

CS-5590 PARALLEL ALGORITHMS

ASSIGNMENT 2

Submitted By : Team 7 - Rama Charan Pavan, Sri Charan , Pratap Rao

1.

Hash Table – Hashing- Hash function $O(1)$

Collision in Hashing

1. Separate Chaining
2. Open addressing
 - a. Linear probing
 - b. Quadratic Probing
 - c. Double hashing

Linear Hashing:

	Serial Approach	Parallel Approach
Lookup :	Best case : $O(n)$ Worst Case : $O(n^2)$	Best case : $O(1)$, Worst case : $O(n)$
Insert	Best Case : $O(n)$, Worst Case : $O(n^2)$	Best case : $O(1)$ Worst Case : $O(n)$
Removal :	Best case : $O(n)$, Worst Case : $O(n^2)$	Best case : $O(1)$ Worst case : $O(n)$

Cuckoo hashing – $O(1)$

Parallel cuckoo hashing

	Serial Approach	Parallel Approach
Lookup :	Best case : $O(n)$ Worst Case : $O(n^2)$	Best case : $O(1)$, Worst case : $O(n)$
Insert :	Best Case : $O(n)$, Worst Case : $O(n)$,	Best case : $O(1)$ Worst Case : $O(n)$,
Removal :	Best case : $O(n)$, Worst Case : $O(n^2)$	Best case : $O(1)$ Worst case : $O(n)$

Lock-free hashing

	Serial Approach	Parallel Approach
Lookup	Best Case : $O(n)$ Worst Case : $O(d*n)$	Best Case : $O(1)$ Worst Case : $O(d)$
Insert	Best Case : $O(n)$, Worst Case : $> O(d*n)$,	Best case : $O(1)$ Worst case : $> O(d)$,
Removal	Best Case : $O(n)$ Worst Case : $O(d*n)$	Best Case : $O(1)$ Worst Case : $O(d)$

Bloom Filter

Complexity of Searches: best case : $O(1)$ worst case : $O(n)$

Inverted Index

Time Complexity: $O(n)$ for n -byte text

Binary Search tree:

Time Complexity: $O(\log n)$

Query: $O(\log n + k)$

Space: $O(n)$

Range Tree Running time: $O(n \log n)$

PRAM

Time Complexity: $O(\log n)$

Total Number Of Steps = $n * \log n = O(n \log n)$

EREW PRAM: $\log p + n/p$

CREW PRAM: $\log p + n/p$

CRCW PRAM:

Running Time : $O(1)$

Parallel techniques work best for algorithms that involve the indexing parameters and can be closely associated with negating the time complexity which is different when compared to that of a sequential algorithm. However, in the searching algorithms sequential would be better as in parallel case, the communication between the different elements can lapse which might totally cause an erroneous situation. In order to overcome this searching algorithms implement a more sequential technique. Some ambiguity can exist but the odds of its being sequential are higher.

2.

Inverted Index Using Spark

```
1 import util.CommandLineOptions
2 import util.FileUtil
3 import org.apache.spark.{SparkConf, SparkContext}
4 import org.apache.spark.SparkContext._
5
6 /**
7  * Created by pavan on 10/2/2016.
8  */
9
10 object InvertedIndexSort {
11   def main(args: Array[String]): Unit = {
12     System.setProperty("hadoop.home.dir", "C:\\Users\\bach\\Downloads\\Pavan\\winutils")
13     val options = CommandLineOptions(
14       this.getClass.getSimpleName,
15       CommandLineOptions.inputPath("output/crawl"),
16       CommandLineOptions.outputPath("output/inverted-index"),
17       CommandLineOptions.master("local"),
18       CommandLineOptions.quiet)
19
20     val argz = options(args.toList)
21     val master = argz("master")
22     val quiet = argz("quiet").toBoolean
23     val out = argz("output-path")
24     if (master.startsWith("local")) {
25       if (!quiet) println(s" **** Deleting old output (if any), $out:")
26       FileUtil.rmrf(out)
27     }
28
29     val name = "Inverted Index (5b)"
30     val conf = new SparkConf().
31       setMaster(master).
32       setAppName(name).
33
34     val sc = new SparkContext(conf)
35     val in = sc.textFile(argz("input-path"))
36     val words = in.flatMap(_.split(" "))
37     val wordGroups = words.map(word => (word, 1))
38     val wordGroups2 = wordGroups.reduceByKey(_+_).mapValues(_.toString)
39     val out2 = wordGroups2.map(word => (word, wordGroups2.get(word).toString))
40     out2.saveAsTextFile(out)
41     if (!quiet) println(s" **** Wrote output to $out")
42   }
43 }
```

IntelliJ IDEA 2016.2.4

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Project: HelloWorld.scala x WordCount.scala x InvertedIndexSort.scala x CommandLineOptions.scala x FileUtil.scala x build.sbt x InvertedIndex.scala x

Project Structure: Hello [hello] C:\Users\bach\IdeaProjects\Hello\src\main\scala\InvertedIndexSort.scala

```
32  setAppName(name).
33  set("spark.app.id", name) // To silence Metrics warning.
34  val sc = new SparkContext(conf)
35  try { val lineRE = """^s*\s*([^\s]+),(.*)\s*s*$"""
36  val input = sc.textFile(argz("input-path")).map {
37    case lineRE(name, text) => (name.trim, text.toLowerCase)
38    case badLine =>
39      Console.err.println(s"Unexpected line: $badLine")
40  }
41  ("", "")
42  }
43  if (!quiet) println(s"Writing output to: $out")
44  input
45  .flatMap {
46    case (path, text) =>
47      text.trim.split("""^[^\sIsAlphabetic]]+""").map(word => (word, path))
48  }
49  .map {
50    case (word, path) => ((word, path), 1)
51  }
52  .reduceByKey(
53    (count1, count2) => count1 + count2
54  )
55  .map {
56    case ((word, path), n) => (word, (path, n))
57  }
58  .groupByKey
59  .sortByKey(ascending = true)
60  .mapValues { iterable =>
61    iterable.toSeq.sortBy { case (path, n) => (-n, path) }.mkString(", ")
62  }
63  .saveAsTextFile(out)
64  finally {
65    if (!quiet) {
66      println("""
67      |
68      | Before closing down the SparkContext, open the Spark Web Console
69      | http://localhost:4040 and browse the information about the tasks
70      | run for this example.
71      |
72      | When finished, hit the <return> key to exit.
73      |
74      |""")
75    }
76    Console.in.read()
77  }
78  sc.stop()
79  }
```

Run: InvertedIndexSort InvertedIndexSort InvertedIndexSort

Compilation completed successfully in 11s 505ms (4 minutes ago)

IntelliJ IDEA 2016.2.4

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

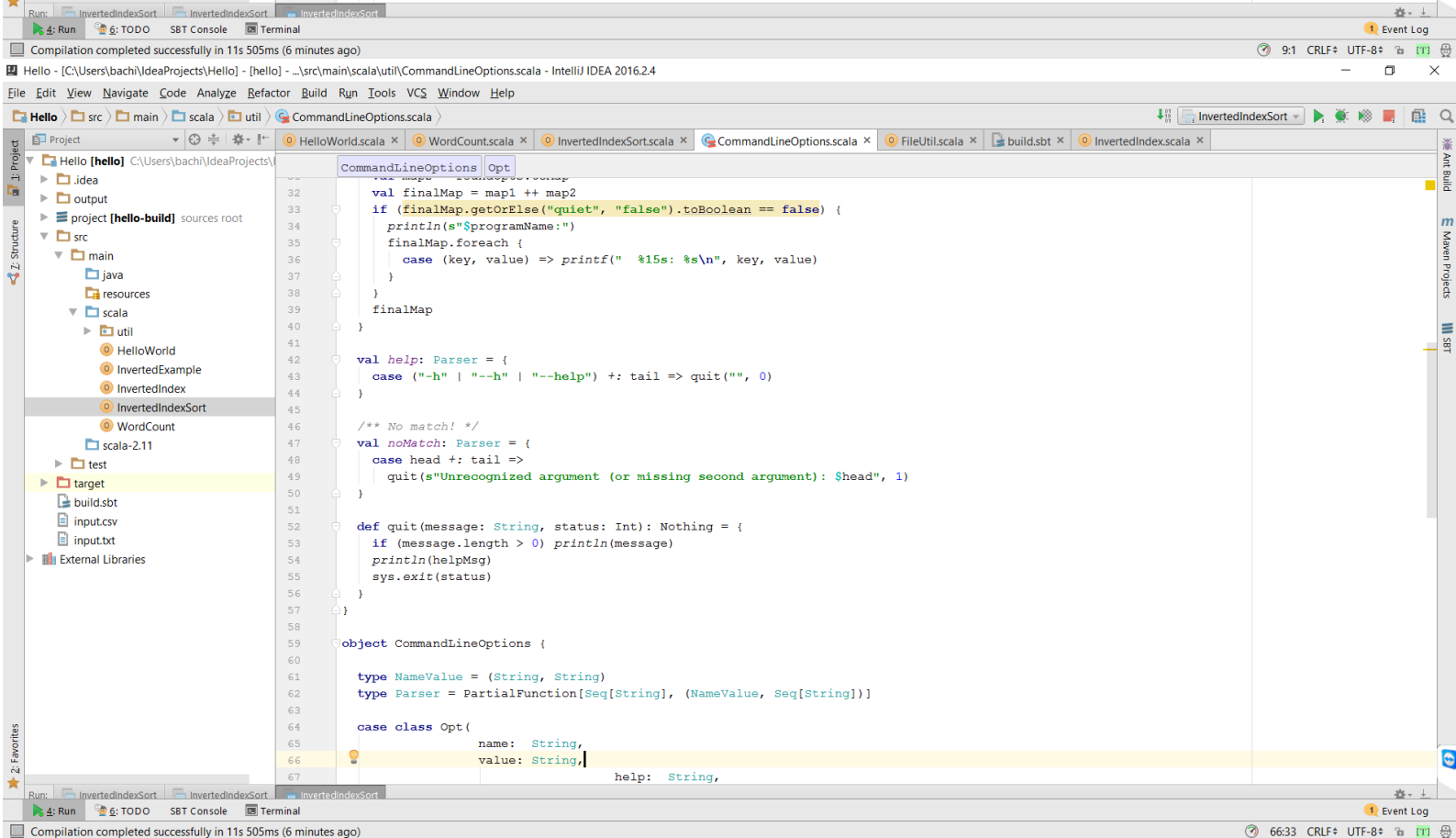
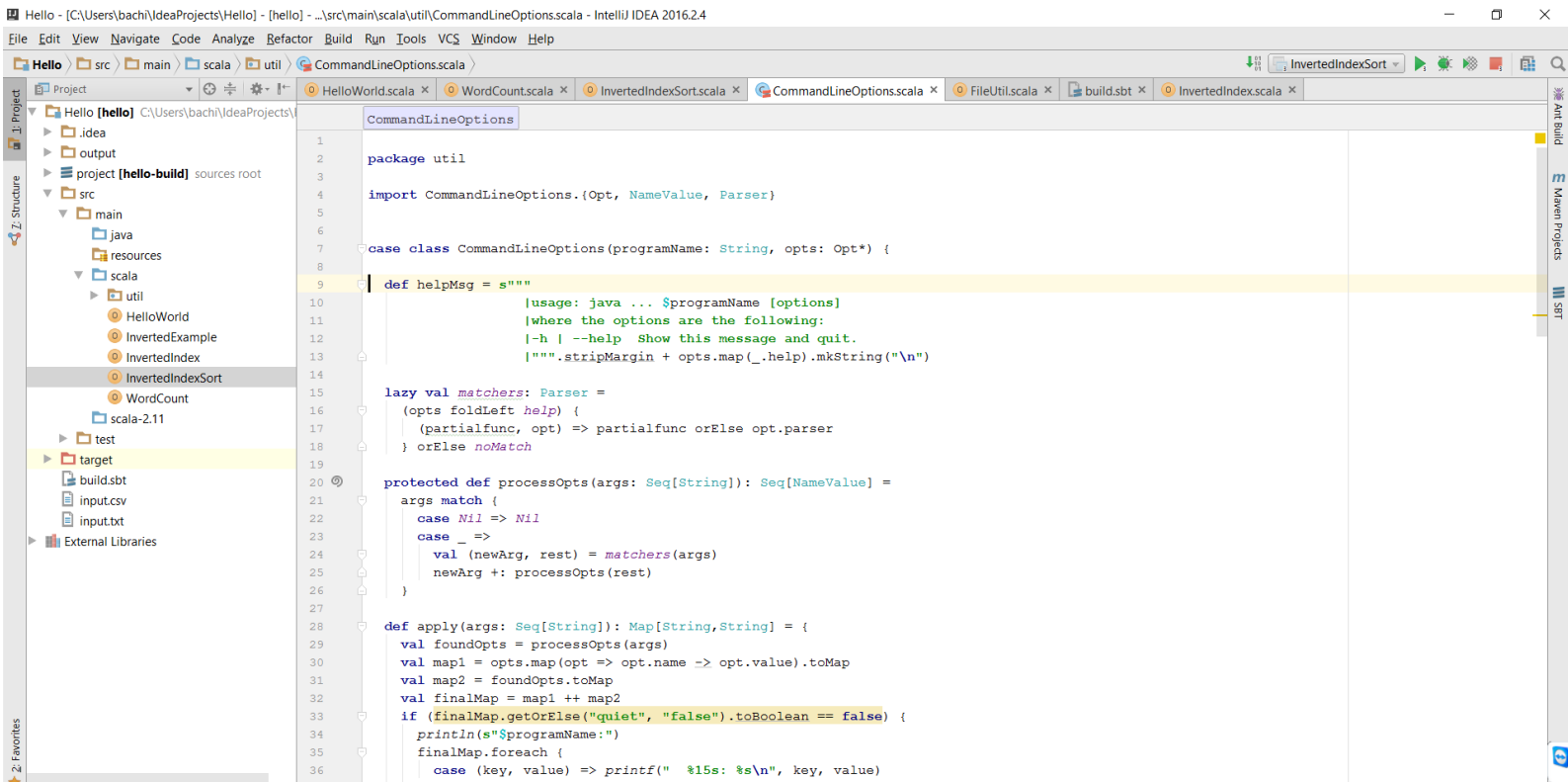
Project: HelloWorld.scala x WordCount.scala x InvertedIndexSort.scala x CommandLineOptions.scala x FileUtil.scala x build.sbt x InvertedIndex.scala x

Project Structure: Hello [hello] C:\Users\bach\IdeaProjects\Hello\src\main\scala\InvertedIndexSort.scala

```
48  }
49  .map {
50    case (word, path) => ((word, path), 1)
51  }
52  .reduceByKey(
53    (count1, count2) => count1 + count2
54  )
55  .map {
56    case ((word, path), n) => (word, (path, n))
57  }
58  .groupByKey
59  .sortByKey(ascending = true)
60  .mapValues { iterable =>
61    iterable.toSeq.sortBy { case (path, n) => (-n, path) }.mkString(", ")
62  }
63  .saveAsTextFile(out)
64  finally {
65    if (!quiet) {
66      println("""
67      |
68      | Before closing down the SparkContext, open the Spark Web Console
69      | http://localhost:4040 and browse the information about the tasks
70      | run for this example.
71      |
72      | When finished, hit the <return> key to exit.
73      |
74      |""")
75    }
76    Console.in.read()
77  }
78  sc.stop()
79  }
```

Run: InvertedIndexSort InvertedIndexSort InvertedIndexSort

Compilation completed successfully in 11s 505ms (4 minutes ago)



```
66         value: String,
67         help: String,
68         parser: Parser)
69 def inputPath(value: String): Opt = Opt(
70     name = "input-path",
71     value = value,
72     help = s"-i | --in | --inpath path The input root directory of files to crawl (default: $value)",
73     parser = {
74         case ("-i" | "--in" | "--inpath") +: path +: tail => (("input-path", path), tail)
75     })
76 def outputPath(value: String): Opt = Opt(
77     name = "output-path",
78     value = value,
79     help = s"-o | --out | --outpath path The output location (default: $value)",
80     parser = {
81         case ("-o" | "--out" | "--outpath") +: path +: tail => (("output-path", path), tail)
82     })
83 def master(value: String): Opt = Opt(
84     name = "master",
85     value = value,
86     help = s"-m | --master M The 'master' argument passed to SparkContext, 'M' is one of:
87         | local, local[N], 'mesos://host:port', or 'spark://host:port'
88         | (default: $value).""".stripMargin,
89     parser = {
90         case ("-m" | "--master") +: master +: tail => (("master", master), tail)
91     })
92 def socket(hostPort: String): Opt = Opt(
93     name = "socket",
94     value = hostPort,
95     help = s"-s | --socket host:port Listen to a socket for events (default: $hostPort unless --inpath used)",
96     parser = {
97         case ("-s" | "--socket") +: hp +: tail => (("socket", hp), tail)
98     })
99
100
101
```

Run: InvertedIndexSort InvertedIndexSort InvertedIndexSort

Compilation completed successfully in 11s 505ms (7 minutes ago)

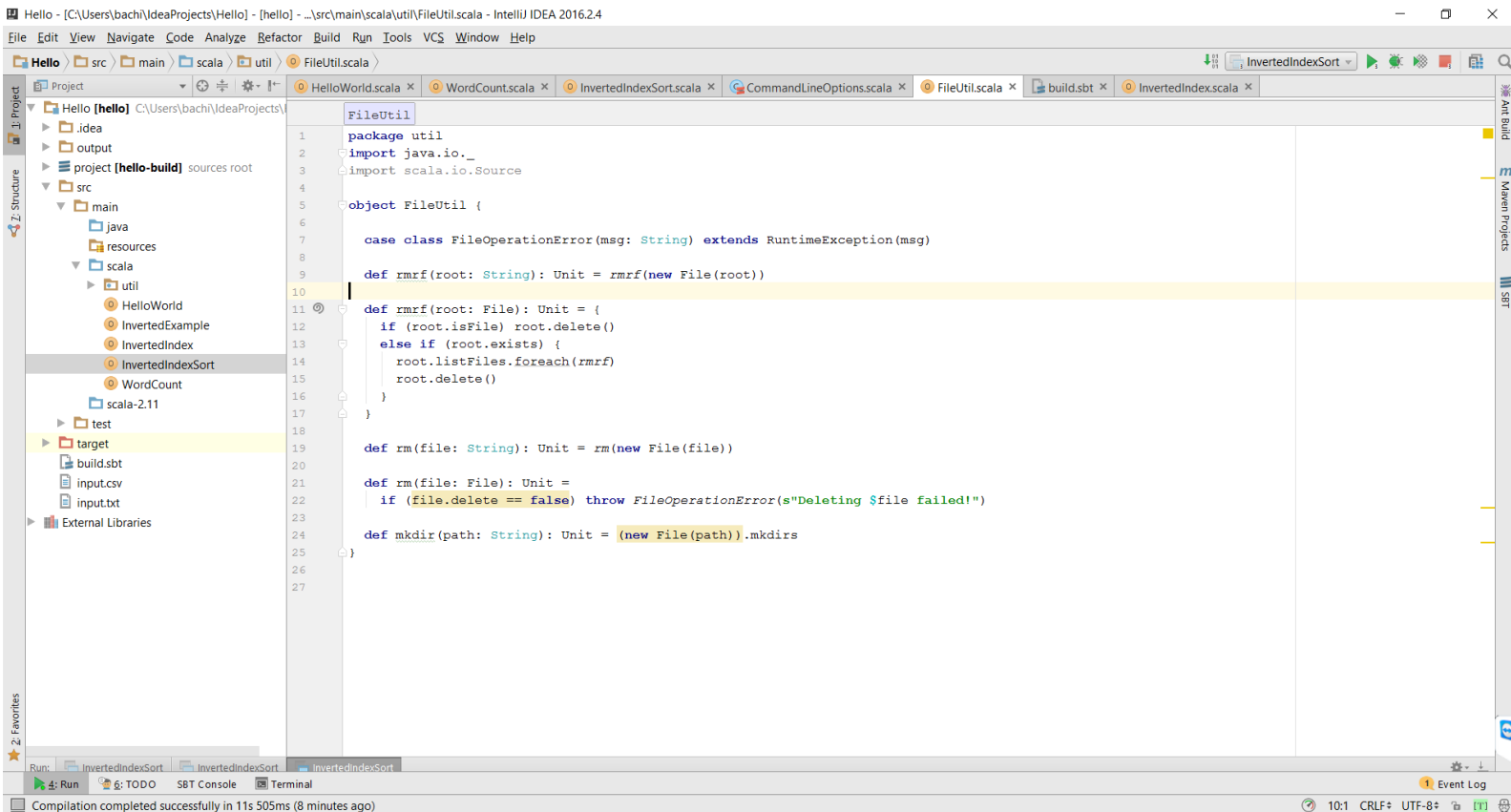
100:7 CRLF UTF-8

```
83 def master(value: String): Opt = Opt(
84     name = "master",
85     value = value,
86     help = s"-m | --master M The 'master' argument passed to SparkContext, 'M' is one of:
87         | local, local[N], 'mesos://host:port', or 'spark://host:port'
88         | (default: $value).""".stripMargin,
89     parser = {
90         case ("-m" | "--master") +: master +: tail => (("master", master), tail)
91     })
92 def socket(hostPort: String): Opt = Opt(
93     name = "socket",
94     value = hostPort,
95     help = s"-s | --socket host:port Listen to a socket for events (default: $hostPort unless --inpath used)",
96     parser = {
97         case ("-s" | "--socket") +: hp +: tail => (("socket", hp), tail)
98     })
99 def terminate(seconds: Int): Opt = Opt(
100     name = "terminate",
101     value = seconds.toString,
102     help = s"--term | --terminate N Terminate after N seconds. Use 0 for no termination.",
103     parser = {
104         case ("--term" | "--terminate") +: n +: tail => (("terminate", n), tail)
105     })
106 def quiet: Opt = Opt(
107     name = "quiet",
108     value = "false",
109     help = s"--q | --quiet Suppress some informational output.",
110     parser = {
111         case ("--q" | "--quiet") +: tail => (("quiet", "true"), tail)
112     })
113
114
115
116
117
```

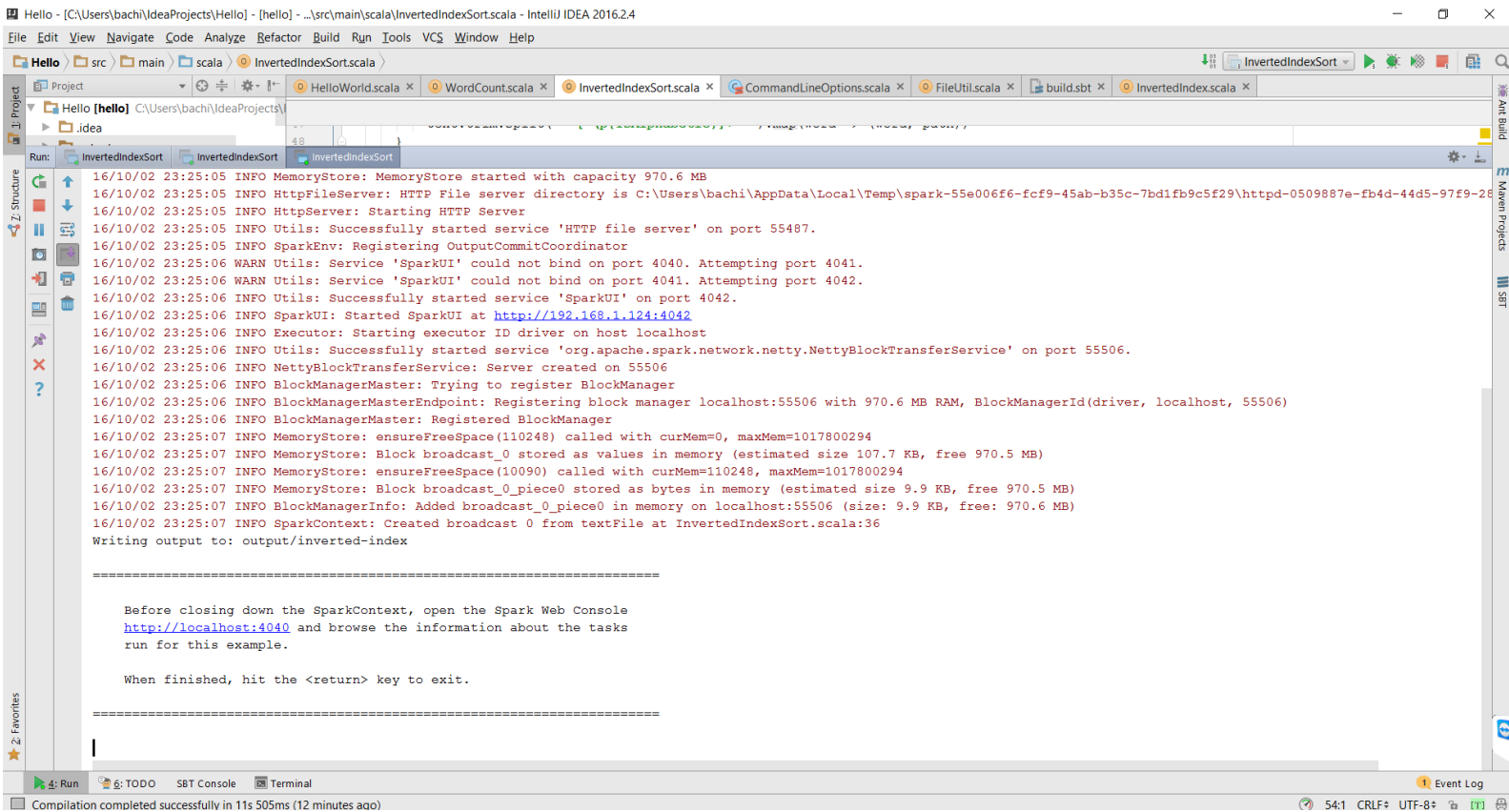
Run: InvertedIndexSort InvertedIndexSort InvertedIndexSort

Compilation completed successfully in 11s 505ms (7 minutes ago)

117:2 CRLF UTF-8



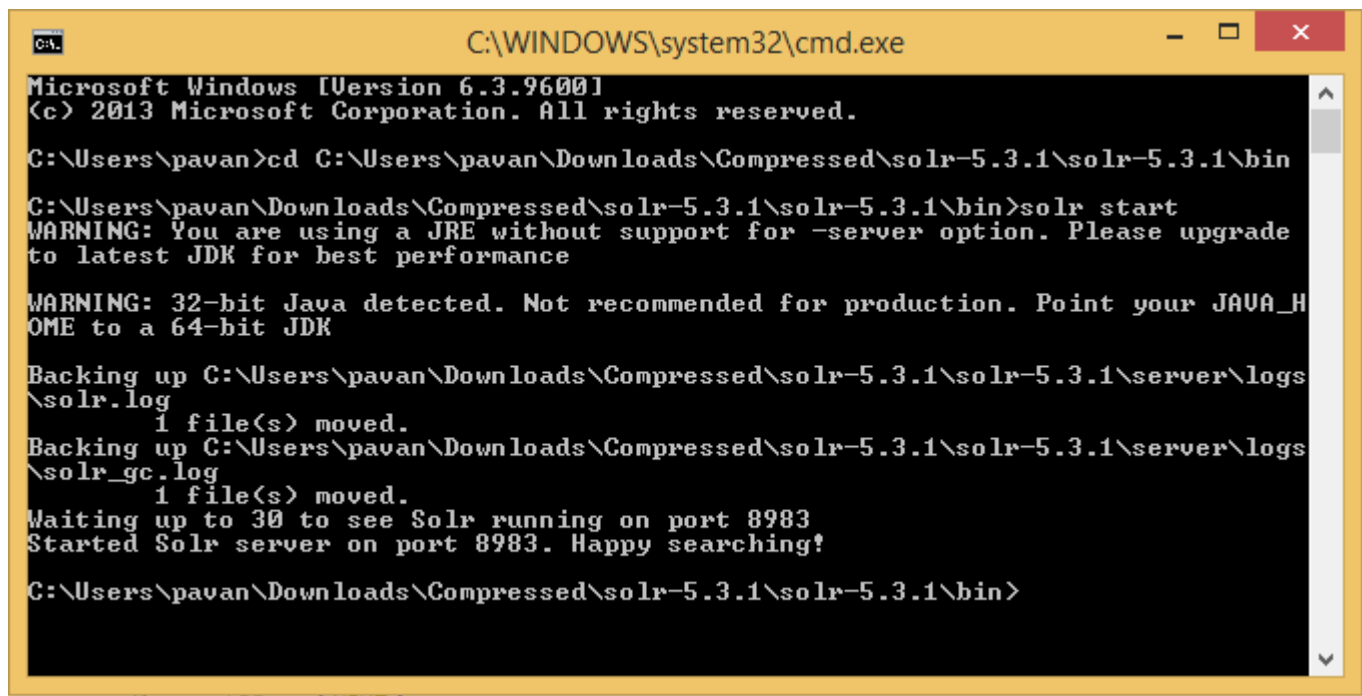
Output:



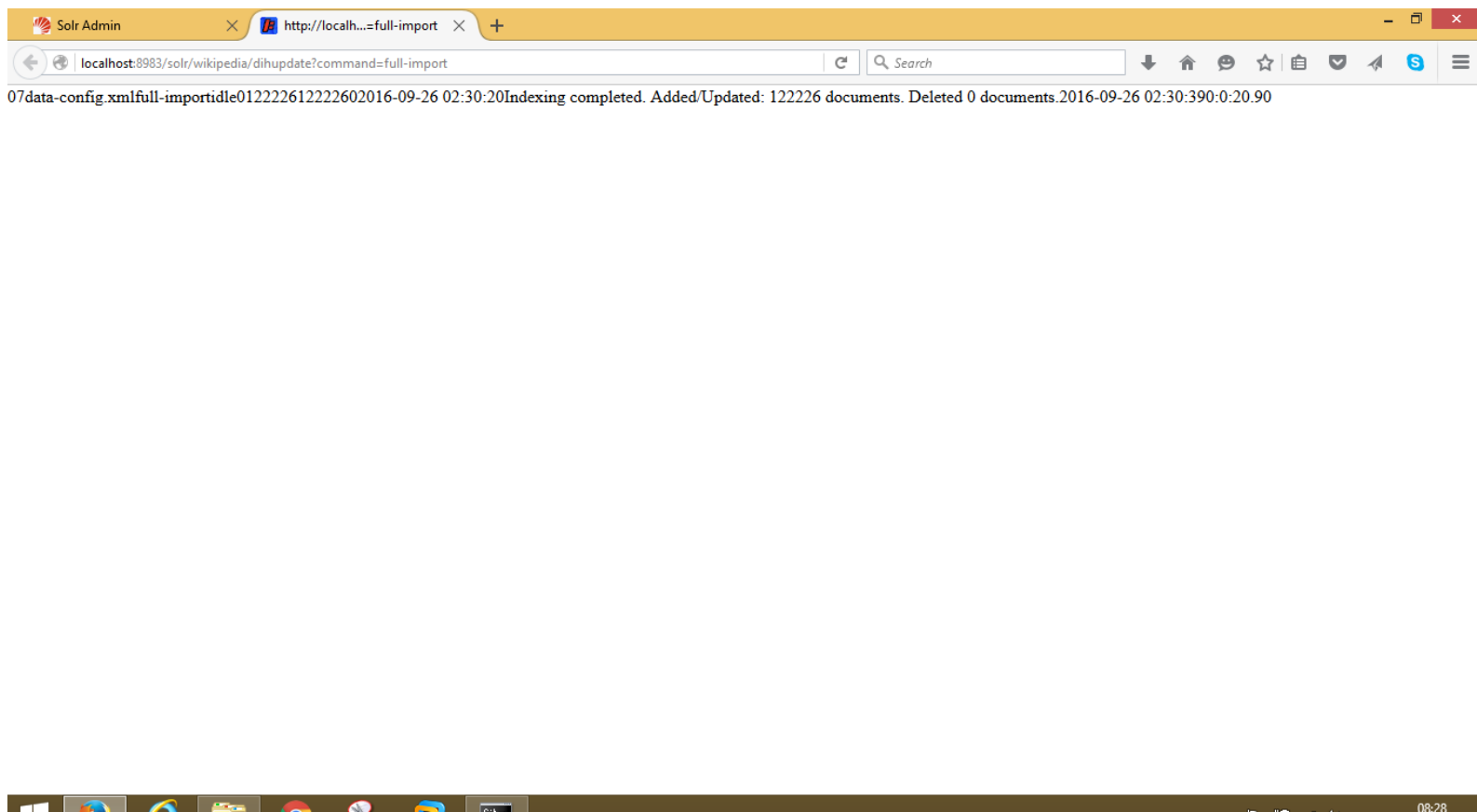
3.

Installation and Queries run on Apache Solr 5.3.1

Running Successfully... !!



Indexing Done Successfully ..!!



Number Of Docs Indexed in Wikipedia 122226 .. !!

The screenshot shows the Solr Admin web interface in a browser. The address bar indicates the URL is `localhost:8983/solr/#/~cores/wikipedia`. The left sidebar contains navigation links: Dashboard, Logging, Core Admin (selected), Java Properties, and Thread Dump. Below these is a 'Core Selector' dropdown menu. The main content area displays the configuration for the 'wikipedia' core. At the top, there are buttons for 'Add Core', 'Unload', 'Rename', 'Swap', 'Reload', and 'Optimize'. The configuration is divided into two sections: 'Core' and 'Index'. The 'Core' section shows 'startTime: about a minute ago', 'instanceDir: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\'', and 'dataDir: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\''. The 'Index' section shows 'lastModified: about 12 hours ago', 'version: 44', 'numDocs: 122226', 'maxDoc: 122226', 'deletedDocs: -', 'optimized: ✓', 'current: ✗', and 'directory: org.apache.lucene.store.NRTCachingDirectory:NRTCachingDirectory(SimpleFSDirectory@C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\index lockFactory=org.apache.lucene.store.NativeFSLockFactory@14f40f4; maxCacheMB=48.0 maxMergeSizeMB=4.0)'. At the bottom of the interface, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax. The Windows taskbar at the bottom shows various application icons and the system clock indicating 20:14 on 26-09-2016.

Solr Admin | `localhost:8983/solr/#/~cores/wikipedia`

Core Configuration:

- Core**
 - startTime: about a minute ago
 - instanceDir: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\
 - dataDir: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\
- Index**
 - lastModified: about 12 hours ago
 - version: 44
 - numDocs: 122226
 - maxDoc: 122226
 - deletedDocs: -
 - optimized: ✓
 - current: ✗
 - directory: org.apache.lucene.store.NRTCachingDirectory:NRTCachingDirectory(SimpleFSDirectory@C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\index lockFactory=org.apache.lucene.store.NativeFSLockFactory@14f40f4; maxCacheMB=48.0 maxMergeSizeMB=4.0)

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

Overview Of Indexed Data ...!!

The screenshot shows the Solr Admin web interface for the 'wikipedia' core, specifically the 'Statistics' and 'Instance' tabs. The left sidebar is similar to the previous screenshot, but with 'Overview' selected under the 'Core Admin' section. The 'Statistics' tab displays various metrics: 'Last Modified: about a minute ago', 'Num Docs: 122226', 'Max Doc: 122226', 'Heap Memory: -1', 'Usage: ', 'Deleted Docs: 0', 'Version: 56', 'Segment Count: 3', 'Optimized: ✓', and 'Current: ✓'. Below this is a 'Replication (Master)' table showing two entries: 'Master (Searching)' and 'Master (Replicable)', both with version 1474901056258 and generation 11. The 'Instance' tab shows the following details: 'CWD: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server', 'Instance: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia', 'Data: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data', 'Index: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\index', and 'Impl: org.apache.solr.core.NRTCachingDirectoryFactory'. A 'Healthcheck' section at the bottom of the instance tab states: 'Ping request handler is not configured with a healthcheck file.' At the bottom of the interface, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax. The Windows taskbar at the bottom shows various application icons and the system clock indicating 20:15 on 26-09-2016.

Solr Admin | `localhost:8983/solr/#/wikipedia`

Statistics

- Last Modified: about a minute ago
- Num Docs: 122226
- Max Doc: 122226
- Heap Memory: -1
- Usage:
- Deleted Docs: 0
- Version: 56
- Segment Count: 3
- Optimized: ✓
- Current: ✓

Replication (Master)

	Version	Gen	Size
Master (Searching)	1474901056258	11	8.23 MB
Master (Replicable)	1474901056258	11	-

Instance

- CWD: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server
- Instance: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia
- Data: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data
- Index: C:\Users\pavan\Downloads\Compressed\solr-5.3.1\solr-5.3.1\server\solr\wikipedia\data\index
- Impl: org.apache.solr.core.NRTCachingDirectoryFactory

Healthcheck

Ping request handler is not configured with a healthcheck file.

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

Overall Data of Xml File in single Query ...!! q= “.”

Solr Admin

localhost:8983/solr/#/wikipedia/query

Search

Solr

Dashboard

Logging

Core Admin

Java Properties

Thread Dump

wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema Browser

Segments info

Request-Handler (qt)

/select

common

q

.*

fq

sort

start, rows

010

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

☒ indent

☐ debugQuery

☐ dismax

http://localhost:8983/solr/wikipedia/select?q=.%3A&wt=json&indent=true

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 532,
    "params": {
      "q": ".*",
      "indent": "true",
      "wt": "json",
      "_": "1474858700744"
    }
  },
  "response": {
    "numFound": 122226,
    "start": 0,
    "docs": [
      {
        "id": "352690",
        "title": "Uncontrolled airspace",
        "_version_": 1546499704324358100
      },
      {
        "id": "352691",
        "title": "Deep house",
        "_version_": 1546499704327504000
      },
      {
        "id": "352693",
        "title": "Fast, Cheap & Out of Control",
        "_version_": 1546499704327504000
      }
    ]
  }
}
```

Searching query q = title : “Shah Rukh Khan”..!!

Solr Admin

localhost:8983/solr/#/wikipedia/query

Search

Solr

Dashboard

Logging

Core Admin

Java Properties

Thread Dump

wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema Browser

Segments info

Request-Handler (qt)

/select

common

q

title : "Shah Rukh Khan"

fq

sort

start, rows

010

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

☒ indent


☐ debugQuery

☐ dismax

http://localhost:8983/solr/wikipedia/select?q=title+.%3A+.%22Shah+Rukh+Khan%22&wt=json&indent=true

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 268,
    "params": {
      "q": "title : \"Shah Rukh Khan\"",
      "indent": "true",
      "wt": "json",
      "_": "1474901420439"
    }
  },
  "response": {
    "numFound": 1,
    "start": 0,
    "docs": [
      {
        "id": "352699",
        "title": "Shah Rukh Khan",
        "_version_": 1546545838075936800
      }
    ]
  }
}
```

Searching id using fl=id ,q= title : "Shah Rukh Khan" ...!!



Dashboard

Logging

Core Admin

Java Properties

Thread Dump

wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema Browser

Segments info

Request-Handler (qt)

/select

common

q

title : "Shah Rukh Khan"

fq

sort

start, rows

010

fl

id

df

Raw Query Parameters

key1=val1&key2=val2

wt

xml

☒ indent

☐ debugQuery

☐ dismax

http://localhost:8983/solr/wikipedia/select?q=title+%3A+%22Shah+Rukh+Khan%22&fl=id&wt=xml&indent=true

<?xml version="1.0" encoding="UTF-8"?>

<response>

<lst name="responseHeader">

<int name="status">0</int>

<int name="QTime">1</int>

<lst name="params">

<str name="q">title : "Shah Rukh Khan"</str>

<str name="indent">true</str>

<str name="fl">id</str>

<str name="wt">xml</str>

<str name="_">1475429202442</str>

</lst>

</lst>

<result name="response" numFound="1" start="0">

<doc>

<str name="id">352699</str></doc>

</result>

</response>