

SolrUseCases

Introduction

Solr can do a lot, and no single application uses all of the available features. Sharing your stories of how you implemented Solr helps other users understand all of the things Solr can do.

Template

Please copy and paste the template information below into the Your Stories section below. If you can only share partial information, please do what you can.

- Company Name:
- URL:
- Demo/Presentation of application available (please provide URL)?:
- Solr Version:
- Why did you choose Solr?
- Goals of implementation:
- Challenges Overcome:
- Data
 - Number of docs
 - Queries per second (or other time units)
 - Docs per second
 - Avg Doc length
- Hardware
 - CPU/Memory/Disk
 - Number of Nodes/replicas/shards
 - Network capacity

Your Stories

- Company Name: State and University Library, Denmark.
- URL: <http://statsbiblioteket.dk> or <http://sbdevel.wordpress.com/net-archive-search/> for a write up.
- Demo/Presentation of application available (please provide URL)?: Not available due to Danish legislation.
- Solr Version: 4.8.1.
- Why did you choose Solr?: Past positive experience.
- Goals of implementation: Index the complete Danish Net Archive from 2005 to present, currently 500TB of raw data and growing.
- Challenges Overcome: Fast faceting on high-cardinality String fields, solved with <http://tokee.github.io/lucene-solr/>
- Data
 - Number of docs: 7 billion indexed as of 2014-11-14 and growing. Estimated 16 billion by mid 2015.
 - Queries per second (or other time units): Median response time of 1 second for term searches with faceting on 6 fields, one of them URL with ~6 billion unique values. Non-faceted search median response time < 200ms, throughput of non-facet search ~50 searches/sec.
 - Docs per second: Indexing at ~400 docs/sec on a dedicated 24 core indexing machine.
 - Avg Doc length: 3KB indexed (total index size: 20TB).
- Hardware
 - CPU/Memory/Disk: 1 dedicated index machine, 1 search machine: 16 CPU cores, 256GB, 25*932GB SSD (Samsung 840 EVO).
 - Number of Nodes/replicas/shards: 25 separate Solr instances on the search machine, no extra replicas, 1 shard/Solr.
 - Network capacity: 1 GBit (not currently relevant due to only having a single searcher).

Example