

Solr as your Search and Suggest Engine

Karan Nangru

Xebia

6TH IndicThreads.com
Conference On

JAVA

2,3 DEC 2011

PUNE, INDIA

What is Solr.....

- Enterprise search server that takes in data in the form of documents (indexing) - JSON, XML, HTTP binaries
- Query (HTTP GET) to fetch the results
- Internally - It's a wrapper over Lucene



What lured us to Solr

- A wrapper over Lucene with predefined extensible features
- Jet set go..... (Web application Archive (WAR) format)
- Provides cool search and suggest features
- Scalable
- Maven support



Setting up the Solr server....

- the admin console....
- the admin query....
- the maven script....

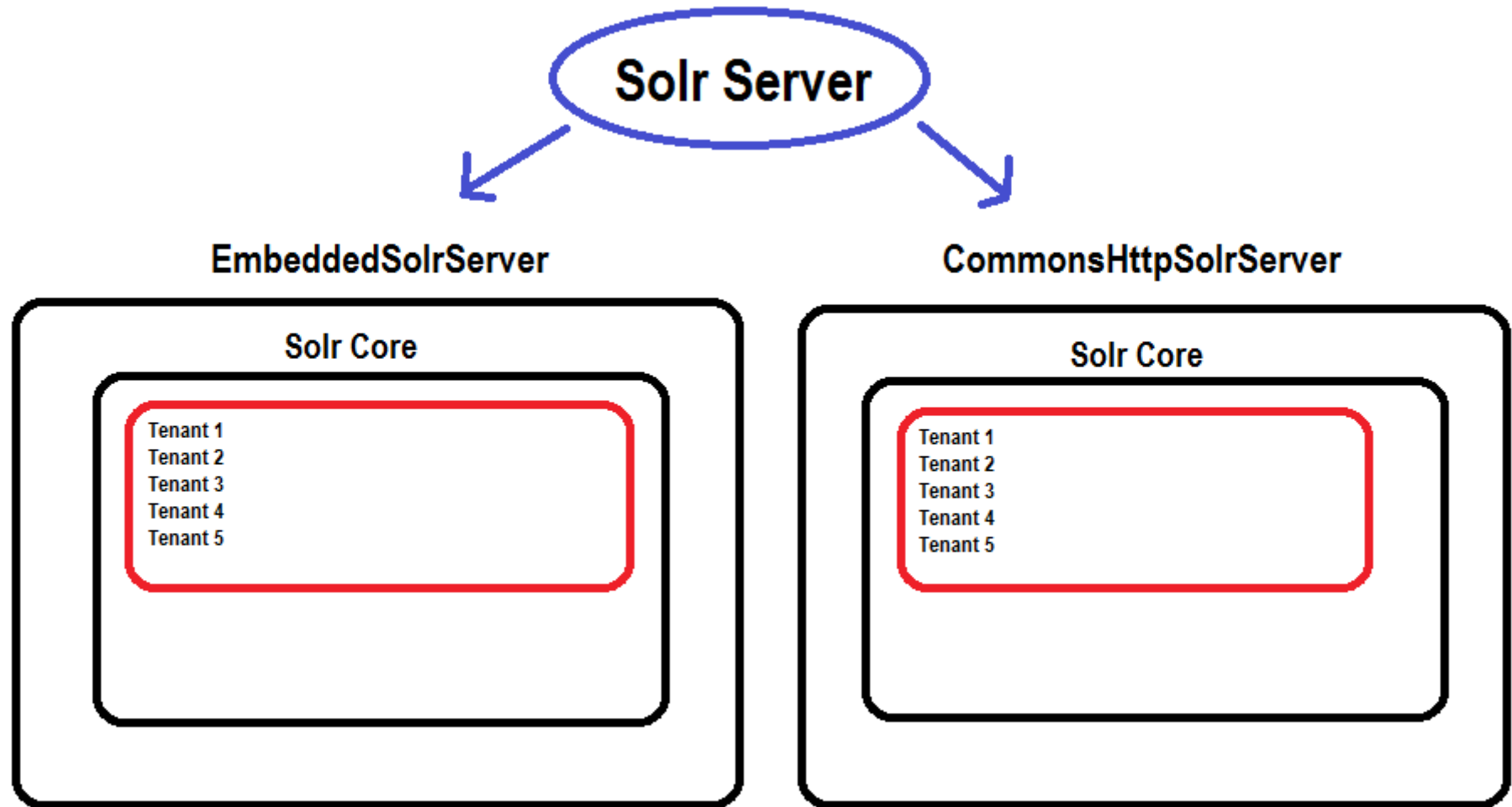


Storage Structure....

		FIELDS		
		A	B	C
DOCUMENTS	1	First Name	Last Name	Age
	2	Karan	Nangru	27
	3	Rahul	Sharma	35
	4	Sameer	Arora	38
	5	Guneet	Sahai	44
	6	Shekhar	Gulati	34



Using Solr @ Java....



Configuration and Schema...

Configuration

```
<requestHandler name="/search" class="solr.SearchHandler"
default="true">
```

Schema

```
<fields>
  <field name="dictionary" type="dictionary" indexed="true"
    stored="false" multiValued="true" />
</fields>

<fieldType name="dictionary" class="solr.TextField"
  positionIncrementGap="100">
  <analyzer type="index">
    <tokenizer class="sdfsd" />
    <filter class="sdflksd"/>
  </analyzer>
  <analyzer type="query">
    <tokenizer class="sdfsd" />
    <filter class="fdssd"/>
  </analyzer>
</fieldType>
```



Tokenizer Factories...

WhitespaceTokenizerFactory

you are listening to Karan at Indic Threads



Token Filter Factories...



Solr for Searching

- Configuration setup....
- Schema setup....
- Java client....
- JUnit....



example....



Solr for Suggestions

- Configuration setup....
- Schema setup....
- Java client....
- JUnit....



example....



Excess of anything is bad

There are various features available but analyze your use case and only use features that you REALLY need.....

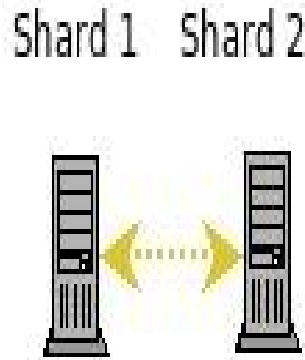
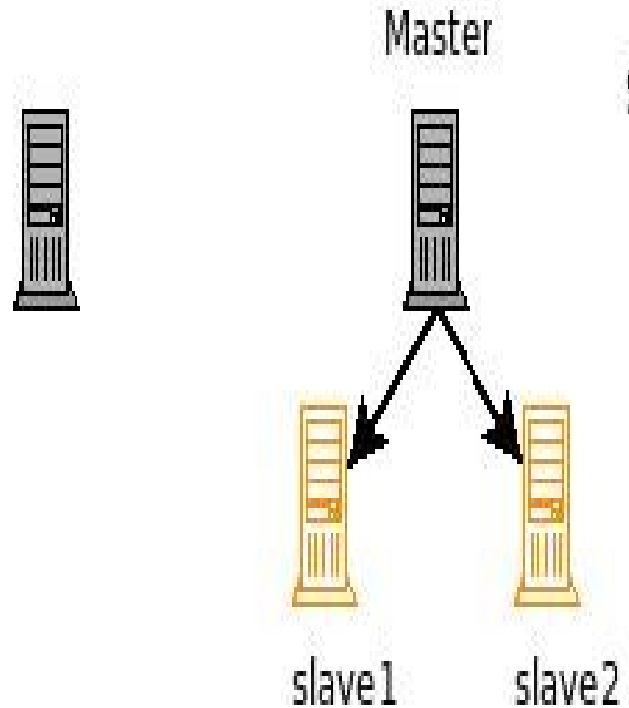
The more you add, the larger the data, the bulkier the operations, the slower the response time



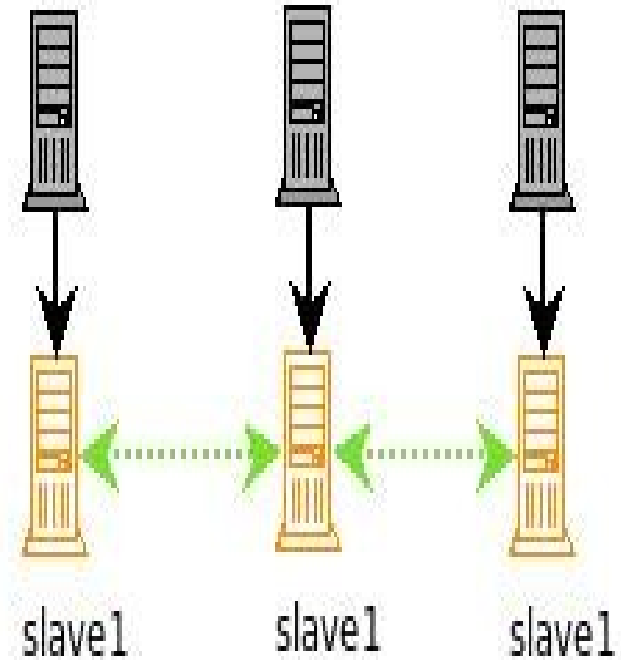
The ease of scaling....

Single Server Replication Distributed

Distributed + Replication



Shard 1 Master Shard 2 Master Shard 3 Master





Karan Nangru

Senior Consultant @ Xebia IT Architects

Twitter ID: KaranNangru

<http://karannangru.wordpress.com>

