Clustering: Ontology and the Dublin Core

Customer's Name

Academic Institution

Clustering: Ontology and the Dublin Core

Introduction

Have you ever wondered how some organizations are able to perform parallel automated operations from different user groups or client locations while at the time ensuring that there are fast, accurate, and highly available information and services? Think of companies such as Google; how does it manage and synchronize all the user needs and focus on all clients at the same time? These issues bring us to clustering: Clustering has been available for at least the last two decades. Let us take a short survey on Google to capture a bigger picture of computer clustering:

Google has spent the past 15 years on systems software studies and research from university laboratories, and established their own proprietary, production value and quality cluster systems. But just what is this policy that Google has been establishing? Simply, it is a circulated computing network which manages;

> Web-scale datasets on 100,000 node-server cluster. It comprises of petabyte, strewn fault tolerant file system, distributed RPC code, possibly network shared memory and process migration, and a datacenter management system which lets a handful of operating system engineers effectively run 100,000 servers.
>
> Skrentha (2004)

The secret of Google's power according to Skrentha (2004), is simply system clustering! Google is a company which has established a single, incredibly huge, custom computer cluster system. It runs its own computer cluster systems. The company makes its giant computer cluster bigger and quicker each month, while reducing CPU cycles cost.

The questions that we may be asking ourselves right now are; ‘what then is this clustering and how does it operate? What benefits does it have over a single operator system? Thankfully, clustering is not as complicated as many people think and the next section, we are going to look into these questions and also consider how ontology and Dublin Core are linked to clustering.

About Clustering

Using simple layman’s language, clustering is simply linking two or more computers together, so that they can act like one computer. However, in computing, clustering refers to the use of multiple computers, commonly the UNIX workstations or PCs, multipart storage devices as well as redundant interconnections to create what appears as one vastly available server system. Cluster computing is used for high availability, fault tolerance, parallel processing and load balancing.

What does Clustering Include?

A server cluster contains a network of servers (such as the one used by Google), referred to us nodes. Nodes usually communicate with each other to generate sets highly available services to users. They are designed to serve functions with long-running in-memory rate or state updated data. They comprise of print servers, file servers, messaging servers and database servers: most of us have certainly come across these servers in the course day to day activities—even though we did not acknowledge the concept of clustering at that time—true?

How does Clustering Work?

Since clustering is terminology broadly used, hardware configuration of system clusters usually depend on the networking machinery chosen and the motive of the system: Three basic kinds of clustering hardware are involved: mirrored disk, shared nothing, and shared disk configurations. Using simplified terminology, let us look into each of these clusters, one at time, beginning with the Shared Disk Clusters.

Shared disk clusters. Mitchell (2010), points out that this method of clustering utilizes fundamental centralized I/O devices available to all network nodes (computers) within the system cluster. They are referred to as shared-disk clusters because the I/O implicated is normally disk storage for ordinary files and databases. Good examples of technologies using Shared-disk cluster are OPS (Oracle Parallel Server) and the HACMP of IBM. Shared-disk clusters depend on a universal I/O bus to access disks but they do not need a shared memory. Since all nodes might concurrently cache data from or write to the centralized disks, harmonization machinery has to be used to safeguard consistency of the system.

What about the Shared –Nothing- Clusters? Ordinarily, we many tend to think that this group of clusters operates like the single server systems, because if taken literally, they share nothing! However, as explained by Mitchell (2002), the term is used to mean that this form does not entail parallel disk accesses from multiple computers. This implies that they do not call for circulated lock manager and they include MSCS (Microsoft Cluster Server). MSCS clusters utilize a common SCSI link connecting nodes that logically leads some people to consider it as shared-disk solution. However, only a single server (owning quorum resource) requires the disks any time, hence no parallel data access takes place. This makes the other forms of clusters more preferable to many people as compared to this one.

Finally, we have the Mirrored Disk Clusters, which, according to Staab and Hotho (2005) consists of Legato's Vinca. Mirroring entails duplicating all application records from primary to secondary backup (possibly at a remote position) for high availability rationale. Duplication takes place whilst the prime system is active, even though the mirrored backup scheme as like Vinca -- normally does not execute any work beyond of its responsibility as a reactive standby. If a breakdown happens in the primary scheme, a failover process reassigns run over to the secondary scheme. Failover may take time, and applications could lose state data once they are reset; however mirroring facilitates a fairly quick recovery system requiring slight operator involvement. Mirrored-disk clusters usually include two nodes only.

Advantages of clustering

Having understood how clustering works, we may also want to clarify the benefits of implementing clustering in our operating systems, although it is apparently clear that clustering is good for your computer systems. Those of us using the single server system may have already noticed that clustering makes work easier; it is basically a tool for making work easier and faster, through allowing several tasks to be performed at the same time, for example, serving many clients concurrently (Staab, Hotho, 2005).

It is also an effective strategy popularly applied in the implementation of parallel process applications since it allows large organizations to leverage the existing workstation and PC investments. Clustering also makes it easy to put in new CPUS into the system by only adding the new PC to the existing network (Clustering, 2010). Shared-disk clusters hold up advanced intensity of system availability: if one computer/ node fault, other node is not affected.

Link clustering to the issues involving ontology and the Dublin Core

Clustering cannot be adequately understood without the mention of ontology; Ontology in computing refers to some kind of formal translation of knowledge via sets of concepts in a domain, and the correlation amongst such concepts.

There are various types of computing ontology which are used to provide machine-process-able semantics of information sources. It is important for us know the different types of ontology since not all can be applicable, depending on individual suitability or user needs.

Such include: representational ontology, task ontology, domain ontology, common sense ontology, and metadata ontology (such as Dublin core).In clustering systems, metadata ontology support shared communication between various workstations and PCs (Lacy, 2005). Therefore you realize that Dublin core is a form of metadata language/ontology. Actually, more than fifty percent of organizations have at one time used Dublin Core. The use of Dublin core as a form of metadata ontology to bring together correlated services and information across organizational clusters which are more appropriate to clients represents one major issue into which ontology is linked to clustering (Brodie, 2001).

The whole strategy starts with “an electronic cluster blueprint” (Brodie, 2001) whereby, sets of subject clusters are generated, each in lieu of a comprehensive set of services or information on a given subject. This system has been applied in some governments (e.g. Canada) where the services/ information referenced at cluster sites supply thousands of connections to federal agencies, departments, provinces as well as a mass of private (not profit making organizations). In the earlier section, you realized that concepts were more easily understood when applied in real life situations; thus I will use the same criteria to explain the use of clustering in metadata.

The real content exists in the organizational websites, along with cluster sites offering a subject-based perspective for clients to get information despite their host organization. Lacy 2005) explains that every cluster site is an entry which provides background information derived from the metadata. I agree with this statement because metadata assists clients locate information they are searching and aids cluster managers in administering and retaining content at the entry portal.

In the contemporary industry, cluster managers need a supple and vibrant metadata set. We can thus foresee in the near future central metadata depository where all content suppliers will add metadata via a universal procedure which will be applied by the diverse clusters. The production will depict the function metadata and how Dublin Core as metadata ontology meets these needs.

Conclusion

In conclusion, network clusters present a high-performance computing options to SMP and extremely parallel operating systems. Combined system performance not considered, cluster systems provide to more dependable computer systems. For every organization that considers competitive advantage in technology and client satisfaction is key strengths, then clustering is not an option. Still, for the individual business people who desire to overcome system overload and redundancy, clustering will be a step in the right direction.

References

Brodie, N. (2001). *Use if Dublin Core in a Portal Environment* [Abstract]. Retrieved from DCMI Conference Papers.

Clustering. (2010). In *Webopedia online.* Retrieved from http://www.webopedia.com/TERM/C/clustering.html

Lacy. W.L. (2005). *OWL: Representing information using the wed ontology language: Dublin Core.* Victoria: Trafford Publishing.

Mitchell, B. (2010). Cluster network computing architecture. Retrieved from http://compnetworking.about.com/od/networkdesign/l/aa041600a.htm

Staab. A., Hotho, S. (2005). Ontology-based text clustering. Retrieved from http://www.loria.fr/~skaf/pmwiki/uploads/Main/ontosurvey.pdf

Skrentha (2004, April 4). The secret source of Google's power [Web log message]. Retrieved from http://blog.topix.com/archives/000016.html