SOLRCLOUD: BEST PRACTICES FOR SITECORE. DESIGN, BUILD AND DEVOPS CONSIDERATIONS

Los Angeles User Group

Akshay Sura
Partner | Konabos Consulting

Sameer Maggon
Founder & Search Ninja | SearchStax

Los Angeles
User Group

sitecore
Own the experience™

AGENDA

1. Get you familiar with Apache Solr
2. Sitecore > Azure vs. Apache Solr
3. Overview & Comparison of Solr Topologies
4. Architecting a Production Grade Solr & Zookeeper
5. Backups and Recovery in Solr
6. Cross Region Solr deployment
7. Demo of Solr-as-a-Service making it happen in 30 mins
8. Q&A

ABOUT AKSHAY SURA & SAMEER MAGGON

Akshay Sura

- Six-time Sitecore MVP
- Founder of SUGCON North America 2015, SUGCON India 2018 & 2019, Unofficial Sitecore Training and the Sitecore Slack

Sameer Maggon

- University of Southern California Alum – Viterbi Engineering School (Masters in Computer Science)
- Built platforms and teams for Enterprise Search and Search-based Consumer companies.
- Spent 15+ years working with Open Source Search Engines (started with Lucene 1.1)
- Currently Work at SearchStax, Inc. – Founded the company in 2015 and primary focus is Product & Technology

Los Angeles
User Group

sitecore
Own the experience™

Sitecore and Azure Search works great until **you don't ask for much**

COMMON AZURE SEARCH LIMITATIONS RE. SITECORE

- High Frequency or Heavy Content Indexing Needs
- Performance Bottlenecks
- Large number of fields (Field Limitation)
- Multiple Languages
- Not Extensible – Lack of further customizability using custom or community plugins (Content Extraction - PDF, word, etc)
- As you scale up, cost increases drastically
- Not available in all regions – Does not meet Disaster Recovery Compliance needs
- Feature gap
- Proprietary API

Solr is the popular, blazing-fast, **open source** enterprise **search platform** built on Apache Lucene

ABOUT SOLR

- Trusted by 90%+ Fortune 500 Companies
- True Open Source with Strong Community
- Massive Scalability – used by Salesforce, Walmart, Iron Mountain, etc.
- Feature Rich
  - Faceted Search, Fuzzy Search, Spellcheck, Geospatial, Multi-language, and more.
  - Auto-Complete / Search Suggest
  - Content Extraction from Filetypes
  - Highly Customizable to specific needs
  - Customizable Relevance Models
  - Machine Learning Support

Los Angeles
User Group

sitecore
Own the experience™

Sitecore Practitioner's experience with Sitecore and Solr

Los Angeles
User Group

sitecore
Own the experience™

WORLD-CLASS SOLR / SITECORE IMPLEMENTATION CYCLE

Design

- Topology
- Security
- Log management
- Monitoring
- Reporting
- Alerting
- Backup
- DR strategy
- Capacity planning strategy

Build

- Hi-availability Solr
- Hi-availability ZK
- Security at rest
- Security in transit
- Log management
- Backup management
- Monitoring system
- Reporting system
- Alerting system
- DR (cold, hot)
- Multi-region
- Autoscaling

Integrate

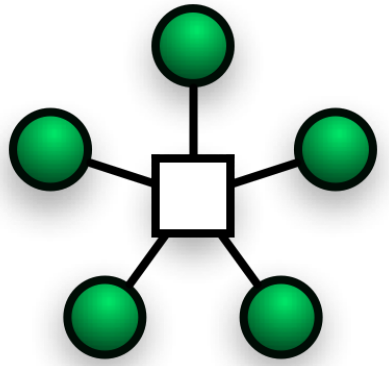
- **Integrate Sitecore with Solr**

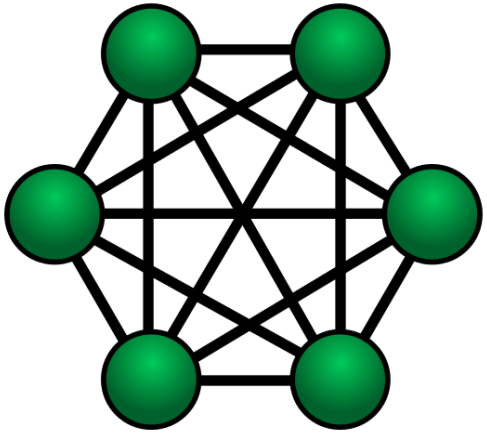
Operate

- Incident management
- OS patch management
- Solr security patches
- ZK security patches
- AWS node replacements
- Cloud node maintenance
- Periodic vulnerability scanning
- Solr error / log triage & analysis
- Solr performance optimization
- Solr minor/major version upgrades
- Scale-up
- Scale-down

SOLR TOPOLOGIES

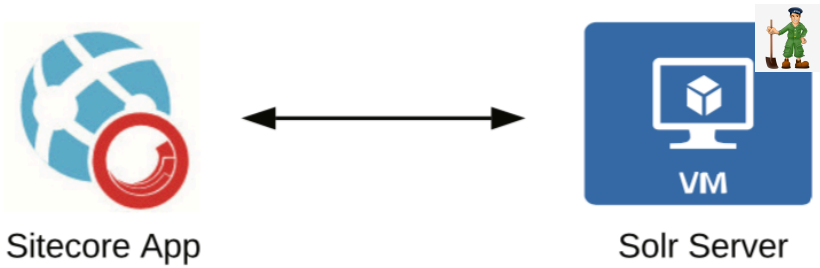
- Standalone/SolrCloud Server
- Master-Slave
- High Availability SolrCloud
- SolrCloud with CDCR (Cross Data Center Replication)



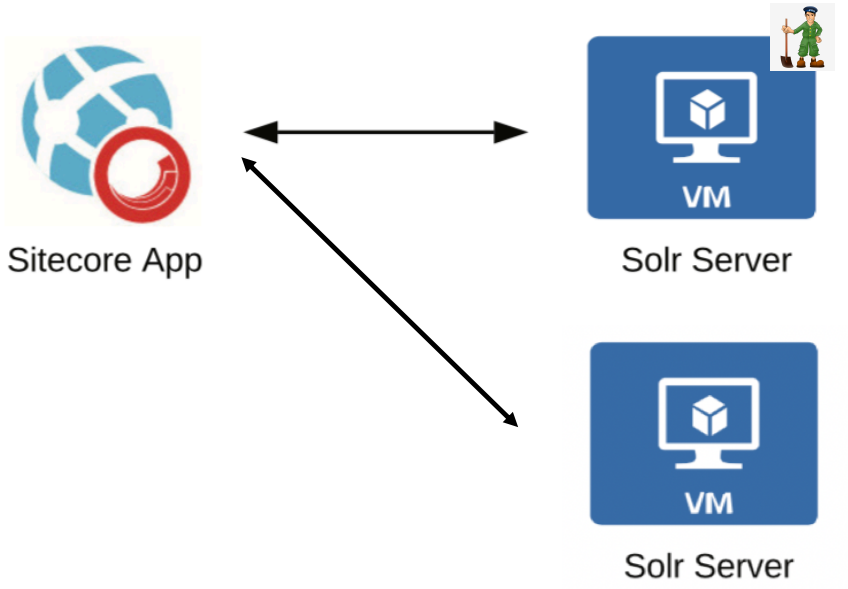


STANDALONE SERVER OR SOLRCLOUD

- Zookeeper and Solr run on same server
- Ideal for Non-Production (Development or QA Environments)
- Can be run in a “Cloud” mode or “Standalone” mode
- No High Availability or Fault Tolerance

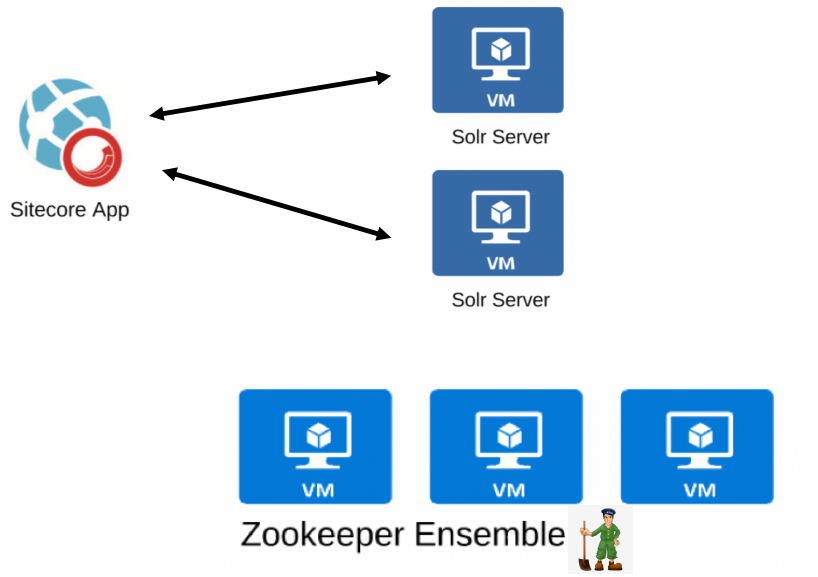


SOLRCLOUD JOURNEY TO HIGH AVAILABILITY



- Zookeeper is single point of failure

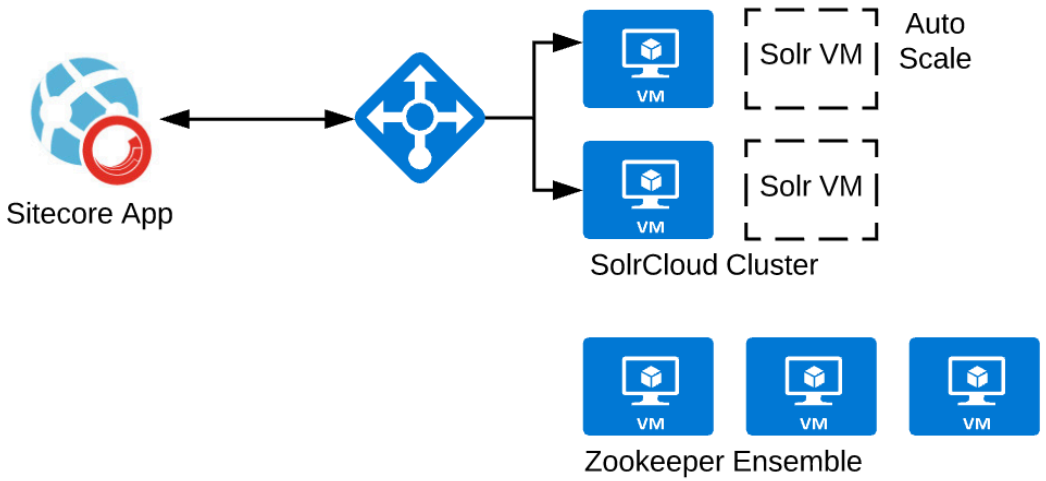
SOLRCLOUD JOURNEY TO HIGH AVAILABILITY



- In Cloud Environments, IPs can change and nodes will go down or get recycled.
- Don't want your application to directly be tied to the individual Solr nodes

SOLRCLOUD (HIGH-AVAILABILITY)

- High-Availability Zookeeper Ensemble
- High-Availability Solr Servers (with replication)
- Load Balancer / App Gateway
- Ability to Scale up without re-architecting



Los Angeles User Group | sitecore Own the experience

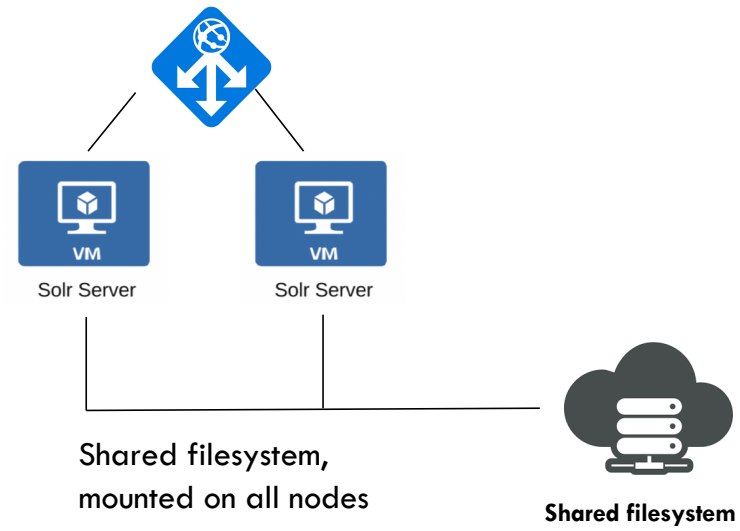
BACKUPS AND RECOVERY: SOLRCLOUD

SolrCloud Backups

Support for backups when running SolrCloud is provided with the Collections API. This allows the backups to be generated across multiple shards, and restored to the same number of shards and replicas as the original collection.

SolrCloud Backup/Restore requires a shared file system mounted at the same path on all nodes, or HDFS.

BACKUPS AND RECOVERY: SOLRCLOUD



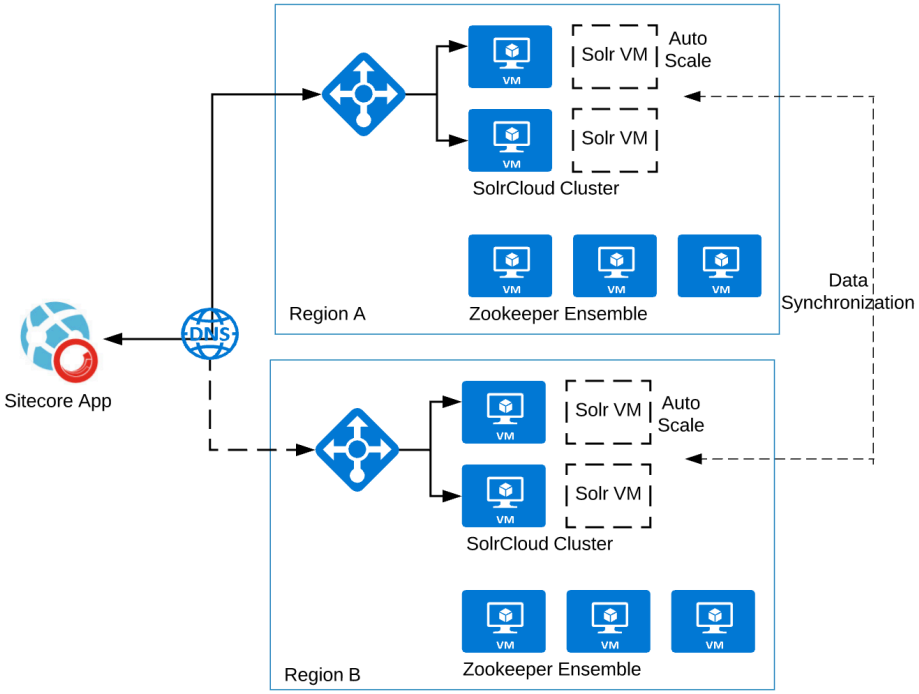
- SolrCloud Backups requires shared filesystem mounted on same path on all nodes
- Sitecore developers responsibility or implementation of:
  - backup schedules
  - retention logic and cleanup
  - backup/restore verification procedures

Los Angeles User Group

sitecore Own the experience

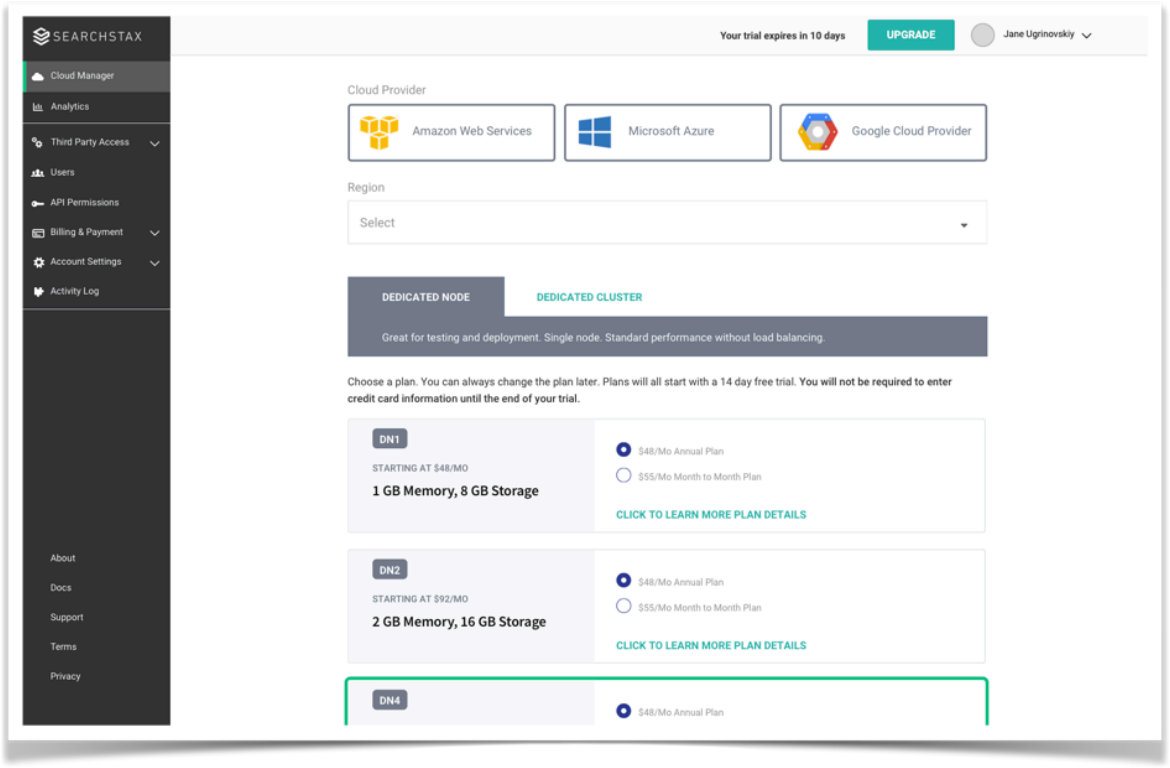
SOLR TOPOLOGIES: SOLRCLOUD WITH CDCR

- Highly Resilient (Entire Data Center can become unavailable without impacting availability)
- Localized Traffic Routing is possible
- Can be scaled to multiple regions



There is an easier way > **Solr-as-a-Service**

SEARCHSTAX DEMO



THANK YOU — Q&A

Akshay Sura
as@kanabos.com
tw: @akshaysura13

Sameer Maggon
sameer@searchstax.com
tw: @maggon