SolrCloud Questionaire

Content

SOLR AND ZOOKEEPER RELATIONS
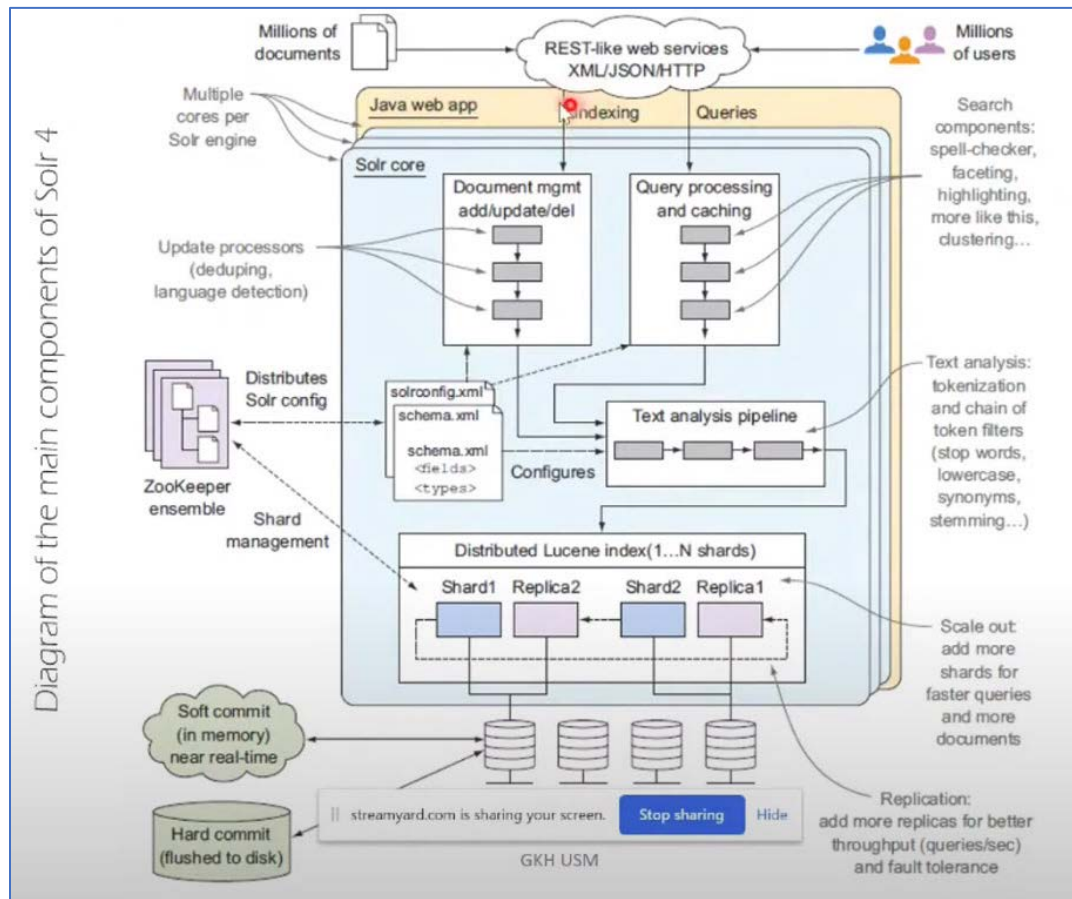
Diagram of the main components of Solr 4

SOLR SYSTEM REQUIREMENTS

https://solr.apache.org/guide/8_4/solr-system-requirements.html
https://cwiki.apache.org/confluence/display/solr/solrperformanceproblems

HARDWARE REQUIREMENTS

Solr runs on a 64-bit Java.
A 64-bit Java requires a 64-bit operating system.

SUPPORTED OPERATING SYSTEMS

Solr is tested on several versions of Linux, macOS and Windows.

JAVA REQUIREMENTS

You will need **the Java Runtime Environment (JRE) version 1.8 or higher.**

```
$ java -version
java version "1.8.0_60"
Java(TM) SE Runtime Environment (build 1.8.0_60-b27)
Java HotSpot(TM) 64-Bit Server VM (build 25.60-b23, mixed mode)
```

LUCENE/SOLR 8.X

- Requires **Java 8** or higher.

- This version has continuous testing with Java 9, 10, 11, 12 and the pre-release version of Java 13.

- There were known issues with Kerberos with Java 9+ prior to Solr 8.1. If using 8.0, you should test in your environment.

- Be sure to test with SSL/TLS and/or authorization enabled in your environment if you require either when using Java 9+.

SITECORE 9.3 SYSTEM REQUIREMENTS - XC

https://doc.sitecore.com/en/developers/93/sitecore-experience-commerce/system-requirements.html
(Sitecore Version 10 does not have system requirements information)

HARDWARE REQUIREMENTS

- 4 core processor
- 16 GB of RAM

HOSTING ENVIRONMENT REQUIREMENTS

The following table lists the software requirements for Sitecore XC 9.3 hosting environment.

Operating system	- Windows Server 2016 - Windows Server 2019 - Windows 10 Professional (64-bit)
.NET Framework	.NET Framework 4.7.1 Developer Pack (including available updates) - ASP.NET MVC 5.2 or later - ASP.NET Core 2.1 .NET Core Windows Server Hosting 2.1.6 or later versions of 2.1 - OData 7.2.0
Database	- Microsoft SQL Server 2019 - Microsoft SQL Server 2017 - Microsoft SQL Server 2016 SP2 - Microsoft Azure SQL - Redis data store - Note Sitecore recommends using Redis for Linux for production installations. The version for Windows should only be used for development or non-production single machine deployments.
Web server	- IIS 10.0 FC: (XP0 9.3) - Microsoft Web Deploy 3.6

	• URL Rewriter
Development	• Visual Studio 2017 + latest updates • Visual Studio 2017 Tooling • Visual Studio 2019 • Visual Studio 2019 Tooling • .NET Core 2.1.7 Windows Hosting module • MSBuild Microsoft Visual Studio Web targets (available from Nuget) • Git client
Deployment	PowerShell 6.0 or later
Runtime	• .Net Core Runtime 2.1 • Node.js • OData Connected Service (available from Visual Studio Marketplace)
Search indexing	Solr 8.1.1 → 8.4.0 for Sitecore 10.1
Sitecore software	• Sitecore Experience Platform 9.3 (available on dev.sitecore.net) • Sitecore Experience Accelerator (SXA) 9.3 (available on dev.sitecore.net) • Sitecore PowerShell Extensions 6.0 for Sitecore 9.3 (available with SXA dowloads on dev.sitecore.net) • Sitecore Identity 4.0.0 • Note Sitecore SXA and Sitecore PowerShell Extensions are installed when you run the Sitecore Experience Commerce deployment script, as part of the XC installation process.
Web browsers	• Microsoft Edge • Internet Explorer 11 • Mozilla Firefox • Google Chrome • Apple Safari

SITECORE HA BACKEND DEPLOYMENT

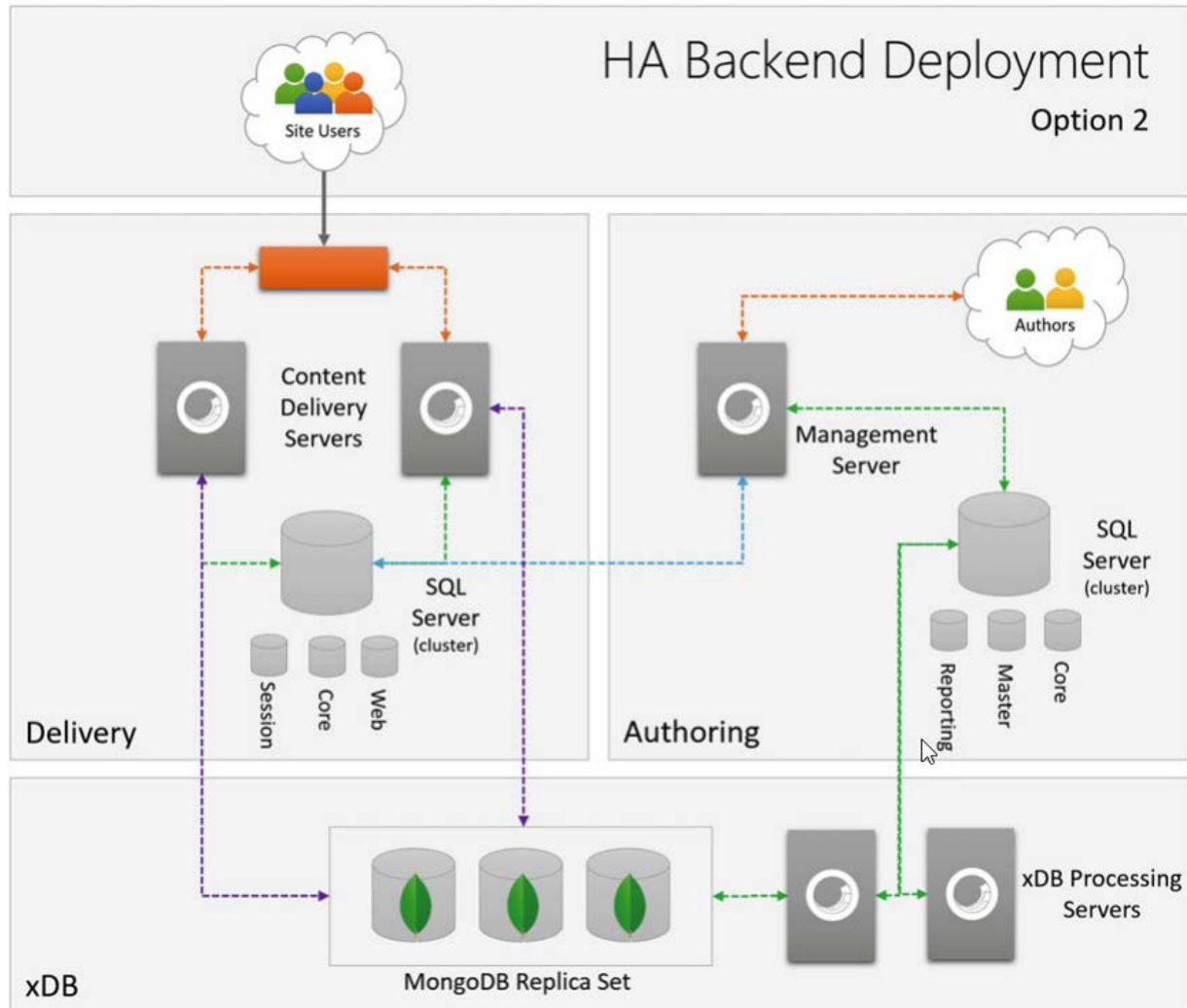Reference : Professional Sitecore 8 Development.pdf, Phil Wicklund & Jason Wilkerson

Figure 1-6. *Option 2 features high availability through the solution, save the management server*

SOLR PERFORMANCE DATA

Created by ASF Infrabot on Jun 28, 2019
https://cwiki.apache.org/confluence/display/SOLR/SolrPerformanceData

CNET SHOPPER.COM - STANDALONE

- **2.4GHz dual opteron (DL385) w/ 16GB memory Linux (2.6.9).**

- index used in these tests contained ~400K records, took up ~900MB of disk

- a cron job forcibly triggered a commit (even though the index hadn't changed) every 15 minutes to force a new searcher to be opened and autowarmed while the queries were being processed

DISCOGS.COM - STANDALONE

- **2.66Ghz P4, with 1 gig ram**

- index has about 1.2 million documents, 1.2 gigs in size.

- This machine handles 250,000 queries per day

HATHITRUST - SOLRCLOUD

- Response time is around 200 ms

- 5 million full-text books indexed

- **production index is spread across 10 shards on 4 machines**

- total index size of over 2 Terabytes

ZVENTS - SOLRCLOUD

- 8 millions users monthly with engaging local contents

- **queries are served by a cluster of 8 machines, each having 16Gigs of memory and 4 cores**

- Our search index contains over 4 million documents

- total index size of over 2 Terabytes

DANISH WEB ARCHIVE – SINGLE SOLRCLOUD

- Archive with 500TB+ harvested web resources

- A single 24 core 256GB CentOS machine builds the shards

- Solr indexer has 32GB heap

- **All instances are in a single SolrCloud**

JAVA PLATFORM - RUNTIME

https://java.com/en/download/help/sysreq.html
https://www.oracle.com/java/technologies/javase/products-doc-jdk8-jre8-certconfig.html

WINDOWS

- Windows 10 (8u51 and above)
- Windows 8.x (Desktop)
- Windows 7 SP1
- Windows Vista SP2
- Windows Server 2008 R2 SP1 (64-bit)
- Windows Server 2012 and 2012 R2 (64-bit)
- **RAM: 128 MB**
- **Disk space: 124 MB for JRE; 2 MB for Java Update**
- **Processor: Minimum Pentium 2 266 MHz processor**
- Browsers: Internet Explorer 9 and above, Firefox

MAC OS X

- Intel-based Mac running Mac OS X 10.8.3+, 10.9+
- Administrator privileges for installation
- 64-bit browser
- A 64-bit browser (Safari, for example) is required to run Oracle Java on Mac.

LINUX

- Oracle Linux 5.5+
- Oracle Linux 6.x (32-bit), 6.x (64-bit)
- Oracle Linux 7.x (64-bit)2 (8u20 and above)
- Red Hat Enterprise Linux 5.5+1 6.x (32-bit), 6.x (64-bit)
- Red Hat Enterprise Linux 7.x (64-bit)2 (8u20 and above)
- Suse Linux Enterprise Server 10 SP2+, 11.x
- Suse Linux Enterprise Server 12.x (64-bit)2 (8u31 and above)
- Ubuntu Linux 12.04 LTS, 13.x
- Ubuntu Linux 14.x (8u25 and above)
- Ubuntu Linux 15.04 (8u45 and above)
- Ubuntu Linux 15.10 (8u65 and above)
- Browsers: Firefox

SolrCloud Questionaire

SOLRCLOUD MODEL - EXAMPLE

Solr and ZooKeeper is just "Service" and has minimal requirement of Java and could be installed on only a server and also installed multiple servers.

- Core : 4 Core, 64-bit operating system – Java 8+, JRE 1.8+
- Ram : 12 to 16 GB
- ZooKeeper : At least 3 ZK service and 5 quorum is recommended
- Solr Server : 2 Servers and above
- May be considered RAM size - documents and index size: Soft & Hard Commit
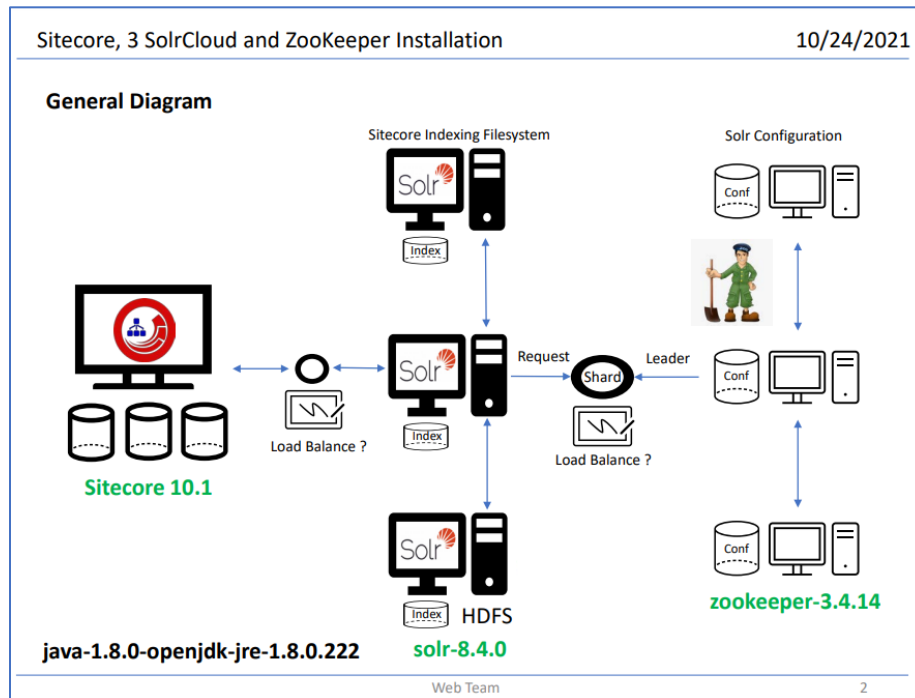
Configuration : 1 Sitecore 10.1, 3 Solr Indexing Service, 3 ZooKeeper Service
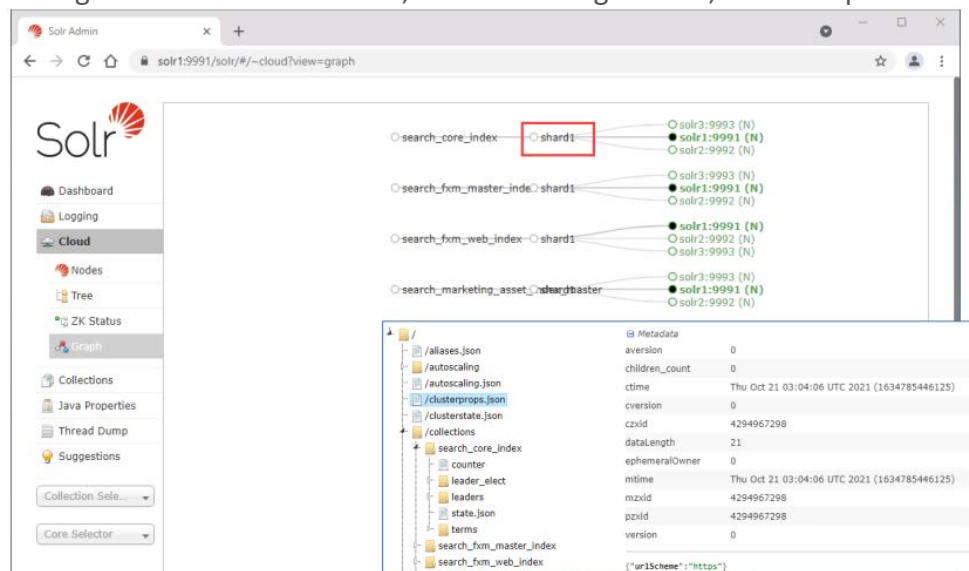
PROS & CONS FOR SOLR

Source :
https://www.realdecoy.com/2018/10/04/solr-powered-search-should-you-look-into-it/

pros	cons	Weighing
Open source = affordable	The need for external monitoring	need to add monitoring tools - such as SPM and AppDynamics
Large community = active support	Open-source means no guarantees	need to address indexing - ideal memory of 12GB
Scalability = flexibility and growth		open-source
Multi-faceted faceting		

Bottom line
You will need :
- a fully capable development team
- adequate resources in terms of memory and hardware for the release systems
- the right documentation for both the business requirements and the solution itself

SOLR VS. SOLRCLOUD

https://www.quora.com/What-factors-should-I-take-into-consideration-in-choosing-between-plain-vanilla-Solr-vs-SolrCloud

- **How large is your data**, does it fit on a single machine and can a single server respond to your queries in a time that is acceptable by your use case? The larger the data, the easier it will be to handle that using SolrCloud, especially when you need to divide it into multiple nodes, handle replication and so on.
- **Is a single Solr node able to index my data** in a acceptable time? If no than the SolrCloud is something that you will probably benefit from as it is easy to distribute a single collection between multiple nodes.
- How near real time you need or want to be? The closer to near real-time the more likely SolrCloud would be a better choice.
- Is Solr a crucial part of the whole application stack - so do we need it to be fault tolerant and highly available? Yes - - SolrCloud is probably a good choice.
- Do I need SolrCloud only features?
- **Is my infrastructure dynamic?** Yes - SolrCloud is probably a good choice because of ease of replica management and autoscaling framework.
- **Will my data grow in foreseeable future and how much it will grow?** If it will grow in a way that it may be a need to distribute - use SolrCloud.

SOLRCLOUD QUESTIONAIRE

Q1. Does SolrCloud appropriate to us ? Why not Standalone ?

'CO' does not use Solr directly but via Sitecore and all Indexing and manage of the data is reply on Sitecore.

Most of search is targeted to Sitecore content information that Sitecore webpage

'CO' do not provide real-time indexing service to public. Which means 'CO' do not need indexing and reindexing often.

Why 'CO' should have big Solr system ?

Q2. How many documents on us exclude duplicates ?

'CO' has about 13 GB of index file size.

Does this really correct ?

Did we remove duplicated documents ?

Q3. Does ZooKeeper Load Balancing is same to Traffic Load balancer ?

Basically, ZooKeeper receive requests from multiple Solr Servers and ZK has load balancing and fail-over. But Is this for SolrCloud topology or Network load balancing ?

I never see the load balancer that we have and No idea about it. But most inbound website traffics are handled by load balancer and it is normally switch or router like system that has software or hardware load balancing algorithm on layer 4 or layer 7and is not same as Solaris, Linux or Windows system. I am not sure we can install ZooKeeper on website traffic Load Balancer.

Q4. How many Shard should we have for our data size ?

The Shard is the Replica management unit and Replicas belong to Shard.

If there is 2 Shard then total 6 replica will be created when a Shard has 3 replica.

Should we have multiple Shard ? How to calculate this number of Shard for us ?

Do we need distributed Replica groups to divide our documents into multiple sets ?

Q5. Who will make solrconfig.xml and What is the rule to make and add/delete/change ?

Solr is the Out Shell of Russian doll. Lucene stay inside of Solr and run separately and also Do not reply on Solr configuration. How do we know Lucene configuration and Who will manage this configuration ?

Q6. What is Sitecore role for Solr System ?

Is there any guideline or help from Sitecore ?

How do we know how many documents we handle per day or peak time ?

SolrCloud Questionaire

How to decide which spec and number of server that we need ?

    What will be 'CO' Topology ?

Does Solr can be shared between CD and CM servers ?

- Core index is on CM and CD
- Master index is on CM
- Web index is on CD
- Sitecore 10.1 has 12 indexes : Core, Master, Web, 2 fxm, 4 Marketing, personalization, Suggested, Testing

Who has responsible for Solrs and ZooKeepers topology – Build, Operation, Troubleshooting?

Is there a maintenance contract and support scope ?

    – software patched, backup/restore, tuning, major/minor upgrades