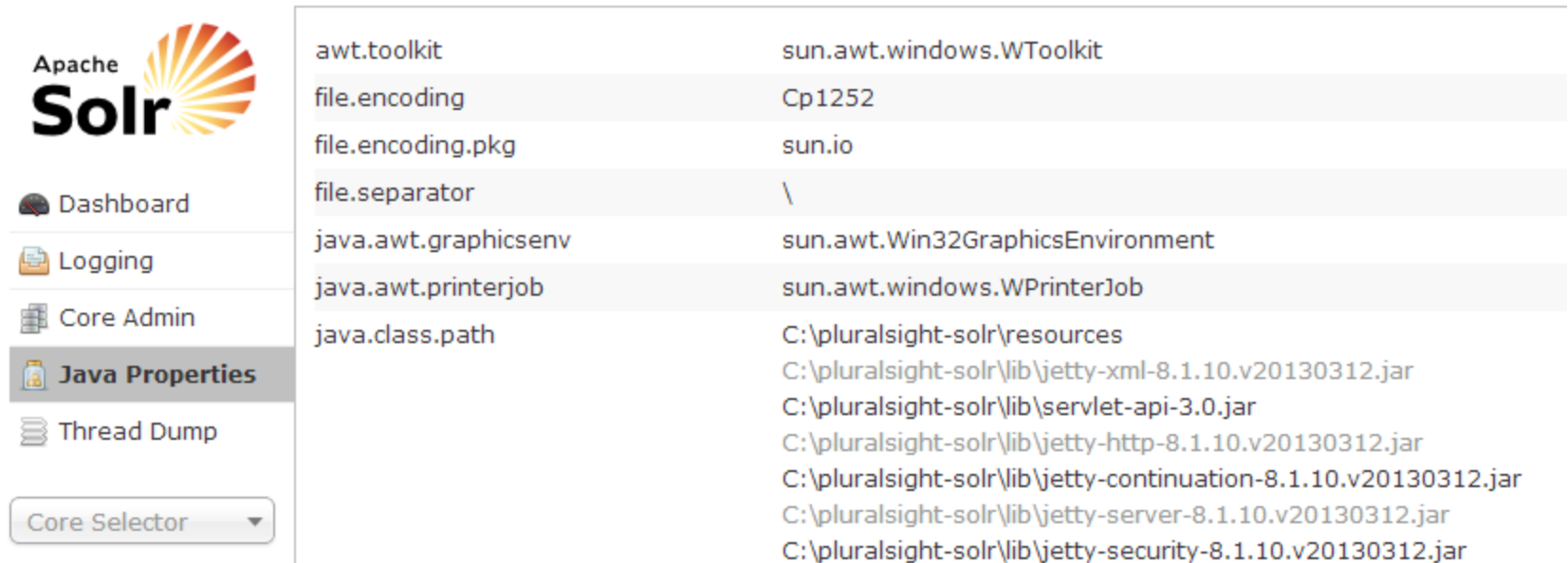
Left sidebar navigation:
- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump

Core Selector ▼

Java Properties

- **Easy access to essential components of a top-performing Solr system**

Thread Dump

- **Inspect currently active threads in your server**

Core Selector

- **Select a Core**
  - Or find it

- **Perform Core specific functions**
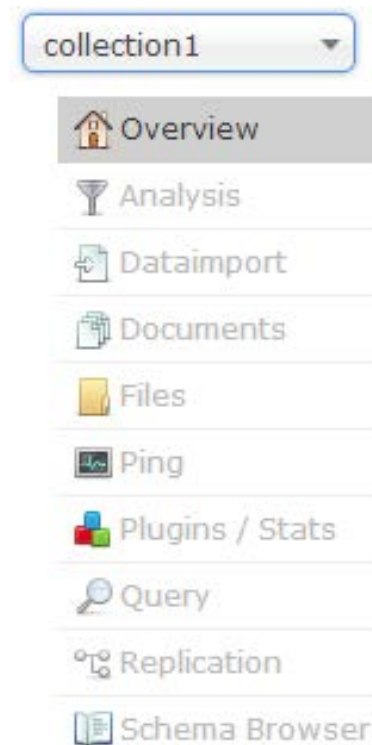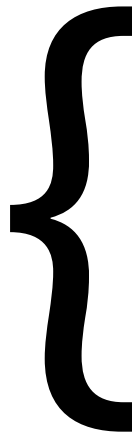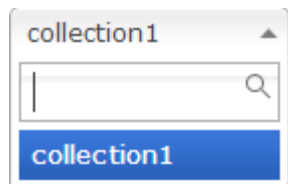
Core: Analysis

- **Analyze data in specific fields**

- **Inspect how data will be handled according to the field/field type and dynamic configurations**

- **Give it a shot. This will be VERY useful**

Field Value (Index)

test with some utf-8

Analyse Fieldname / FieldType: name

ST	text	test	with	some	utf	8
	raw_bytes	[74 65 73 74]	[77 69 74 68]	[73 6f 6d 65]	[75 74 66]	[38]
	start	0	5	10	15	19
	end	4	9	14	18	20
	type	<ALPHANUM>	<ALPHANUM>	<ALPHANUM>	<ALPHANUM>	<NUM>
	position	1	2	3	4	5

Core: Dataimport

- **Status of the Data Import Handler.**

- **Allows you to start indexing data**

Core: Documents

- **Execute Solr indexing commands**

- **Copy or upload documents (JSON, CSV, XML) and submit to the index**

- **Construct documents by selecting fields and field values**

- **Always start by defining RequestHandler to use (aka qt)**

- **By default /update will be defined**

Request-Handler (qt)

/update

Document Type

JSON

Document(s)

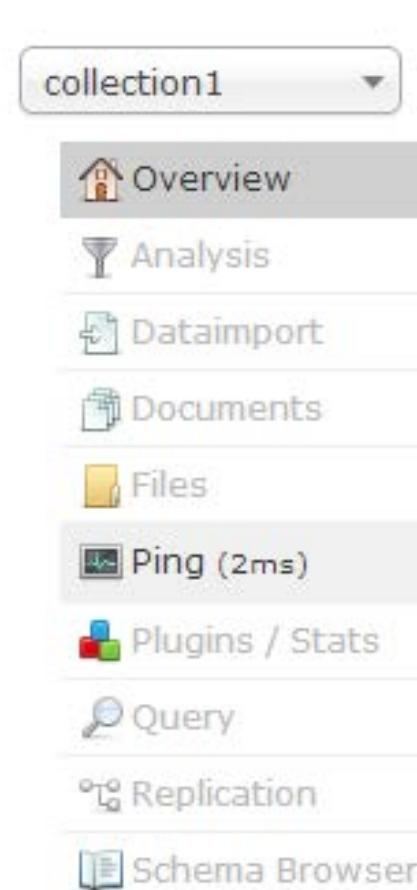{"id":"change.me","title":"change.me"}

Commit Within

1000

Overwrite

true

Boost

1.0

Submit Document

Ping

- **Ping a named core and determine if active.**

- **Configured using a requestHandler in the solrconfig.xml file:**

collection1

🏠 Overview
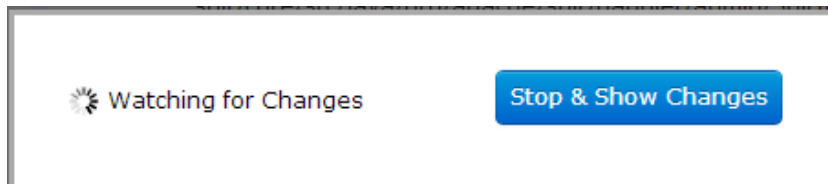
🔻 Analysis

📥 Dataimport

📄 Documents

📁 Files

📊 Ping (2ms)

🔲 Plugins / Stats

🔍 Query

🔀 Replication

📖 Schema Browser

Core: Plugins/Stats

- **Statistics for plugins and installed components**

- **Status and performance of caches, searches and configuration of searchHandlers and requestHandlers**

- **Snapshot taken when page is loaded**
  - Watch Changes or Refresh Values

documentCache

class:	org.apache.solr.search.LRUCache
version:	1.0
description:	LRU Cache(maxSize=512, initialSize=512)
src:	$URL: https://svn.apache.org/repos/asf/luce solr/core/src/java/org/apache/solr/search/LR

stats:		
	lookups:	3
	hits:	0
	hitratio:	0
	inserts:	6
	evictions:	0
	size:	6
	warmupTime:	0
	cumulative_lookups:	3
	cumulative_hits:	0
	cumulative_hitratio:	0
	cumulative_inserts:	3
	cumulative_evictions:	0

- fieldCache
- fieldValueCache
- filterCache
- perSegFilter
- queryResultCache

Watching for Changes   **Stop & Show Changes**

Core: Query

- **Submit a structured query and analyze results**

- **Results have at least two sections:**
    - responseHeader: status, Qtime and params
    - response: documents that match the query in doc subsections

- **Experiment (have some fun!)**

Request-Handler (qt)

/select

— common —

q

:

fq

sort

start, rows

0   10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

Core: Query → Parameters

Request-handler	Specify request handler or use standard if not specified
q	The query
fq	Filter query
sort	Sort ascending or descending

Core: Query → Parameters

start, rows	Start is offset of returned results, default value is 0. Rows controls how many results are returned
fl	Stored fields should be returned. Solr 4 and above can include function results.
wt	Response Writer to be used to format the query response. Defaults to XML if not specified.
indent	Return results with proper indentation
debugQuery	Use to augment debug information for programmer or administrator

Core: Query → Parameters

dismax	Enable the Dismax query parser.
edismax	Enable the Extended query parser.
hl	Enable highlighting in the query response
facet	Enables faceting, the arrangement of search results into categories based on indexed terms.
spatial	Enable using location data for use in spatial or geospatial searches.
spellcheck	Click this button to enable the Spellchecker, which provides inline query suggestions based on other, similar, terms.

Core: Replication

- **Enable/disable replication.**

- **Show current replication status.**

Core: Schema Browser

- **Displays schema data**

- **Loads a specific field if opened from Analysis window**

- **Else, select field or field type**

Solr Configuration Tomcat Installation
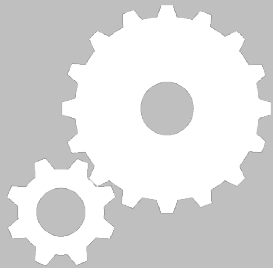
Installation

D
Demo

Admin UI

Tomcat

Tomcat

- **Solr can run in multiple web containers**

- **Tomcat being one of the favorites**

- **Install Tomcat as a service**

- **Deploy Solr.war**

- **Set up SLF4J and log4j**

- **Configure log4j**
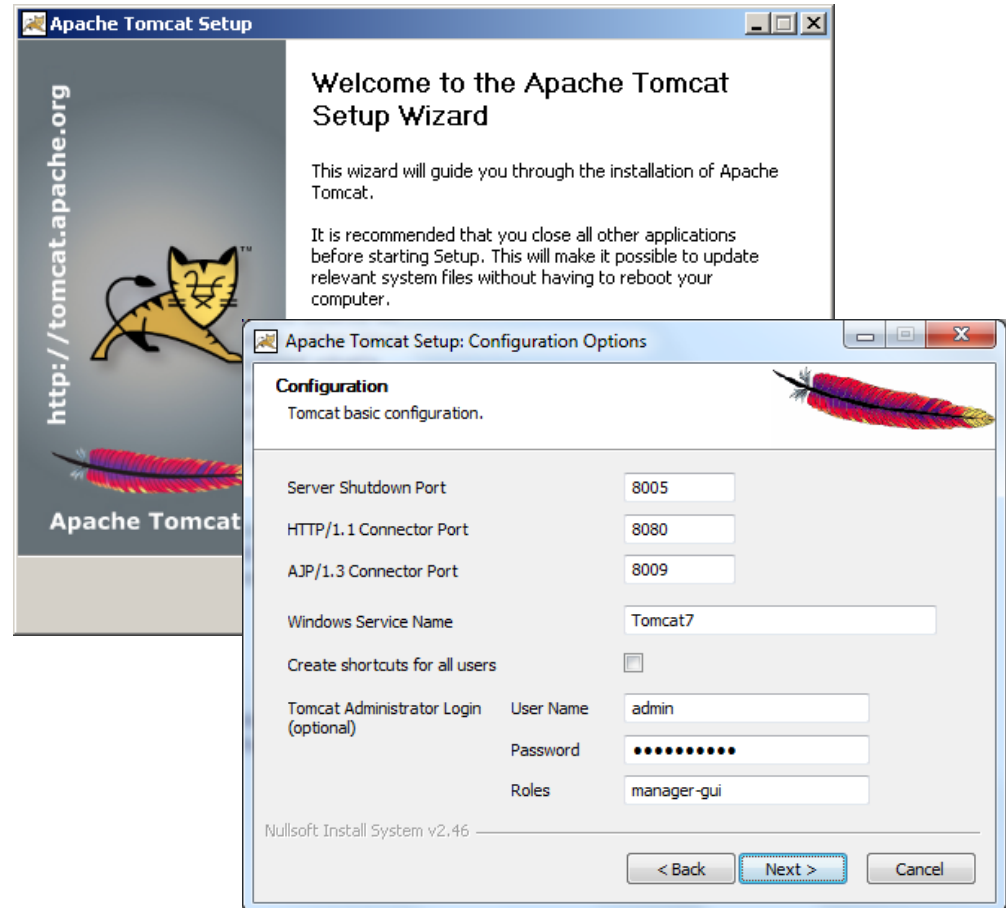
- **Changes in configuration**

- **Confirm all OK!**
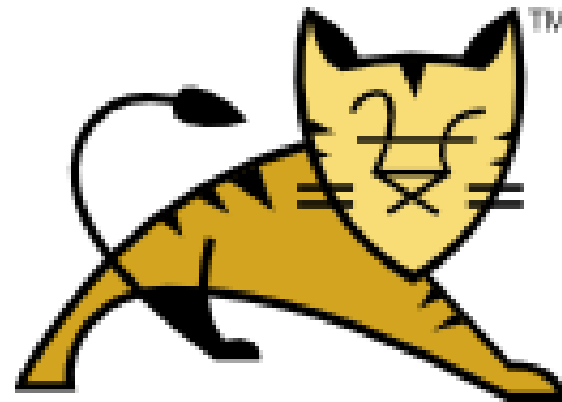
Installing Tomcat

- **Download and install**
  - Service

- **Select desired port**

- **Create administrator login**

Deploy Solr.war

- **Open Solr distribution and look for \solr-4.7.0\example\webapps**

- **Copy the war file to tomcat/webapps**
  - C:\Program Files (x86)\Apache Software Foundation\Tomcat 7.0\webapps

- **If Tomcat is running it will automatically deploy**
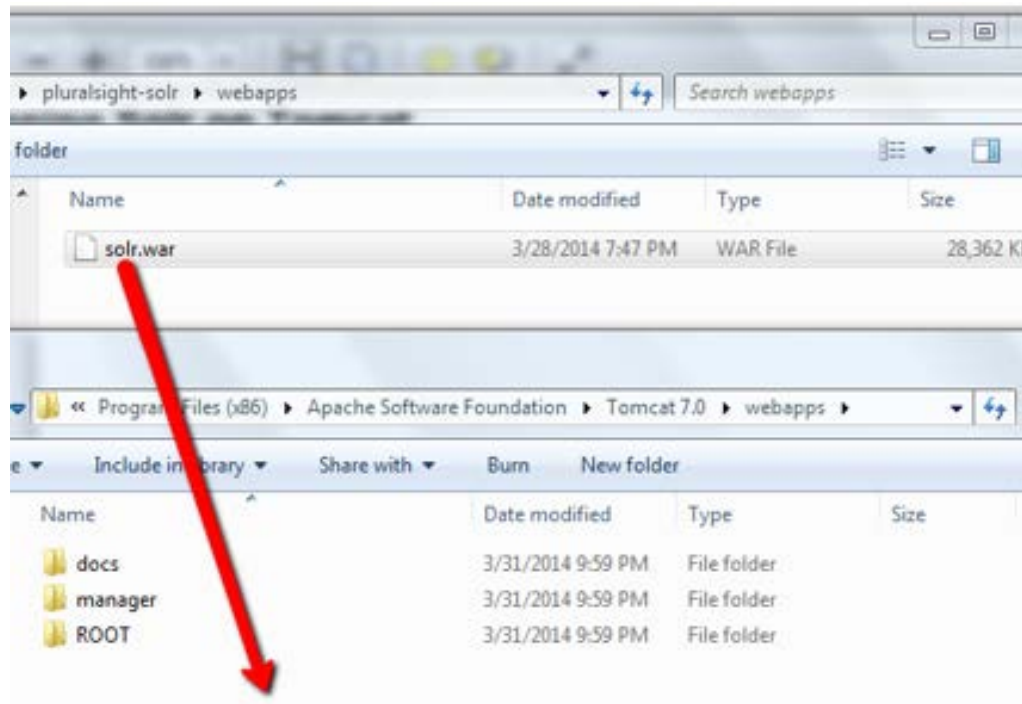
Set Solr Home

- **Run Tomcat7w.exe from**
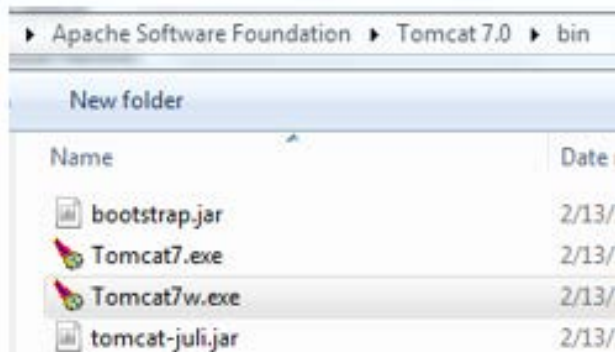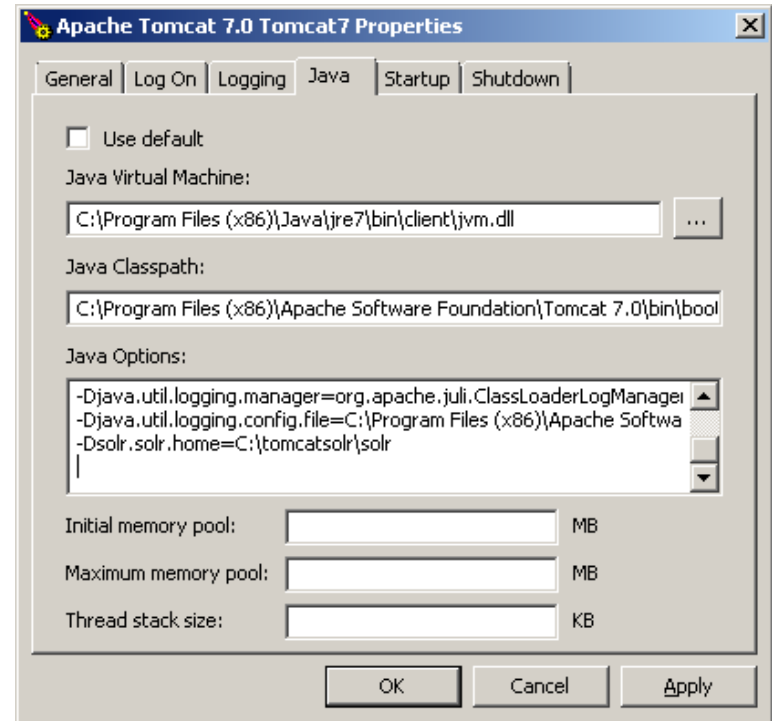 - /Tomcat 7.0/bin

- **Copy all Solr files to preferred folder**

- **Add Java option to set Solr home**
  - -Dsolr.solr.home=C:\tomcatsolr\example\solr
  - Hint: where your collections are

SLF4J & log4j

- **Copy jars from**
  - □ \solr-4.7.0\example\lib\ext

- **To**
  - □ \Tomcat 7.0\lib

solr-4.7.0 ▾ example ▾ lib ▾ ext	
...are with ▾ Burn New folder	
Name ▲	Date modified
jcl-over-slf4j-1.6.6.jar	6/11/2012 4:43 PM
jul-to-slf4j-1.6.6.jar	6/11/2012 4:43 PM
log4j-1.2.16.jar	3/31/2010 6:25 AM
slf4j-api-1.6.6.jar	6/11/2012 4:41 PM
slf4j-log4j12-1.6.6.jar	6/11/2012 4:42 PM

| ...tware Foundation\Tomcat 7.0\lib	
...gram Files (x86) ▾ Apache Software Foundation ▾ Tomcat 7.0 ▾ lib	
...hare with ▾ Burn New folder	
Name ▲	Date modified
annotations-api.jar	2/13/2014 1:29 AM
catalina.jar	2/13/2014 1:29 AM
catalina-ant.jar	2/13/2014 1:29 AM
catalina-ha.jar	2/13/2014 1:29 AM
catalina-tribes.jar	2/13/2014 1:29 AM

Log4j Properties File

- **Copy log4j.properties to a location in the classpath**

- **Or to the same location**
  - □ And add java option

  -Dlog4j.configuration=//path/log4j.properties

And restart Tomcat....
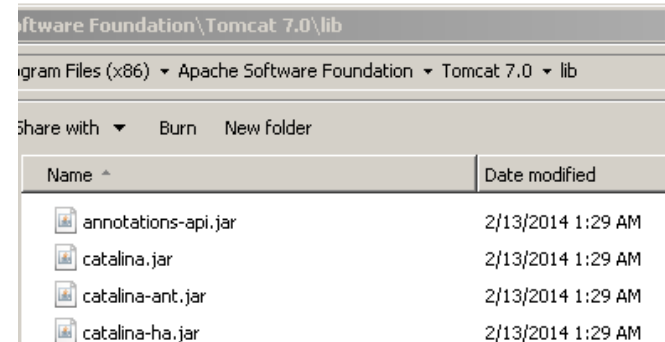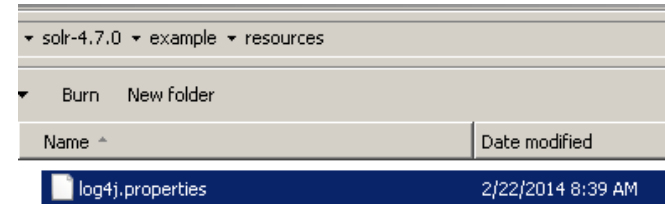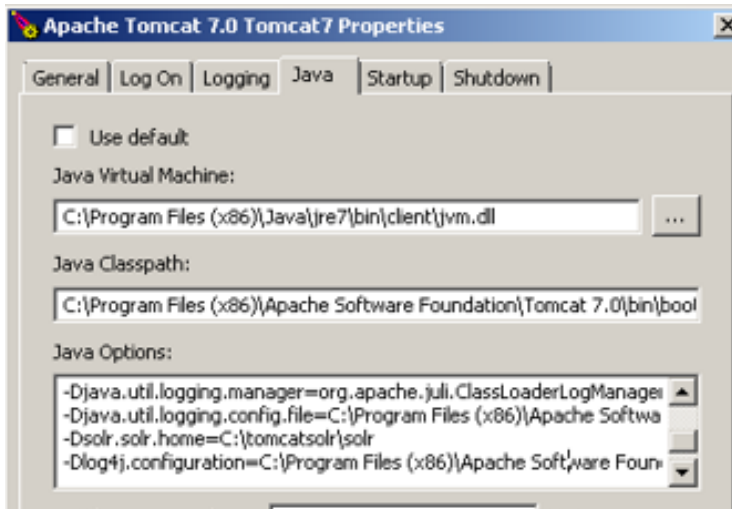
Solr on Tomcat

- **It works!**

  - Check Logging to confirm no warnings or errors
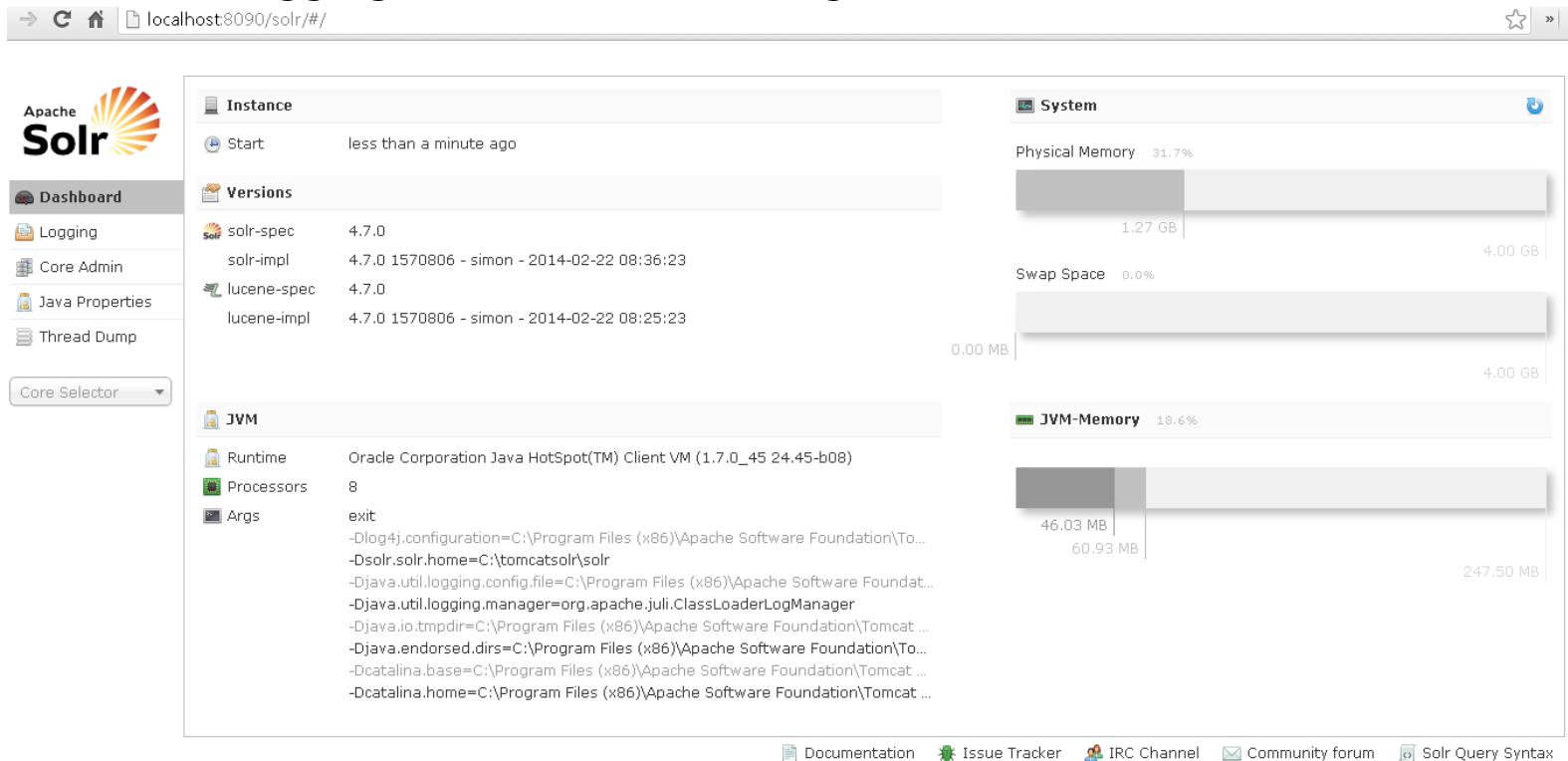
Security

- **Solr does not have security!**

- **Secured at container level**

- **How do I do it?**
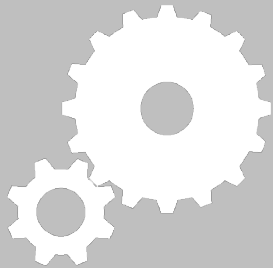 - AWS with rules
  - Many other ways
    - Investigate a few!

Solr Configuration Takeaway

Installation

D
Demo

Admin UI

Tomcat

Takeaway

- **Download & install Solr**

- **Reviewed what's included**

- **Started Solr & indexed data**

- **Ran queries & analyzed a document**

- **Changed port configuration**

- **Reviewed Admin section**

- **Installed in Tomcat**

- **Security at container level**