Solr Learning

Monday, March 17, 2014 10:42

1. First of all: Solr is based on Lucene, which is another important project of text searching.
2. Solr version I learned is Version 4
3. Install Solr:
    a. Download tar package of Solr
    b. Un-pack Solr package and move it to some path
    c. Set SOLR_INSTALL. (Note the difference of SOLR_INSTALL and SOLR_HOME)
    d. Run Solr on Jetty, which is a slight-weight web container like Tomcat. We can also set Solr based on Tomcat.
    e. Java -jar start.jar….. (Start Solr commond)
4. Each Solr home can have multiply core. (Default Solr home is ${start.jar's path}/solr. But we can set it by setting -Dsolr.solr.home)
5. Each core have two important folder and file:
    a. Solr.xml: which contains all the core information in it.
    b. Data: the real index file in it
    c. Conf: all the configuration files in it
        i. Solrconfig.xml: which is specific configuration based on each Core
        ii. Schema: which is related to the schema of this Core
6. Simple administrator website: http://localhost:8983/solr (port can be set in the solr.xml)
7. Solr configuration
    a. Solrconfig.xml
        i. Request handler. (like query/browse/search/get in example)
        Which will be add to the request if user doesn't specify it in their request.
            1) Set some other feature.
                a) Wt - Writer Type. Possible value: xml, jason, python, velocity
                b) Facet
                c) Highlight
                d) Spell checking
                e) …
            2) Searcher.
                a) Only commit document can be seen by the user.
                b) Solr will create a new searcher once commit.
                c) In order to optimism, Solr support a warm-search.
    b. Schema.xml
        i. Field: simple field
            1) Store field
            2) Index field
        ii. Unique key
        iii. Multi-valued fields
        iv. Dynamic-fields
        v. Copy field
        vi. Field type
8. Analysis Text
    a. Raw text must be indexed by Solr, which is the point of text analysis
    b. LowerCaseFilterFactory
    c. StopFilterFactory (remove "a", "in"… and some unimportant words in raw text)
    d. SynonymFilterFactory
    e. …