

APACHE SOLR

Open Source Search Platform

Background

- Six years of enterprise search consulting experience
- Search platforms are typically deployed within a company firewall
 - File Shares, Intranet Sites
 - SharePoint, Documentum
 - SAP, PLM, Legacy Applications
- Experience with several enterprise search commercial products



Agenda

- Introduce Apache Solr
 - Terminology, Concepts, History, Architecture and Features
- Index Population
 - Schema Design (schema.xml)
 - Feed Payloads
 - Apache Tika
- Index Query
 - Search Protocol
 - Response Payloads
 - Request Handlers (solrconfig.xml)
 - Search Components
- Search-Based Applications

Concepts & Terminology

Apache Lucene – is a full text search engine library written entirely in Java. Lucene is embedded with Solr.

Apache Solr – is an enterprise search platform written in Java. It exposes web services that can manage the lifecycle of documents in the index.

Document – is Lucene/Solr's primary unit of storage – representing a flat collection of fields (no nesting).

Field – definition consists of a name and configurable type (text, integer, double, date).

Core – separate index and configuration. A single server can support multiple cores and it is used for data partitioning. Supports multitenant applications.

Shard – Is a chunk of a larger index. They are created to scale an index horizontally across machines.

SolrCloud – refers to a set of features that enable your search index to be scaled across a cluster of nodes.

Concepts & Terminology

Synonyms – is a query expansion feature where (e.g. MB => megabyte)

Stop Words – are words that should be filtered from index storage and queries

Structured Content – refers to content that has been richly tagged with metadata.

Unstructured Content – MS Office, PDF documents, emails, instant messages, etc.

ACL – access control list used to capture document permissions

Early Binding – an authorization enforcement model where the document ACLs are stored in the index.

Late Binding – an authorization enforcement model where document authorization is not determined until query time.

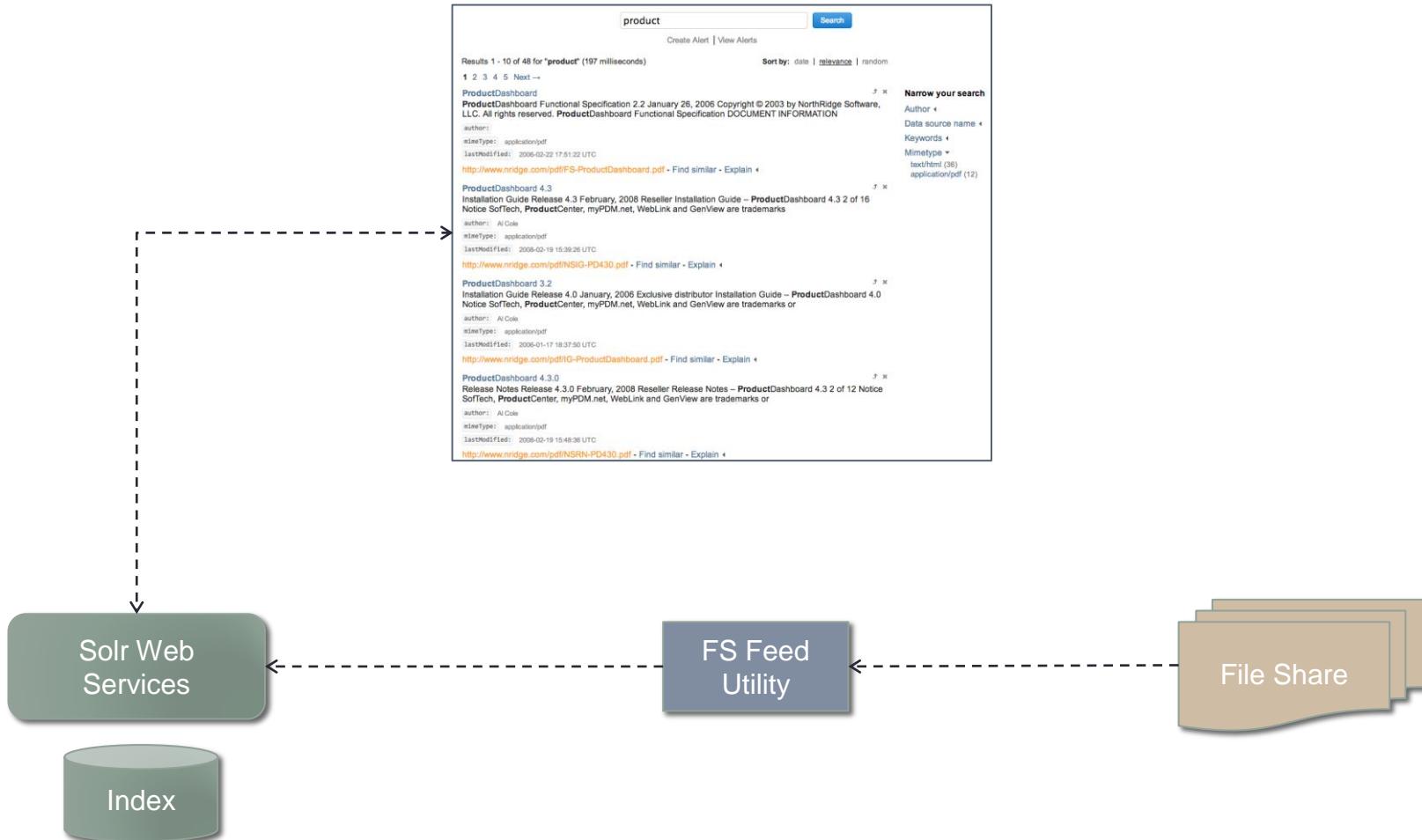
ETL – extract (content source), transform (normalize the data), load (into index)

Search Based Application – built on top of search platforms and they are designed to deliver unified information access.

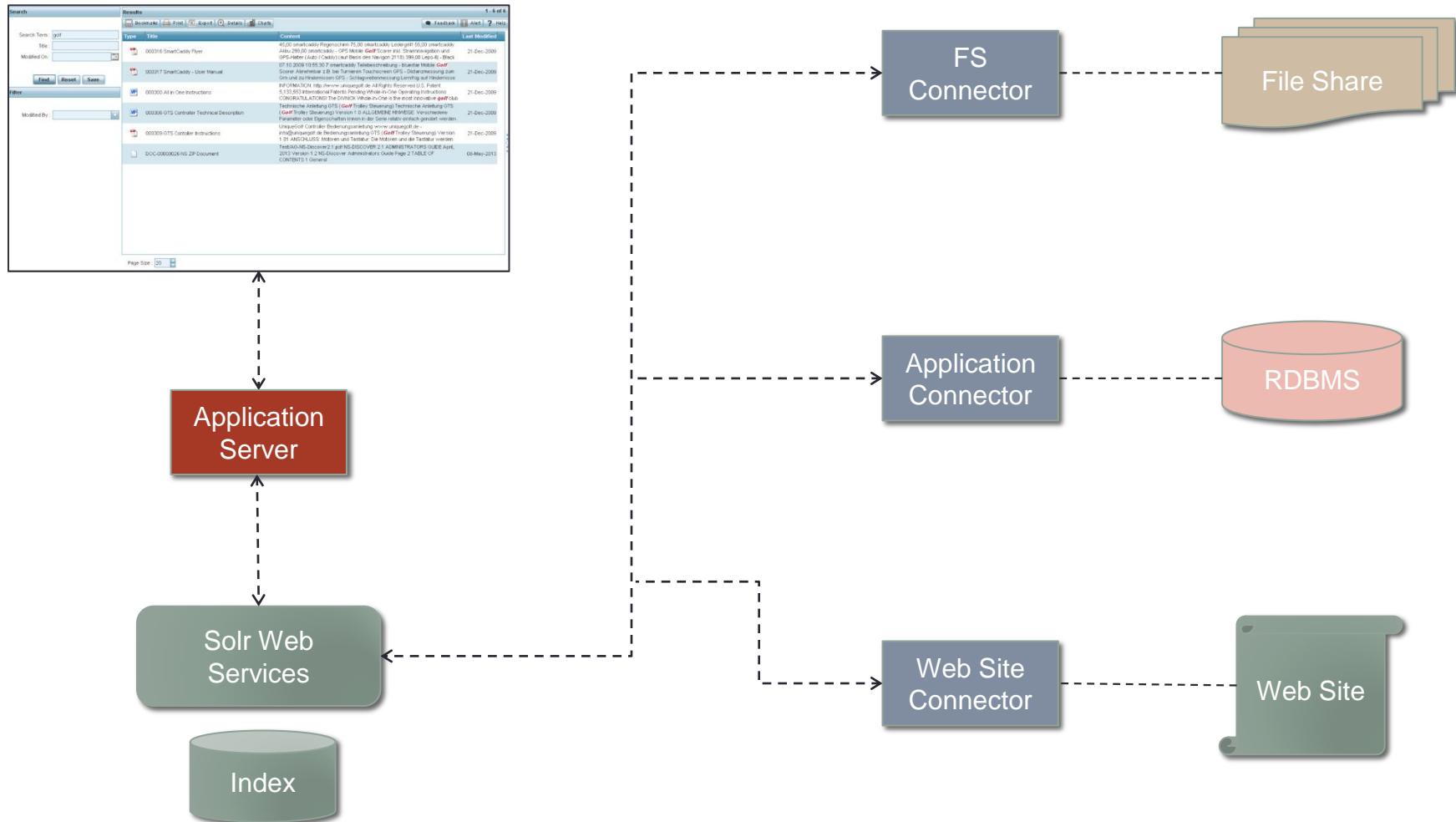
Lucene/Solr History

- Doug Cutting created Lucene in 1999
 - Recognized as a top level Apache Software Foundation project in 2005
- Yonik Seeley created Solr in 2004
 - Recognized as a top level Apache Software Foundation project in 2007
- Apache Lucene and Solr projects merge in 2010
 - Apache Lucene/Solr Release 1.4 in 2011
 - Apache Lucene/Solr Release 3.x in 2012
 - Apache Lucene/Solr Release 4.x in 2013

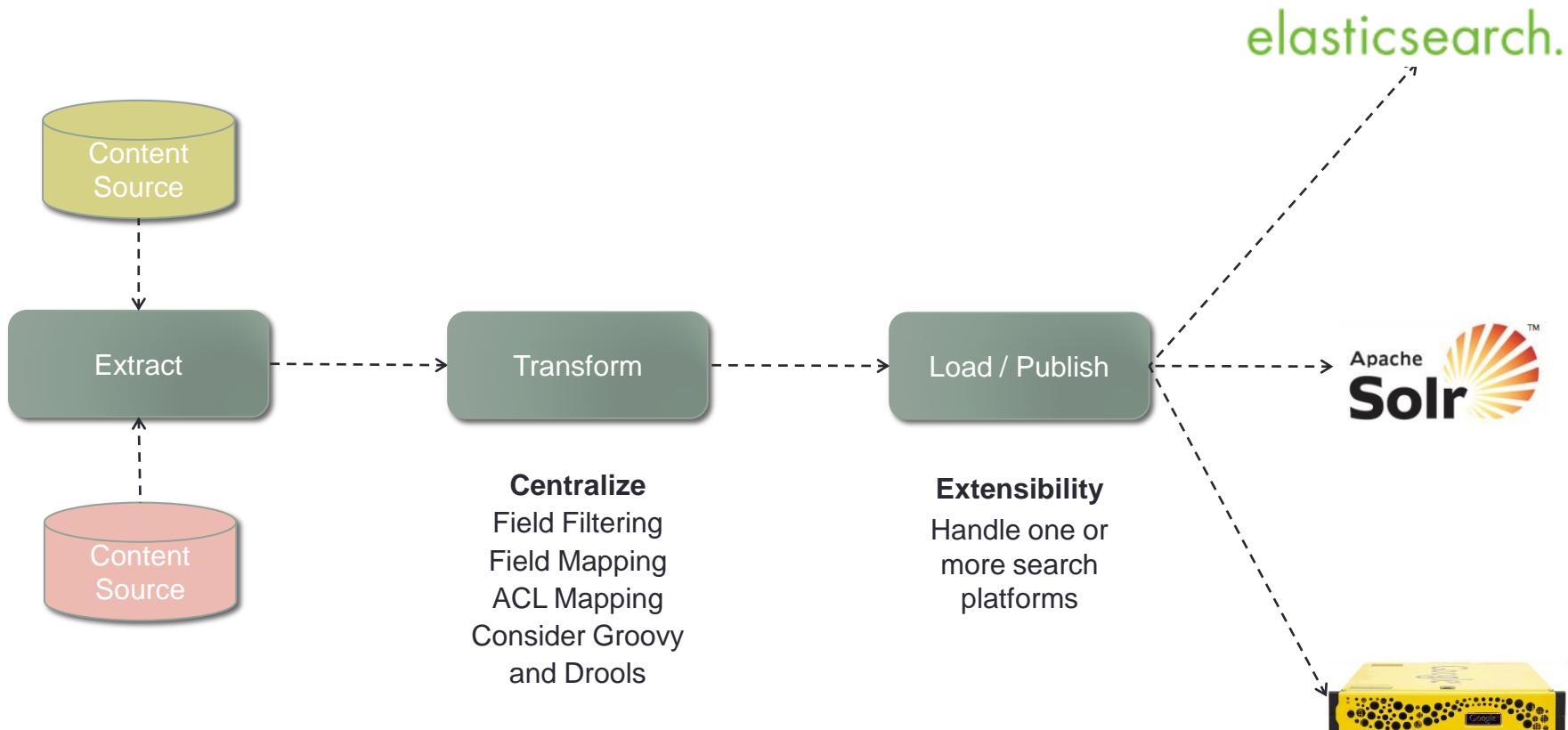
Simple Search Architecture



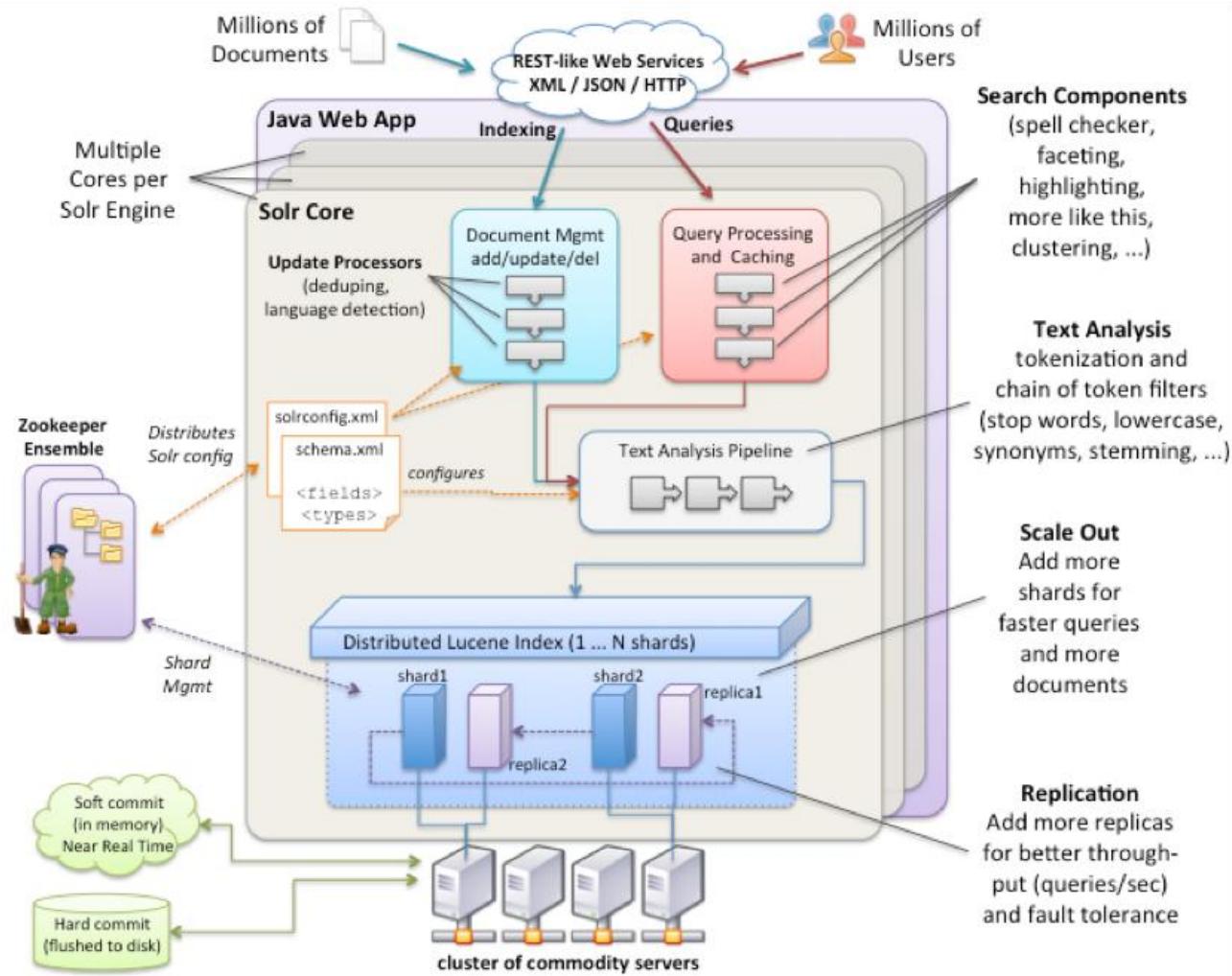
Enterprise Search Architecture



ETL Process



Solr Architecture



Solr Features

Keyword Searching – queries of terms and boolean operators

Ranked Retrieval – sorted by relevancy score (descending order)

Snippet Highlighting – matching terms emphasized in results

Faceting – ability to apply filter queries based on matching fields

Paging Navigation – limits fetch sizes to improve performance

Result Sorting – sort the documents based on field values

Solr Features

Spelling Correction – suggest corrected spelling of query terms

Synonyms – expand queries based on configurable definition list

Auto-Suggestions – present list of possible query terms

More Like This – identifies other documents that are similar to one in a result set

Geo-Spatial Search – locate and sort documents by distance

Scalability – ability to break a large index into multiple shards and distribute indexing and query operations across a cluster of nodes

Solr Feature Example

discover Search

Create Alert | View Alerts

Filters Keywords: Discover Enterprise Remove All

Results 1 - 10 of 32 for "discover" (11 milliseconds)

Sort by: date | relevance | random

1 2 3 4 Next →

NS Discover License Request Form ✖
-> NS Discover License Request NS Discover License Registration Form You must complete the form below in order to receive a valid license file. (required fields are in bold)

mimeType: text/html
<http://www.nr ridge.com/nsd-license.php> - Find similar - Explain ←

Discover Innovator - Search Made Simple ✖
Overview Discover Enterprise Discover Innovator — Data Sheet — User Guide — Release Notes — Administration Guide

mimeType: text/html
<http://www.nr ridge.com/discover-innovator.php> - Find similar - Explain ←

Discover Enterprise - Search Made Simple ✖
Overview Discover Enterprise — Data Sheet — Release Notes — Administration Guide Discover Innovator

mimeType: text/html
<http://www.nr ridge.com/discover-enterprise.php> - Find similar - Explain ←

NorthRidge Software, Enterprise Search Solutions ✖
News Discover Innovator Release 3.0 Discover Enterprise Release 3.0 Aras ACE 2013 Conference Sponsor

lastModified: 2013-12-18 20:13:16 UTC
mimeType: text/html
<http://www.nr ridge.com/> - Find similar - Explain ←

NorthRidge Software, Enterprise Search Solutions ✖
News Discover Innovator Release 3.0 Discover Enterprise Release 3.0 Aras ACE 2013 Conference Sponsor

lastModified: 2013-12-18 20:13:16 UTC
mimeType: text/html
<http://www.nr ridge.com/index.htm> - Find similar - Explain ←

NorthRidge Software, Enterprise Search Solutions ✖
-mail : Product Name : Discover Enterprise Discover Innovator Product

mimeType: text/html
<http://www.nr ridge.com/contact-sales.php> - Find similar - Explain ←

Narrow your search

Data source name ↗

Keywords ↗

- ✖ Discover Enterprise (32)
- Discover Innovator (32)
- Enterprise Search (32)
- Lucene (32)
- LucidWorks Search (32)
- PDM (32)
- PLM (32)
- Product Data Manage... (32)
- Product Lifecycle M... (32)
- Solr (32)

Mimetype ↗

Solr Installation

- Tutorial Available
 - https://lucene.apache.org/solr/4_6_1/tutorial.html
- Download
- Installation
- Index Population
 - Sample Documents
 - Feed Upload
 - Document Updates
 - Document Deletion
- Querying
 - Keywords
 - Facets

Solr Tutorial

Overview

This document covers the basics of running Solr using an example schema, and some sample data.

Requirements

To follow along with this tutorial, you will need...

1. Java 1.6 or greater. Some places you can get it are from [Oracle](#), [Open JDK](#), or [IBM](#).
 - Running `java -version` at the command line should indicate a version number starting with 1.6.
 - Gnu's GCJ is not supported and does not work with Solr.
2. A [Solr release](#).

Getting Started

Please run the browser showing this tutorial and the Solr server on the same machine so tutorial links will correctly point to your Solr server.

Begin by unzipping the Solr release and changing your working directory to be the "`example`" directory. (Note that the base directory name may vary with the version of Solr downloaded.) For example, with a shell in UNIX, Cygwin, or MacOS:

```
user:~/solr$ ls  
solr-nightly.zip  
user:~/solr$ unzip -q solr-nightly.zip  
user:~/solr$ cd solr-nightly/example/
```

Solr can run in any Java Servlet Container of your choice, but to simplify this tutorial, the example index includes a small installation of Jetty.

To launch Jetty with the Solr WAR, and the example configs, just run the `start.jar` ...

```
user:~/solr/example$ java -jar start.jar  
2012-06-06 15:25:59.815:INFO:oejs.Server:jetty-8.1.2.v20120308  
2012-06-06 15:25:59.834:INFO:oejdp.ScanningAppProvider:Deployment monitor .../solr/example/webapps at interval 0  
2012-06-06 15:25:59.839:INFO:oejd.DeploymentManager:Deployable added: .../solr/example/webapps/solr.war  
...  
Jun 6, 2012 3:26:03 PM org.apache.solr.core.SolrCore registerSearcher  
INFO: [collection1] Registered new searcher Searcher@7527e2ee main[StandardDirectoryReader(segments_1:1)]
```

This will start up the Jetty application server on port 8983, and use your terminal to display the logging information from Solr.

You can see that the Solr is running by loading <http://localhost:8983/solr/> in your web browser. This is the main starting point for Administering Solr.

Indexing Data

Your Solr server is up and running, but it doesn't contain any data. You can modify a Solr index by POSTing commands to Solr to add (or update) documents, delete documents, and commit pending adds and deletes. These commands can be in a [variety of formats](#).

The `exampledocs` directory contains sample files showing of the types of commands Solr accepts, as well as a java utility for posting them from the command line (a `post.sh` shell script is also available, but for this tutorial we'll use the cross-platform Java client. Run `java -jar post.jar -h` so see its various options).

Schema Document Design

- Information is captured in a document container.
- Each document consists of a list of fields.
- One field must uniquely identify each document in the index.
- Which fields will your users want to search on?
- What fields should be displayed in your search results?
- Structured versus unstructured content.
- Security model – public, ACLs, early versus late binding.

The New York Times @nytimes · 1h
The Monuments of Tech nyti.ms/1gKgseh

[Reply](#) [Retweet](#) [Favorite](#) [More](#)

[Hide summary](#)

The New York Times

The Monuments of Tech

The workplaces of Facebook, Twitter and Google may look whimsical. But each design is calculated to mirror its company's values and culture.

[View on web](#)

RETWEETS 28 FAVORITES 24



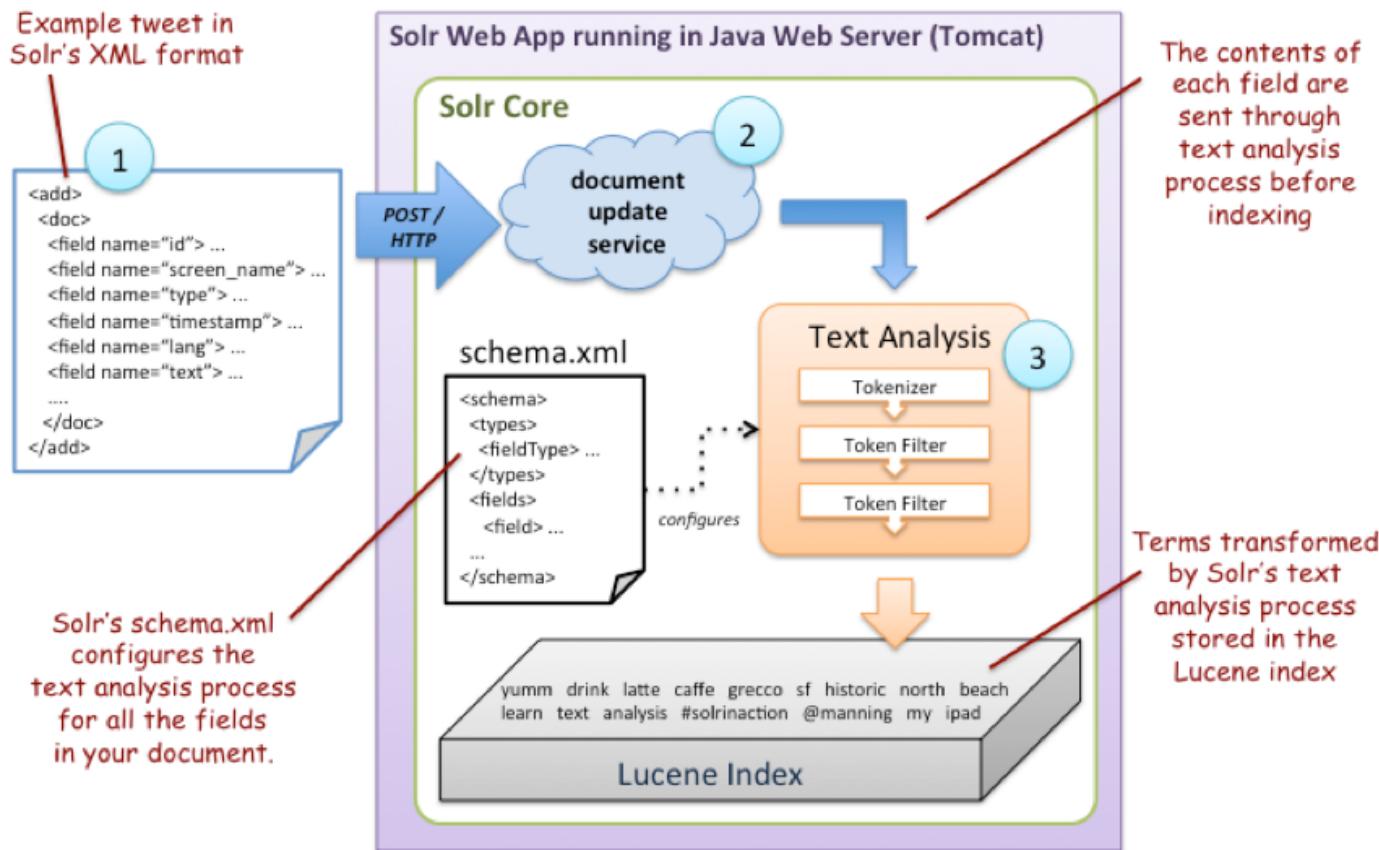
TE Connectivity - 206044-1
206044-1, 14 position Circular Plastic Connector Plug Shell Size 17 UL 94V-0

Distributor	SKU	Stock	MOQ	x1	x100
* Newark	46F320	4,005	1	USD \$ 4.220	3.310
* Avnet Express	206044-1	33,850	1	USD \$ 5.877	3.671
* Verical	206044-1	8,700	..	USD \$
* Digi-Key	A1358-ND	401	1	USD \$ 6.120	4.845
* Arrow	206044-1	553	1	USD \$ 2.890	..

[show more...](#)

* [authorized - manufacturer info](#)

Indexing Process



Inverted Index

Original documents		Lucene's inverted index		
Doc #	Content field	Term	Doc #	(Continued)...
1	A Fun Guide to Cooking	a	1,3,4,5,6,7,8	...
2	Decorating Your Home	becoming	8	guide
3	How to Raise a Child	beginner's	6	home
4	Buying a New Car	buy	9	house
5	Buying a New Home	buying	4,5,6	how
6	The Beginner's Guide to Buying a House	car	4	new
7	Purchasing a Home	child	3	owner
8	Becoming a New Home owner	cooking	1	purchasing
9	How to Buy Your First House	decorating	2	raise
		first	9	the
		fun	1	to
		your

Schema Configuration (schema.xml)

```
<schema name="example" version="1.5">
  <fields>
    <field name="id" type="string" indexed="true" stored="true" required="true" multiValued="false" />

    <field name="title" type="text_general" indexed="true" stored="true" multiValued="true"/>
    <field name="subject" type="text_general" indexed="true" stored="true"/>
    <field name="description" type="text_general" indexed="true" stored="true"/>
    <field name="comments" type="text_general" indexed="true" stored="true"/>
    <field name="author" type="text_general" indexed="true" stored="true"/>
    <field name="category" type="text_general" indexed="true" stored="true"/>
    <field name="last_modified" type="date" indexed="true" stored="true"/>
    <field name="links" type="string" indexed="true" stored="true" multiValued="true"/>
    <field name="content" type="text_general" indexed="false" stored="true" multiValued="true"/>
    <field name="text" type="text_general" indexed="true" stored="false" multiValued="true"/>

    <field name="weight" type="double" indexed="true" stored="true"/>
    <field name="price" type="float" indexed="true" stored="true"/>
    <field name="popularity" type="int" indexed="true" stored="true" />
    <field name="inStock" type="boolean" indexed="true" stored="true" />
    <field name="store" type="location" indexed="true" stored="true"/>

    <dynamicField name="*_s" type="string" indexed="true" stored="true" />
    <dynamicField name="*_dt" type="date" indexed="true" stored="true"/>
  </fields>

  <uniqueKey>id</uniqueKey>

  <copyField source="title" dest="text"/>
  <copyField source="author" dest="text"/>
  <copyField source="description" dest="text"/>
  <copyField source="keywords" dest="text"/>
  <copyField source="content" dest="text"/>
```

Schema Configuration (schema.xml)

```
<types>
  <fieldType name="string" class="solr.StrField" sortMissingLast="true" />
  <fieldType name="boolean" class="solr.BoolField" sortMissingLast="true"/>
  <fieldType name="int" class="solr.TrieIntField" precisionStep="0" positionIncrementGap="0"/>
  <fieldType name="float" class="solr.TrieFloatField" precisionStep="0" positionIncrementGap="0"/>
  <fieldType name="long" class="solr.TrieLongField" precisionStep="0" positionIncrementGap="0"/>
  <fieldType name="double" class="solr.TrieDoubleField" precisionStep="0" positionIncrementGap="0"/>

  <fieldType name="text_general" class="solr.TextField" positionIncrementGap="100">
    <analyzer type="index">
      <tokenizer class="solr.StandardTokenizerFactory"/>
      <filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt" />
      <filter class="solr.LowerCaseFilterFactory"/>
    </analyzer>
    <analyzer type="query">
      <tokenizer class="solr.StandardTokenizerFactory"/>
      <filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt" />
      <filter class="solr.SynonymFilterFactory" synonyms="synonyms.txt" ignoreCase="true" expand="true"/>
      <filter class="solr.LowerCaseFilterFactory"/>
    </analyzer>
  </fieldType>
</types>
</schema>
```

Schema Design: [Solr Unleashed Tutorial](#)

Analyzers, Tokenizers and Filters: [Solr Reference Documentation](#) [Solr Unleashed Tutorial](#)

Document Text Extraction

NS-Discover Administrator's Guide

TABLE OF CONTENTS

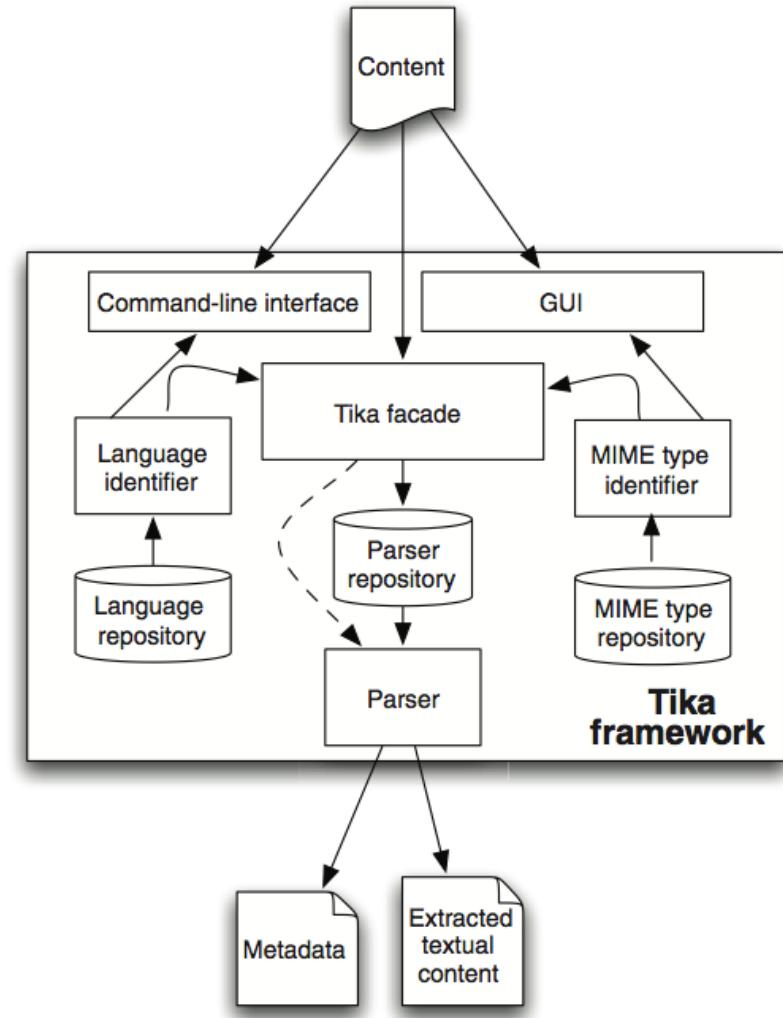
1 General Information.....	4
1.1 Objective	4
1.2 Supported File Formats.....	4
2 Installation	5
2.1 Platform Support	5
2.2 Server Sizing.....	5
2.3 Preparation	6
2.3.1 Identify Host Servers.....	6
2.3.2 Register and Download Software.....	7
2.4 Upgrade Installation Process	8
2.4.1 Shutdown NSD Services.....	8
2.4.2 Backup the Current Install Folders	8
2.4.3 Uninstall NS-Discover	9
2.4.4 Install NS-Discover	9
2.4.5 Update Configuration Files	9
2.4.6 IE Browser	10
2.5 Installation Process	11
2.5.1 NSD-RDBMS Install	11
2.5.2 NSD-Enterprise Install	14
2.5.3 Index Server Install	17
2.5.4 UI Server Install (Optional)	21
2.5.5 Discover Innovator Connector Install	26
2.6 License Management	31
3 Post Installation Configuration.....	32
3.1 Starting Connector Services	32
3.2 Aras Innovator Configuration.....	33
3.3 Discover Innovator Connector Configuration	36
3.3.1 Adding Innovator Item Types	37
3.3.2 Adding Innovator Item Properties	38
3.4 Reset Index and Run Full Feed.....	42
3.4.1 Reset the Index	42
3.4.2 Run a Full Feed.....	42
4 Troubleshooting.....	43

GENERAL INFORMATION 1.1 OBJECTIVE The objective of this document is to provide the NS-Discover administrator with all the information required to install and maintain all NS-Discover products. This guide includes instruction for the installation and maintenance of: Discover Enterprise NS-Discover Index Server Solr 4.1 Data Source Discover Innovator Connector Discover Innovator Client NS-Discover RDBMS Server 1.2 SUPPORTED FILE FORMATS The following table identifies all the file formats whose content will be included as searchable text from NS-Discover clients. FileType File Extension Notes MS Word .doc .docs MS Excel .xls , .xlsx MS PowerPoint .ppt , .ppbx MS Visio .vsd MS Project .ppt , .pptx Extracts the Title property only MS Publisher .pub HTML .htm XML .xml Indexes arbitrary text Adobe Acrobat .pdf Plain Text .txt Rich Text Format .rtf Comma Separated Values .csv OpenXML .xml Microsoft's XML format ZIP Archive .zip Indexes archive file names only NS-Discover Administrator's Guide Page 5.2 INSTALLATION The Administrator is advised to review the entirety of this document in order to properly understand and be prepared for a successful installation. 2.1 PLATFORM SUPPORT The following table identifies the platforms upon which you can deploy and run the NS-Discover suite of products. Component OS Version Index Server Windows Server 2003, 2008 Windows XP, 7 Discover Innovator Connector Windows Server 2003, 2008 Windows XP, 7 Discover Innovator Client Internet Explorer 8, 9* * Please reference section 4.1.2 if you are running IE 9. 2.2 SERVER SIZING The Index Server is typically sized based on the number of documents it will process. However, there can be many other variables involved (e.g. queries per second, average document sizes, etc.), so the table below should be considered a starting point for hardware resource planning. Documentation Disk Space (GB) RAM (GB) CPU Cores 10,000 5.2 2 100,000 5 4 2 1,000,000 50 6 4 Although the number of users concurrently searching against the index will affect the performance, the estimates provided above should be sufficient to support hundreds of users. In addition, you should note that scaling vertically and horizontally is also possible with this architecture. Given this flexibility, we recommend keeping the initial configuration simple and deferring the use of a more sophisticated approach (e.g. master / slave machines, load balancers, etc.) until the production usage patterns justify that level of investment. NS-Discover Administrator's Guide Page 6.2.3 PREPARATION The following steps must be taken before beginning the installation process. 2.3.1 IDENTIFY HOST SERVERS NS-Discover is comprised of four server components, which can be installed on the same server or can be distributed amongst multiple servers. The following section provides guidance for installing the server components. 2.3.1 INDEX SERVER The NS-Discover Index Server is based on the Apache open source Lucene/Solr search platform. This component of the architecture is responsible for providing a repository for all the content in a centralized index, optimized for the search and retrieval of results. This platform is scalable to millions of documents and can be configured to meet specific customer requirements. The default installation of the Index Server includes the UI Server. If you are installing the NS-Discover Index Server on the same machine that hosts Aras Innovator, then there is no need to install the separate UI Server. 2.3.1.2 UI SERVER The NS-Discover UI Server hosts the client interface for Discover Innovator and must be installed on the Aras Innovator host machine. If the Index Server is installed on the same host as Aras Innovator, then the UI Server that is delivered with that installer will be used. If you install the Index Server on a host other than where Aras Innovator is installed, then a separate UI Server installation will be required on the Aras Innovator host machine. 2.3.1.3 DISCOVER INNOVATOR CONNECTOR The Discover Innovator Connector is responsible for extracting content from a populated Aras Innovator instance and feeding the data into the NS-Discover Index. The connector is also responsible for retrieving and opening files that have been requested by the user from the Discover Innovator Client. The Discover Innovator Connector must be installed on the same machine where the Innovator Vault Server resides. An Innovator account is required to install the Discover Innovator Connector. You must ensure that the Innovator account used for the installation has read access to all files stored within the Aras Innovator vault folders you wish to index. 2.3.1.4 RDBMS SERVICE The Discover RDBMS Service is required to support the following NS-Discover features. Discover User Portal Saved Searches Document Bookmarks Transaction monitoring and reporting NS-Discover Administrator's Guide Page 7.2.3.2 REGISTER AND DOWNLOAD SOFTWARE A license file for the Discover Innovator Connector can be obtained by completing the license request form on the NorthRidge Software web site. You can locate the license request form within the "Contact Us" menu on the main web page. Simply select the "License" menu entry and you will be directed to a request form like the one shown below. After you populate the form and submit it, you will receive a reply email message that contains: license file links to download the software links to the installation instructions NS-Discover Administrator's Guide Page 8.2.4 UPGRADE INSTALLATION PROCESS Follow these steps before running the NS-Discover installation program if you are upgrading from a previous version. This process will remove your existing index and repopulate the index with a full feed. Contact NorthRidge Software if you wish to preserve your existing index while upgrading to this release. 2.4.1 SHUTDOWN NSD SERVICES Open the Windows Services dialog from the server on which each of these services are currently running. They can be located on a single server or may be distributed across servers. Stop each of the services before proceeding. NSD-AI NSD-Enterprise NSD-Index NSD-RDBMS The number of services running in your environment may vary depending on the products you currently have installed. 2.4.2 BACKUP THE CURRENT INSTALL FOLDERS Navigate to the current NS-Discover installation folder (`NS_HOME`). Copy the following folders to a separate backup folder. [NS_Home] NSD-AI [NS_Home] NSD-Index [NS_Home] NSD-Enterprise [NS_Home] NSD-RDBMS NS-Discover Administrator's Guide Page 9.2.4.3 UNINSTALL NS-DISCOVER

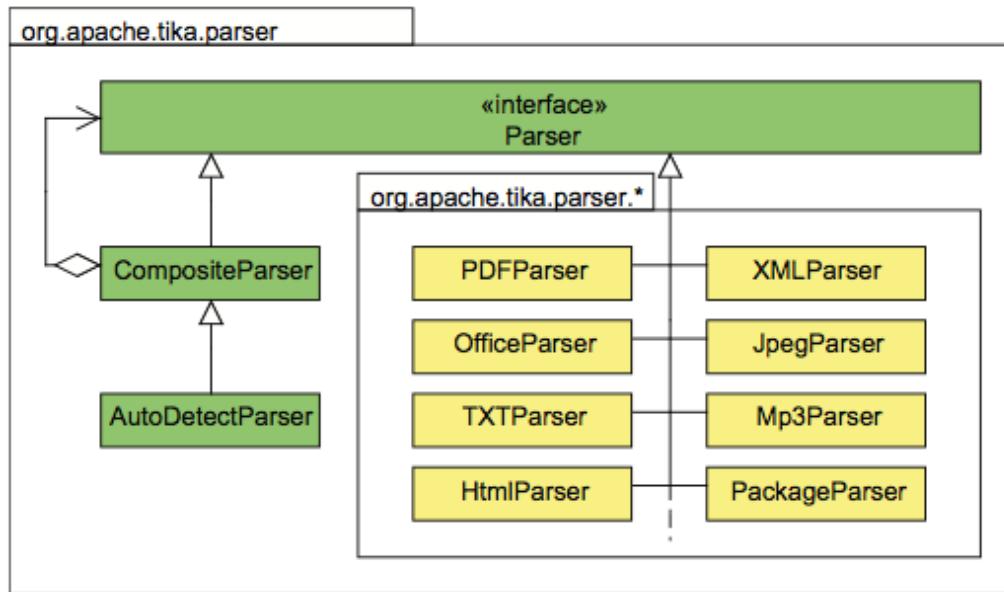
Apache Tika Framework

Supported Document Formats

- HyperText Markup Language
- XML and derived formats
- Microsoft Office document formats
- OpenDocument Format
- Portable Document Format
- Electronic Publication Format
- Rich Text Format
- Compression and packaging formats
- Text formats
- Audio formats
- Image formats
- Video formats
- Java class files and archives
- The mbox format



Apache Tika Framework



```
File document = new File("example.doc");
String content = new
Tika().parseToString(document);
System.out.print(content);
```

```
Parser tikaParser = new AutoDetectParser();
ParseContext parseContext = new ParseContext();
Parser recursiveMetadataParser = new RecursiveMetadataParser(new AutoDetectParser());
parseContext.set(Parser.class, recursiveMetadataParser);
```

```
WriteOutContentHandler writeOutContentHandler = new WriteOutContentHandler(aWriter, mMaxContentSize);
tikaParser.parse(inputStream, writeOutContentHandler, tikaMetaData, parseContext);
```

Solr Document

```
<add>
<doc>
  <field name="id">SP2514N</field>
  <field name="name">Samsung SpinPoint P120 SP2514N - hard drive - 250 GB - ATA-133</field>
  <field name="manu">Samsung Electronics Co. Ltd.</field>
  <field name="cat">electronics</field>
  <field name="cat">hard drive</field>
  <field name="features">7200RPM, 8MB cache, IDE Ultra ATA-133</field>
  <field name="features">NoiseGuard, SilentSeek technology, Fluid Dynamic Bearing (FDB) motor</field>
  <field name="price">92</field>
  <field name="popularity">6</field>
  <field name="inStock">true</field>
  <field name="manufacturedate_dt">2006-02-13T15:26:37Z</field>
  <!-- Near Oklahoma city -->
  <field name="store">35.0752,-97.032</field>
</doc>

<doc>
  <field name="id">6H500F0</field>
  <field name="name">Maxtor DiamondMax 11 - hard drive - 500 GB - SATA-300</field>
  <field name="manu">Maxtor Corp.</field>
  <field name="cat">electronics</field>
  <field name="cat">hard drive</field>
  <field name="features">SATA 3.0Gb/s, NCQ</field>
  <field name="features">8.5ms seek</field>
  <field name="features">16MB cache</field>
  <field name="price">350</field>
  <field name="popularity">6</field>
  <field name="inStock">true</field>
  <!-- Buffalo store -->
  <field name="store">45.17614,-93.87341</field>
  <field name="manufacturedate_dt">2006-02-13T15:26:37Z</field>
</doc>
</add>
```

SolrJ Library – Document Add

```
private void addSamplePart()
    throws IOException, SolrServerException
{
    HttpSolrServer solrServer = new HttpSolrServer("http://localhost:8983/solr/ce");
    solrServer.setParser(new XMLResponseParser());

    SolrInputDocument solrInputDocument = new SolrInputDocument();

    solrInputDocument.addField("id", "A1538104802");
    solrInputDocument.addField("datasheetTitle", "A 4342, DATASHEET (2011/12/13)");
    solrInputDocument.addField("datasheetUrlLatest", "http://images.ihcontent.net/TXIIIS133592-1.pdf");
    solrInputDocument.addField("objectId", "1538104802");
    solrInputDocument.addField("partCategory", "Operational Amplifier");
    solrInputDocument.addField("partDescription", "QUAD OP-AMP, 6000 uV OFFSET-MAX, 1 MHz BAND WIDTH");
    solrInputDocument.addField("partType", "Analog IC");
    solrInputDocument.addField("pubDate", "2011-12-13T00:00:00Z");
    solrInputDocument.addField("status", "ACTIVE");
    solrInputDocument.addField("manufacturerName", "EXAS INSTRUMENTS INC");
    solrInputDocument.addField("packageShape", "RECTANGULAR");
    solrInputDocument.addField("packageStyle", "SMALL OUTLINE");
    solrInputDocument.addField("technology", "CMOS");
    solrInputDocument.addField("terminalFinish", "NICKEL PALLADIUM GOLD");

    UpdateResponse updateResponse = solrServer.add(solrInputDocument);
    if (updateResponse.getStatus() == 0)
        solrServer.commit();
}
```

Solr Dashboard

Apache Solr

Dashboard

Core Selector

Instance
Start about 6 hours ago

Versions

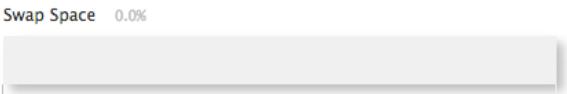
Module	Version	Build Date
solr-spec	4.6.1	
solr-impl	4.6.1	1560866 - mark - 2014-01-23 20:21:50
lucene-spec	4.6.1	
lucene-impl	4.6.1	1560866 - mark - 2014-01-23 20:11:13

JVM
Runtime Oracle Corporation Java HotSpot(TM) 64-Bit Server VM (1.7.0_51 24.51-b...
Processors 8

System 1.00 1.21 1.29

Physical Memory 71.6%

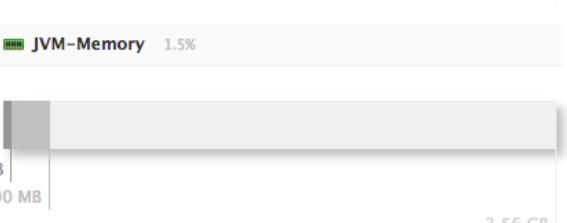

Used	Total
11.46 GB	16.00 GB

Swap Space 0.0%


Used	Total
0.50 MB	1.00 GB

File Descriptor Count 1.5%


Used	Total
150	10240

JVM-Memory 1.5%


Used	Total
56.14 MB	310.00 MB

[Documentation](#) [Issue Tracker](#) [IRC Channel](#) [Community forum](#) [Solr Query Syntax](#)

Query Parameters

Parameter	Description
q	Main query parameter; documents are scored by their similarity to terms in this parameter.
fq	Filter query; restricts the result set to documents matching this filter but doesn't affect scoring.
start	Specifies the starting offset for a page for results; uses 0-based indexing. Start should be incremented by the page size to advance to the next page.
rows	Page size; restricts the number of results returned per page.
sort	Specifies the sort field and sort order; supports ascending (asc) and descending (des).
fl	List of fields to return for each document in the result set.
wt	Response-writer type; governs the format of the response.

Query Syntax Examples

Equal

title:discover title:"discover enterprise"

Not Equal

-title:discover

In Set

id:(100 OR 200 OR 300)

Not In Set

-id:(100 OR 200 OR 300)

String Data Type

Starts With

title:discover*

Contains

title:*discover*

Ends With

title:discover*

Numeric Data Type

Greater Than

price:[100 TO *]

Less Than

price:[* TO 100]

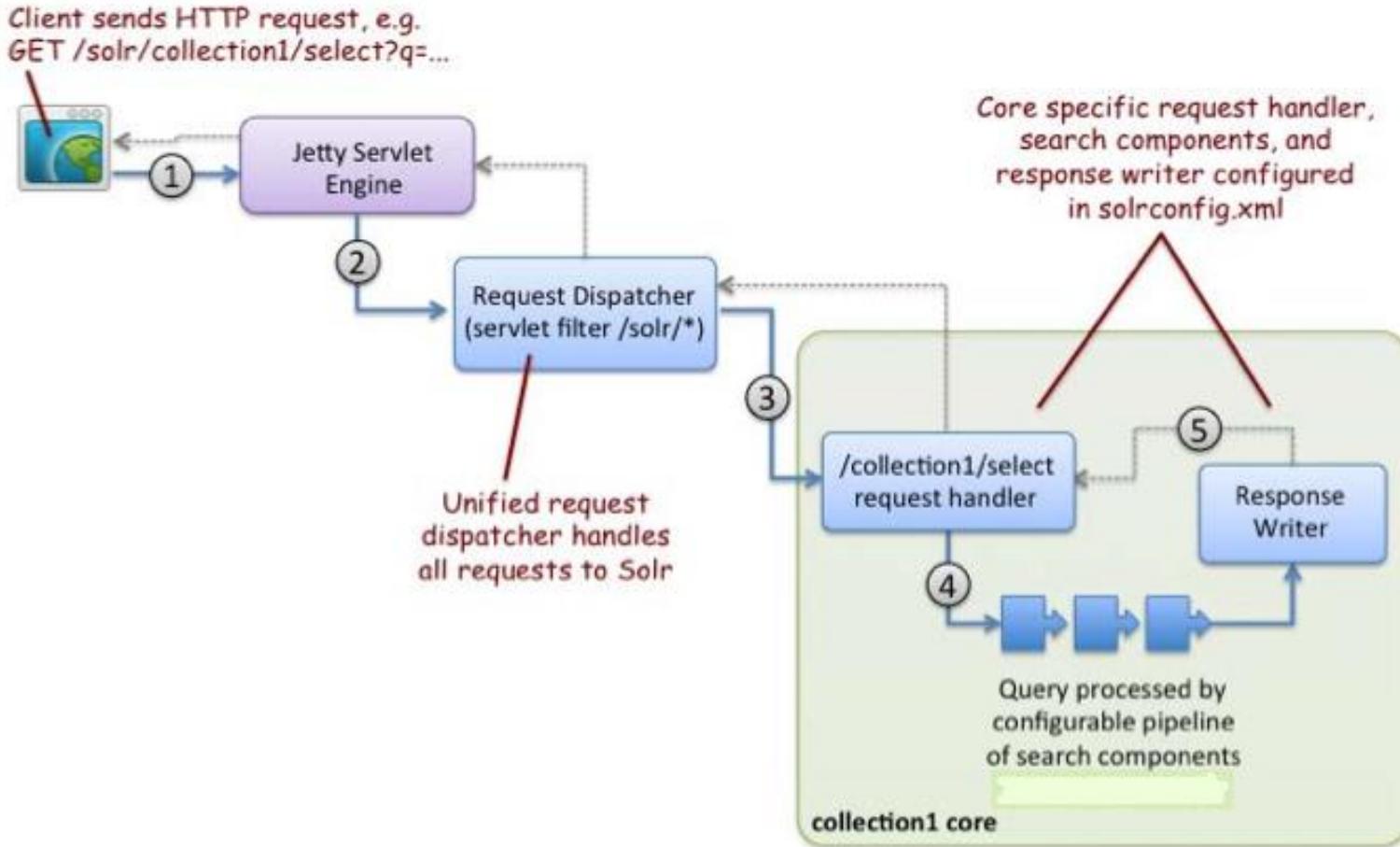
Between

price:[100 TO 500]

Not Between

-price:[100 TO 500]

Index Query



Request Configuration (solrconfig.xml)

```
<requestHandler name="/browse" class="solr.SearchHandler">
  <lst name="defaults">
    <str name="echoParams">explicit</str>

    <!-- VelocityResponseWriter settings -->
    <str name="wt">velocity</str>
    <str name="v.template">browse</str>
    <str name="v.layout">layout</str>
    <str name="title">Solritas</str>

    <!-- Query settings -->
    <str name="defType">edismax</str>
    <str name="qf">
      text^0.5 features^1.0 name^1.2 sku^1.5 id^10.0 manu^1.1 cat^1.4
      title^10.0 description^5.0 keywords^5.0 author^2.0 resourcename^1.0
    </str>
    <str name="df">text</str>
    <str name="mm">100%</str>
    <str name="q.alt">*:*</str>
    <str name="rows">10</str>
    <str name="fl">*,score</str>

    <str name="mlt.qf">
      text^0.5 features^1.0 name^1.2 sku^1.5 id^10.0 manu^1.1 cat^1.4
      title^10.0 description^5.0 keywords^5.0 author^2.0 resourcename^1.0
    </str>
    <str name="mlt.fl">text,features,name,sku,id,manu,cat,title,description,keywords,author,resourcename</str>
    <int name="mlt.count">3</int>
```

Request Configuration

```
<str name="mlt.qf">
  text^0.5 features^1.0 name^1.2 sku^1.5 id^10.0 manu^1.1 cat^1.4
  title^10.0 description^5.0 keywords^5.0 author^2.0 resourcename^1.0
</str>
<str name="mlt.fl">text,features,name,sku,id,manu,cat,title,description,keywords,author,resourcename</str>
<int name="mlt.count">3</int>

<!-- Faceting defaults -->
<str name="facet">on</str>
<str name="facet.field">cat</str>
<str name="facet.field">manu_exact</str>
<str name="facet.field">content_type</str>
<str name="facet.mincount">1</str>

<!-- Highlighting defaults -->
<str name="hl">on</str>
<str name="hl.fl">content features title name</str>
<str name="hl.encoder">html</str>
<str name="hl.simple.pre">&lt;b&gt;</str>
<str name="hl.simple.post">&lt;/b&gt;</str>

<!-- Spell checking defaults -->
<str name="spellcheck">on</str>
<str name="spellcheck.extendedResults">false</str>
<str name="spellcheck.count">5</str>
</lst>

<!-- append spellchecking to our list of components -->
<arr name="last-components">
  <str>spellcheck</str>
</arr>
</requestHandler>
```

SolrJ Library – Document Query

```
private void querySamplePart()
    throws IOException, SolrServerException
{
    HttpSolrServer solrServer = new HttpSolrServer("http://localhost:8983/solr/ce");
    solrServer.setParser(new XMLResponseParser());

    SolrQuery solrQuery = new SolrQuery();
    solrQuery.setRequestHandler("getPart");

    solrQuery.setQuery("QUAD 6000");
    solrQuery.addFilterQuery("technology:CMOS", "status:ACTIVE");
    solrQuery.setFields("id", "datasheetTitle", "partCategory", "partDescription", "manufacturerName");
    solrQuery.setStart(0);
    solrQuery.set("defType", "edismax");

    QueryResponse queryResponse = solrServer.query(solrQuery);
    SolrDocumentList docResults = queryResponse.getResults();
    for (int i = 0; i < docResults.size(); ++i)
        System.out.println(docResults.get(i));
}
```

Solritas



Type of Search: [Simple](#) [Spatial](#) [Group By](#)

Find:

[Submit](#)[Reset](#)

Boost by Price

Field Facets

cat

[electronics](#) (14)

[currency](#) (4)

[memory](#) (3)

[connector](#) (2)

[graphics card](#) (2)

[hard drive](#) (2)

[monitor](#) (2)

[search](#) (2)

[software](#) (2)

[camera](#) (1)

[copier](#) (1)

[multifunction printer](#) (1)

[music](#) (1)

[printer](#) (1)

[scanner](#) (1)

manu_exact

[Apache Software Foundation](#) (2)

[Belkin](#) (2)

[Canon Inc.](#) (2)

[Corsair Microsystems Inc.](#) (2)

[A-DATA Technology Inc.](#) (1)

[ASUS Computer Inc.](#) (1)

[ATI Technologies](#) (1)

[Apple Computer Inc.](#) (1)

[Bank of America](#) (1)

[Bank of Norway](#) (1)

[Dell, Inc.](#) (1)

[European Union](#) (1)

[Maxtor Corp.](#) (1)

[Samsung Electronics Co.](#)

32 results found in 17 ms Page 1 of 4

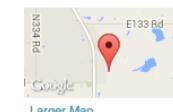
Test with some GB18030 encoded characters [More Like This](#)

Id: GB18030TEST

Price: 0,USD

Features: No accents here 这是一个功能 This is a feature (translated) 这份文件是很有光泽 This document is very shiny (translated)

In Stock: true



Samsung SpinPoint P120 SP2514N - hard drive - 250 GB - ATA-133 [More Like This](#)

Id: SP2514N

Price: 92,USD

Features: 7200RPM, 8MB cache, IDE Ultra ATA-133 NoiseGuard, SilentSeek technology, Fluid Dynamic Bearing (FDB) motor

In Stock: true



Maxtor DiamondMax 11 - hard drive - 500 GB - SATA-300 [More Like This](#)

Id: 6H500F0

Price: 350,USD

Features: SATA 3.0Gb/s, NCQ 8.5ms seek 16MB cache

In Stock: true



Belkin Mobile Power Cord for iPod w/ Dock [More Like This](#)

Id: F8V7067-APL-KIT

Price: 19.95,USD

Features: car power adapter, white

In Stock: false

Search-Based Applications

Federated Client

- Search across all content
- Authorized access only
- Simplified presentation
- Document viewing

Intranet Portal

- Easy access to search
- News and event notification
- Single sign-on authentication
- Application launching

Search Based Applications

The screenshot shows the eureka! search interface with a sidebar for navigation (protein, Classification, Instrument, Program, Year) and buttons for Search, Reset, and Export. The main area displays a grid of results with columns for Type, Name, Summary, Date, and Size. A filter icon is present in the header. The results list includes various protein assays and total protein assays from different dates.

Type	Name	Summary	Date	Size
	Protein Inv.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays ...	20-Oct-10	43520
✓	85-BCA-893-G.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays ...	20-Nov-10	43520
	85-BCA-898-G.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays ...	01-Dec-10	43520
	1-MAX-86.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays 1 ...	01-Jan-11	43520
	1720C-1-99-888.PDA	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays 1 ...	20-Feb-11	43520
	84-BCA-18.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays ...	20-Mar-11	43520
	Tween 8_18_98C	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays T_Prot ...	20-Apr-11	43520
	2-MAX-259.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays 2 ...	20-May-11	43520
	03-BCA-44.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays ...	20-Jun-11	43520
	2210C-8-99-8826.PDA	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays ...	20-Jul-11	43520
	8-MAX-52.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays 0 ...	20-Aug-11	43520
	2-MAX-298.pda	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3383-QCS Assays 2 ...	20-Sep-11	43520
	1720C-1-99-8851.PDA	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays ...	20-Oct-11	43520
	147QC388814aGAGAL	corp.shire.com HOT Scientific Quality Archive Biochemistry Spectramax Plus E-123 BOA TOTAL PROTEIN Total Protein Assays 3380-QCS Assays ...	20-Nov-11	43520

Instrument Datasets

- Optimized for scientists
- Data dependent menus
- Specialized grid filters

Regulatory Documents

- Designed for researchers
- Rich meta-data access
- Spreadsheet exports
- View document accelerator

The screenshot shows the eureka! search interface with a sidebar for navigation (Keyword, Product, Document Name, Title, Document Subtype, Document Unit, Document Status, Authors, Object ID) and buttons for Search, Reset, and Export. The main area displays a grid of results with columns for Type, Name, Summary, Date, and Size. A filter icon is present in the header. The results list includes various medical plausibility documents and clinical studies.

Type	Name	Summary	Date	Size
✓	Medical plausibility	... manifestations of patients inflicted with either type 3 or type 2 diseases ... patients with a confirmed diagnosis of type 1 Gaucher disease (patients with...) ...	15-Jan-10	43520
	Weinreb et al 2002	... Type 1 Gaucher disease is a multi-system disease caused by a genetic deficiency of lysosomal glucocerebrosidase(1). Its clinical manifestations ...	24-Jul-09	143480
	Cho ME 2004.pdf	... The presence of foam cells, however, is not diagnostic for Fabry disease, as it is characteristic for other lysosomal storage diseases (where lipidoses ...)	07-Apr-09	525389
	Einstein_2004_2.pdf	... Gaucher disease. In: Scriver CR ,Valle D ,Beaudet A ,Sly WS , editors. The metabolic and molecular bases of inherited diseases. Vol II. ...	06-May-10	123779
	Mignot 2006	... neurological involvement is sufficient to suspect the diagnosis at the onset of the disease. Type 2 and perinatal-lethal Gaucher diseases are easily ...	23-Sep-08	176698
	Component Clarke 2007	... Prague, Czech Republic, National Institute of Neurological Disease and Stroke ... the main agent of lysosomal storage diseases from Genzyme ...	10-Aug-10	496495
	Eng et al 2007	... Taskins D (1999) Rare diseases in renal replacement therapy in the ERA ... term safety and efficacy of enzyme replacement therapy for Fabry disease ...	15-Jan-10	247806
	Dehouf 2004	... enzyme replacement therapy in patients with Fabry disease ... Fabry disease usually has a deleterious effect on quality of life ...	14-Jan-10	125500
	Guest_Managing_Fabry_Italy_...	... in 117 000 live births for males [5,6] and is the second most common of the 40 lysosomal storage diseases (after Gaucher disease) and	06-Jun-11	238307

Search Based Applications

The screenshot shows the Aras Innovator search interface. The title bar reads "Discover Innovator". The main area is a grid showing search results. The columns are Type, Title, Content, Last Modified, and File Name. The results include various documents and files such as "NSD-002 MS-Visio Test", "NSD-003 NSD Snippet HTML", and "NSD-004 Test Document". On the left, there is a sidebar with a navigation tree and a search form. The search form includes fields for Search Term, Title, Modified On, and a toolbar with Find, Reset, and Save buttons. Below the search form is a filter section for Item Type, Classification, Status, Modified By, Revision, and Author.

Embedded in PLM Application

- Substantially better search experience than an RDBMS could provide
- Late-binding security model
- Document actions exposed on toolbar

Solr Resources

<http://wiki.apache.org/solr/FrontPage>

<http://wiki.apache.org/solr/SolrResources>

<https://cwiki.apache.org/confluence/display/solr/>

Apache Solr 3 Enterprise Search Server

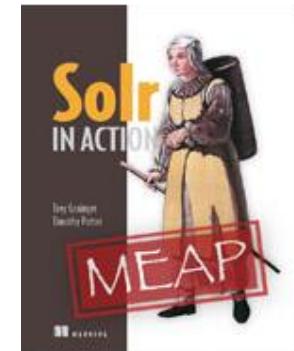
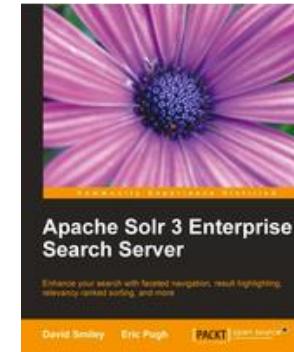
David Smiley and Eric Pugh

Packt Publishing

Solr In Action

Trey Grainger and Timothy Potter

Manning Publications



Thank You!

Al Cole

acole@nridge.com

www.linkedin.com/in/coleal