

Configuring and Running a Solr Search Server



Xavier Morera

PASSIONATE ABOUT TEACHING

@xmorera www.xaviermorera.com



Getting Solr up & Running



Download



Unzip



Run

(Prerequisites)



Running Solr

Can run in Windows/Mac/Linux

Solr 5.x is now a standalone server

- 4.x and older was distributed as a .WAR

You need Java (64 bit recommended for production)

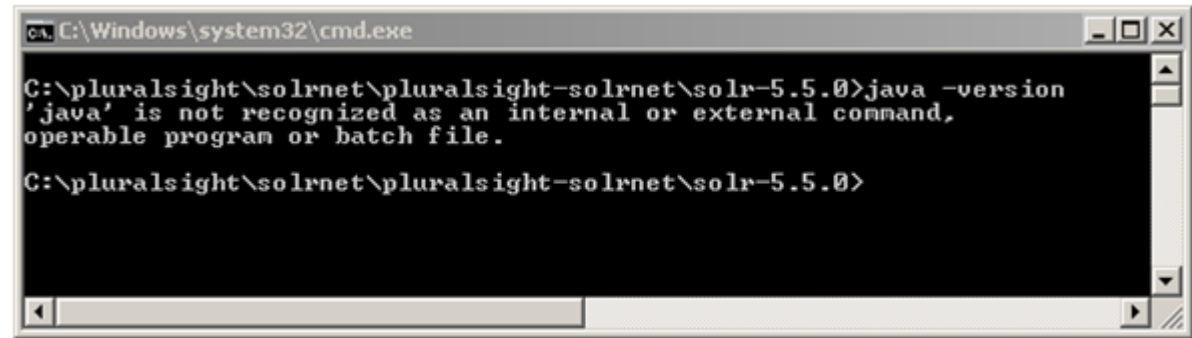
Use Solr 5.5 for this course

Let's get this party started!



Setting up Your Environment

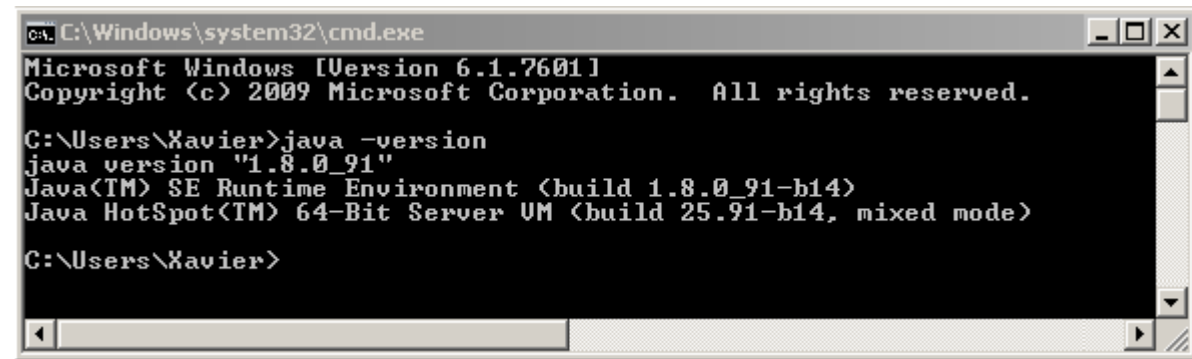
> java -version



```
C:\Windows\system32\cmd.exe

C:\pluralsight\solrnet\pluralsight-solrnet\solr-5.5.0>java -version
'java' is not recognized as an internal or external command,
operable program or batch file.

C:\pluralsight\solrnet\pluralsight-solrnet\solr-5.5.0>
```



```
C:\Windows\system32\cmd.exe

Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Xavier>java -version
java version "1.8.0_91"
Java(TM) SE Runtime Environment (build 1.8.0_91-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.91-b14, mixed mode)

C:\Users\Xavier>
```



Prerequisites



Install JRE

- Java Runtime Environment

Search for 'java runtime'

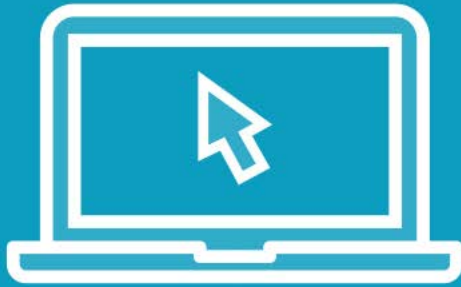
or

<https://java.com/en/download/>

Follow instructions to install



Demo

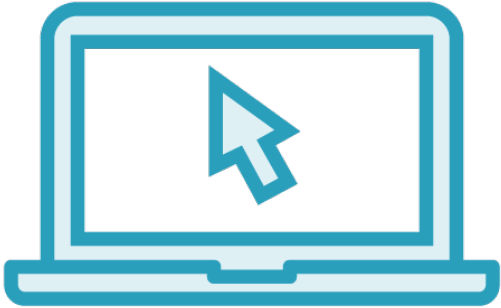


Starting Solr

M3 D1



Demo Summary



We have Solr up and running



It was easy!

We need to add data to be able to search

“Indexing documents”

What is a document? Let's define it

Defining a Document



Xavier
xavier@familiamoreira.com



Using Git with a GUI

by Xavier Morera


Getting started with Git is hard but I will help you smooth out the learning curve, understand Git's mechanics, branching model, and in general be productive much quicker with the aid of a great free tool: Atlassian SourceTree.

[▶ Resume Course](#)

[Table of contents](#)[Description](#)[Transcript](#)[Exercise files](#)[Discussion](#)[Recommended](#)

There is no doubt Git is taking over the world of source control management, mainly in open source but growing rapidly in enterprise as well. The problem usually lies in that Git is not for the faint of heart as it comes with a steep learning curve - it is hard to get started. In this course, I will help you smooth out the learning curve, understand Git's mechanics and branching model, and in general be productive much quicker with the aid of a great free tool: Atlassian SourceTree.

Course author



Xavier Morera

Xavier is an entrepreneur, project manager, technical author, trainer, Certified Scrum Professional & Scrum Master, and Certified Microsoft Professional who has spent a great deal of his career wor...

Course info

Level	Beginner
Rating	★★★★★ (37)
My rating	★★★★★
Duration	2h 38m
Released	8 Mar 2016

[Bookmark](#)



Defining a Document

```
{
  "coursetitle": "Using Git with a GUI",
  "releasedate": "2016-03-08T00:00:00Z",
  "description": "There is no doubt Git is taking over the world of source control management, mainly in open source but growing rapidly in enterprise as well. The problem usually lies in that Git is not for the faint of heart as it comes with a steep learning curve - it is hard to get started. In this course, I will help you smooth out the learning curve, understand Git's mechanics and branching model, and in general be productive much quicker with the aid of a great free tool: Atlassian SourceTree.",
  "tags": [
    "git",
    "github",
    "sourcetree",
    "bitbucket",
    "developer",
    "version-control",
    "source-control",
    "gui"
  ]
}
```

A self contained piece of data made up of a collection of one or more fields that is sent to Solr

The screenshot shows a course page for 'Using Git with a GUI' by Xavier Morera. The page has a dark theme with a blue header. The title 'Using Git with a GUI' is prominently displayed in white. Below the title, it says 'by Xavier Morera'. A brief description follows: 'Getting started with Git is hard but I will help you smooth out the learning curve, understand Git's mechanics, branching model, and in general be productive much quicker with the aid of a great free tool: Atlassian SourceTree.' A 'Resume Course' button is visible. On the right, there's a sidebar with the author's profile (Xavier Morera), course info (Level: Beginner, Rating: 5 stars (37), My rating: 5 stars, Duration: 2h 38m), and a table of contents with tabs for Description, Transcript, Exercise files, Discussion, and Recommended. The Description tab is active, showing the start of the course description.

What do you want to learn?

Xavier
xavier@familiamorera.com

Course author

Xavier Morera

Xavier is an entrepreneur, project manager, technical author, trainer, Certified Scrum Professional & Scrum Master, and Certified Microsoft Professional who has spent a great deal of his career wor...

Course info

Level Beginner

Rating ★★★★★ (37)

My rating ★★★★★

Duration 2h 38m

Table of contents Description Transcript Exercise files Discussion Recommended

Getting started with Git is hard but I will help you smooth out the learning curve, understand Git's mechanics, branching model, and in general be productive much quicker with the aid of a great free tool: Atlassian SourceTree.



Document

Document contains one or more fields

- Not nested

Each field is of a certain type

- Properties: indexed, stored, ...

Single valued or multivalued

Field analysis

- Tokenize, normalize, cleanse, linguistic processing and more

Can use these documents from our app via SolrNet (mapping)

How do you define them in Solr?



Schema When You Want, Schemaless When You Don't

Data Driven Schema

Known as schemaless mode

Construct schema simply by adding data

Solr can guess the type

Explicitly create fields with Schema API

Easy to get started

ManagedIndexSchemaFactory

managed-schema

Reindex after changes!

Schema

Define fields, types and unique key

Reload cores after changes

Error if field does not exist

Still can use dynamic fields

Suitable for production

ClassicIndexSchemaFactory

schema.xml

Reindex after changes!



Schema API

Read and Write access to the schema for each collection or core

Read access to all schema elements

To enable API schema needs to be managed and mutable

Outputs JSON or XML

- Complete schema is XML

Core reload is automatic

- Applied to newly indexed documents

Reindexing required for previously indexed documents



Both Have Pros and Cons

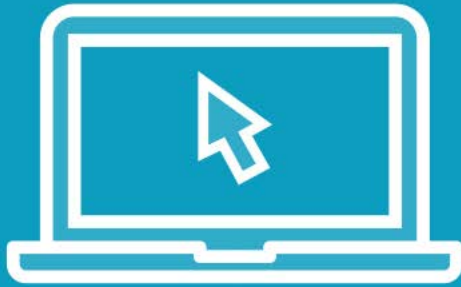
Schema
When You Want

Schemaless
When You Don't

Pick the One That Works for You



Demo

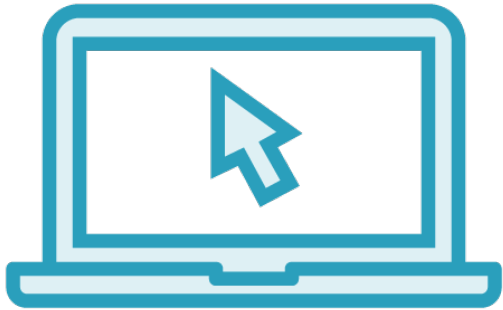


Schemaless Mode and Schema.xml

M3 D2



Demo Summary



Schemaless mode is great for rapid prototyping

Create fields by adding sample data

Or can create via Schema API

Schema.xml suitable for production

- Define each field
- Reload core

Always reindex after making changes

- Don't underestimate planning your schema

Single Instance or Bigger Needs?



SolrCloud

Enables creating durable & highly available clusters

- Using commodity hardware

Layman terms → Scaling up Solr

- “Similar” to master-replica
- Automates lots of manual work using Apache Zookeeper

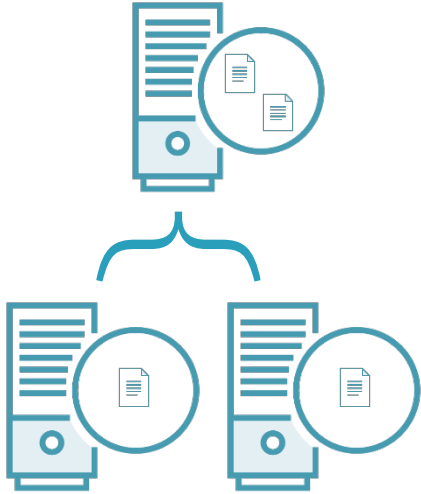
SolrCloud cluster looks like regular Solr to applications

Shards

- Divides the index



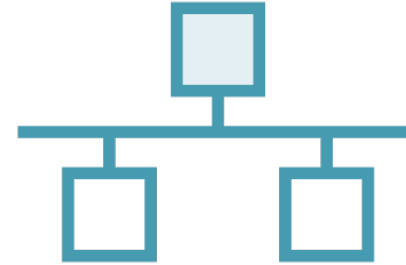
Why Do I Need SolrCloud?



Sharding



Near Real Time
Search
Incremental
Indexing



Query
Distribution
Load Balancing



High Availability



Demo

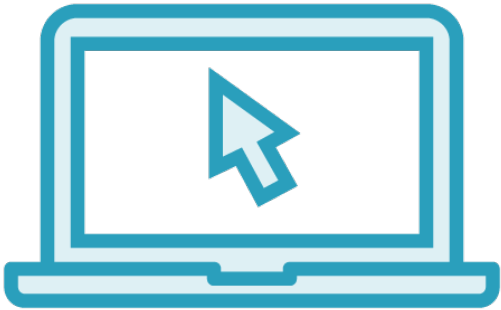


SolrCloud

M3 D3



Demo Summary



Starting SolrCloud is easy

Uses Apache Zookeeper for configuration

It comes with embedded Zookeeper

- Recommended to use standalone

SolrCloud feels like regular Solr to applications

Topic for an entire course

Takeaway



Starting Solr is fast and easy

JRE as prerequisite

Download, unzip and run

Define your documents

- Data driven schema or schema.xml
- Reindex documents after changes

SolrCloud for bigger needs

- Scalability, high availability and more

