Solr for Developers

Course Summary

Description

This course introduces students to the Solr platform. Through a combination of lecture, discussion and labs students will gain hands on experience configuring effective search and indexing.

The class begins with basic Solr installation and configuration then teaches the attendees the search features of Solr. Students will gain experience with faceting, indexing and search relevance among other features central to the Solr platform. The course wraps up with a number of advanced topics including spell checking, suggestions, Multicore and SolrCloud.

Format: Lectures and hands on labs. (50% lecture + 50% labs). Pace of the class is determined by the students.

Topics

- Fundamentals
- Searching
- Indexing
- Schema Updating
- Relevance
- Extended features
- Multicore
- SolrCloud
- Developing with Solr API
- Developing with Lucene API
- Conclusion

Audience

This course is designed for developers, business users, and administrators.

Prerequisites

All attendees should be experienced technical staff with a background in web application operations and, preferably, development.

Duration

Three days
Solr for Developers

Course Outline

I. Fundamentals
   A. Solr Overview
   B. Installing and running Solr
   C. Adding content to Solr
   D. Reading a Solr XML response
   E. Changing parameters in the URL
   F. Using the browse interface
   G. Labs: install Solr, run queries

II. Searching
    A. Sorting results
    B. Query parsers
    C. More queries
    D. Hardwiring request parameters
    E. Adding fields to default search
    F. Faceting
    G. Result grouping
    H. Labs: advanced queries, experiment with faceted search

III. Indexing
     A. Adding your own content to Solr
     B. Deleting data from Solr
     C. Building a bookstore search
     D. Adding book data
     E. Exploring the book data
     F. Dedupe update processor
     G. Labs: indexing various document collections

IV. Schema Updating
     A. Adding fields to the schema
     B. Analyzing text
     C. Labs: customize Solr schema

V. Relevance
   A. Field weighting
   B. Phrase queries
   C. Function queries
   D. Fuzzier search
   E. Sounds-like
   F. Labs: implementing queries for relevance

VI. Extended features
    A. More-like-this
    B. Geospatial
    C. Spell checking
    D. Suggestions
    E. Highlighting
    F. Pseudo-fields
    G. Pseudo-joins
    H. Multilanguage
    I. Labs: implementing spell checking and suggestions

VII. Multicore
     A. Adding more kinds of data
     B. Labs: creating and administering cores

VIII. SolrCloud
      A. Introduction
      B. How SolrCloud works
      C. Commit strategies
      D. ZooKeeper
      E. Managing Solr config files
      F. Labs: administer SolrCloud

IX. Developing with Solr API
     A. Talking to Solr through REST
     B. Configuration
     C. Indexing and searching
     D. Solr and Spring
     E. Labs: code to read and write Solr index, exercise in Spring with Solr

X. Developing with Lucene API
    A. Building a Lucene index
    B. Searching, viewing, debugging
    C. Extracting text with Tika
    D. Scaling Lucene indices on clusters
    E. Lucene performance tuning
    F. Labs: coding with Lucene

XI. Conclusion
    A. Other approaches to search
       1. ElasticSearch
       2. DataStax Enterprise: Solr+Cassandra
       3. Cloudera Solr integration
       4. Blur
    B. Future directions

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically.