

# November 23<sup>rd</sup>, 2013

## Search Topology and Optimization



Mike Maadarani  
SharePoint Architect  
[mike@maadarani.com](mailto:mike@maadarani.com)

SHAREPOINT **SATURDAY**  
■■■■■ ■■■■  
OTTAWA

# Thank you to all of our Sponsors!!



## ➤ Mike Maadarani

- App Dev and Architecture for over 18 years (15 Years Microsoft, 3 Years with the “Other Guys”)
- Business focused on Enterprise Content Management & Publishing Sites
- Technology focused on SharePoint, SQL Server and SharePoint Integration
- Architect, trainer, and presenter
- Blog: [www.maadarani.com](http://www.maadarani.com)
- [mike@maadarani.com](mailto:mike@maadarani.com); @mikemaadarani

# Agenda

SharePoint 2013  
Search Overview

Architecture and Resource  
Utilization

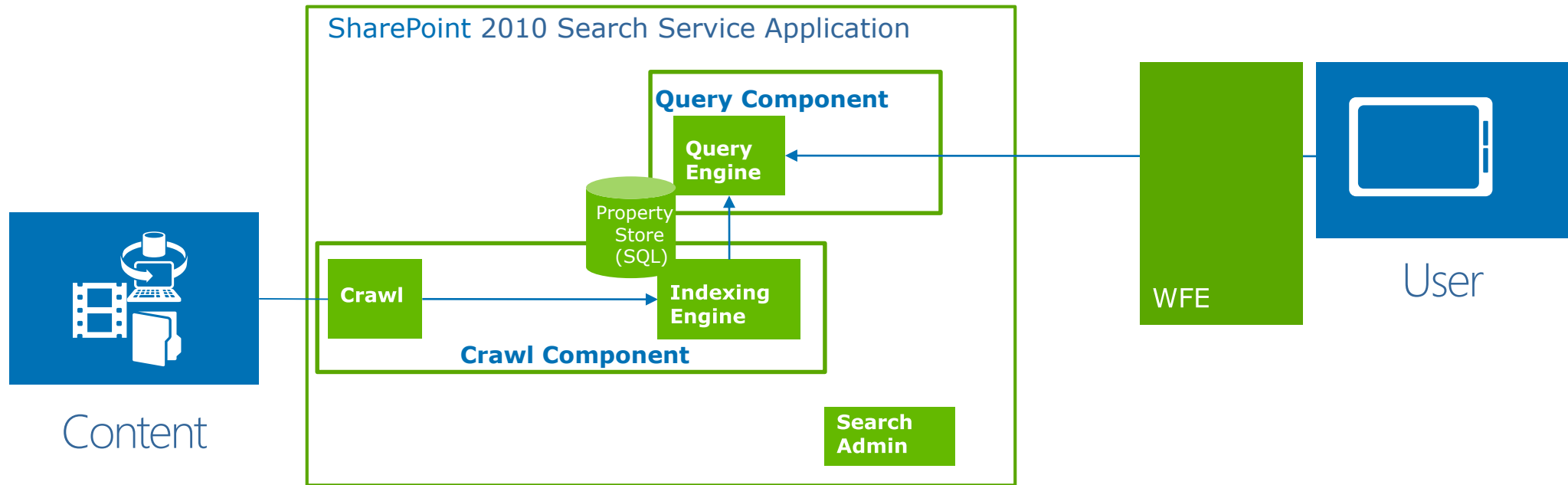
Configuring SSA  
and PS

Topology  
Scenarios

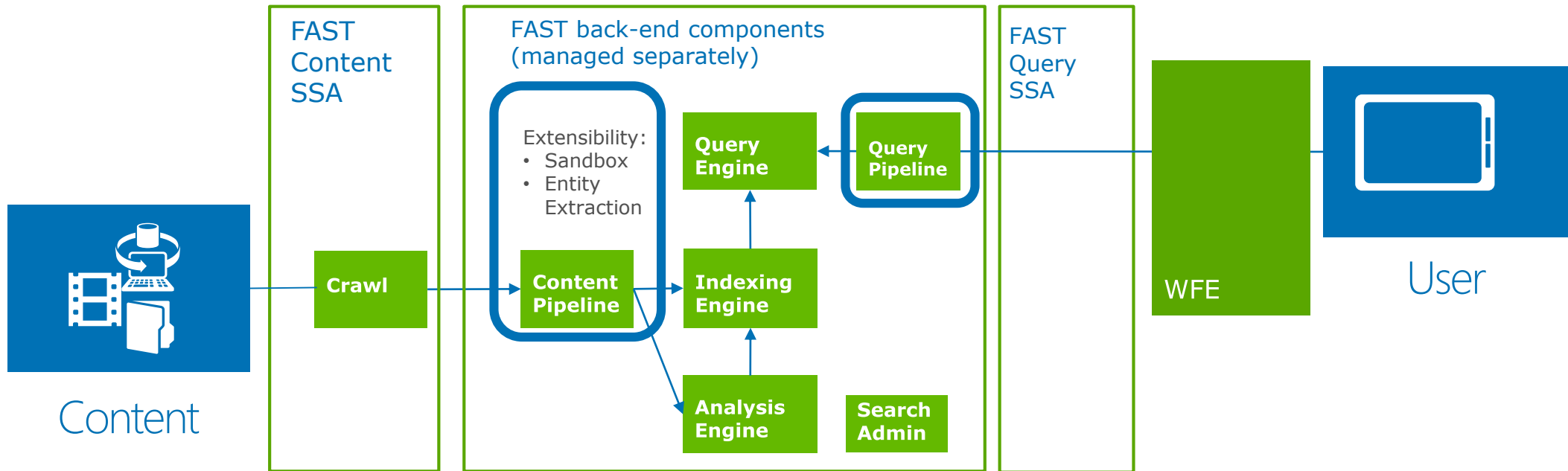
Relevancy, Query Builder, &  
Optimization

Closing and Q&A

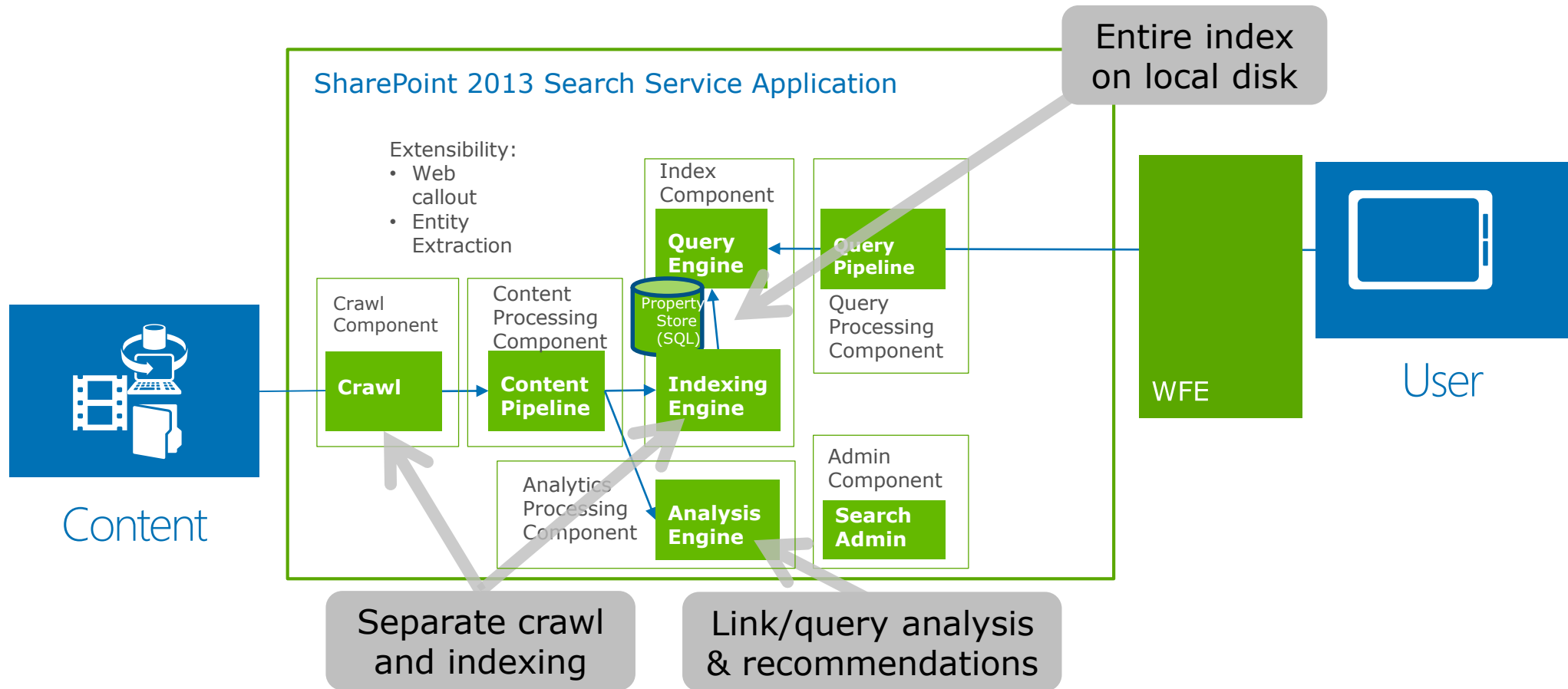
# Search in 2010



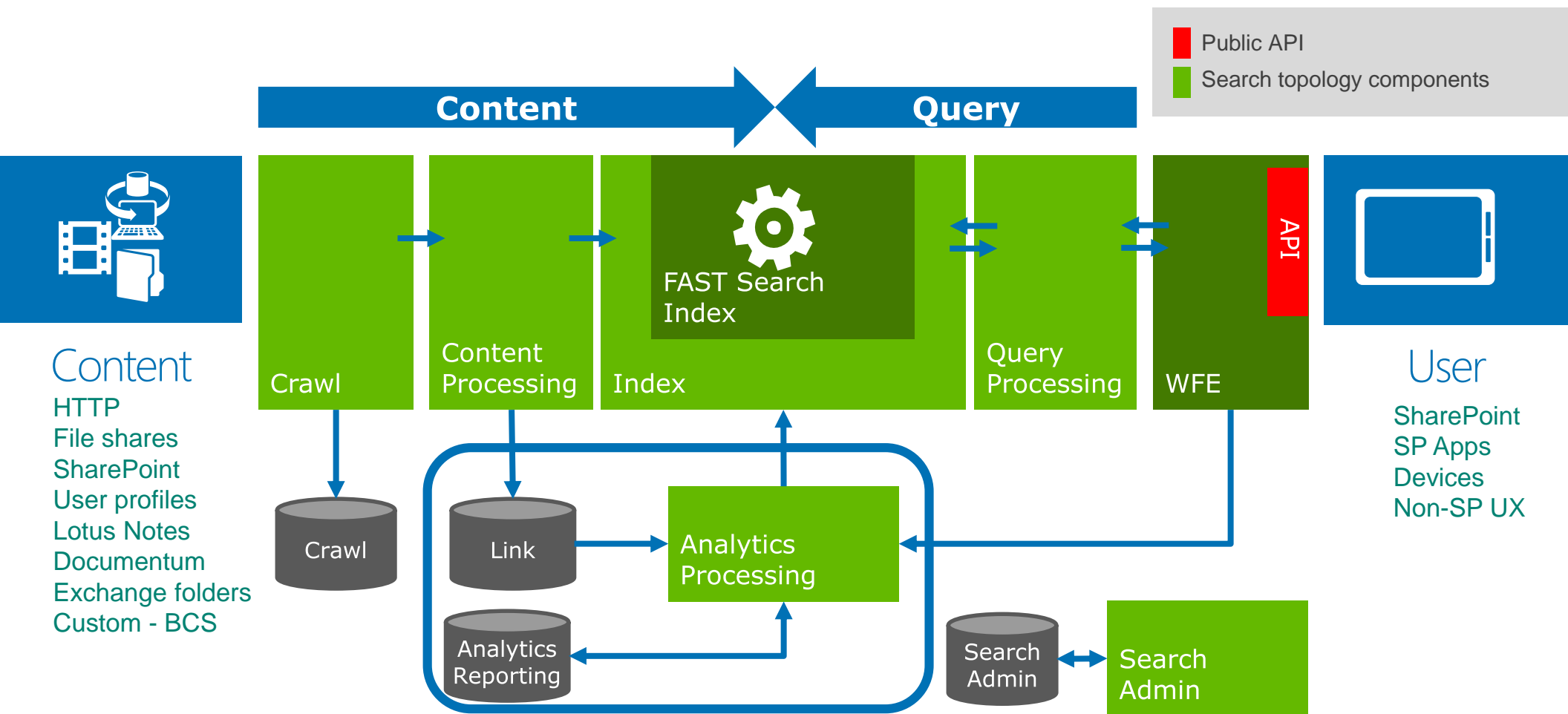
# FAST Search for SharePoint 2010



# ... In SharePoint 2013



# SharePoint 2013 Search Architecture





# Why Search is so important?

## FAST



I just uploaded a  
document.  
Make it searchable, quick!

# Why Search is so important?

## EASY



Everything People Conversations Videos

You've seen this result before

[Contoso Electronics](#)

[Contoso Electronics](#)

? Sales Shape Color Words Symbols Other 48 26 13 7 6 ... It's about the person, not the packaging ... Market BreakdownChart Title Games Video Audio Mobile 4.8 3.2 1.4 ...  
[pmg095.sharepoint.com/.../Contoso 2010 Market Segment Per...](#)



[Contoso Marketing Campaign Overview](#)

[pmg095.sharepoint.com/.../Contoso Marketing Campaign Over&#8230;](#)



[Home](#)

Together we strive for excellence every day ... Our goal is to provide great products and services to our customers ... Building long term relationships by listening to our customers and ...  
[pmg095.sharepoint.com/sites/contoso](#)



[“ Non-English Market ”](#)

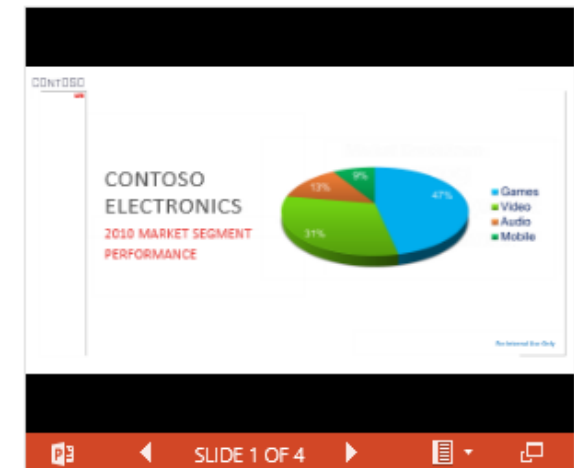
Posted by Alex Darrow October 30

What presentations, documentation or guidance do we have in terms of selling in non-English-speaking markets?  
[pmg095.sharepoint.com/.../Non-English Marke...](#)

4 replies  
2 likes  
✓ best reply

Contoso Electronics

25 views



Take a look inside

[Contoso Electronics](#)

[Design considerations](#)

[Fit the person—not the Box](#)

[Analysis – market breakdown](#)

Changed by Sara Davis on 7/31/2012 6:05 AM

[EDIT](#) [FOLLOW](#) [SEND](#) [VIEW LIBRARY](#)

Result type

Discussion

Excel

Newsfeed post

PowerPoint

Web page

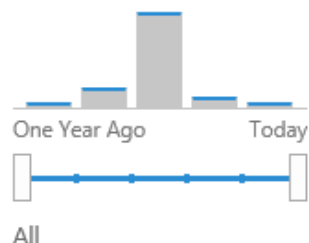
Word

SHOW MORE

Author



Modified date



# Why Search is so important?

## EASY



### Department

Finance

Sales & Marketing

### Job Title

Administrative Assistant

Product Manager

### Keywords

Cooking

Budget Creation and Trac...

Budgeting Process & Wo...

Competitive Swimming

Creative Thinking

SHOW MORE

### Office Location

London

pa



**P**avel Banský

Anne **P**allace

Dorena **P**aschke

Fabrice **P**anel

Janet **P**aschorr

Legal, Sales and Marketing, and now in Finance. I'm ve...



**Molly Dempsey**

Product Manager

Sales & Marketing

Ask me about: **Project Falcon designs** | Product **Design** Roundtable...

I have been with Contoso for five years, the last three as a Product Manager. I like to develop ideas for creating and market...

### People who've written about "project falcon designs"



**Alex Darrow**

Marketing Assistant

Sales & Marketing

I have been at Contoso for five years, the last three as a Marketing Assistant. I like to help marketing teams plan and execu...



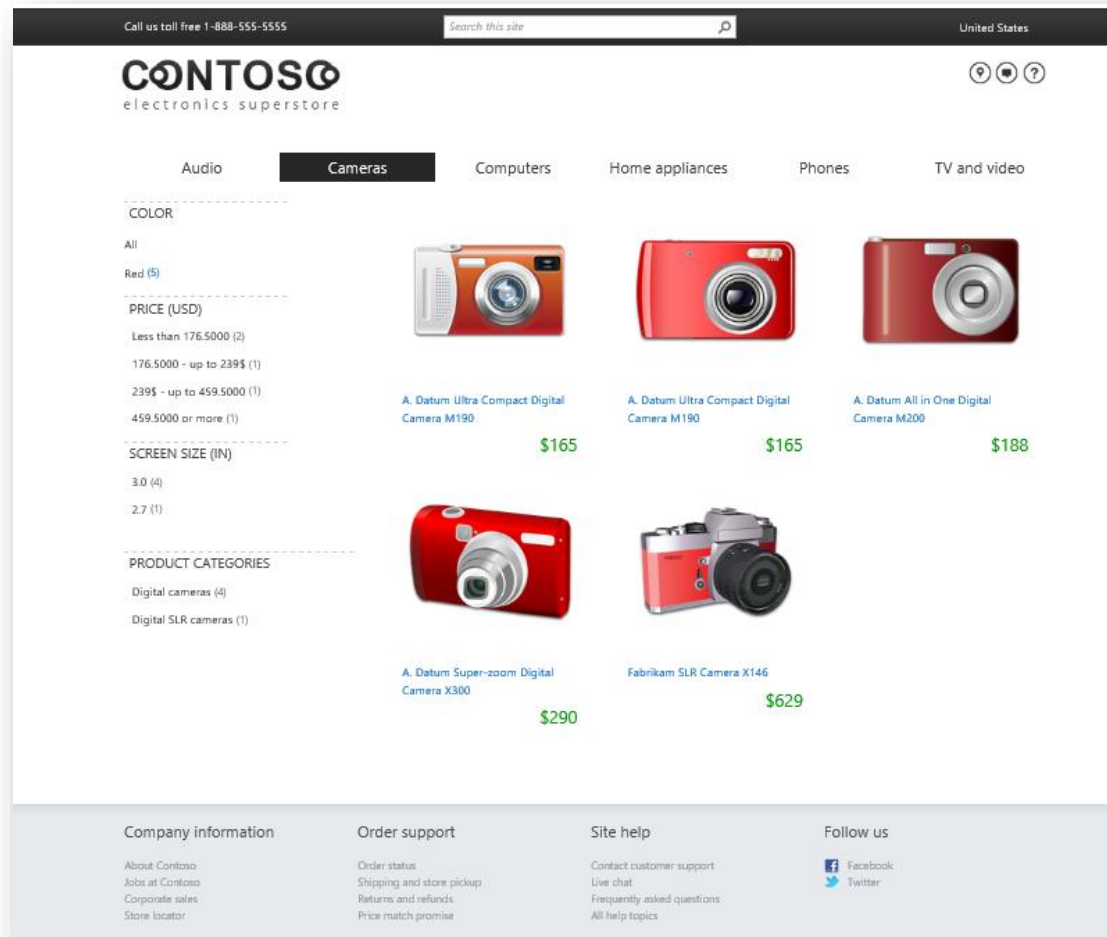
**Sara Davis**

Product Manager

Sales & Marketing

# Why Search is so important?

## Search Driven Applications



# Why Search is so important?

## Search Everything



I can find ALL of Rob  
Ford's hidden videos!

# Where does Search live in the farm?

## Windows services

SharePoint Search Host Controller service

Runtime/lifecycle control of search components (except crawler)

→ hostcontrollerservice.exe

SharePoint Server Search service

Crawl Component

→ mssearch.exe

→ mssdmn.exe

Still there, but only  
Crawl Component

## Processes

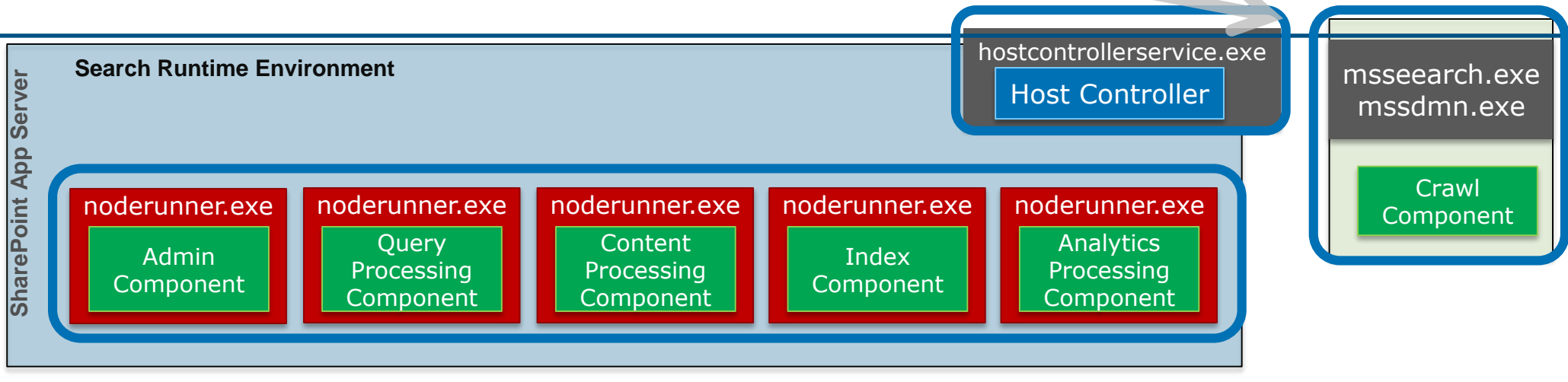
Noderunner.exe

Runtime environment for search components (except crawler)

## Admin entities

Search Service Instance: Provisioning of the search service on each box

Search Service Application: SharePoint Configuration entity



**Where do I host my  
components?**

# Query processing component (QPC)

## CPU load

### Driving factors

QPS

Query transformations

## Network load

### Driving factors

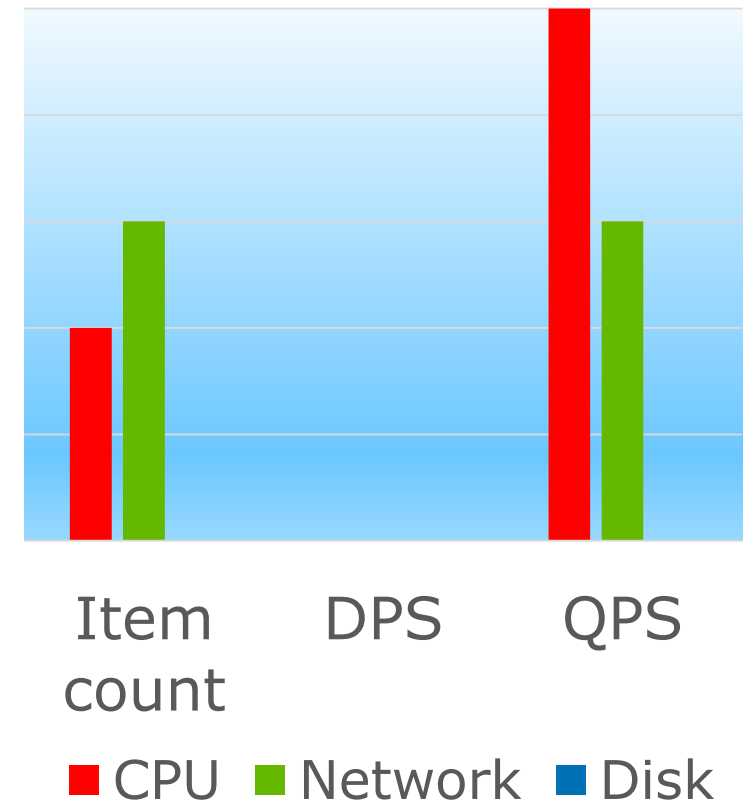
Number of index partitions

Size of queries and results

### Example:

20 index partitions @ 20 qps => 200/100 Mbit/s  
in/outbound

Load impact (relative)





# Index component

## CPU load

### Driving factors

QPS and item count

### Guidelines per index component @ 2 GHz CPU

1M items: 5 QPS per CPU core

5M items: 2 QPS per CPU core

10M items: 1 QPS per CPU core

## Disk load

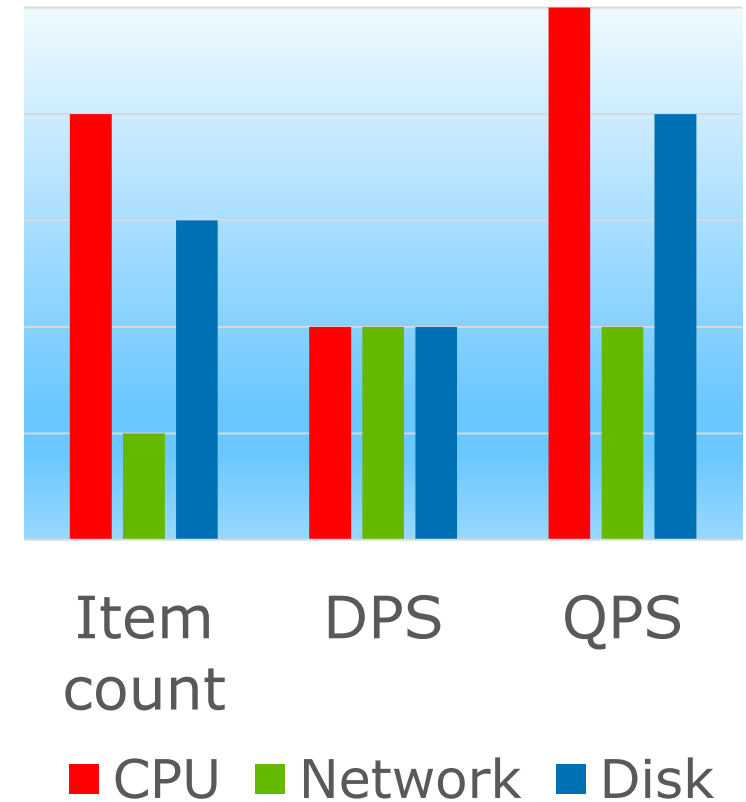
### Driving factors

QPS and item count

New content invalidates caches

Disk size: 500GB @ 10M items per index component

Load impact (relative)



# Crawl component

## CPU load

### Driving factors

- Documents per second

- Link discovery

- Crawl management

## Network load

### Driving factors

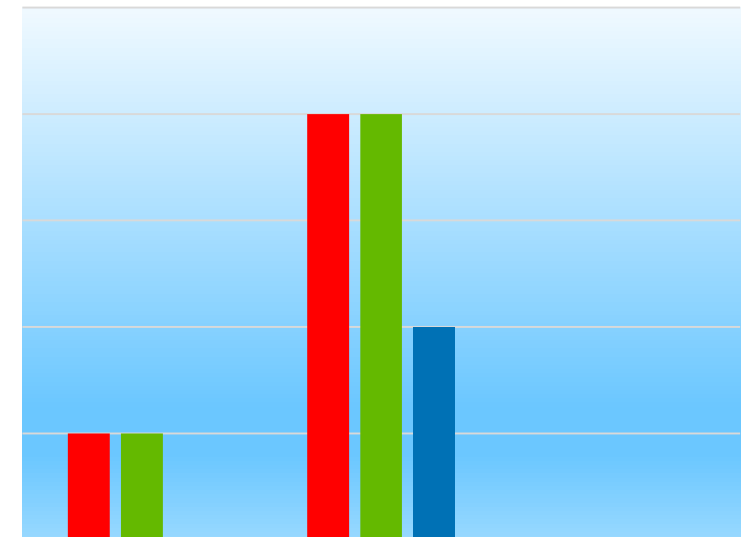
- Downloading items from content sources

- Passing items on to CPC

## Disk load

All documents are temporarily stored in data folder

Load impact (relative)



Item  
count

DPS

QPS

■ CPU ■ Network ■ Disk

# Content processing component (CPC)

## CPU load

### Driving factors

- Documents per second

- Document size and complexity

- Feature extraction

Estimate: 5-10 DPS per CPU core

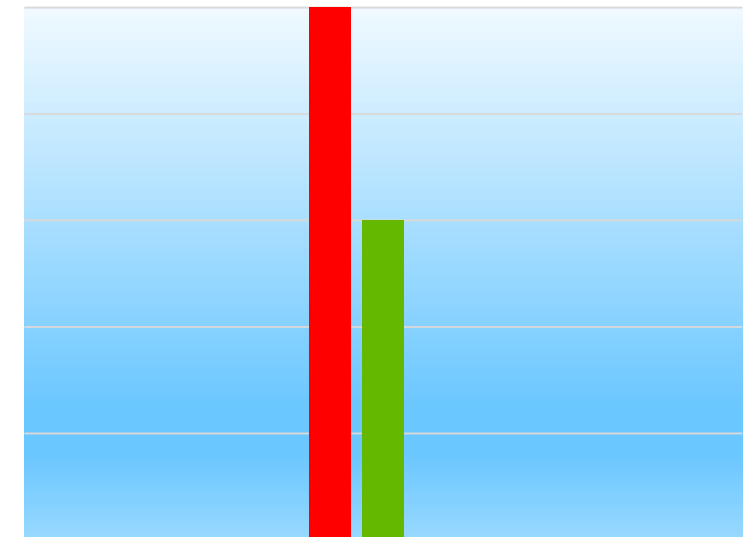
## Network load

### Driving factors

- Documents per second

- Document size

Load impact (relative)



Item  
count

DPS

QPS

■ CPU ■ Network ■ Disk

# Analytics processing component (APC)

## CPU load

### Driving factors

- Number of items

- Site activity

## Disk load

Local disk used for temporary storage

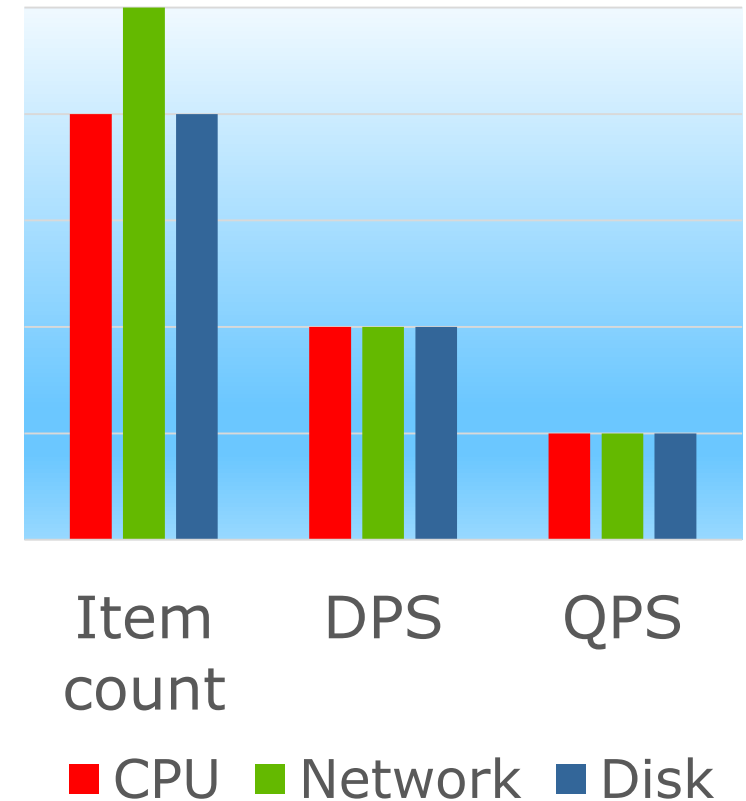
Bulk load, primacy concern is load isolation

## Network load

Same as for CPU load

PLUS: Network traffic increases when distributing APC across multiple machines

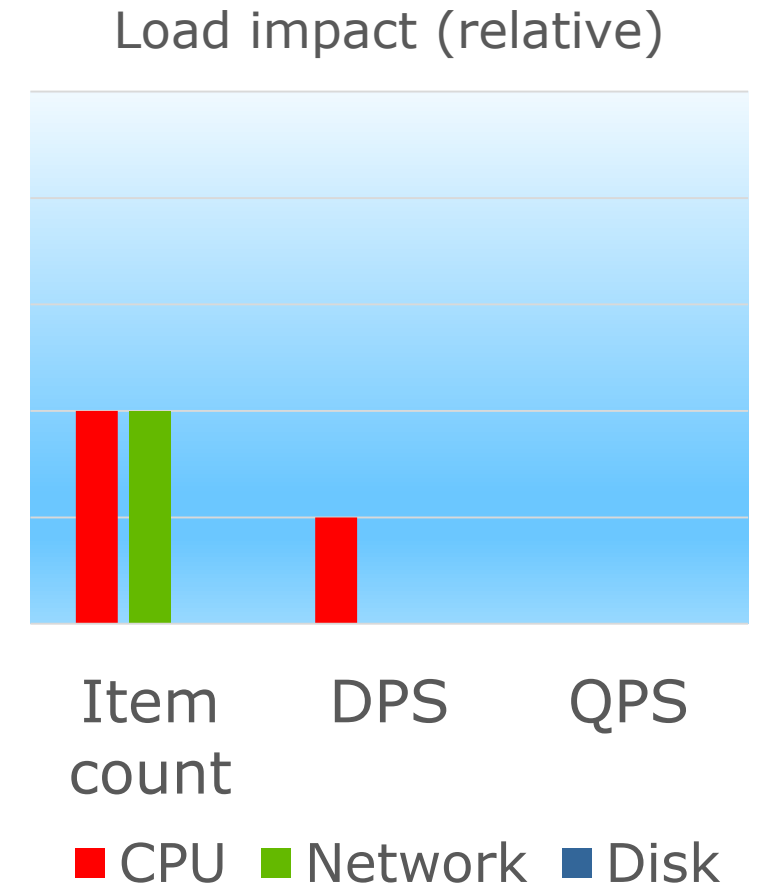
Load impact (relative)



# Search administration component

Low CPU and network load

Load increase with more components in the search topology



# Create your SSA

```
$SSADB = "SharePoint_Demo_Search"
```

```
$SSAName = "Search Service Application Ottawa"
```

```
$SVCAcct = "mcm\sp_search"
```

```
$SSI = get-spenterprisesearchserviceinstance -local
```

## #1. Start the search services for SSI

```
Start-SPEnterpriseSearchServiceInstance -Identity $SSI
```

## #2. Create the Application Pool

```
$AppPool = new-SPServiceApplicationPool -name $SSAName"-AppPool" -account $SVCAcct
```

## #3. Create the search application and set it to a variable

```
$SearchApp = New-SPEnterpriseSearchServiceApplication -Name $SSAName -applicationpool $AppPool -databaseserver SQL2012 -databasename $SSADB
```

## #4. Create search service application proxy

```
$SSAProxy = new-SPEnterpriseSearchServiceApplicationProxy -name $SSAName" Application Proxy" -Uri $SearchApp.Uri.AbsoluteURI
```

## #5. Provision Search Admin Component

```
Set-SPEnterpriseSearchAdministrationComponent -searchapplication $SearchApp -searchserviceinstance $SSI
```

## #6. Create the topology

```
$Topology = New-SPEnterpriseSearchTopology -SearchApplication $SearchApp
```

## #7. Assign server(s) to the topology

```
$hostApp1 = Get-SPEnterpriseSearchServiceInstance -Identity "SPWFE"
```

```
New-SPEnterpriseSearchAdminComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchCrawlComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchContentProcessingComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchAnalyticsProcessingComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchQueryProcessingComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchIndexComponent -SearchTopology $Topology -SearchServiceInstance $hostApp1 -IndexPartition 0
```

## #8. Create the topology

```
$Topology | Set-SPEnterpriseSearchTopology
```

# Small Search Topology

## Search Application Topology

Server Name	Admin	Crawler	Content Processing	Analytics Processing	Query Processing	Index Partition
						0

SPWFE



Database Server Name

Database Type

Database Name

SQL2012

Administration Database

SharePoint\_Demo\_Search

SQL2012

Analytics Reporting Database

SharePoint\_Demo\_Search\_AnalyticsReportingStore

SQL2012

Crawl Database

SharePoint\_Demo\_Search\_CrawlStore

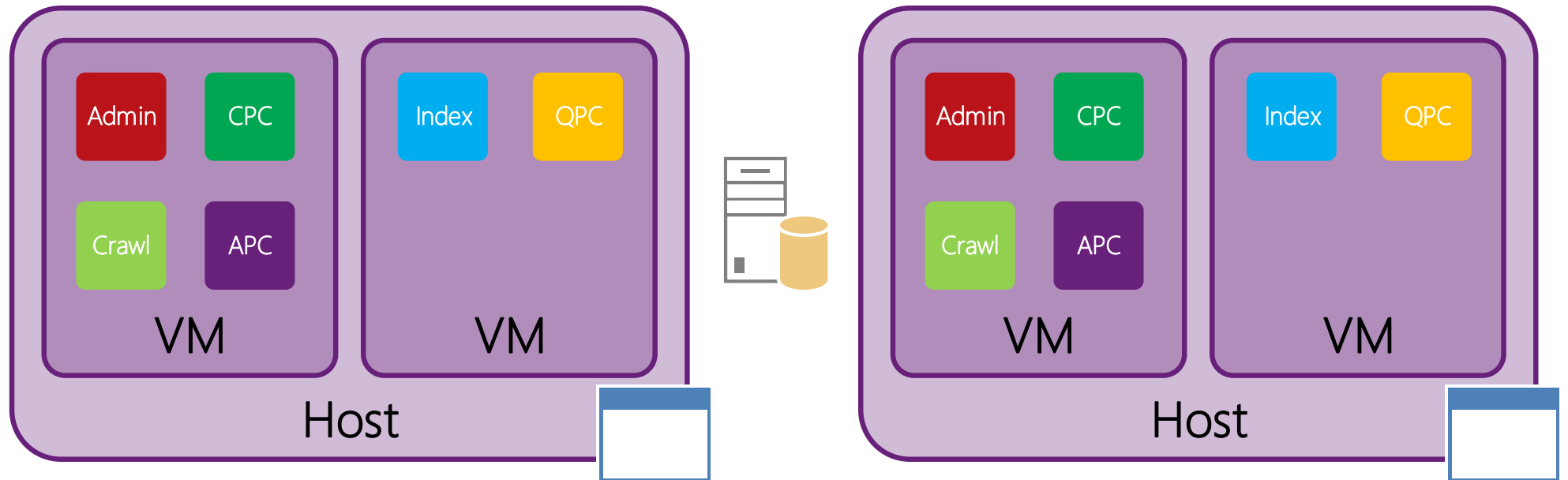
SQL2012

Link Database

SharePoint\_Demo\_Search\_LinksStore

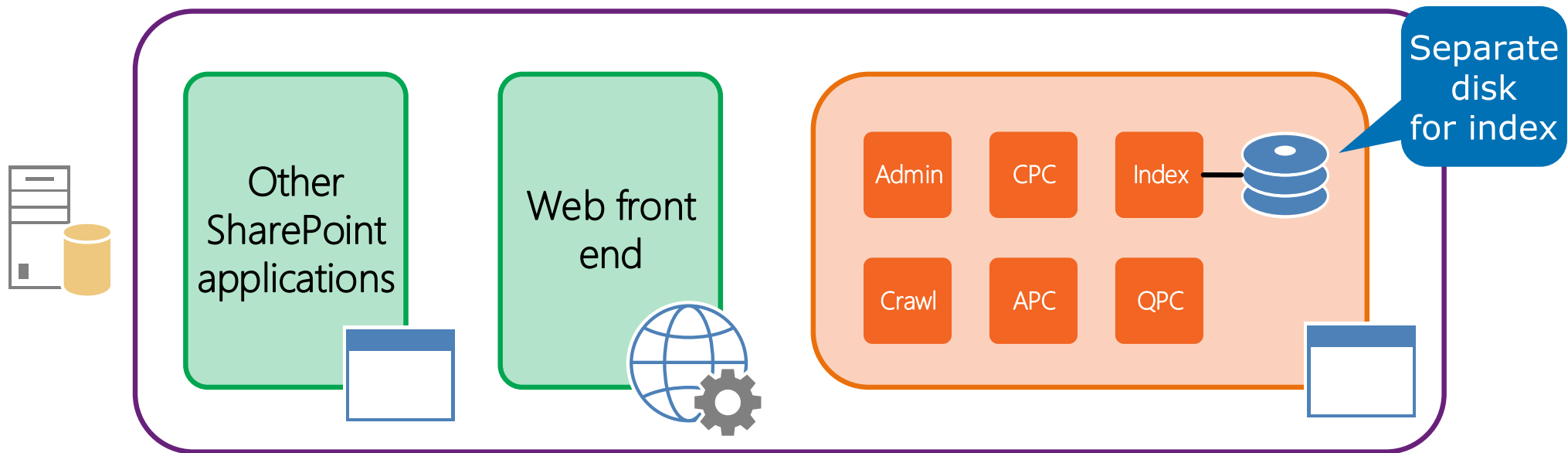


# Fault tolerant small search topology





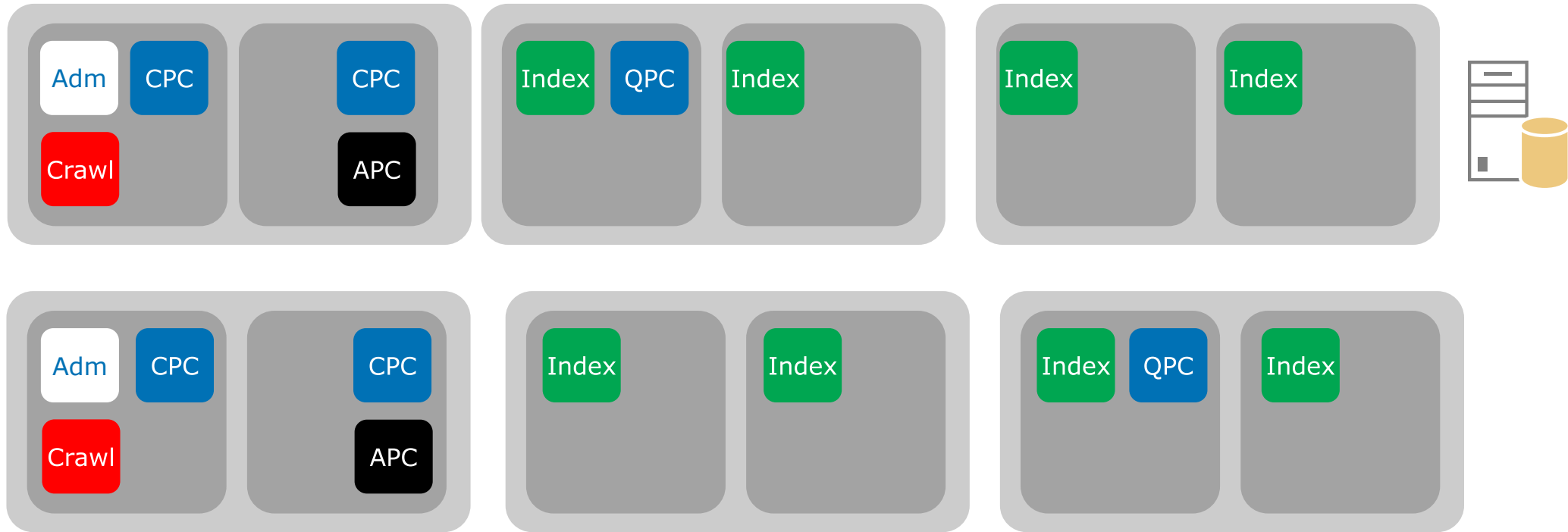
# Small search farm (up to 10M items)



Sized independently

Resources @ 10M items  
8x CPU cores  
24 GB RAM  
800 GB disk

# Scaling from small to medium search topology



# Extend your SSA

## #2. Extend the Search Topology:

```
$hostApp1 = Get-SPEnterpriseSearchServiceInstance -Identity  
"SPWFE"
```

```
$hostApp2 = Get-SPEnterpriseSearchServiceInstance -Identity  
"SPSearch"
```

```
Start-SPEnterpriseSearchServiceInstance -Identity $hostApp1
```

```
Start-SPEnterpriseSearchServiceInstance -Identity $hostApp2
```

## #3. Keep running this command until the Status is Online:

```
Get-SPEnterpriseSearchServiceInstance -Identity $hostApp1  
Get-SPEnterpriseSearchServiceInstance -Identity $hostApp2
```

## #4. Once the status is online, you can proceed with the following commands:

```
$ssa = Get-SPEnterpriseSearchServiceApplication  
$active = Get-SPEnterpriseSearchTopology -SearchApplication  
$ssa -Active
```

```
$newTopology = New-SPEnterpriseSearchTopology -  
SearchApplication $ssa
```

```
New-SPEnterpriseSearchAdminComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchCrawlComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp1
```

```
New-SPEnterpriseSearchContentProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp1
```

```
New-SPEnterpriseSearchAnalyticsProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp1
```

```
New-SPEnterpriseSearchQueryProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp1
```

```
New-SPEnterpriseSearchIndexComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp1 -IndexPartition  
0
```

```
New-SPEnterpriseSearchAdminComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp2
```

```
New-SPEnterpriseSearchCrawlComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp2
```

```
New-SPEnterpriseSearchContentProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp2
```

```
New-SPEnterpriseSearchAnalyticsProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp2
```

```
New-SPEnterpriseSearchQueryProcessingComponent -  
SearchTopology $newTopology -SearchServiceInstance  
$hostApp2
```

```
New-SPEnterpriseSearchIndexComponent -SearchTopology  
$newTopology -SearchServiceInstance $hostApp2 -IndexPartition  
1
```

## #5. Activate the topology:

```
Set-SPEnterpriseSearchTopology -Identity $newTopology
```

# Medium Search Topology

## Search Application Topology

Server Name	Admin	Crawler	Content Processing	Analytics Processing	Query Processing	Index Partition	
						0	1
SPSEARCH	✓	✓	✓	✓	✓		✓
SPWFE	✓	✓	✓	✓	✓	✓	

Database Server Name	Database Type	Database Name
SQL2012	Administration Database	SharePoint_Demo_Search
SQL2012	Analytics Reporting Database	SharePoint_Demo_Search_AnalyticsReportingStore
SQL2012	Crawl Database	SharePoint_Demo_Search_CrawlStore
SQL2012	Link Database	SharePoint_Demo_Search_LinksStore



# **Tweaking Your results**

# Challenges: Intent



Infrastructure  
Project

Where is my talk  
Project Plan?

Are Documents held at  
the same place?

I wonder if there are  
references from  
previous projects?

There is rarely a single right  
answer

Different people have different  
intent

Query Rules help you handle  
intent

# Configuration in the Conceptual Relevance Flow

## Query:

*HR Employment  
quarterly  
report*

(WORDS *HR*, Human Resources) AND  
(WORDS *employees*, employed) AND  
(WORDS *quarterly*, quarterlies) AND  
(WORDS *report*, reports, reported)

Mixed Results for:

- HR Employment best bet
- HR Employment quarterly report
- HR Employment  
ContentType=reports

Search  
Web Part

Query Processing

Engine

Thesaurus: HR → Human Resources

Best bets: HR Employment → /HR/employment

Dynamic Reordering Rules:

Quarterly Report →  
{prefer docs from http://reports}

Query Rule:

{Terms} Quarterly Report →  
{Terms} ContentType="reports"

Document  
Collection

For all queries:

Authorities: Level 1: http://employment

Ranking model: {incorporate user ratings}

# Authorities: SSA-level configuration



## Search Service Application: Specify Authoritative Web Pages

Use this page to specify authoritative Web pages. Search uses these lists to enhance the overall ranking of results.

Central Administration

Farm Search Administration

Search Administration

Diagnostics  
Crawl Log  
Crawl Health Reports  
Query Health Reports  
Usage Reports

Crawling  
Content Sources  
Crawl Rules  
Server Name Mappings  
File Types  
Index Reset

\* Indicates a required field

### Authoritative Web Pages

Authoritative pages are those that link to the most relevant information. Search uses this list to calculate the rank of every page in the index. There might be many authoritative Web pages in your environment. Type the most valuable in the topmost text box. You also have the option of specifying second and third-level authorities in the lower text boxes.

Enter one full URL per line. Example: `http://intranet/site-directory.html`.

Sites that are important

Most authoritative pages: \*

`http://jib151/`

Second-level authoritative pages:

### Non-authoritative Sites

You can specify that all content from certain sites be ranked lower than that of all other sites. Type the URLs of those sites here, one per line.

URLs typed in this section are treated as prefix matches. Example: entering `http://archive/` will demote the rank of all URLs that begin with `http://archive/`.

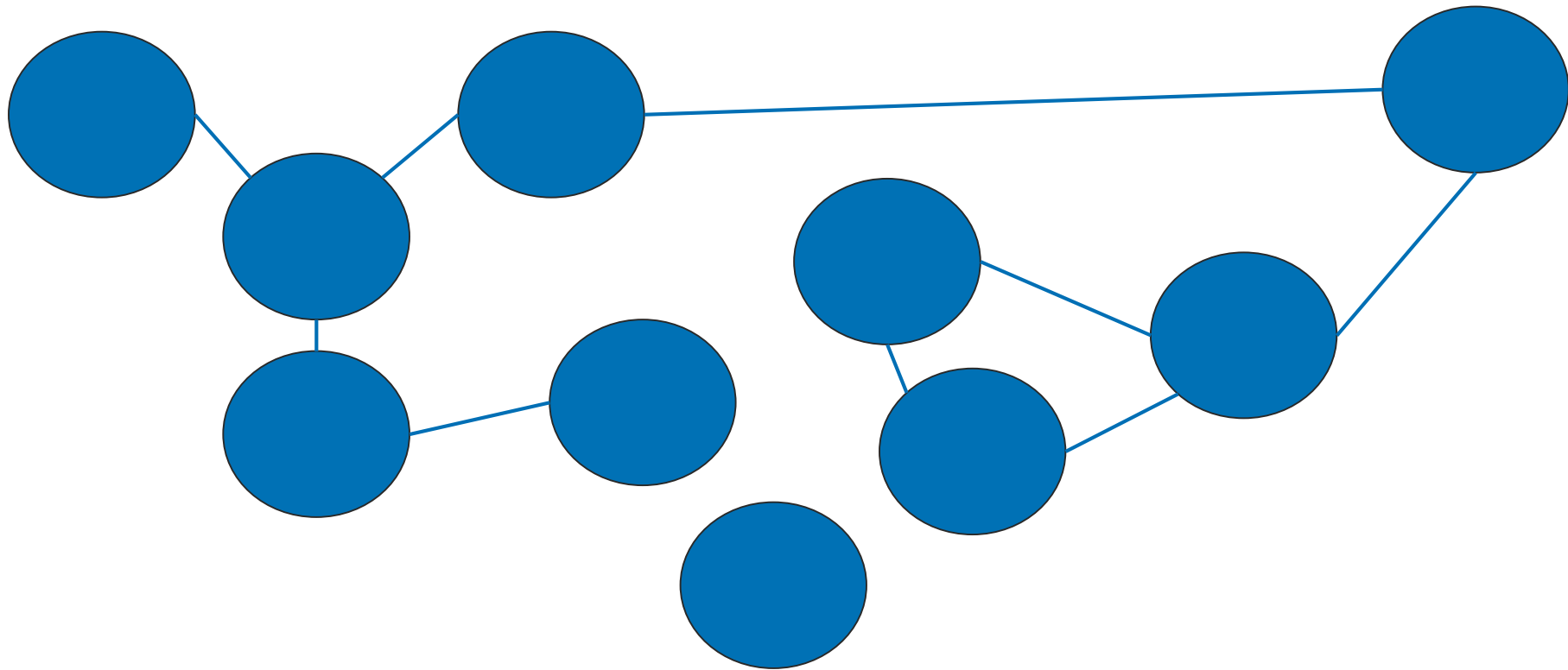
Sites with low intrinsic relevance

Sites to demote:

Takes ~24hrs to propagate

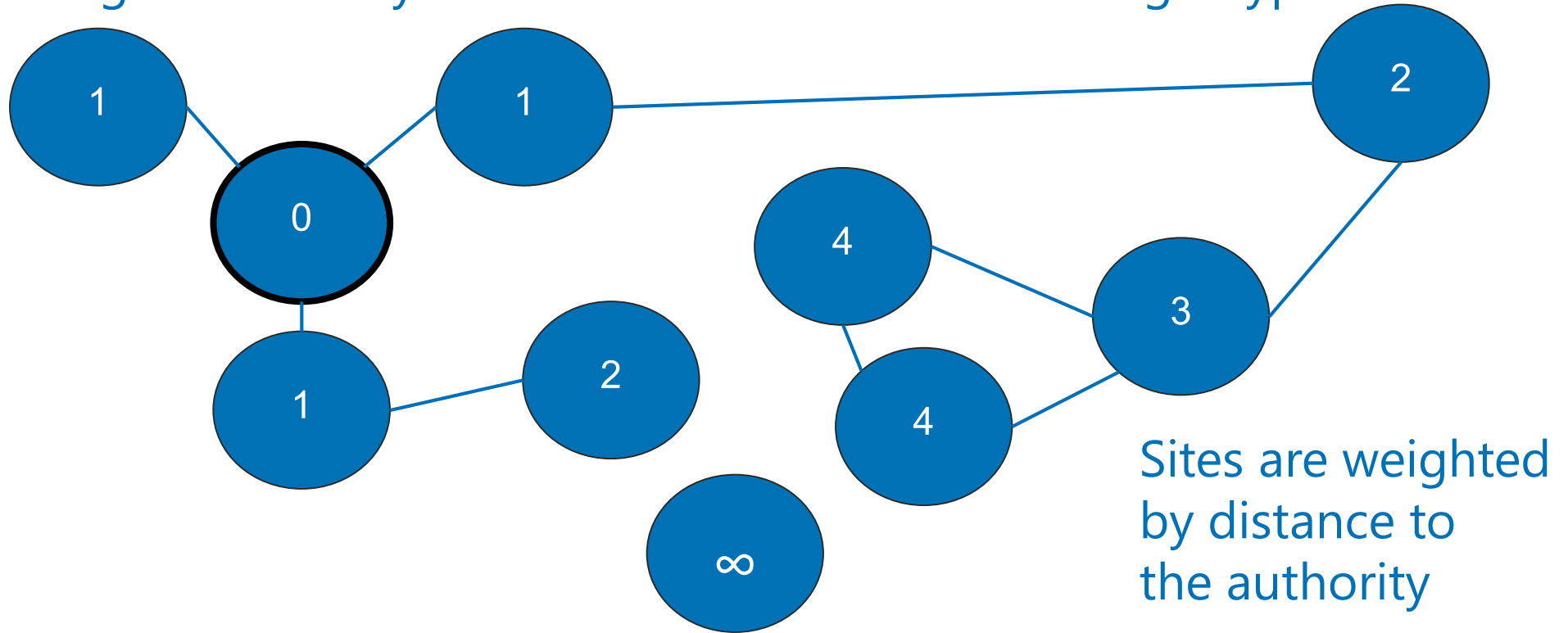


# Authorities: Connected



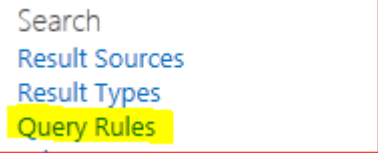
# Authorities: Connected

Setting an authority affects all sites connected through hyperlinks



# Query Rules

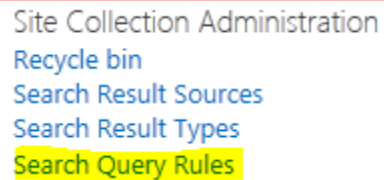
- Tune Search Results
- Created at the SSA, Tenant, Site Collection or Site
  - SSA



A screenshot of a navigation menu for the Search Service Administration (SSA) interface. The menu items are listed vertically: 'Search', 'Result Sources', 'Result Types', and 'Query Rules'. The 'Query Rules' item is highlighted with a yellow background.

- Search
- Result Sources
- Result Types
- Query Rules

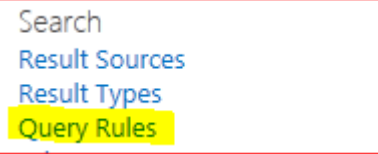
- Site Collection



A screenshot of a navigation menu for the Site Collection interface. The menu items are listed vertically: 'Site Collection Administration', 'Recycle bin', 'Search Result Sources', 'Search Result Types', and 'Search Query Rules'. The 'Search Query Rules' item is highlighted with a yellow background.

- Site Collection Administration
- Recycle bin
- Search Result Sources
- Search Result Types
- Search Query Rules

- Site



A screenshot of a navigation menu for the Site interface. The menu items are listed vertically: 'Search', 'Result Sources', 'Result Types', and 'Query Rules'. The 'Query Rules' item is highlighted with a yellow background.

- Search
- Result Sources
- Result Types
- Query Rules

# Query Rules

- ▶ Condition

- When Do I apply the rule?

- ▶ Action

- What to do when the rule is matched?

- ▶ Publishing

- When should the rule be active?

# Query Rules

## Conditions

- Exact match, beginning or end
- Ad-hoc or term store dictionary
- Match a regex (advanced)
- Is this query more likely aimed at the following source...?
- Do people mostly click on result of the following type...?

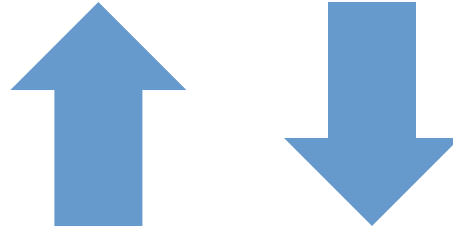
## Actions

- Show a promoted result
- Show a block of results
- Replace the core results with a different query

- › Dynamically Ranking Change

- › Part of the query

- › Results Ranking



# Query Builder

BASICS **SORTING** TESTS

Sort results

You can have several levels of sorting

Sort by:

Rank

Ranking Model

Choose a way of ranking results.

Select mode

Dynamic ordering

On top of the ranking model, dynamic ordering allows you to promote or demote items in the search results.

Change ranking when:

File extension

Dynamic ordering

On top of the ranking model, dynamic ordering allows you to promote or demote items in the search results.

Change ranking when:

Title contains key monthly report

Promote to top Remove

Promote to top

Promote by 10

Promote by 9

Promote by 8

Promote by 7

Promote by 6

Promote by 5

Promote by 4

Promote by 3

Promote by 2

Promote by 1

No change

Demote by 1

Demote by 2

Demote by 3

Demote by 4

Demote by 5

Demote by 6

Demote by 7

Demote by 8

Demote by 9

Demote by 10

Demote to bottom

Dynamic ordering rule

# Session Objective and Takeaways

High Availability and Performance

Better Search Quality

Better management

Friendly results and tools



# Q & A

Thank You / Merci

[www.maadarani.com](http://www.maadarani.com), [mike@maadarani.com](mailto:mike@maadarani.com) , [@mikemaadarani](https://twitter.com/mikemaadarani)

[www.slideshare.net/maadarani](http://www.slideshare.net/maadarani)

**Remember to fill out your evaluation forms to win some great prizes!**

**&**

**Join us for SharePint today!**

Date & Time: Nov 23<sup>rd</sup>, 2013 @6:00 pm

Location: The Observatory Pub,  
Algonquin Student's Association

Address: A-170 on Algonquin Campus

Parking: No need to move your car!\*

Site: <http://www.algonquinsa.com/ob.aspx>

\*Please drive responsibly! We are happy to call you a cab 😊