

Search Environment Planning for Microsoft SharePoint Server 2010

Search model 2 of 4

1 Where are your users and content?

Are your users and content spread across multiple geographic regions?

The location of your users and content will influence your overall plan for enterprise search. Global architecture refers to how Microsoft® SharePoint® Products are deployed to meet the needs of users within a large region or around the globe. This section presents several common global architectures based on geographic location of users and content.

Key concepts to consider include:

- **Central site** — The location that hosts the majority of company data and employee computers. In many cases, a SharePoint solution deployed to a central site can serve the needs of users that are spread across a region or around the globe.
- **Regional sites** — Locations that hosts a subset of corporate data and employee computers connected together by using a combination of LAN and WAN links. If your organization includes regional sites, you will have to decide how to include these locations in the overall global architecture. Several options are presented (right).
- **WAN performance** — If you have users spread across a large region or located around the globe, performance considerations to factor into your planning include:
 - **User response times** — the time it takes users to receive results from common SharePoint operations, including viewing a Web page, opening a document, and receiving results from queries.
 - **Content crawl times** — the time it takes to crawl content across a WAN.
- **Bandwidth and latency** — Results of bandwidth and latency tests for the previous versions of SharePoint Products and Technologies indicate that latency—more than bandwidth—affects WAN performance until bandwidth becomes restricted at 512 kilobits per second and lower. For example, a T1 line with a latency of 500 milliseconds provides about the same performance as a T3 line with the same amount of latency. For best performance over a WAN, we recommend that you target a bandwidth range of 3 megabits per second (Dual T1) or greater.
- **Scope of enterprise search** — A key driver in architecture design is the breadth of results you want to include in enterprise-wide search and how integrated you want the results to be. Along with the decision of how to deploy globally, you have to decide whether to crawl all content, federate to some content sources, or provide a more limited search scope that does not include all content in the organization.

Single region



Users and content are located within a single geographic region.

If users and content are contained within a single geographic region, you can deploy SharePoint Server 2010 in a central location and optimize sites and services based on the scale and capacity that you want to achieve at the central site.

Central farm with users around the world

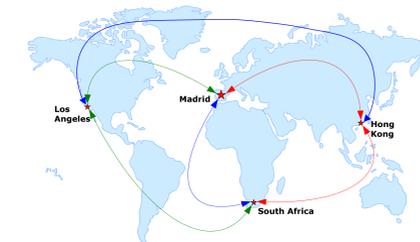


All SharePoint Server 2010 farm components and services are hosted at a central site. Users at remote sites access all sites and services across the WAN.

This solution is the recommended solution for environments in which bandwidths and latencies between WAN connections provide a reasonable user experience. This solution typically works well when work sites and users span a single continent. If you support users across low-bandwidth connections or users who are located on multiple continents, it is important to evaluate the bandwidth and latency combinations for these connections along with the expected use of SharePoint Server 2010 (page sizes, file sizes, and usage patterns) to gauge how well a central deployment will serve your organization.

If users are unable to use the sites and services because performance over the WAN is too slow, consider deploying multiple server farms to regional sites. However, before scaling beyond the central solution, evaluate the ways in which you can optimize a central solution to improve performance over the WAN.

Central farm with regional SharePoint deployments



In addition to a central site, regional sites also run SharePoint Server 2010. Users collaborate with local teams on the local server farm. Users access all enterprise-wide features across the WAN.

Deploying SharePoint Server 2010 at regional sites is recommended if the performance of WAN connections makes it difficult for regional users to efficiently use the central site and you are unable to optimize the central solution to improve WAN performance to acceptable levels.

The central with regional sites solution provides local access to services that regional users are likely to use most—for example, collaboration sites and My Sites. It also provides local search for local content. With this solution, you must decide how to incorporate regional content in search results. Options include crawling regional farms from the central farm, using search federation across farms, or not including all regional content in enterprise-wide search.

Deploying multiple farms increases the complexity and operations costs of the overall solution. This solution also requires greater organizational coordination to build effective governance of content that is authored in multiple geographic locations.

Central farm with many distributed SharePoint deployments



Regional sites operate autonomously from other regional sites. This solution includes a centralized portal site, but the solution is not optimized for collaboration across regional sites. Regional sites can take advantage of SharePoint Foundation 2010 instead of deploying SharePoint Server 2010.

This solution is recommended for organizations that have many offices distributed geographically, such as branch offices. This solution provides local access to collaboration sites at the regional office as well as access to the central site. It also provides the ability to share content between a regional site and the central site.

With this solution, you must decide how unified the search experience is. You can provide enterprise-wide search in which the central site crawls content on each of the regional sites. Or, you can provide a more limited experience. Be aware that if SharePoint Foundation 2010 is deployed at regional sites, search at regional sites is scoped to the site-collection level.

2 What content do you plan to include in search results?

Where is your content located and what format is it in?

There are a variety of ways you can include content in search results. Factoring in the overall volume of content that you plan to include and where it is located can help you understand which options are appropriate for your environment. It is also important to know what format the content is in. This can help you decide about whether to crawl or to federate. Several options for including content in search results are described to the right.

Do you plan to federate search results from multiple content stores?

You can federate search results to reduce the amount of content that is crawled by any one farm. Federation also lets you take advantage of content that has already been indexed by a different farm or technology.

In general, plan to limit federated locations, especially if you plan to display federated search results in different Web Parts on the same page. For example, two or three federated locations are reasonable.

If you plan to use federation, decide on federated choices as part of initial architecture. This will help determine what content remains for crawling. For example, you can use federation to pull together content that is indexed by farms at regional locations. Or, you can use federation to search across subscription news services or to take advantage of the built-in search functionality provided by other applications such as e-mail or line-of-business applications.

Do you plan to crawl content over the WAN?

SharePoint Server 2010 provides the ability to configure separate crawler components to crawl content at remote locations. You can put these crawler components on separate crawl servers. Consequently, crawling over WAN connections can take place continuously without affecting the performance or time it takes to crawl other content. For more information, see "Plan for bandwidth requirements" in the TechNet library.

Federate search results from other content locations



Federated search enables users to run a query that searches multiple locations and displays results in separate Web Parts on a single search results page. These locations can be other farms in your environment, other content repositories, subscription news sites, or other search engines.

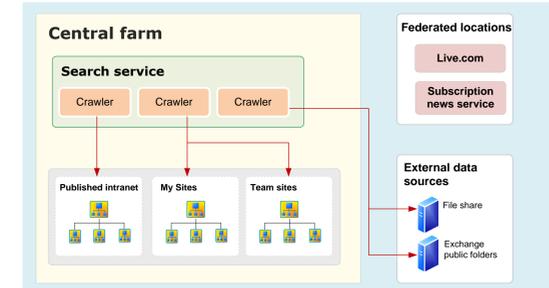
Using federation enables you to provide more extensive query results for your users without devoting server resources to crawling content.

Key questions to answer include:

- Will search federation be a part of the architecture?
- What is the best use of federation across the organization?
- Which content locations are the best federation candidates?

In a distributed environment with server farms in different regions, federated search can be used to provide search over central and regional content without crawling over the WAN. The user will see search results from each region in a different Federated Results Web Part. The results can be displayed as soon as they are received. For example, search results from the local server farm will most likely be returned before search results that are received over WAN connections.

Crawling content within a single region

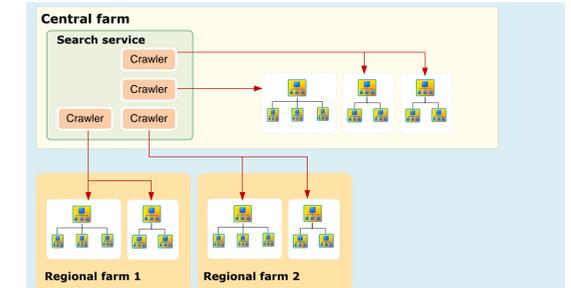


Within a single region, planning for crawling can be straightforward. This illustration shows an implementation in which crawling is used to incorporate internal data in SharePoint sites, file shares, and Exchange public folders. Federation is used to incorporate external data sources.

Within a single region, key information to gather during the environment planning phase includes:

- What SharePoint sites to crawl, both on the central farm and on other farms that might be in the environment.
- Other content locations to crawl (such as file shares or other types of Web sites). This will help determine what search connectors you will need.
- Approximate volume of content in each external data source.

Crawling content at regional sites



If your environment includes SharePoint sites that are hosted at regional locations, this will be a primary factor driving the architecture of your search environment. Key decisions to make include how widely to include search results. Several options include:

- Enterprise-wide search in which results from all locations are included and relevance ranked together (requires crawling all content from one farm).
- Enterprise-wide search in which results from different regions are listed separately (federated).
- Central search and regional search as separate search experiences (each location crawls its own content).

The example illustrated above shows a central farm crawling content at regional farms. In this example, search results include all content in the environment and relevance ranking is applied uniformly. Regional farms can also crawl local content to provide more local search of local content.

3 How many farms do you have to plan for?

This section prompts you to think about the number of farms that you might need to plan for and how the search service can be deployed to serve these farms. With any farm, FAST™ Search Server 2010 for SharePoint can be used in place of SharePoint Server 2010 search. To get ideas about how to architect and size each farm, review the fourth model in this series: *Designing Search Architectures for Office SharePoint Server 2010*.

Do you need regional farms?

After accounting for your users and content, you have to decide how many farms to deploy. The first question to answer is whether you will deploy SharePoint farms to regional locations. Given the distribution of user and content and the options for including content in search results, you can weigh the options and make an informed decision for your environment.

A reasonable approach is to implement a central farm, optimize this farm for WAN connections, and then decide if additional farms for regional locations are warranted.

If you decide to deploy regional farms, the most common implementation is to deploy a single farm to regional locations that provides full SharePoint Server 2010 functionality (as opposed to a dedicated search farm). The topology and sizing of regional farms depend on the needs of each individual region.

Do you need a dedicated services farm?

Within the central region, where the majority of company computers and data are located, services such as search are optimized when these are hosted by a dedicated services farm. This is especially true in environments that include multiple SharePoint farms. A dedicated services farm enables all farms to consume services that are managed centrally.

Consider deploying a dedicated services farm if:

- Your search environment includes 20 million items or more.
- There are two or more farms in the environment.
- You want query results to be as fresh as possible.

How about a dedicated search farm?

If a dedicated services farm makes sense for your environment, consider whether your environment will be better served with a dedicated search farm. In this case, all search services are hosted on a dedicated search farm and all other shared services are hosted on a separate services farm. This allows you to optimize the hardware and configuration of search components specifically for search.

Consider deploying a dedicated search farm if:

- Your search environment includes 40 million items or more.
- There are two or more farms in the environment.
- You want query results to be as fresh as possible.
- You are crawling content over the WAN.
- You want to implement a different SLA (failover, backup and restore, etc.) for the search service than other services hosted by SharePoint.

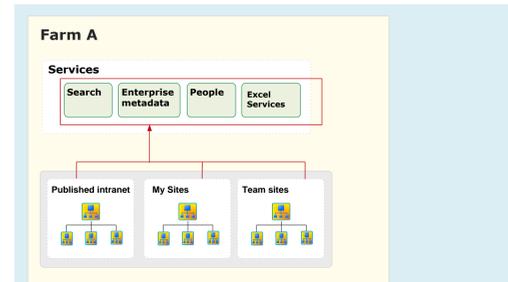
How many other farms are required for the central location?

If you decide to deploy a dedicated services or search farm, you have to deploy at least one more farm to host content. Also, other groups within an organization might require separate and dedicated farms.

Common reasons to deploy multiple farms to a central location include:

- To host services and content on separate farms.
- To host published intranet content.
- To host a records management solution.
- To meet organizational needs. For example:
 - Physical isolation of data — A group might require physical isolation of sensitive data.
 - Delegate management of one or more services — A group might need to directly manage a specific service (rather than sharing the service across the organization).

All-purpose farm with both services and content



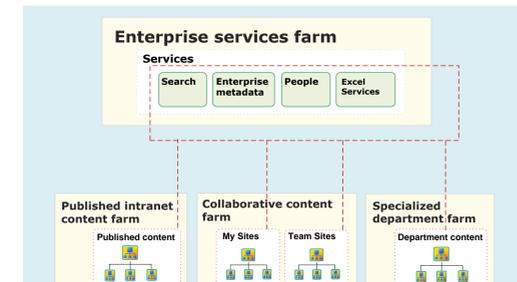
For many environments, a single farm hosting the full set of functionality that you want is all that is necessary.

The illustration shows a farm that hosts both content and services.

Considerations when deploying an all-purpose farm at regional locations where there is also a central farm at the central location include the following:

- Some services can be shared over the WAN, but not all. Performance of services shared over the WAN will vary.
- If you plan to crawl content locally, host the search service locally.
- You can host a local instance of the Enterprise Metadata service while also connecting to this service at a central farm. The Enterprise Metadata service manages taxonomy, social tagging, and other features. In this scenario, you must designate which service instance is the primary service which hosts the corporate taxonomy.

Dedicated services farm



For environments where there is a lot of content (20 million items or more) or where multiple farms are deployed, we recommend a dedicated services farm.

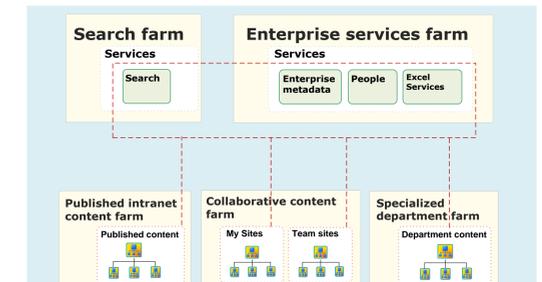
The illustration shows an environment with three different types of farms consuming services that are hosted by a centrally managed services farm. Architecture details to consider include the following:

- Some services cannot be shared across farms, such as client-related services. These services will need to be hosted locally on each farm.
- Services are configured at the Web application level. Each Web application can be configured to use a different set of services.

For more information about how services are configured, see the following models:

- [Services in SharePoint Products and Technologies](#)
- [Cross-Farm Services in SharePoint Products and Technologies](#)

Dedicated search farm



In large environments where there is a lot of content (40 million items or more) or where search performance and freshness is a high priority, we recommend a dedicated search farm.

This illustration shows an environment with two different services farms:

- Dedicated search farm
- Dedicated services farm for all other cross-farm services

Although the illustration shows how the search farm relates to the three other farms in the same region, a dedicated search service farm will likely also crawl external data sources and perhaps content at regional farms.

DRAFT