

Connect ▶ Find ▶ Act

Distributed Search and SharePoint Deployments

Technical Article

December 2007

Guy Mounier
CEO

Abstract

A distributed organization has two main options to deploy a single global SharePoint farm: “Centralized” vs. “Distributed” option, while ensuring that the content is globally shared for collaboration and retrieval regardless of the option selected. Some Architects often refer to a third option which is “Centralized with Regional Distributed Sites”.

This technical articles summarizes optimal strategies for distributed deployment of Microsoft Search and SharePoint, as well as discusses a customer case with distributed deployment of Microsoft Search.

The information contained in this document represents the current view of BA-Insight on the issues discussed as of the date of publication. Because BA-Insight must respond to changing market conditions, it should not be interpreted to be a commitment on the part of BA-Insight, and BA-Insight cannot guarantee the accuracy of any information presented after the date of publication. This document is for informational purposes only. BA-INSIGHT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, AS TO THE INFORMATION IN THIS DOCUMENT. BA-Insight may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from BA-Insight, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

Contents

Centralized vs. Distributed SharePoint Deployment Which Option is Right for My Organization?	4
Does my Organization Need Content Acceleration or Replication?	5
Where Does Enterprise Search Fit into my Distributed SharePoint Deployment?	6
Customer Case Study	9

Centralized vs. Distributed SharePoint Deployment Which Option is Right for My Organization?

A distributed organization has two main options to deploy a single global SharePoint farm: “Centralized” vs. “Distributed” option, while ensuring that the content is globally shared for collaboration and retrieval regardless of the option selected. Some Architects often refer to a third option which is “Centralized with Regional Distributed Sites”. This last option is in fact the result of geographical partitioning of content, where global content shared across all offices is centralized, and local content access is mainly targeted to a local audience, in essence partitioned off from the centralized content into an isolated SharePoint farm instance. The result is the co-existence or multiple global and local “Centralized” SharePoint farms throughout the organization targeted to various audiences and partitioned by content type. The rest of this article will focus primarily on the technical implications of a truly “Centralized” vs. “Distributed” global SharePoint and Search deployment.

Centralized Deployments

The “Centralized” option is often picked in single-continent deployment because it is the easiest option to maintain and has the lowest hardware cost of ownership. It carries however certain risks and stress points to your network topology. First of all, you need to evaluate the number of users and related expected load that each distributed office will put on your network. As a rule of thumb, in a “Centralized” environment, a distributed office should have a network line of 3Mbps. Network latency will typically range from 100ms to 250ms depending on number of users at the site. Any lower data line speed will increase the network latency to a point where it will become quite noticeable to the end-user and hurt the user experience.

Distributed Deployment

The “Distributed” option is the recommended approach for widely distributed organizations. It does also offer the greatest flexibility in sharing content between the regional SharePoint farms. With geographically distributed content, each regional SharePoint farm becomes essentially a “central” hub hosting that content. The bandwidth requirements for such a Hub and Spoke architecture is half: 1.5 Mbps with network latency ranging from 100ms to 500ms depending on number of users expected to hit that regional SharePoint farm. The table below summarizes the bandwidth requirements and expected network latency by number of users.

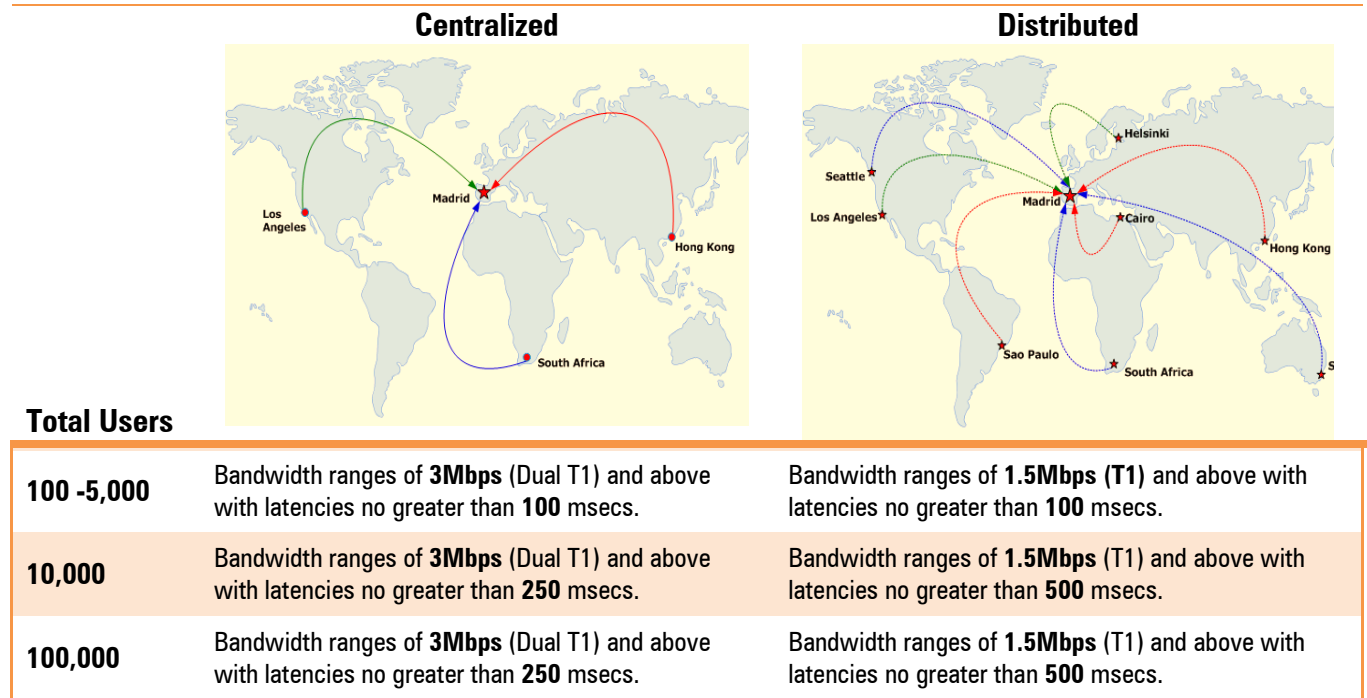


Figure 1: Recommended Bandwidth Ranges

Does my Organization Need Content Acceleration or Replication?

As discussed previously, many organizations have a hybrid environment, where some subset of content is meant to be shared globally, while managed centrally in a single SharePoint farm. If the bandwidth speed connecting the remote office to the central farm is sufficient, i.e. 3Mbps or more, there isn't a need to use any form of optimization solution. In most cases though, the organization will look into either "Application Acceleration" or "Content Replication" to speed up information retrieval. The table below describes the Pros and Cons of both information retrieval acceleration options.

	Acceleration	Replication
Description	Method to compress content, or transform it into a light-weight preview, that takes a fraction of the time to be retrieved by the Client.	Method to replicate and synchronize content across multiple SharePoint farms so that local users can retrieve a local copy of content.
Pros	Easy to deploy Provides faster access to WAN content	Content closer to the user Provides LAN speed for WAN content Resilient to network outages
Cons	Dependent on constant connectivity Not as effective on dynamic environments	Duplication of data Higher cost of ownership

Figure 2: Content Acceleration vs. Replication

Where Does Enterprise Search Fit into my Distributed SharePoint Deployment?

Enterprise Search, and in this particular context, SharePoint Search, does require some special attention by the Architecture Design team. Between the regularly scheduled incremental and full index crawls, and the volume of cross-site search queries, Enterprise Search often triggers a significant surge into network bandwidth utilization. It's a good thing for the Organization as a whole as the increase in utilization means more content sharing and reuse, the primary objective of an Enterprise Search initiative. However, it does impact the network topology if the Technical team does not prepare for it accordingly. It is also something that can catch a Technical team by surprise. As user adoption increases, a network topology that initially appeared adequately designed, quickly increases in latency as the user load increases.

The table below illustrates the four options for deploying SharePoint Search in a Distributed environment. It emphasizes the advantages, as well as the tradeoffs, of each option.

Option 1

Centralized Search

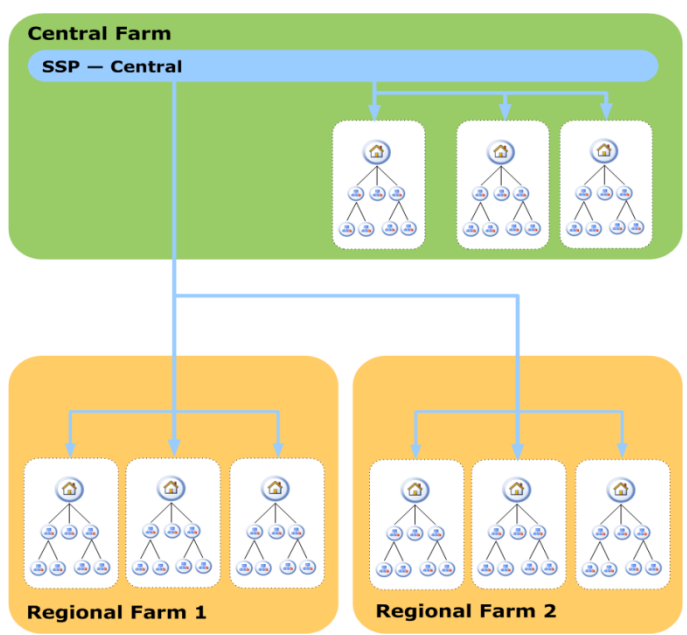
- Search service at the central farm crawls content at all regional farms.
- Search queries of regional users are sent to the central farm.

Advantages:

- Search relevancy is retained.
- Shared Services Provider (SSP) management is centralized.

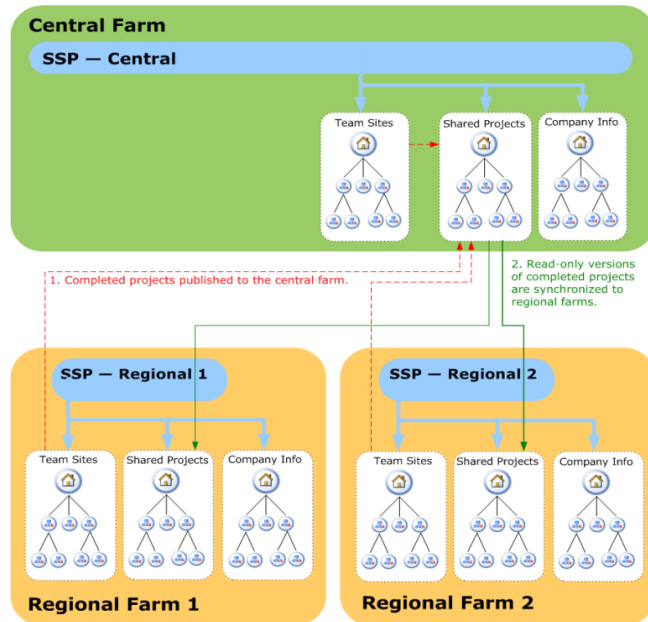
Tradeoffs:

- Content crawling over the WAN uses bandwidth.
- Keeping indexes current can be difficult in environments with high volumes of data and high rates of change.
- Query performance is subject to performance of WAN links.



Option 2 Regional SSPs with Synchronized Content Search

Each region hosts its own SSP. Because content is synchronized, there is no need to crawl content remotely over the WAN.



Advantages:

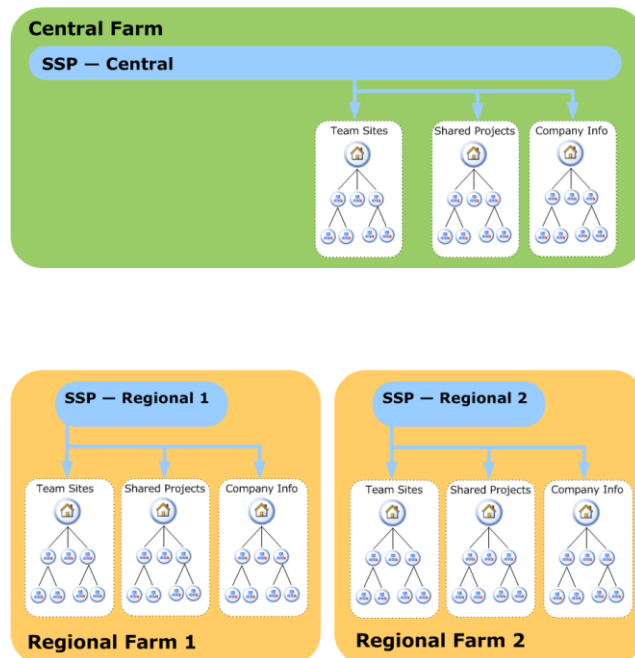
- Content is crawled locally.
- Search query performance is not subject to the performance of WAN links.
- Search relevancy is retained within each farm.

Tradeoffs:

- Multiple SSPs increase administrative costs.
- Synchronizing content across an organization increases the complexity of the solution.

Option 3 Distributed Search

Each regional farm hosts its own SSP, and the search service that is provided by each regional SSP crawls only local content.



Advantages:

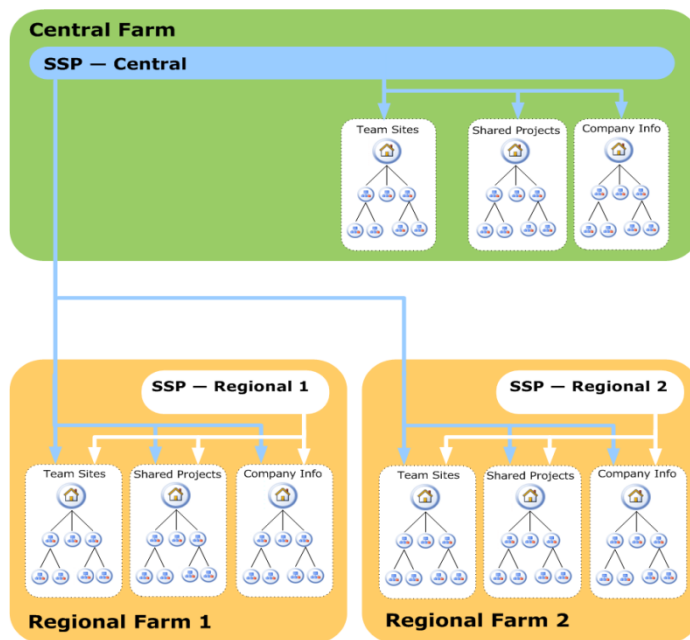
- Search relevancy is retained.
- Content is not crawled over WAN links.

Tradeoffs:

- Search is not enterprise-wide.
- Users at regional farms must connect to the central farm to search centralized content.

Option 4 Centralized Search + Distributed Search

- Each region hosts its own SSP that crawls local content.
- The central farm also crawls Web applications on other farms.



Advantages:

- Query performance is optimized for local content.
- Reduces the amount of queries over the WAN.
- Search relevance is optimized, based on the scope of the search (local or global).

Tradeoffs:

- Multiple SSPs increase administrative costs.
- Crawling content over the WAN uses bandwidth.
- For regional users, global query performance is affected by performance of WAN links.

In the real-world, we find that the hybrid option “**Option 4: Centralized Search + Distributed Search**” is most utilized for distributed SharePoint deployments. It makes sense given the fact that most Organizations have global content they want indexed, searchable, and accessible from a single unified index, while at the same time, local content that is retrieved mostly by local audiences. In such scenarios, the global search option can be supplemented by the Microsoft Federated Search framework, which will pull results from Regional SSP’s (i.e. indices). This architecture only works if most of the relevant content to a query is expected to come from the unified global index. Otherwise, the network latency issues will kick in when retrieving remote Regional Content frequently, causing issues of search user experience and above-average download duration.

Customer Case Study

The case study below illustrates a real-world example of a hybrid option SharePoint deployment. The result of the Bandwidth Consumption Analysis is summarized in the chart below for one of the farms (Central MOSS farm based in the U.S.). The details of the analysis can be found in the in the Bandwidth Requirements Analysis Tool in MS Excel which can be downloaded from the BA-Insight Website.

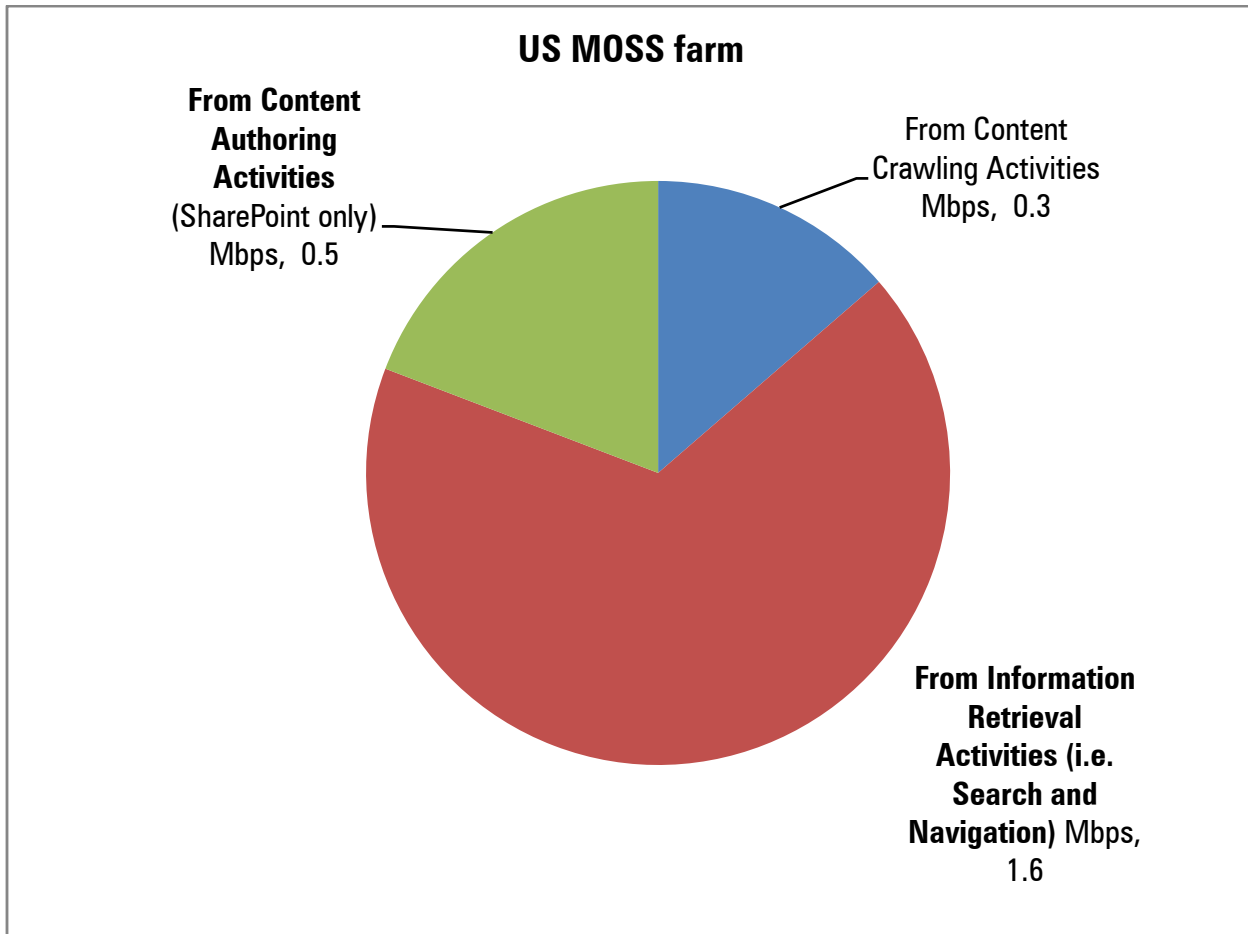


Figure 4: Customer example of Centralized + Distributed SharePoint Farm and Search Deployment

The Customer had the following SharePoint farm topology illustrated in Figure 5, with content synchronization enabled between the U.S. SharePoint Farm and the two additional Regional farms, one in Asia and the other in Europe. As a result of the analysis, the Customer resorted to invest in Content Replication and Synchronization technology to facilitate the dissemination of global content from the US MOSS farm to the Asia and Europe farms. The average 2.4Mbps total bandwidth consumption for the US MOSS farm was well-above their network topology capabilities, with certain remote sites well-under the 500 Kbps.

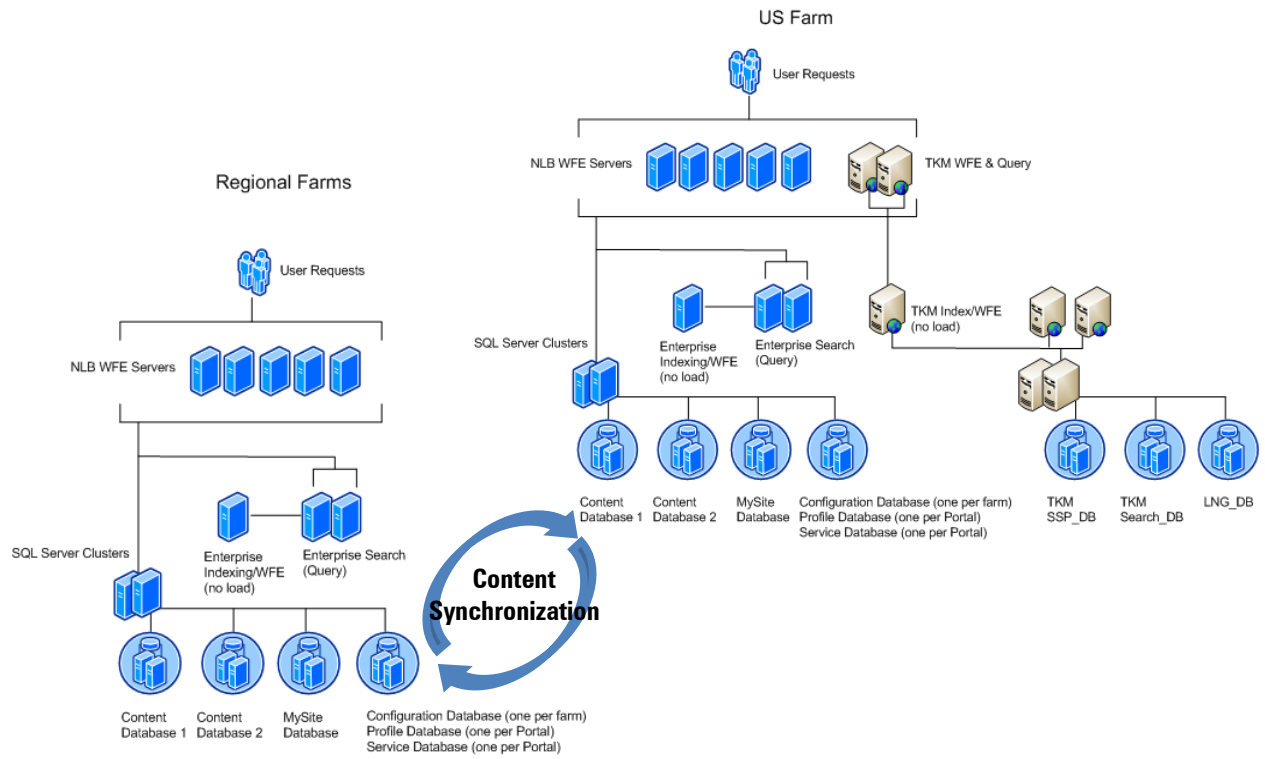


Figure 5: Customer example of Centralized + Distributed SharePoint Farm and Search Deployment



United States

445 Hamilton Avenue, SUITE 1102

White Plains, New York 10601

United States of America

Tel | 914.220.8395

Europe

Vester Voldgade 106,

DK-1552 København V

Denmark

Tel | +45 2161 6500

Web | www.BA-Insight.net

Email | info@ba-insight.net