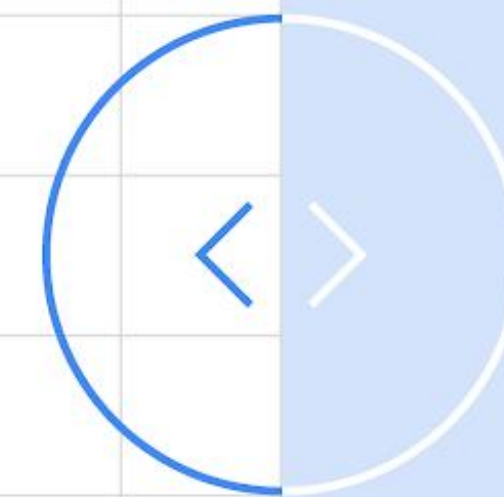Developer Student Clubs

# Support Search Using Apache Solr

Indexing and Querying

Jigar M. Pandya
Assistant Professor
Computer Engineering Dept., FoT.
D.D.University, Nadiad.

Google Developers

# Search

## The most used computing operation

Basic Algorithms : Linear Search, Binary Search, Hashing, etc.

Search Engines : Altavista, Google, Bing, etc.

Enterprise Search: Apache Lucene/Solr (Open Source)

# About Lucene

https://lucene.apache.org

Lucene Core is a Java library providing powerful indexing and search features, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities. The PyLucene sub project provides Python bindings for Lucene Core.

# About Solr

https://lucene.apache.org

Solr is a high performance search server built using Lucene Core. Solr is highly scalable, providing fully fault tolerant distributed indexing, search and analytics. It exposes Lucene's features through easy to use JSON/HTTP interfaces or native clients for Java and other languages.

# About Solr …

https://lucene.apache.org/solr/features.html

Solr is a standalone enterprise search server with a REST-like API. You put documents in it (called "indexing") via JSON, XML, CSV or binary over HTTP. You query it via HTTP GET and receive JSON, XML, CSV or binary results.
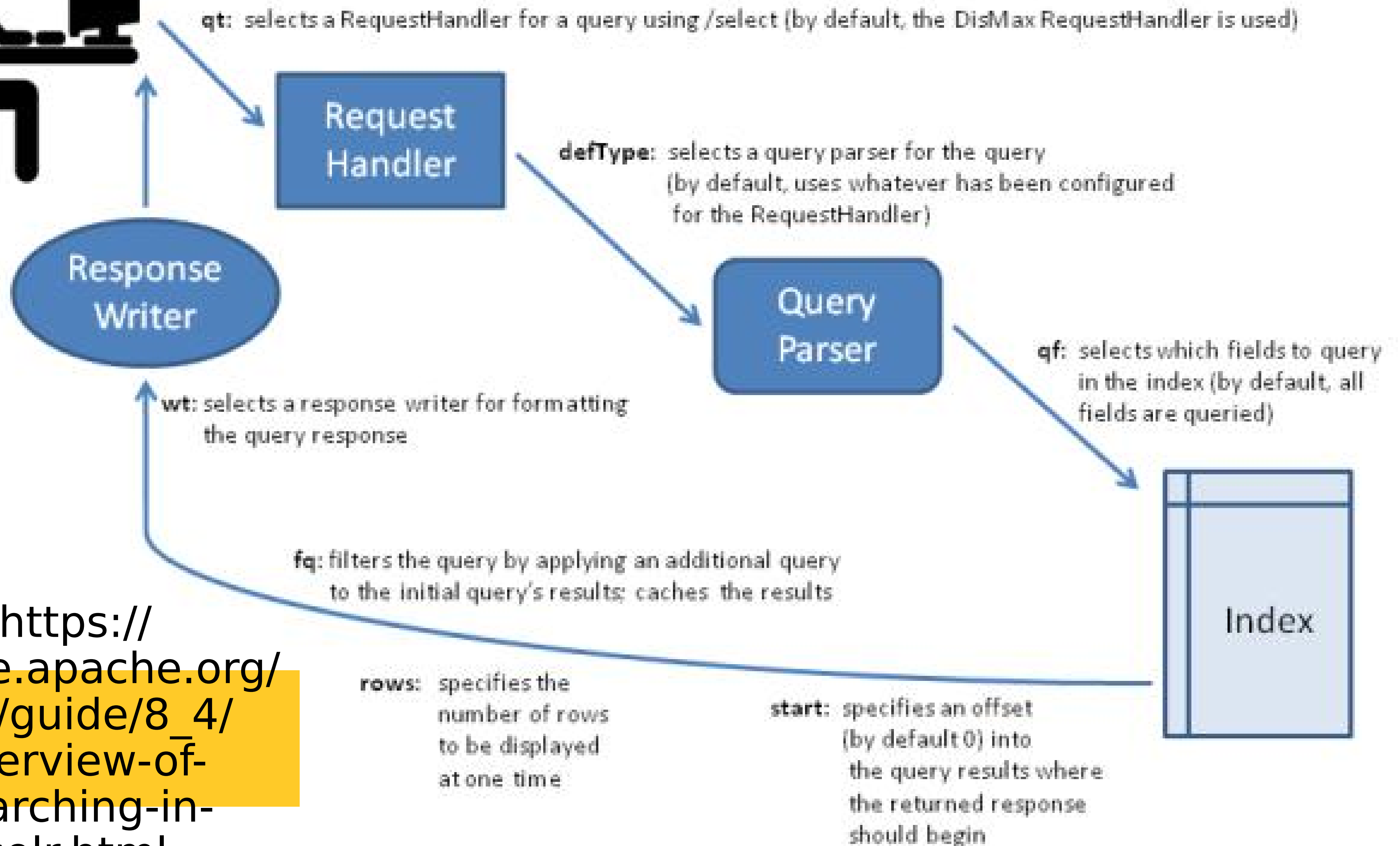
# How Solr Sees the World

https://lucene.apache.org/solr/guide/8_4/overview-of-documents-fields-and-schema-design.html

Solr's basic unit of information is a **document**, which is a set of data that describes something. Documents are composed of **fields**, which are more specific pieces of information.

When you add a document, Solr takes the information in the document's fields and adds that information to an **index**.

When you perform a **query**, Solr can quickly consult the index and return the matching documents.

**qt**: selects a RequestHandler for a query using /select (by default, the DisMax RequestHandler is used)

**defType**: selects a query parser for the query (by default, uses whatever has been configured for the RequestHandler)

**qf**: selects which fields to query in the index (by default, all fields are queried)

**wt**: selects a response writer for formatting the query response

**fq**: filters the query by applying an additional query to the initial query's results; caches the results

**rows**: specifies the number of rows to be displayed at one time

**start**: specifies an offset (by default 0) into the query results where the returned response should begin

Request Handler

Response Writer

Query Parser

Index

https://lucene.apache.org/solr/guide/8_4/overview-of-searching-in-solr.html

# High Level Steps

| Configure | | Index | | Query | | Response |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | ←→ | | ←→ | | ←→ | |
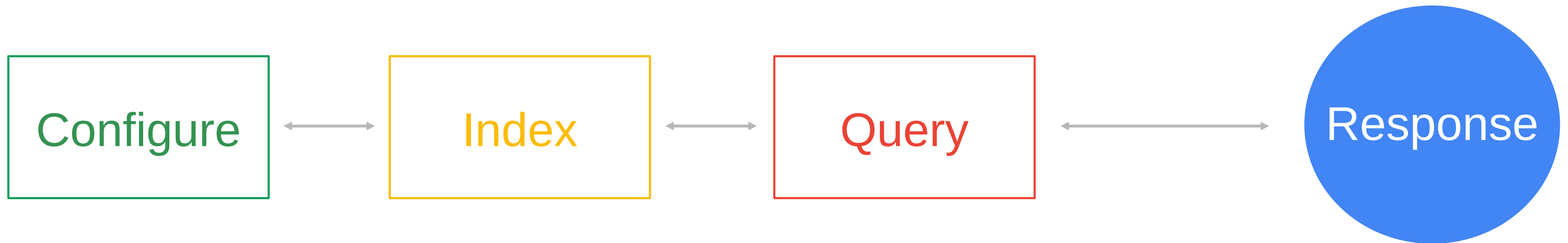
*Full Indexing first time*

*Delta indexing nightly*

*or Near Real Time integration by SolrJ like APIs.*

# Installing Solr

## Administration

sudo tar xvzf solr-8.4.1.tgz —directory=/opt/

sudo ln -s /opt/solr-8.4.1/bin/solr /usr/local/bin

sudo chown -R JigarPandya:staff /opt/solr-8.4.1


solr start

solr status

solr stop –all

solr -e techproducts

# Indexing Structured Data

https://lucene.apache.org/solr/guide/8_4/uploading-structured-data-store-data-with-the-data-import-handler.html#uploading-structured-data-store-data-with-the-data-import-handler

full-import

A Full Import operation can be started with a URL

http://localhost:8983/solr/dih/dataimport?command=full-import.

delta-import

For incremental imports and change detection. Only the SqlEntityProcessor supports delta imports.

http://localhost:8983/solr/dih/dataimport?command=delta-import.

 db-data-config.xml

last_modified > '${dataimporter.last_index_time}'

Developer Student Clubs

Google Developers

# Accessing Solr

## Over HTTP

- Admin
  - http://localhost:8983/solr/#/

- Select
  - http://localhost:8983/solr/
    techproducts/select?q=DDR

- Browse
  - http://localhost:8983/solr/
    techproducts/browse?q=DDR
- Schema
  - http://localhost:8983/solr/
    techproducts/schema
  - http://localhost:8983/solr/
    techproducts/schema/fields

Developer Student Clubs

Google Developers

Solr

**Dashboard**

Logging

Core Admin

Java Properties

Thread Dump

Core Selector ▾

🖥 **Instance**

🕐 Start            about 7 hours ago

📄 **Versions**

🔴 solr-spec  8.4.1

   solr-impl  8.4.1 832bf13dd9187095831caf69783179d41059d013 –

🐊 lucene-spec 8.4.1

   lucene-impl 8.4.1 832bf13dd9187095831caf69783179d41059d013 –

🔒 **JVM**

🔒 Runtime    Oracle Corporation Java HotSpot(TM) 64–Bit Server VM 1.8

🟩 Processors 2

🖥 Args       –DSTOP.KEY=solrrocks
              –DSTOP.PORT=7983
              –Djetty.home=/opt/solr-8.4.1/server
              –Djetty.port=8983
              –Dlog4j.configurationFile=file:/opt/solr-8.4.1/server/reso

3.91 GB
4.00 GB

Swap Space   38.4%

393.50 MB
1.00 GB

File Descriptor Count   2.3%

235
10240

▬ **JVM-Memory**   61.6%

315.14 MB
512.00 MB
512.00 MB

...**/select**

```
{
  "responseHeader":{
    "status":0,
    "QTime":4,
    "params":{
      "q":"DDR"}},
  "response":{"numFound":3,"start":0,"docs":[
      {
        "id":"VS1GB400C3",
        "name":"CORSAIR ValueSelect 1GB 184-Pin DDR SDRAM Unbuffered DDR 400 (PC 3200)
System Memory - Retail",
        "manu":"Corsair Microsystems Inc.",
        "manu_id_s":"corsair",
        "cat":["electronics",
          "memory"],
        "price":74.99,
        "price_c":"74.99,USD",
        "popularity":7,
        "inStock":true,
        "store":"37.7752,-100.0232",
        "manufacturedate_dt":"2006-02-13T15:26:37Z",
        "payloads":"electronics|4.0 memory|2.0",
        "_version_":1661044917874458624,
        "price_c____l_ns":7499},
      {
        "id":"VDBDB1A16",
        "name":"A-DATA V-Series 1GB 184-Pin DDR SDRAM Unbuffered DDR 400 (PC 3200) System
Memory - OEM",
        "manu":"A-DATA Technology Inc.",
```

.../**browse**

*(Velocity Templates)*

# Solr

Type of Search: | **Simple** | Spatial | Group By

Find: **DDR**    Submit    Reset

☐ Boost by Price

## Field Facets

3 results found in 22 ms Page 1 of 1

**cat**

electronics (3)

memory (3)

*missing* (0)

**manu_exact**

Corsair Microsyst... (2)

A-DATA Technology... (1)

*missing* (0)

**content_type**

*missing* (3)

**author_s**

*missing* (3)

## Query Facets

**CORSAIR ValueSelect 1GB 184-Pin DDR SDRAM Unbuffered DDR 400 (PC 3200) System Memory - Retail** More Like This

Id: VS1GB400C3

Price: 74.99,USD

Features:

In Stock: true

**A-DATA V-Series 1GB 184-Pin DDR SDRAM Unbuffered DDR 400 (PC 3200) System Memory - OEM** More Like This

Id: VDBDB1A16

Price:

Features: CAS latency 3, 2.7v

.../schema/fields

```
{
  "responseHeader":{
    "status":0,
    "QTime":0},
  "fields":[{
      "name":"_root_",
      "type":"string",
      "docValues":false,
      "indexed":true,
      "stored":false},
    {
      "name":"_src_",
      "type":"string",
      "indexed":false,
      "stored":true},
    {
      "name":"_version_",
      "type":"plong",
      "indexed":false,
      "stored":false},
    {
      "name":"author",
      "type":"text_gen_sort",
      "multiValued":false,
```

# Response Write to be used

The wt parameter selects the Response Writer to be used.

Json (Default) - http://localhost:8983/solr/techproducts/select?q=DDR

Xml - http://localhost:8983/solr/techproducts/select?q=DDR&wt=xml

Csv - http://localhost:8983/solr/techproducts/select?q=DDR&wt=csv

javabin

velocity

and more …

```java
[Source: https://lucene.apache.org/solr/6_5_1/solr-solrj/index.html ]
package com.example;
import org.apache.solr.client.solrj.SolrServer;
import org.apache.solr.client.solrj.SolrServerException;
import org.apache.solr.client.solrj.impl.CommonsHttpSolrServer;
import org.apache.solr.client.solrj.response.QueryResponse;
import org.apache.solr.common.params.ModifiableSolrParams;
import java.net.MalformedURLException;

public class SolrJExample {
  public static void main(String[] args) throws MalformedURLException,
SolrServerException {
    SolrServer solr = new
CommonsHttpSolrServer("http://localhost:8983/solr/techproducts");
    ModifiableSolrParams params = new ModifiableSolrParam
    params.set("q", "DDR");
    QueryResponse response = solr.query(params);
    System.out.println(response.getResults().getNumFound());
```

Java Program to retrieve count of matching documents

```
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ cat ./compile.sh
javac -cp lib/apache-solr-solrj-3.3.0.jar com/example/SolrJExample.java
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ ./compile.sh
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ tree .
.
├── com
│   └── example
│       ├── SolrJExample.class
│       └── SolrJExample.java
├── compile.sh
├── lib
│   ├── apache-solr-solrj-3.3.0.jar
│   ├── logback-classic-0.9.jar
│   ├── logback-core-0.9.6.jar
│   ├── org-apache-commons-codec.jar
│   ├── org-apache-commons-logging.jar
│   ├── org.apache.commons.httpclient.jar
│   ├── slf4j-api-1.5.5.jar
│   └── solr-solrj-8.4.1.jar
└── runme.sh

3 directories, 12 files
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ cat ./runme.sh
java -cp lib/logback-classic-0.9.jar:lib/logback-core-0.9.6.jar:lib/org.apache.commons.httpclient.jar:lib/apache-s
olr-solrj-3.3.0.jar:lib/slf4j-api-1.5.5.jar:lib/org-apache-commons-logging.jar:lib/org-apache-commons-codec.jar:.
com.example.SolrJExample
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ ./runme.sh
3
Jigar-Pandyas-MacBook:JavaProgram JigarPandya$ █
```

*Interacting to Solr instance using Java Program*

# Know your solr

Start with README.txt

server/README.txt
server/solr/README.txt

https://lucene.apache.org/solr/
guide/8_4/installing-
solr.html#directory-layout

Developer Student Clubs

```
sh-3.2# pwd
/opt/solr-8.4.1
sh-3.2# ls -l
total 3336
-rw-r--r--@  1 JigarPandya  staff  909075 Jan 10 12:24 CHANGES.txt
-rw-r--r--@  1 JigarPandya  staff   12646 Jan 10 12:20 LICENSE.txt
-rw-r--r--@  1 JigarPandya  staff  742863 Jan 10 12:24 LUCENE_CHANGES.txt
-rw-r--r--@  1 JigarPandya  staff   27717 Jan 10 12:20 NOTICE.txt
-rw-r--r--@  1 JigarPandya  staff    7490 Jan 10 12:24 README.txt
drwxr-xr-x@ 11 JigarPandya  staff     374 Mar 13 16:06 bin
drwxr-xr-x@ 12 JigarPandya  staff     408 Jan 10 13:40 contrib
drwxr-xr-x  18 JigarPandya  staff     612 Mar 10 03:25 dist
drwxr-xr-x   4 JigarPandya  staff     136 Mar 10 03:26 docs
drwxr-xr-x@  8 JigarPandya  staff     272 Mar 13 16:06 example
drwxr-xr-x@601 JigarPandya  staff   20434 Jan 10 12:24 licenses
drwxr-xr-x@ 13 JigarPandya  staff     442 Mar 13 14:18 server
sh-3.2# 
```

*Directory Structure*

# Technical Words

## https://lucene.apache.org/solr/guide/8_4/solr-glossary.html

- Core
  - An individual Solr instance (represents a logical index). Multiple cores can run on a single node.

- Collection
  - In Solr, one or more Documents grouped together in a single logical index using a single configuration and Schema.
  - In SolrCloud a collection may be divided up into multiple logical shards, which may in turn be distributed across many nodes, or in a Single node Solr installation, a collection may be a single Core.

Developer Student Clubs

Google Developers

# Technical Words

https://lucene.apache.org/solr/guide/8_4/solr-glossary.html

- Facet
  - The arrangement of search results into categories based on indexed terms
- Replica
  - A Core that acts as a physical copy of a Shard in a SolrCloud Collection.

- Shard
  - In SolrCloud, a logical partition of a single Collection. Every shard consists of at least one physical Replica, but there may be multiple Replicas distributed across multiple Nodes for fault tolerance. See also SolrCloud.
- Suggester
  - Functionality in Solr that provides the ability to suggest possible query terms to users as they type.

Developer Student Clubs

Google Developers

# Important Files

## Configuration

bin/solr.in.sh (Set solr java home and more)

server/solr/solr.xml (Host context, port and more)

server/logs/solr.log (Runtime Query log and more)

example/techproducts/solr/techproducts/conf/solrconfig.xml (Lucene Match Version, Index Data Directory Location and more)

example/techproducts/solr/techproducts/conf/managed-schema (schema.xml)

example/example-DIH/solr/db/conf/db-data-config.xml (Structured Data Importer)

# Important Links

## Official as of Solr 8.4

https://lucene.apache.org/solr/guide/8_4/index.html

Solr Tutorial

https://lucene.apache.org/solr/guide/8_4/solr-tutorial.html

Solr Quick Overview

https://lucene.apache.org/solr/guide/8_4/a-quick-overview.html#a-quick-overview

# Important Links

## Official as of Solr 8.4

https://lucene.apache.org/solr/guide/8_4/searching.html#searching
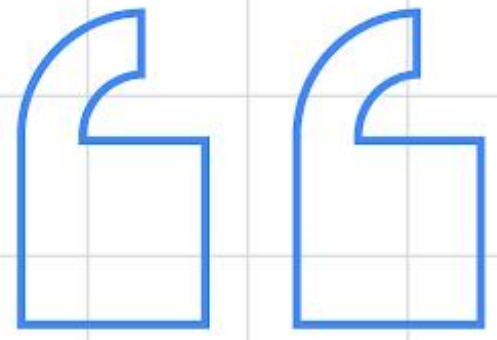
"Enjoy the journey called LIFE"

- Jigar Pandya

Developer Student Clubs

# Thank You.

Jigar M. Pandya

Assistant Professor

Computer Engineering Dept., FoT.

D.D.University, Nadiad.

jigarpandya.ce@ddu.ac.in