GartnerConsulting

The Emergence of Distributed Content Management and Peer-to-Peer Content Networks

GartnerConsulting

Engagement #010022501

January 2001

GartnerGroup

251 River Oaks Parkway, San Jose, CA 95134
+1-408-468-8811

Table of Contents

Executive Summary

There are three key trends driving the need for, and emergence of, distributed content management solutions: explosion of unstructured data; the critical need to formally manage content; and internetworking and collaboration within and between enterprises. These trends are converging to produce two key requirements—the need to create superior online user experiences and the need to work collaboratively.

Distributed content management systems address the need to access content wherever it resides, produce content while maintaining control over it, and collaborate efficiently by sharing data real-time within a distributed network of stakeholders. These systems create virtual content repositories that eliminate the need for structured storage. In fact, with these systems, data structure becomes irrelevant because information is accessed at its source, in its native format, expanding the reach and participation of stakeholders.

Gartner sees immediate synergies between distributed content management and Web content management (WCM) solutions. Together, distributed content management and WCM solutions provide access to potentially all enterprise and interenterprise content, and allow that content to be effectively managed and distributed via the Web.

Of course, distributed content management solutions can also stand on their own. Enterprises that need to share real-time information across geographically dispersed knowledge workers can benefit from such solutions immediately. When distributed content management solutions provide a virtual content repository without distracting from the strengths of other process-dependent systems, they can form the backbone of any platform that requires real-time, efficient information sharing.

In the short term, distributed content management solutions may complement enterprise portal solutions, which are less affected by wider e-business processes. Longer-term, distributed content management solutions may complement supply chain management (SCM), customer relationship management (CRM) and e-commerce solutions.

Peer-to peer computing (P2P) is currently touted as the next “killer application” for the Internet. Because P2P encourages a distributed architecture, it will create significant challenges in security, policy and workflow, but these are not insurmountable. Gartner believes that by developing formal P2P content networking solutions based on the Data Centered P2P model[1], enterprises will be able to create significant new dimensions of competitive advantage by leveraging real-time and otherwise unavailable content and content stakeholders.

[1] Data Centered is accomplished by creating a dynamic index server: as clients connect to the network, specific storage areas of clients are scanned and relevant data indexes added to the server. Users seeking data can search via the server, then connect directly to other clients for direct access to the data.

Given e-business trends, market drivers and the importance of partnership strategies, Gartner expects that P2P content networks will become prevalent within the next five years. Gartner believes that by the year 2003, 30 percent of corporations will have experimented with Data Centered P2P applications for content distribution (0.7 probability). Gartner also believes that half of the current server-based content management vendors will add Data Centered P2P functionality to their product offerings by 2005 (0.7 probability).

The emergence of P2P content networks will further blur the lines between organizational boundaries and empower users. Internetworking, where connections are made between enterprises to share information and collaborate, will create value for end customers. For users, search requests can be set to run at the server level, and as new content is indexed, alerts will be sent to the users, thereby greatly increasing the ease of searching. Users will be able to create their own personal index that will reduce information overload, and will be able to share this index with other users who have similar requirements.

In a nutshell, Gartner believes that P2P content networks will be an important part of an enterprise's ability to work collaboratively, although security and control issues will need to be addressed before adoption becomes widespread. Gartner also believes that P2P content networks will be an important part of creating superior online user experiences, particularly when successfully integrated into WCM solutions.

Introduction

In this paper, Gartner explores the emergence of distributed content management and P2P content networking and its importance to enterprises which seek to benefit from collaborative, efficient and real-time sharing of data within a distributed network of stakeholders within and outside of their enterprise.

Distributed content refers to content that lies not only on centralized servers, but also to content which resides on individual computers within an enterprise and to content that resides in other enterprises. A distributed content management system addresses all stakeholders' needs to access content wherever it resides, and produce content while maintaining control over it.

Three key trends are converging to create the need for distributed content management:

- Data explosion
- The critical need to manage content
- Internetworking and collaboration.

We will explore the trends and consider the opportunities that emergent distributed content management solutions can fill. Given that P2P computing is currently touted as the next "killer application" for the Internet, we will analyze its impact on distributed content management.

The Need for Distributed Content Management

Driving Trends

Data Explosion

The first key trend contributing to the need for distributed content management is data explosion. Complicating individuals' and enterprises' attempts to filter, manage and share information is the fact that much of that information is unstructured. Structured data is stored in data warehouses, data marts and application databases. Unstructured data can not be readily stored in a relational database and is, therefore, difficult for users to access.

Other solutions will explore removing the need for a central repository of information and, therefore, the removal of the issue of having to structure data in order to store it. Distributed content management solutions, that create a virtual content repository by being able to tap into any data in its native format wherever it resides are not affected by the existence of structured and unstructured data.

The Need for Content Management Solutions

Web Content Management

One of the key areas of focus has been the management of content for Web sites, or WCM. WCM systems have been created to:

- Manage and tap directly into decentralized content sources while maintaining a centralized repository.
- Combine content components.
- Separate content from design.
- Provide flexibility and speed in the combination and delivery of content.
- Manage workflow of content creation.
- Provide scalability.

Gartner believes that by year-end 2001, more than 60 percent of Web sites that have more than 1,000 pages will rely on commercial packages for management of Web site content and components.

Despite their ability to help manage content for Web site delivery, WCM solutions rely on a centralized repository for the storage and management of content, and this introduces constraints in terms of access to content and the ability to keep the content dynamically updated. It is possible that distributed content management solutions could complement WCM solutions by providing a decentralized or virtual content repository on which WCM solution could be based.

Global Web Content Management

As enterprises tap into the Internet to expand their global reach, they are faced with an even greater content management challenge—global Web content management (GWCM). GWCM solutions address the requirement for content in multiple versions to be optimized for reuse based on culture, language and commerce. To create a global presence, enterprises are likely to need multiple, potentially replicated and centralized content repositories in different locations.

Distributed content management solutions could provide a global virtual content repository to GWCM solutions and alleviate the need for replication of centralized content repositories across the global enterprise. GWCM, however, puts stress on the areas of control and versioning, and distributed content management solutions will need to work with GWCM solutions to make sure that the necessary processes and controls for localizing language, culture, currency and business logic are maintained.

Internetworking and Collaborative Commerce

As enterprises have started to mature in their application of e-business, internetworking—where connections are made between enterprises to create recombinant business models and communities—has become increasingly prevalent and extremely powerful for efficiently creating superior value for the end customer through information sharing and collaboration.

The collaborative commerce (c-commerce) vision includes internetworking, but it goes a step further by enabling multiple enterprises to work interactively online to find ways to save money, make money and solve business problems—often by dynamically restructuring their relationships.

While distributed content management solutions are unlikely to deal with the transactional data created as a result of c-commerce, they could support collaboration in the sharing, and ultimately the creation, of information required to achieve and sustain recombinant business models and communities.

Enterprise Portals

While traditionally enterprise portals may have been focused on employee productivity, their reach is extending beyond the enterprise to embrace customers and trading-partner communities. Portals are unique in their ability to deliver quick and personalized access to a massive array of subject-related information.

Distributed content management will enhance the success of enterprise portals because it will provide an efficient and effective way for employees and external stakeholders to collaboratively share information whether it is unstructured or structured. Distributed content management solutions, however, are not reliant on a portal interface to allow networked stakeholders to view content.

Gartner believes that distributed content management will be applied to enterprise portal environments. Enterprise portals will leverage the ability of distributed content management solutions to provide real-time, personalized networks of information to all internetworked users.

Distributed Content Management Emerges

Distributed content refers to content that lies not only on centralized servers, but also to content which resides on individual computers within an enterprise and to content that resides in other enterprises. A distributed content management system addresses all stakeholders' needs to access content wherever it resides, and produce content while maintaining control over it.

The combination of the explosion of unstructured data, the need for content management, internetworking and collaboration, and distributed networks has created the need for distributed content management. By creating a virtual content repository, distributed content management solutions remove the need for physical and structured storage and allow real-time, efficient information sharing between multiple, geographically spread stakeholders within and outside of an enterprise.

Characteristics of Distributed Content Management

Distributed content management systems should ideally allow:

- Searching of distributed sources as if they exist on a centralized server. This is an absolute must-have item.
- Stakeholders to provide content from distributed sources, with the content residing where owners maintain control of it and owners can deliver real-time updates and changes.
- Preservation of secure access based on the rights from the original location.
- Administration of the entire network from a single point of access.
- Delivery of real-time content to the network that is not replicated.
- Authentication of users.
- Personalization of the display of content based on a user's profile.
- Creation of dynamic tables of contents.
- Consistent presentation of content regardless of its native format. Currently, the differing formats of distributed content act as a barrier to users attempting to collaborate with other enterprises.

Distributed Content Management Complements WCM

WCM systems currently employ a centralized data model—content is stored in a central repository even though content creators can be decentralized and connected through workflow. Even with the advent of WCM systems, some limitations exist:

- Enterprise content held outside of the central WCM repository is invisible to content creators and important knowledge and contextual information is lost.
- WCM systems may not hold the most recent version of content unless a content stakeholder updates content via the WCM system's workflow.

- Content held within the central WCM repository is not available to real-time searching because the workflow process controls new content.
- Expertise of various stakeholders remains untapped, especially when a stakeholder is outside of the enterprise's firewall.
- Stakeholders must work within WCM system workflow, which should be designed around business processes, but which may still not naturally fit with multiple work environments.

Together, distributed content management and WCM solutions provide access to potentially all enterprise and interenterprise content, provided that all stakeholders are adequately networked and allow that content to be effectively managed and distributed via the Web. WCM solutions provide the more rigorous infrastructure required to manage and publish content assets and processes, while distributed content management solutions contribute the ability to access, search and share information across all potential content stakeholders in real-time.

Content management grew out of the needs for security, control and auditing of content. To be successful, distributed content management and WCM models must be integrated without breaching the fundamental tenants of content management.

A Solution for Distributed Content Management: P2P Content Networks

An Introduction to P2P

P2P networking is currently touted as the next "killer application" for the Internet and, in particular, for distributed content management. However, it is vital to understand that P2P is not a solution in its own right, but a collection of five very different application architectural models that exhibit very different strengths and weaknesses. While P2P will have a major impact on how content management is deployed in the near future, it also poses some new challenges of its own.

By their very nature, all five P2P models encourage a distributed architecture. With this come challenges in terms of security, policy and workflow governance. However, it is incorrect to assume that all approaches to P2P encourage wild and uncontrolled access to users and data. All P2P models can be designed to be either:

- *Informal*: Uncontrolled access to index or directory and their related resources. Not governed by formal corporate policy (e.g., Napster and Gnutella).
- *Formal*: Strictly controlled access to index and directory services, with control residing at both server and client levels, and related resources. Also, pre-planned workflow for application usage. Protocols easily defined and monitored at the network level (e.g., NextPage and Groove).

Corporate content management will require a highly focused *formal* P2P approach.

The Five Models of P2P

The five models are described in full in the Appendix. This section gives a brief explanation of the five models.

Atomistic

The *Atomistic model* is the earliest and simplest approach to P2P. To many, this is the "true" P2P architecture since it involves direct client-to-client connectivity without any mediation by servers. However, because it lacks a server component, it has no method for establishing communication links based on data availability or user identity.

User Centered

User Centered P2P applications utilize a *directory* (distributed or on a single server) to provide an efficient way for users to make connections with other users on a network.

Data Centered

Data Centered (data-focused P2P application) allows users to search and access data and content held on other users' systems. This is the area of greatest excitement, if not promise.

Web Mk 2

Web Mk 2 is a convergence of all the above models with current Web architectures and infrastructure. In this model, today's browsers will evolve into user-configurable workspace managers, which integrate the three types of P2P interactions previously described into task-specific work environments. Multiple directory services will link users together on an ad-hoc basis. Concurrently, multiple indexes will allow access to different forms of data, whether on servers (the Web, FTP, application servers) or on client systems.

Compute Centered—Distributed Processing

Distributed Processing is also seen as a type of P2P computing. Here, instead of using a single large processor, an application's processing is split between multiple clients. A server is used to coordinate the split processing. The distinction from traditional parallel processing is that the nodes are distributed over the Internet and that they are used on an opportunistic basis.

P2P Models that Apply to Content Networks

Data Centered and Web Mk 2 support P2P content network requirements in which disparate resources are connected and accessed as if they were in one central repository. This section examines these models in more detail.

Data Centered

Data Centered (data-focused P2P application) allows users to search and access data and content held on other users' systems. As stated previously, synergies can be created through partnerships of WCM and distributed content management solution providers. This is the area of greatest excitement, if not promise.

How it Works

Data Centered is accomplished by creating a dynamic index server: as clients connect to the network, specific storage areas of the clients are scanned and relevant data indexes added to the server. Users seeking data can search via the server, then connect directly to other clients for direct access to the data. When a client goes offline, the index removes the reference, or availability to the reference, of the client's data.

More advanced implementations of the Data Centered model will place increased intelligence at the index level. In a business environment, this includes rules that govern content access based on security levels or job functions. In addition, certain file or content categories may be blocked entirely, for network security reasons or corporate governance.

Uses

The key *formal* use of Data Centered will be next-generation content management solutions. By using the Data Centered model with strict adherence to corporate content policy, companies will be able to provide search and access to internal and external documents.

Impact

The Data Centered model will spread the reach of the network deep into users' devices. Formal Data Centered applications will emerge to become an important part of enterprise content management. Gartner believes that by the year 2003, 30 percent of corporations will have experimented with Data Centered P2P applications for content distribution (0.7 probability). In some cases, these experiments will have been undertaken without IT management's guidance—just as many Web projects originated as skunkworks, so too will corporate use of Data Centered P2P.

Gartner also believes that half of the current server-based content management vendors will add Data Centered P2P functionality to their product offerings by 2005 (0.7 probability).

Web Mk 2

Web Mk 2 is a convergence of all the above models with current Web architectures and infrastructure. In this model, today's browsers will evolve into user-configurable workspace managers, which integrate the three types of P2P interactions previously described into task-specific work environments. Multiple directory services will link users together on an ad-hoc basis. Concurrently, multiple indexes will allow access to different forms of data, whether on servers (the Web, FTP, application servers) or on client systems.

How it Works

Web Mk 2 will rely heavily upon software "bots" (e.g., small software robots or agents), that will often be packaged into interlocking clusters to provide the desired online capability while masking the connectivity complexity. In advanced applications that are spread over the network as a whole, clusters of software bots residing on multiple clients and servers will interact.

Uses

Informal Web Mk 2 solutions will be used as personal desktop portals, placing much of the personalization power now governed by Web site owners into the hands of individual users. For example, users of an informal Web Mk 2 application would be able to extract and consolidate multiple sources of information on a company for investment purposes from Web servers and other clients. They could leave comments, or even modification notices, on the original content, and even open live discussions about any part of the content with like-minded individuals.

Formal Web Mk 2 applications provide management with the ability to limit content access and user interaction. These applications will enable rapid deployment of corporate intranets, based on ad-hoc and/or predefined information and organizational structures. The architecture will also enable the deployment of *distributed applications*—for example, an application where the logic and process control are spread over an entire supply chain, with each organization on the supply chain using specific bots that negotiate and exchange information with other elements along the supply chain. When used in this manner, the Web Mk 2 model is a platform for executing highly complex and changing community-based applications.

Impact

Unlike Atomistic, User Centered and Data Centered, Web Mk 2 will dramatically change how users interface with the Internet. The inherent complexity of this model will require an entirely new approach to client software, pushing the current Web-browser designs beyond their capacity.

The Web browser as we know it today will give way to informal Web Mk 2 clients that drive the Internet deeper into users' devices. Applications and data will reside in multiple locations, with users' being able to interrogate information held on servers and other users systems, or even bypass server-based computing altogether when desired. The fluid nature of Web Mk 2 will cause an upheaval for existing online services, portals, auction sites, banner advertising and existing B2B transaction models. The traditional print and digital media companies will be faced with a transition even more brutal than the Web.

Informal Web Mk 2 client vendors will face legal challenges from online publishers that see content being stripped of advertising during the search and retrieval process by the year 2004 (0.8 probability).

Corporations will at first resist the highly distributed and ad-hoc nature of Web Mk 2 applications. However, as vendors place greater control options over Web Mk 2 applications, the formal Web Mk 2 model will become a far more cost-effective way of deploying corporate intranets. By 2007, up to 20 percent of new Intranet deployments will include Web Mk 2 application elements (0.7 probability).

P2P Content Networks—Opportunities and Challenges

Blending Internal and External Content—The Permeable Organization

Organizations can no longer afford to be islands of information. While traditional approaches to content management provided a formal structure to organizations' internal data, the increasing amount of information available from public sources (e.g., the Web, newsgroups, chat servers, newsfeeds, etc.) is being left untapped.

There are two ways that P2P content networks based on the Data Centered model can assist in harnessing external information and blend it with internal corporate information for competitive advantage. The first approach to is to use index servers that scan for content both inside the company and also for content from external public sources, based on search criteria set by the organization. These central servers then make all information available as if held within a single repository, regardless of physical network location. The benefit of this solution is its simplicity and ease of deployment. Access to indexed information in this scheme can be via a Web browser, with file formats being moved natively. The downside is scalability, network utilization and auditing of information sources.

The second approach is to consolidate search results at the client. Using existing Web indexes (such as Yahoo and AltaVista) and private internal index servers, client software can locate content from multiple sources. Client-based software will then combine the results of these

indexes. The benefit of this approach is scalability—indexes are distributed. The drawback is bandwidth utilization, as clients will make multiple network requests for a single search.

Persistent Searching

An extension to ad-hoc indexed searching in a Data Centered environment is to provide *persistent searching* via the Web Mk 2 model. Here, a user's search requests (search rules) can be set to run either in real-time or batch at the index server level. As new content is indexed, alerts are sent to the user via a directory server. Because this is a P2P Web Mk 2 model, alerts will be passed to whatever device the user has specified or is connected to at the time.

E-mail news retrieval systems already provide some level of this functionality. However, holding the search rules at the index server level provides a greater level of management and control, and the use of a directory service expands the number of possible alert channels.

Because many P2P content management solutions will be deployed with multiple index servers networked together in a P2P fashion, vendors looking to provide persistent searching will need to deploy search rules as agents or bots that can be replicated between index servers, but without duplicated processing. For this reason, index server management will become a paramount concern.

Personal Indexes

With the ability to locate and draw down content from multiple internal and external sources, and obtain immediate alerts when new content is added or modified, information overload will become an issue. An approach to help solve this will be the creation of *personal indexes*—indexes that use search rules to produce a subset of pointers to content. When users view these indexes, they will already be limited to just information that the user, or corporate rules, have dictated.

Initially, this will be applied at the index server level. However, Gartner believes that the Data Centered model will encourage these personal indexes to be held at the client level. The client-based index will be available online and offline, plus it will distribute the processing load from the server-side index engine. Synchronization based on alerts from persistent searching will ensure that the client-side index is updated. As personal indexes become popular, increasingly less inquiry requests will be demanded of the server-side indexing engines, although the increased demands to process search rules will push up the hardware requirements.

Content Communities—Sharing Personal Indexes

As users build personal indexes that have a strong theme, it is certain they will wish to share these with other users. They will request ways to make their personal indexes available to other users within the corporation. From a technical point of view, it is likely that many users will have similar—if not the same—search rules, creating many duplicate search requests. It therefore will be essential that rules can be consolidated at the index server level. Rules that are sufficiently similar from multiple users will be combined and executed once, with the results being distributed to all participants' personal indexes.

Given the nature of a P2P content environment, it is also possible to allow users to select a subscription to another user's personal index, thereby creating communities around a theme of content.

Content Consolidation—The Personal Desktop Portal

Next generation browsers will meet this challenge by allowing data within a P2P content network to be consolidated. For example, content may be downloaded from the search results from a dozen indexes, compared with each other for duplication, then merged into a single view.

This approach is functionally similar to current server-based personalization engines, where content is consolidated into a single view. However, by placing the rules and intelligence of the personalization at the client level, the content base for personalization is broadened and placed under the control of the user, not the service provider.

The challenge for organizations with external Web sites will be to retain branding when content can be stripped from their site and merged with content from other sites.

Living Documents—Content that Sticks with People

With the Web Mk 2 model, content will no longer be separate from users, but can contain direct links to users that:

a) Authored the content.

b) Formally edited the content.

c) Read the content/held discussions regarding the content.

d) Are currently reviewing the content.

e) Have been assigned as a content owner (personal or corporate user).

This will be accomplished by storing pointers to users and the directory services where the users can be found.

Conclusion

Distributed content management solutions will emerge as a complement to WCM solutions. Trends driving distributed content management are the explosion of unstructured data, the critical need to formally manage content, and internetworking and collaboration within and between enterprises. P2P computing is currently touted as the next “killer application” for the Internet. P2P content networking is particularly exciting because, by allowing access to information that would otherwise not be accessible, it can create competitive advantage either as a standalone solution or in conjunction with a WCM solution.

P2P technology is widely predicted to have significant impact on the emerging distributed content management market. Because P2P encourages a distributed architecture, it will have to address challenges in security, policy and workflow. Gartner believes that by developing formal P2P content network solutions based on the Data Centered P2P architecture, enterprises will be able to create significant new dimensions of competitive advantage by leveraging real-time and otherwise unavailable content and content stakeholders. Gartner believes that by the year 2003, 30 percent of corporations will have experimented with Data Centered P2P applications for content distribution (0.7 probability). Gartner also believes that half of the current server-based content management vendors will add Data Centered P2P functionality to their product offerings by 2005 (0.7 probability).

Enterprises that need to give users access to distributed, business-critical content without attempting to centralize the data should consider distributed content management solutions and, in particular, P2P content networks. Enterprises that are highly dependent on sharing real-time information across geographically spread knowledge workers are likely to benefit immediately from P2P content network solutions.

Appendix: The Five Models of P2P

This appendix describes the five models of peer-to-peer (P2P) in more detail.

The Five Models of P2P

The five Models of P2P computing demand an understanding of five *application dimensions:*

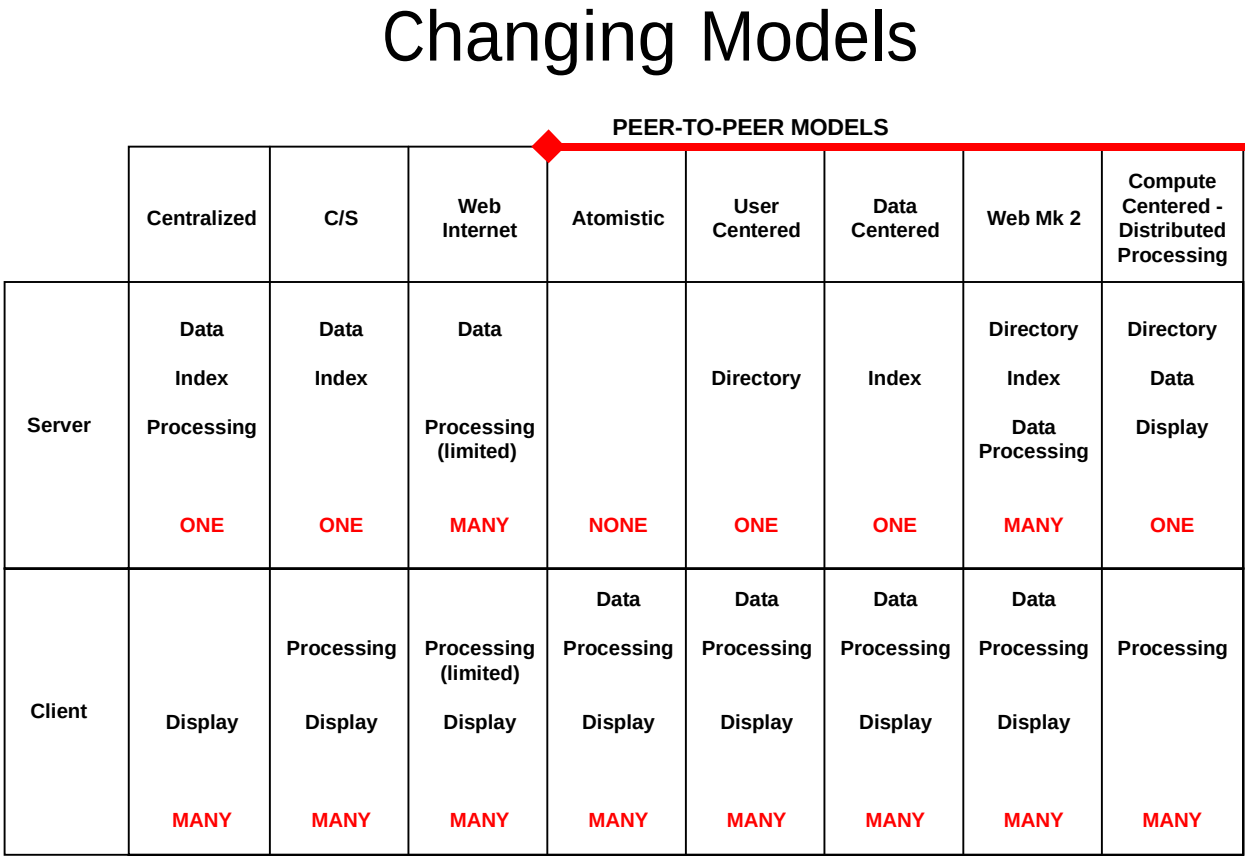
- *Data:* Raw data or information held in either a structured or nonstructured manner.
- *Index:* Collection of logical links to *data elements*, no matter where on the network this is found.
- *Directory:* Collection of logical links to *users*, no matter where on the network they are found.
- *Processing:* The application of computing power to analyze data or transform it.
- *Display:* Delivering data, either in raw form or as a result of processing.

It is important to note the semantic difference between an *index* and a *directory*. *Indexes* explicitly point to data, while *directories* point to users. In many vendor products, such as Novell's NDS and Microsoft's Active Directory, a directory includes users, data, locations and discrete devices on the network. However, for the sake of clarity and simplicity, in the rest of this paper we will refer to directories and indexes as discrete architectural components.

When data resides on a user client, such as with Napster, the *index* may indirectly point to a user. However, one cannot use an *index* to find a user based on identity—only based upon the data they posses. Conversely, a *directory* may point to a user based on profile data, such as age or willingness to accept adult content (as with NetMeeting), but would not be able to locate data over a network.

These five application dimensions may be spread between *servers* and *clients,* and not all forms of computing will utilize all five (see Figure A-1). Traditional centralized approaches, two-tier client/server and the Web can be represented in much the same manner as the five newer forms of P2P computing.

Figure A-1. The Five Models of P2P



Source: Gartner

It should also be noted in that in all P2P models, the definition of a server and a client begin to blur. In some instances, servers themselves may act as clients to other servers, creating a cobweb-like set of applications that trigger events in a distributed manner, or building distributed indexes for enterprise content. In other instances, clients may temporarily act as a server for a specific task, such as hosting an impromptu chat session for multiple users. While purists may see this as a security threat, with some justification, we feel that this blending is a natural evolution for computing and a direct result of Metcalfe's Law[2]. As processing power and bandwidth continue to increase in power, certain applications will evolve into P2P models.

[2] Bob Metcalfe, reasoned that 1,000 people on a network can have roughly one million different conversations, so he said the value of a network grows in proportion to the square of the number of users. The n^2 effect says that, given the choice of joining a large existing network with many users or an incompatible new one with few users, new users will almost always decide that the bigger one is far more valuable. The result is often explosive, accelerating growth once a network establishes dominance.

Atomistic

The *Atomistic model* is the earliest and simplest approach to P2P. To many, this is the "true" P2P architecture since it involves direct client-to-client connectivity without any mediation by servers. However, because it lacks a server component, it has no method for establishing communication links based on data availability or user identity.

How it Works

Network-based multiplayer games utilize this approach by sending out a greeting over the network to announce a new player or user. Once clients have recognized each other, communication between them is direct.

While this negates the need for a *directory,* or centralized server of any type, it is limited to a small number of users (under 200)—rarely the Internet as a whole. The bandwidth and network structure required to allow the announcement of a new client restricts Atomistic to private networks, or to applications where a user's Internet Protocol (IP) address is known.

Uses

An example of using Atomistic on the public Internet is NetMeeting, used when connecting directly to another NetMeeting client at a known IP address. In this case, it is vital that one user already knows the IP address of the recipient to be contacted. In contrast, a more common approach to using NetMeeting is to signal a new client's availability via a central directory, that then displays this information to other NetMeeting users looking at the directory—this is no longer the Atomistic model, but the User Centered mode, described below.

Impact

Atomistic will have minimal impact on the broader Internet. Its primary use will be within small- to medium-sized organizations with limited requirements for P2P applications. It will continue to be used for linking specific multi-user functions in closed network environments.

User Centered

User Centered P2P applications utilize a *directory* (distributed or on a single server) to provide an efficient way for users to make connections with other users on a network.

How it Works

In this model, a client will register itself with a directory or directories. The client will also scan directories for logged-in users that meet specific requirements. By using addressing information found in the directory, a client can request a direct connection with another client or clients. Once two clients are connected, the application becomes strictly client-to-client based.

Uses

Currently, User Centered applications have been limited to personal computer instant messaging applications, such as ICQ, AOL Instant Messenger and MSN Messenger. However, this model will evolve to embrace model devices ranging from pocket PCs and personal digital assistants (PDAs) to mobile phones.

Impact

Companies that run the *directory* servers are in a strong position to monitor the number and types of connections made and effectively know who every user on the network is. This *ownership* of users provides a compelling competitive marketing advantage. For this reason, telecommunications companies and online services will both vie to dominate this space, causing tension between the two entities.

Data Centered

Data Centered (data-focused P2P application) allows users to search and access data and content held on other users' systems. As stated previously, synergies can be created through partnerships of WCM and distributed content management solution providers. This is the area of greatest excitement, if not promise.

How it Works

Data Centered is accomplished by creating a dynamic index server—as clients connect to the network, specific storage areas of the clients are scanned and relevant data indexes are added to the server. Users seeking data can search via the server, then connect directly to other clients for direct access to the data. When a client goes offline, the index removes the reference, or availability to the reference, of the client's data.

More advanced implementations of the Data Centered model will place increased intelligence at the index level. In a business environment, this includes rules that govern content access based on security levels or job functions. In addition, certain file or content categories may be blocked entirely for network security reasons or corporate governance.

Uses

The key *formal* use of Data Centered will be next-generation content management solutions. By using the Data Centered model with strict adherence to corporate content policy, companies will be able to provide search and access to internal and external documents.

Impact

The Data Centered model will spread the reach of the network deep into users' devices.

Formal Data Centered applications will emerge to become an important part of enterprise content management. Gartner believes that formal Data Centered applications, in the form of trial projects, focused on accessing, sharing and reusing internal and external content will be present in 30 percent of enterprises by 2003 (0.7 probability).

Web Mk 2

Web Mk 2 is a convergence of all the above models with current Web architectures and infrastructure. In this model, today's browsers will evolve into user-configurable workspace managers, which integrate the three types of P2P interactions previously described into task-specific work environments. Multiple directory services will link users together on an ad-hoc

basis. Concurrently, multiple indexes will allow access to different forms of data, whether on servers (the Web, FTP, application servers) or on client systems.

How it Works

Web Mk 2 will rely heavily upon software "bots" that will often be packaged into interlocking clusters to provide the desired online capability while masking the connectivity complexity. In advanced applications that are spread over the network as a whole, clusters of software bots residing on multiple clients and servers will interact.

Uses

Informal Web Mk 2 solutions will be used as personal desktop portals, placing much of the personalization power now governed by Web site owners into the hands of individual users. For example, users of an informal Web Mk 2 application would be able to extract and consolidate multiple sources of information on a company for investment purposes from Web servers and other clients. They could leave comments, or even modification notices, on the original content, and even open live discussions about any part of the content with like-minded individuals.

Formal Web Mk 2 applications provide management with the ability to limit content access and user interaction. These applications will enable rapid deployment of corporate intranets, based on ad-hoc and/or predefined information and organizational structures. The architecture will also enable the deployment of *distributed applications*—for example, an application where the logic and process control are spread over an entire supply chain, with each organization on the supply chain using specific bots that negotiate and exchange information with other elements along the supply chain. When used in this manner, the Web Mk 2 model is a platform for executing highly complex and changing community-based applications.

Impact

Unlike Atomistic, User Centered and Data Centered, Web Mk 2 will dramatically change how users interface with the Internet. The inherent complexity of this model will require an entirely new approach to client software, pushing the current Web-browser designs beyond their capacity.

The Web browser as we know it today will give way to informal Web Mk 2 clients that drive the Internet deeper into users' devices. Applications and data will reside in multiple locations, with users' being able to interrogate information held on servers, other users systems, or even bypass server-based computing altogether when desired. The fluid nature of Web Mk 2 will cause an upheaval for existing online services, portals, auction sites, banner advertising and existing B2B transaction models. The traditional print and digital media companies will be faced with a transition even more brutal than the Web.

Informal Web Mk 2 client vendors will face legal challenges from online publishers that see content being stripped of advertising during the search and retrieval process by the year 2004 (0.8 probability).

Corporations will at first resist the highly distributed and ad-hoc nature of Web Mk 2 applications. However, as vendors place greater control options over Web Mk 2 applications, the

formal Web Mk 2 model will become a far more cost-effective way of deploying corporate intranets. By 2007, up to 20 percent of new intranet deployments will include Web Mk 2 application elements (0.7 probability).

Compute Centered—Distributed Processing

Distributed Processing is also seen as a type of P2P computing. Here, instead of using a single large processor, an application's processing is split between multiple clients. A server is used to coordinate the split processing. The distinction from traditional parallel processing is that the nodes are distributed over the Internet and that they are used on an opportunistic basis.

Uses

While this approach is appealing because it can harness massive parallel computing power from under-utilized low-cost clients, it is limited to a specific type of group of applications—calculations that can be broken down into smaller tasks. Examples include calculating large amounts of data acquired from external sources (such as SETI@home) and rendering individual frames of 3-D movies.

Impact

While some Compute Centered applications could exist for corporate users—such as finding patterns in massive amounts of credit card or shipping data—the level of trust with regard to deploying such sensitive tasks over even internal networks will hinder this model.

In the area of content management, a distributed computing approach can be given to indexing large amounts of corporate and/or public data. Rather than providing a simple, central index server, index servers can be scattered over the organization. Each index engine could index primary data from geographic or departmental sources, spreading the processing load of indexing. Then, each index server could cross-link its results with other index servers, either via a centralized or hierarchy P2P Compute Centered model. Index servers may even be personal in nature, residing on an individual's personal computer.