

APACHE
LUCENE
EUROCON



Solr 4 Highlights - Preview

Mark Miller, LucidImagination.com
markrmiller@gmail.com

Presented by

lucid
IMAGINATION



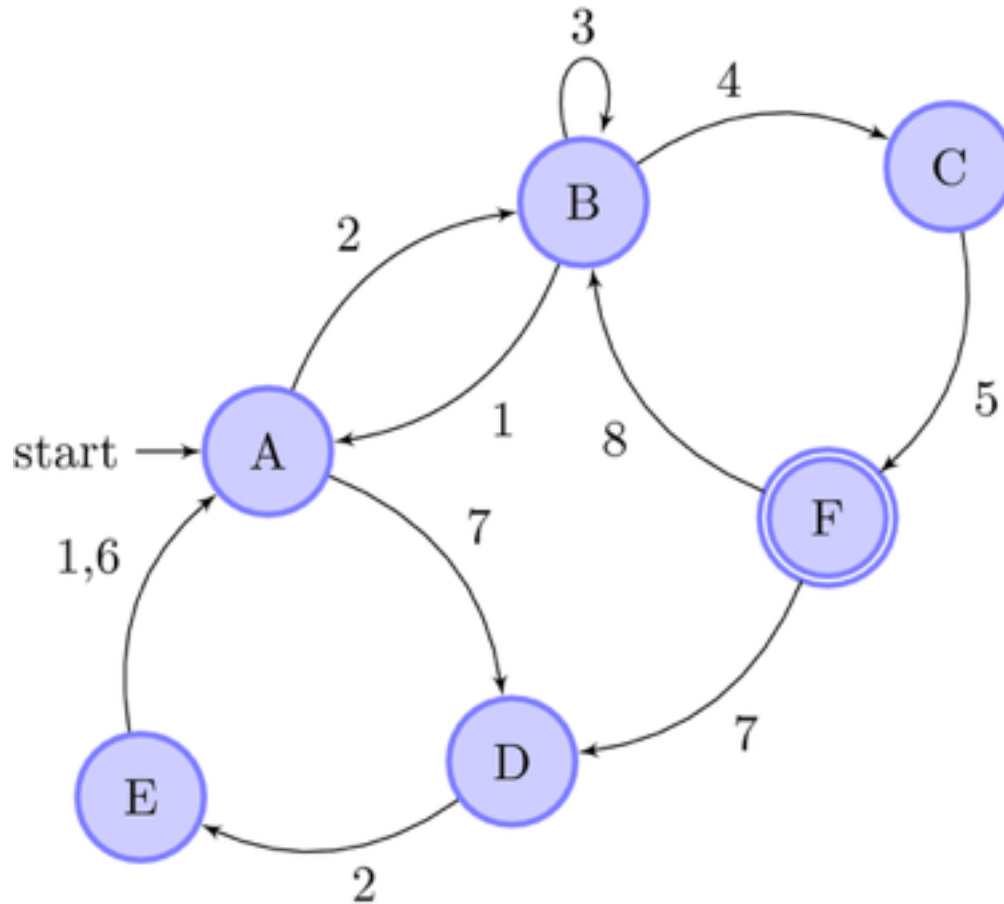
What I Will Cover

- What are some cool features coming in Solr 4?
- Mark Miller - Lucene/Solr committer, PMC member, Apache member, blah blah blah
- This talk is a shortened version of what I will give in Barcelona.
- DirectSolrSpellChecker
- NRT
- Realtime Get
- SolrCloud - search side
- SolrCloud - indexing side (WIP)

DirectSolrSpellchecker

- New spellcheck implementation that works with the new automaton code.
- A practical benefit of this spellchecker is that it requires no additional data structures (neither in RAM nor on disk) to do its work.
- Also, no build phase - suggestions are always up to date.

Automaton



SolrConfig

```
<searchComponent name="spellcheck" class="solr.SpellCheckComponent">
```

```
  <!-- a spellchecker built from a field of the main index -->
```

```
  <lst name="spellchecker">
```

```
    <str name="name">default</str>
```

```
    <str name="field">name</str>
```

```
    <str name="classname">solr.DirectSolrSpellChecker</str>
```

Realtime Get

```
<updateLog class="solr.FSUpdateLog">  
  <str name="dir">${solr.data.dir:}</str>  
</updateLog>
```

<!-- realtime get handler, guaranteed to return the latest stored fields of any document, without the need to commit or open a new searcher. The current implementation relies on the updateLog feature being enabled.

```
-->  
<requestHandler name="/get" class="solr.RealTimeGetHandler">  
  <lst name="defaults">  
    <str name="omitHeader">true</str>  
  </lst>  
</requestHandler>
```

Realtime Get

```
curl 'http://localhost:8983/solr/update/json?commitWithin=10000000'  
-H 'Content-type:application/json' -d '  
[  
  {  
    "id" : "mydoc",  
    "title" : "realtime-get test!"  
  }  
]'
```

<http://localhost:8983/solr/get?ids=mydoc,mydoc>

<http://localhost:8983/solr/get?id=mydoc&id=mydoc>

```
{"doc":{"id":"mydoc","title":["realtime-get test!"]}}
```

NRT

UpdateHandler Re-Architecture

- Fix commit waits for background merges
- Fix reloaded core does not share IndexWriter
- Remove unnecessary locking
- Add NRT
- SoftCommit
- SoftAutoCommit

NRT AutoCommit

```
<updateHandler class="solr.DirectUpdateHandler2">
```

```
  <autoCommit>
```

```
    <maxTime>60000</maxTime>
```

```
  </autoCommit>
```

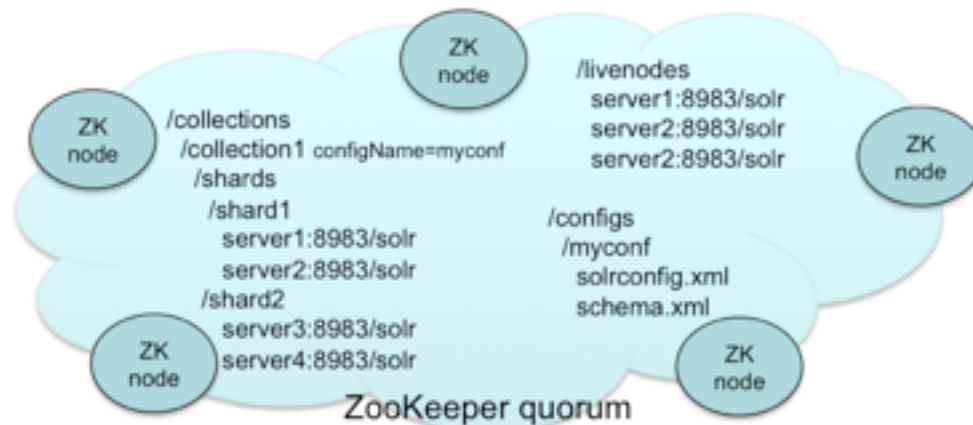
```
  <autoSoftCommit>
```

```
    <maxTime>1000</maxTime>
```

```
  </autoSoftCommit>
```

SolrCloud - Search Side

<http://.../solr/collection1?distrib=true>



ZooKeeper Layout

/solr

/solr/configs

/solr/configs/conf1

/solr/configs/conf1/mapping-ISOLatin1Accent.txt

DATA: ...supressed...

/solr/configs/conf1/protwords.txt

DATA: ...supressed...

/solr/configs/conf1/old_synonyms.txt

DATA: ...supressed...

/solr/configs/conf1/solrconfig.xml

DATA: ...supressed...

/solr/configs/conf1/stopwords.txt

DATA: ...supressed...

/solr/configs/conf1/schema.xml

DATA: ...supressed...

← Centralized Configuration

Cluster Composition



/solr/collections/collection1/shards

/solr/collections/collection1/shards/shard4/localhost:50270_solr_

DATA:

node_name=localhost:50270_solr

url=http://localhost:50270/solr/

Ephemeral Nodes

```
/solr/live_nodes
```

```
/solr/live_nodes/localhost:50272_solr
```

```
/solr/live_nodes/localhost:50268_solr
```

```
/solr/live_nodes/localhost:50274_solr
```

```
/solr/live_nodes/localhost:50270_solr
```

```
/solr/live_nodes/localhost:50266_solr
```

Getting Started

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

```
java -Dbootstrap_confdir=./solr/conf  
-Dcollection.configName=myconf  
-DzkRun  
-jar start.jar
```

Run an internal
ZK server

Upload /solr/conf
to ZK and call it
"myconf"

<http://localhost:8983/solr/collection1/admin/zookeeper.jsp>

SolrCloud - Index Side

Goals

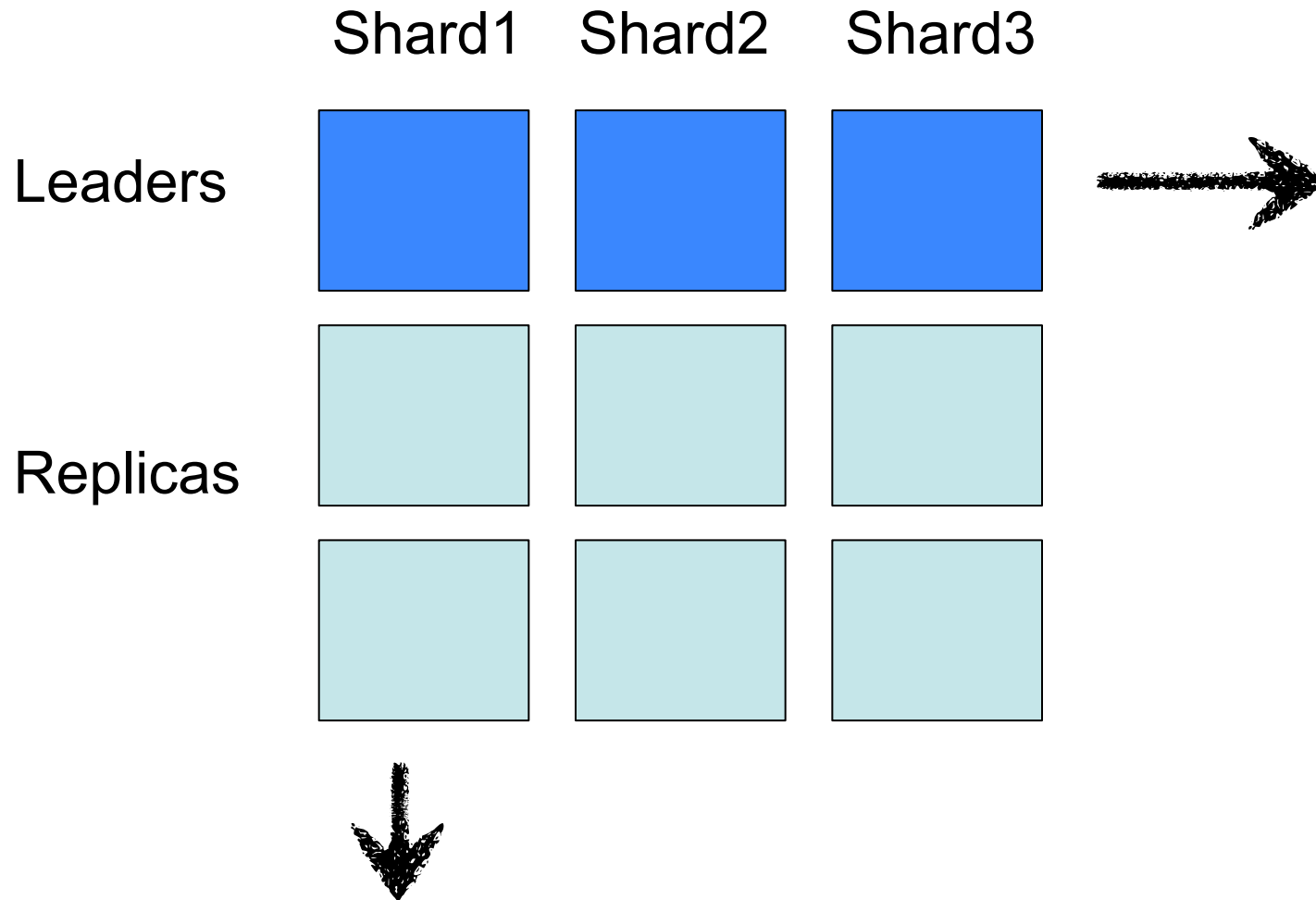
- Distributed Indexing
- Fault Tolerant
- Elastic
- Durable

SolrCloud - Index Side

What's Being worked on this very moment?

- Realtime Get
- Transaction Log
- Leader per Shard
- Auto Assign Shard Names
- Versions

Cluster



Wrap Up

- These are only a few features - there are many more already, and more coming.
- Pseudo Join, Pivot Faceting, Pseudo Fields, Per Segment Faceting, Configure Codec Per Field, Use Alternate Scoring Algorithms...
- When? Your guess is as good as mine ;) I hope sometime early next year.

Sources

- Links

- <http://wiki.apache.org/solr/SpellCheckComponent>
- <http://wiki.apache.org/solr/NearRealtimeSearch>
- <http://www.lucidimagination.com/blog/2011/09/07/realtime-get/>

Contact

- Mark Miller
 - markrmiller@gmail.com / mark.miller@lucidimagination.com
 - <http://www.lucidimagination.com/blog/author/markmiller/>
 - <http://twitter.com/heismark> 