

Practical 1

Report on Open Source Software

SOLR



History

Solr is an open-source enterprise-search platform, written in Java, from the Apache Lucene project.

In 2004, Solr was created by Yonik Seeley at CNET Networks as an in-house project to add search capability for the company website. In January 2006, CNET Networks decided to openly publish the source code by donating it to the Apache Software Foundation. In January 2007, Solr graduated from incubation status into a standalone top-level project (TLP) and grew steadily with accumulated features, thereby attracting users, contributors, and committers. In September 2008, Solr 1.3 was released including distributed search capabilities and performance enhancements among many others. In January 2009, Yonik Seeley along with Grant Ingersoll and Erik Hatcher joined Lucidworks (formerly Lucid Imagination), the first company providing commercial support and training for Apache Solr search technologies. Since then, support offerings around Solr have been abundant. November 2009 saw the release of Solr 1.4. In March 2010, the Lucene and Solr projects merged. Solr became a Lucene sub project. Separate downloads continued, but the products were now jointly developed by a single set of committers. In 2011 the Solr version number scheme was changed in order to match that of Lucene. After Solr 1.4, the next release of Solr was labeled 3.1, in order to keep Solr and Lucene on the same version number. In October 2012 Solr version 4.0 was released, including the new SolrCloud feature. 2013 and 2014 saw a number of Solr releases in the 4.x line, steadily growing the feature set and improving reliability. In February 2015, Solr 5.0 was released, the first release where Solr is packaged as a standalone application, ending official support for deploying Solr as a war. In April 2016, Solr 6.0 was released. In September 2017, Solr 7.0 was released. In March 2019, Solr 8.0 was released. In November 3, Solr 8.7.0 was released. Solr is widely used for enterprise search and analytics use cases and has an active development community and regular releases.

Idea

Solr was designed for providing distributed search, index replication, and for scalability and fault tolerance.

What problems does it solves?

It uses the Lucene Java search library at its core for full-text indexing and search, and has REST-like HTTP/XML and JSON APIs that make it usable from most popular programming languages. Solr's external configuration allows it to be tailored to many types of applications without Java coding, and it has a plugin architecture to support more advanced customization. Solr runs as a standalone full-text search server.

Solr offers powerful features such as distributed full-text search, faceting, near real-time indexing, high availability, NoSQL features, integrations with big data tools such as Hadoop, and the ability to handle rich-text documents such as Word and PDF.

Detailed licensing model

Solr is licensed under Apache License 2.0. The Apache License is permissive; unlike copyleft licenses, it does not require a derivative work of the software, or modifications to the original, to be distributed using the same license. It still requires application of the same license to all unmodified parts. In every licensed file, original copyright, patent, trademark, and attribution notices must be preserved.

In every licensed file changed, a notification must be added stating that changes have been made to that file. The Apache License 2.0 makes sure that the user does not have to worry about infringing any patents by using the software. The user is granted a license to any patent that covers the software. This license is terminated if the user sues anyone over patent infringement related to this software. This condition is added in order to prevent patent litigations.

Popularity

Solr is ranked number three by DB-Engines, which ranks database management systems and search engines according to their **popularity**.

Solr is bundled as the built-in search in many applications such as content management systems and enterprise content management systems. Hadoop distributions from Cloudera, Hortonworks and MapR all bundle Solr as the search engine for their products marketed for big data. DataStax DSE integrates Solr as a search engine with Cassandra. Solr is supported as an end point in various data processing frameworks and Enterprise integration frameworks

Solr exposes industry standard HTTP REST-like APIs with both XML and JSON support, and will integrate with any system or programming language supporting these standards. For ease of use there are also client libraries available for Java, C#, PHP, Python, Ruby and most other popular programming languages.

Impact

Solr introduced enhancements in indexing, searching and faceting along with many other improvements such as rich document processing (PDF, Word, HTML), Search Results clustering based on Carrot2 and also improved database integration.

Solr features many additional plug-ins.

Solr 5.3 featured a built-in pluggable Authentication and Authorization framework.

Solr added support for executing Parallel SQL queries across SolrCloud collections. Includes StreamExpression support and a new JDBC Driver for the SQL Interface.

Solr added support multiple replica types, auto-scaling, and a Math engine.

Solr nodes can now listen and serve HTTP/2 requests and plotting math expressions in Apache Zeppelin is also made possible.

Many bugfixes and component updates are provided.

Solr has both individuals and companies who contribute new features and bug fixes.