



# Update on CERN Search: SharePoint 2013

E. Alvarez, S. Fernandez, A. Lossent, I. Posada, A. Wagner [CERN]

CERN's enterprise Search solution 'CERN Search' provides a central search solution for both users and CERN service providers. Public and protected documents from a wide range of documents are indexed and available for retrieval.

High Availability

Improved design

High Reliability

2000 daily queries

33 million documents

1100 daily unique visitors

29 servers

12000 daily page views

3 server farms



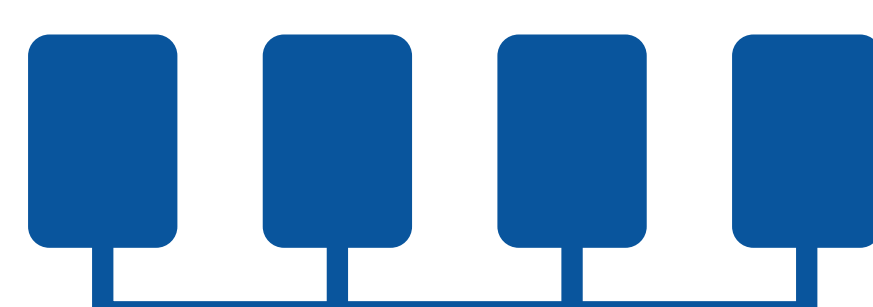
CDS



SharePoint 2013

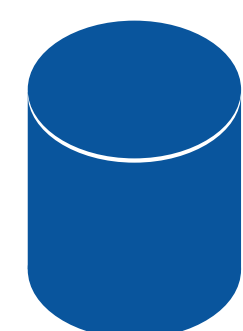


Content Sources



Push web services

Pre-processing



Crawler

- Push model.
- Access control for each source.

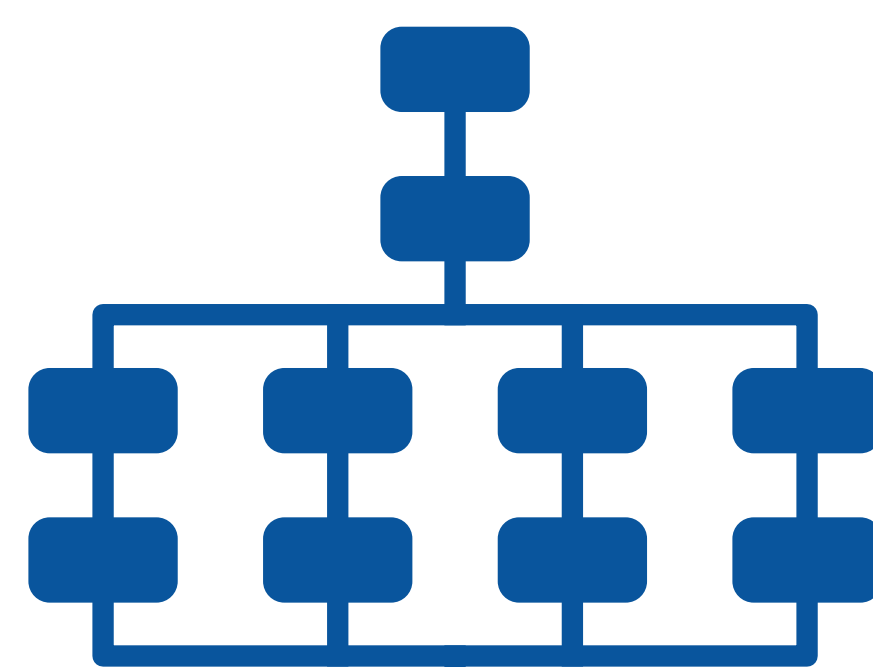
- Pull model.
- Dynamic URI list.

- Stores data in protected space.
- Extracts metadata from binary files.



- Sets ACL for each document.

Indexing



- Analyze and split documents in several stages.
- Custom processing.
- Different processing by content source.
- Dynamic property mapper.
- Property mapping by configuration.
- Auto load-balanced CEWS call.

- Microsoft SharePoint 2013 search engine.

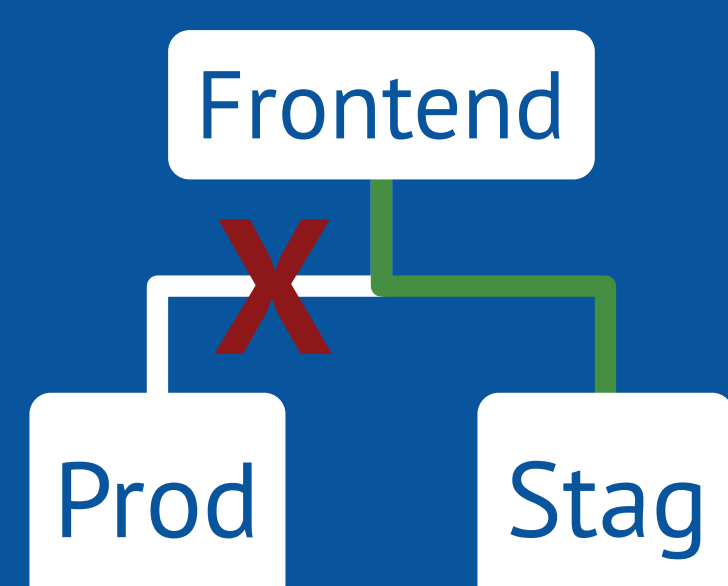


- Redundancy of key components.

A: Admin component.  
C: Crawl component.  
P: Processing component.

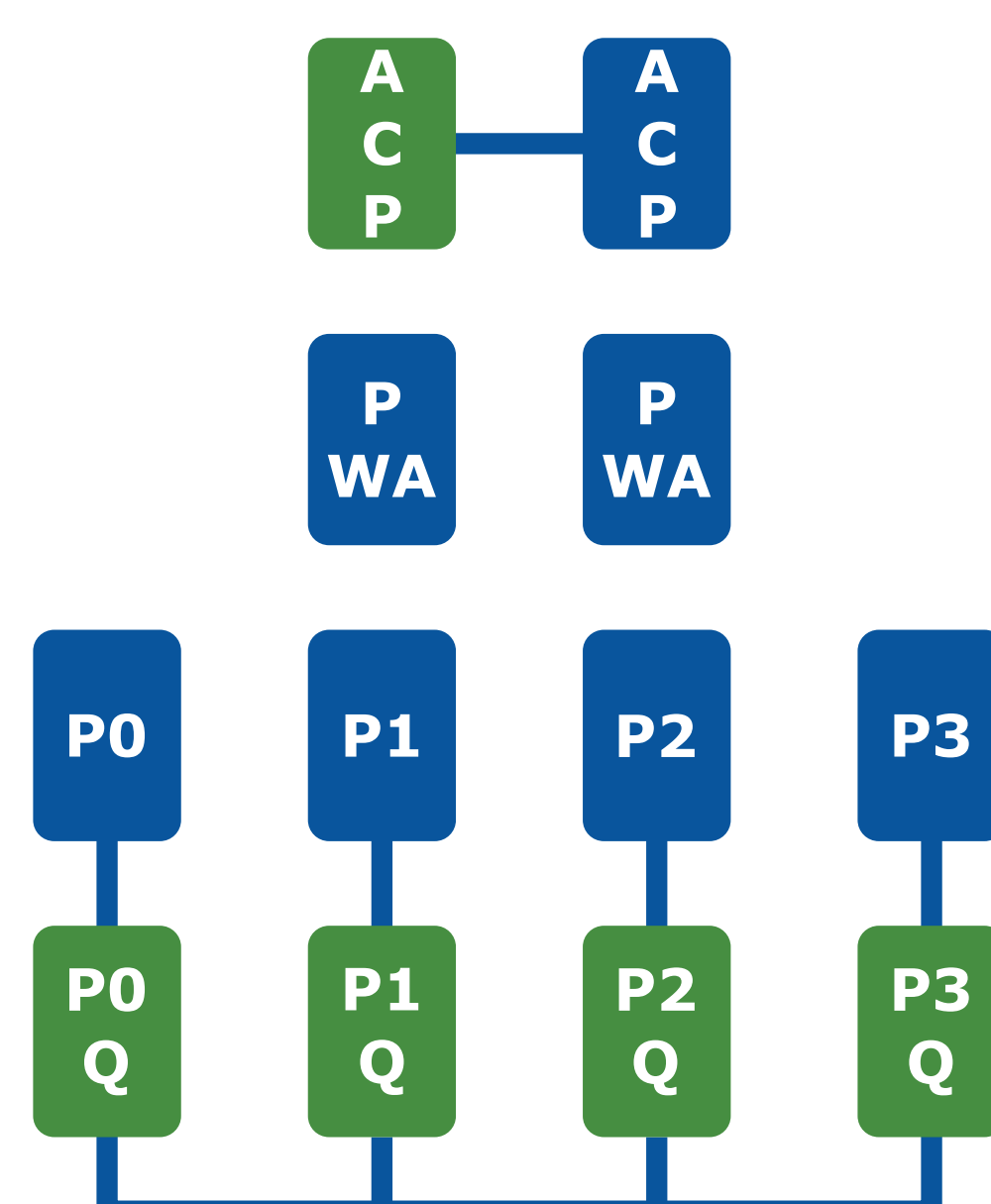
WA: Web Analyzer.  
Px: Index Partition x.  
Q: Query component.

- Staging farm as high availability failover.



- <5 seconds switch.

Search Engine

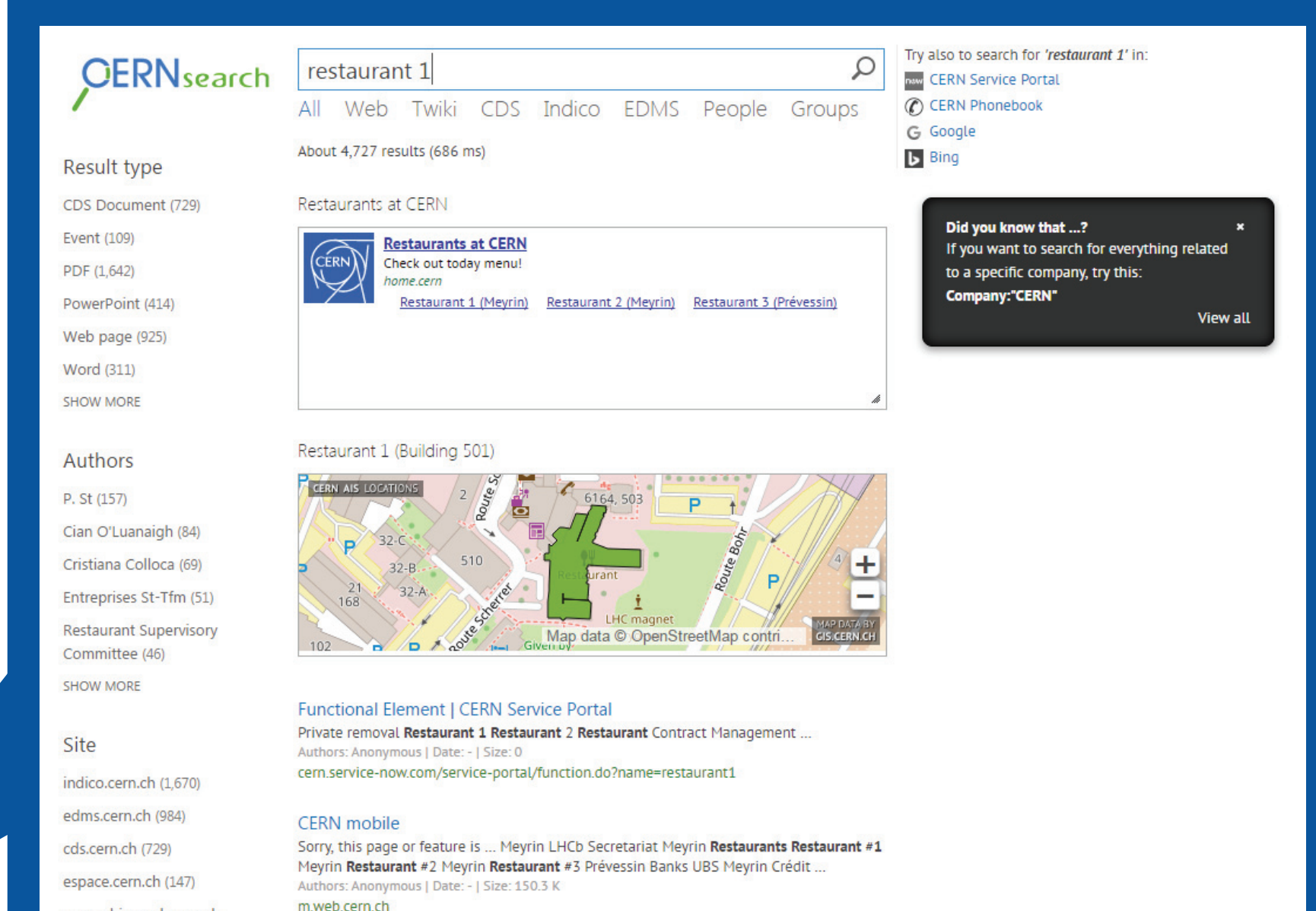


Searching

Frontend

REST

- User interface upgraded.
- Improved user experience.
- Thesaurus for acronyms.
- Promoted CERN most relevant sites.
- Map integration for locations.



- Improved integration as search backend.

