

Virtual Machine Knowledge Discovery and Representation: An Ontology-based Development

Jehad Sabri Alomari
ahal8@yahoo.com

Annette Lerine Steenkamp
steenkamp@ltu.edu

Lawrence Technological University
Southfield, MI 48075-1058 USA

Abstract

A virtual machine environment is dynamic, collaborative and continually evolving to enhance performance, availability, and resource utilization. A virtual client can be migrated to different hosts, and a different host can be migrated to a different cluster to provide high performance and resource allocation. This dynamic environment requires knowledge discovery of operational intelligence that supports critical processes such as backup, recovery and software deployment. The discovery of dynamic changes is a critical process that can be enhanced by an ontology-based approach to provide a high service level, improve the decision making process, and enhance competitive advantage. This paper reports on a knowledge-based conceptual model which utilizes ontology for a virtual environment to support the backup and recovery process. The prototype illustrates the potential support that may be obtained by utilizing Knowledge Management with an ontology-based solution. It offers a fast, efficient, and consistent representation and retrieval of information to increase the knowledge of the virtual environment's dynamic changes.

Keywords: Knowledge Engineering, Knowledge Discovery and Mapping, Knowledge Management, Ontology Development, Virtual Infrastructure.

1. INTRODUCTION

Virtualization is an abstraction of software that provides the capability of collaboration connection between hardware and the operating system (OS). Utilizing virtualizations, multiple machines with heterogenous operating systems and applications can run in isolation and share hardware resources together concurrently. The separation of hardware from the applications and the OS can provide many benefits, providing a scalable computing model which enhances the infrastructure of an organization to meet business challenges.

Since the 1960s, virtualization technology has been gaining momentum in the research

community by adopting several methodologies, such as simulation, emulation, partitioning, and so on. There are two types of virtual techniques for Virtual Machine Monitoring (VMM), illustrated by Gelberg (1973) namely:

- VMM and Virtual Machine (VM) run directly on the hardware; an example of this type is the open source QUEM (Bellard, 2005).
- VMM and VM on a host OS; an example of this type is VMWare (VMware, Inc).

A knowledge mapping method can be used to extract knowledge and identify relationships between concepts in the virtual

environment. Proactive KM for a virtual environment requires dynamic knowledge-intensive discovery of resources to understand the environment and make better decisions. The primary purpose of this paper is to propose a knowledge-based conceptual model of a virtual environment using an ontology, which enables one to gain a sound understanding of the complexities of software deployment. This proposal also considers the potential support that may be provided by utilizing Knowledge Management (KM), as well as the enablement of ontology-based solutions.

Section 2 outlines the importance and benefits of utilizing an ontology-based approach to manage a virtual environment. The research design for this project is focused on developing an enabling KM ontology-based discovery and mapping solution to support the business requirement of the software deployment process. Section 3 focuses on the discovery and representation of data and information to create knowledge. Section 4 presents a high level conceptual solution of an ontology-based discovery and mapping infrastructure architecture to support discovery of knowledge of the business and technical requirements for decision making and deployment. Also Section 5 illustrates the conceptual model for the ontology development. Section 6 presents related work to Knowledge KM and ontology development addressing business requirement of software deployment. Conclusions drawn based on this ongoing research project are presented in Section 6 along with some plans for the remainder part of the project.

2. BENEFETS OF KNOWLEDGE DISCOVERY IN VIRTUAL ENVIRONMENT

Knowledge of virtual environment resources can improve the decision-making process and enhance the organization's competitive advantage. The flood of available information cannot be interpreted by enabling information retrieval techniques alone or by assuming that people understand their information needs (Belkin, 2000). Even if a search for knowledge were successful to solve a problem, it requires

some knowledge to understand its potential relevance (Fischer et al., 1992). An ontology-based system supported by KM will offer many benefits, including:

- Supporting the understanding and management of the virtual enterprise, resource balancing, consolidation, and so on.
- Supporting information infrastructure management and knowledge discovery to provide dynamic business solutions.
- Supporting knowledge related to business requirements such as performance, availability, ease of deployment, and disaster recovery.

Adopting a KM ontology-based approach for a virtual environment based on discovery and mapping would provide many benefits to organizations, especially to increase the return on investment in organization and aid in overcoming business challenges.

3. MANAGEMENT OF KNOWLEDGE DISCOVERY AND REPRESENTATION

KM processes continually interact in a perpetual knowledge enrichment cycle in order to create, preserve, and utilize intellectual assets (Steenkamp and Konda, 2003). Knowledge creation, preservation and application are strategic processes, and regarded as long term success factors of many organizations. The essential purpose of the knowledge creation process is to build the enterprise knowledge repository (EKR). As such knowledge creation must be understood, comprehensively studied, and treated as a process in the context of its domain. The process of knowledge creation focuses on knowledge conversion, as considered by (Nonaka, al, 1995). Knowledge creation as a conversion process describes knowledge flow within the four forms of knowledge (socialization, externalization, combination, and internalization), and the interplay of tacit and explicit knowledge. Steenkamp and Konda (2003) created a knowledge creation life cycle which focuses on building the Enterprise Knowledge Repository (EKR) based on enterprise domain analysis, then building the knowledge asset of EKR based on business value, representing an enterprise standard for all knowledge

enabled projects. The adoption of the ontology life cycle can proactively accomplish the integration across SDLC processes and knowledge creation life cycle. Previous studies of knowledge creation have focused on the interchange of tacit and explicit knowledge, organization learns from the experience of another (Inter-Organizational knowledge transfer), and firms and internal knowledge flow in an industry. However, the lack of empirical studies in the area of knowledge creation has limited our understanding of knowledge creation to certain level. This study is focused on understanding the a process entirely by identifying relationships between environmental and organizational concepts.

Adopting ontology to automate the process of discovery and mapping can utilize knowledge creation life cycle to improve the SDLC processes.

The discovery, mapping, and availability of virtual environment knowledge from this surfeit of information can offer a tremendous amount of value in the knowledge conversion process and increase the level of understanding of the process entirely. The definition of KM presented as ontology in Appendix 1 is given from the perspective of the Software Engineering professional. Data, information, knowledge, and insight are representations of the level of understanding. Data is obtained from facts that have no meaning or context. There are two types of data namely quantitative and qualitative, and they are always explicitly presented. Information is data that has a context, relationship, and meaning. Knowledge is information (a "Thing" (being)) that is transformed to become more valuable or more effective as a basis for action. Insight (wisdom) is the highest level of understanding and is mostly tacit knowledge that provides unique principles and solutions. These are domain assets that create knowledge. Organizations build their assets from a process perspective and therefore process and knowledge has life cycles that can improve the quality and value of knowledge.

4. KNOWLEDGE DISCOVERY AND REPRESENTATION ARCHITECTURE

We propose a knowledge discovery, retrieval, and mapping architecture as illustrated in Figure 1. It is focused on deriving data from the virtual environment via the Management and Retrieval Applications, such as Scripts Application and VM Management Application. The discovered data (reports) are uploaded to the ontology using the Executer. The Executer is software which uploads the retrieved data according to the mapping of the knowledge-based Ontology.

The Virtual Layer provides the capability to transform hardware resources (Memory, CPU, Network, Hard disk) to a VM. Each VM has its own OS and business application and operates just like a physical machine. The VMM is responsible for dynamic allocation of hardware resources.

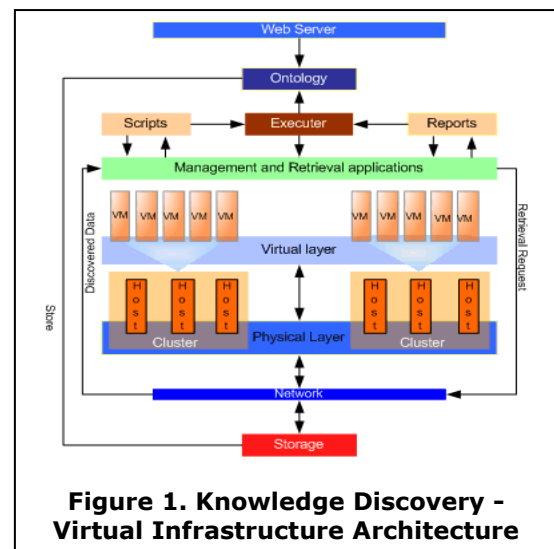


Figure 1. Knowledge Discovery - Virtual Infrastructure Architecture

The Physical Layer consists of hardware resources required to run the VMs. The Cluster represents physical machines (Hosts) that are coupled together to work as a single computer. Sterling defined a cluster as a "computing system comprising a set of independent computers and a network interconnecting them" (Sterling, 2001). Cluster computing provides high availability and load balancing by moving VMs to a different physical host in the case of host failure or host over-utilization to achieve high performance (VMWare, Inc).

The Network is responsible for packet delivery. The virtual network and configuration VMs are the same as the physical machine with Virtual Ethernet Cards, Virtual Switches, IP addresses, MAC Addresses, and so on.

Storage data can be stored locally (on hosts) or on a Storage Area Network (SAN). However, the migrations of VMs among hosts are easier with SAN.

5. CONCEPTUAL MODELS FOR ONTOLOGY REPRESENTATION

The conceptual model is focused on knowledge discovery in a virtual environment to support the decision making process, as well as other daily activities such as system performance, resource utilization and software deployment. These activities are perpetually dependent on knowledge discovery. This class model is the result of a detailed analysis of an Application Service Provider (ASP) organization with two Data Centers. The virtual environment of this case study is the second type of VMM. The conceptual model illustrated in Figure 2 is a class diagram incorporating the following classes: Client, Location, Host, Cluster, Backup, Type, Status, and Platform. The following is description for each class:

- The Client Class is the main building block of knowledge discovery, which connects other classes to support the knowledge discovery process. Each instance (individual) of the Client Class has a unique identity to represent specific client information. The Identity must be unique to the ontology. The Client Class has a recursive relationship to indicate the role of client to purchase other clients. Furthermore, the Client Class has inverse relations for dynamic mapping of data between other classes.
- The Host Class represents a physical computer. One host can be dedicated to -many clients. The Host Class will have an inverse hostedClients relationship with runningOn, providing dynamic reporting of where each client is hosted.
- The Cluster Class represents a group of linked hosts so that they form a single

computer. A host is a partOf one cluster and a cluster can have one-to-many Hosts.

- The Backup Class represents the daily backup process of every virtual client. It is the most important activity of ASP. Therefore, this proposed ontology tracks the daily backup status of every VM. A Client must have one daily backup. Our on going research will include more details about the backup process in future. However, we were able to present the backup status for each virtual client as illustrates in Table 1. Applying inverse relationship for hasBackup object property provides a grouping mechanism for each backup status (passed or failed)

- The Location Class has a locatedAt property relation with a Host and a Client class. Host's location indicates the physical location of a computer, which implies the location of the VMs that run on that host. Clients are physically located in the United States or Canada; however the location of the virtual environment resources may be at either location in the United States (AA RV).

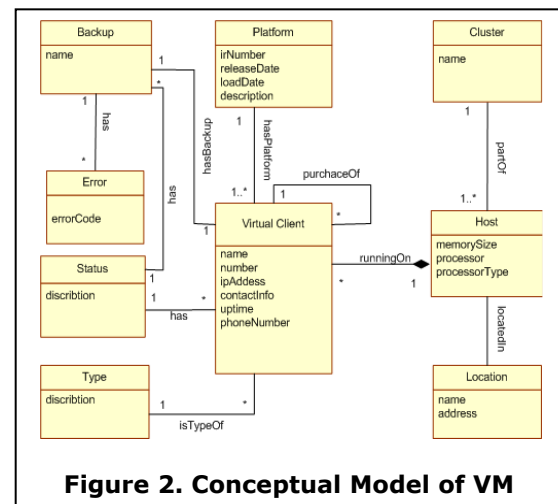


Figure 2. Conceptual Model of VM

Python code has been developed to upload the retrieved data from scripts and reports. Each class is represented by a Python file, which can be executed from the Ontology Editor Script Tape (ProtegeScriptTap).

Appendix 2 displays a visual example of all the hosts on RVCluster04 (rvhost2 and rvhost6) and all the clients *runningOn* on each host from our developed Virtual Environment Ontology. This KM ontology model capable of tracking any changes in the virtual environment such as clusters, hosts, and other resources based on discovery and mapping to support decision making and software deployments.

The ontology defines each concept in the virtual environment domain and the relations. Table 1 is derived information and instances that have been uploaded to the ontology model. This information is valuable to make a quick and knowledgeable business and technical decisions. An example business requirement is "any installation of software on a server must be completed successfully after the daily backup". Therefore, to do a multiple installation it requires some knowledge about the machine, its businesses and technical environments.

Name		Total	Data Center	
			AA	RV
Client		515	199	316
Cluster		7	4	3
Host		26	10	6
Support Location - Canada		3	0	0
Support Location -US		8		
Platform	Windows2003Server	16	7	9
	OtherLinux-32bit	7	2	5
	2.4xLinux-32-bit	490	189	301
	RedHatEnterpriseLinux-32-bit	2	1	1
Server Status	Power On	505	191	314
	Power Off	10	3	7
Backup status	Passed	315	199	316
	Failed	0	0	0
Uptime Status	VM up > 109 days	28	1	27
Memory	Number of hosts with < 6 Gig	1	0	1

Table 1. Findings from the Knowledge based Ontology for Virtual Environment

6. FUTURE WORK

A key goal of this research is to enable the discovery, mapping, and creation of explicit knowledge by developing a semantically interconnected knowledge-based ontology for the software deployment process. Software deployment is a knowledge intensive process involving integrating knowledge from a diverse set of resources. The collaborative knowledge-based ontology is centered on supporting decision-making based on business requirements for the software deployment process. The next step of the research is to enable the discovery, mapping and creation of more detailed knowledge about clients' data storage and tracking software deployment issues.

7. CONCLUSIONS

The knowledge-based discovery, retrieval and representation ontology helps to understand the dynamicity virtual environment and the complexity of many business processes, such as software deployments. Also, it provides a basis for making daily knowledgeable decisions, which helps organizations to obtain innovative solutions thereby advancing their competitive advantage. The business benefits and dynamicity of a virtual environment require the enablement of a dynamic knowledge-based business solution utilizing discovery and representation. The ongoing research is keenly focused on developing an Ontology-based KM enabled system which uses discovery and mapping to improve the decision-making process in terms of quality and speed of the relevant business deployment resources. Knowledge about such relevant resources can be created by understanding the process by enabling an ontology-based discovery, retrieval, and mapping of data and information.

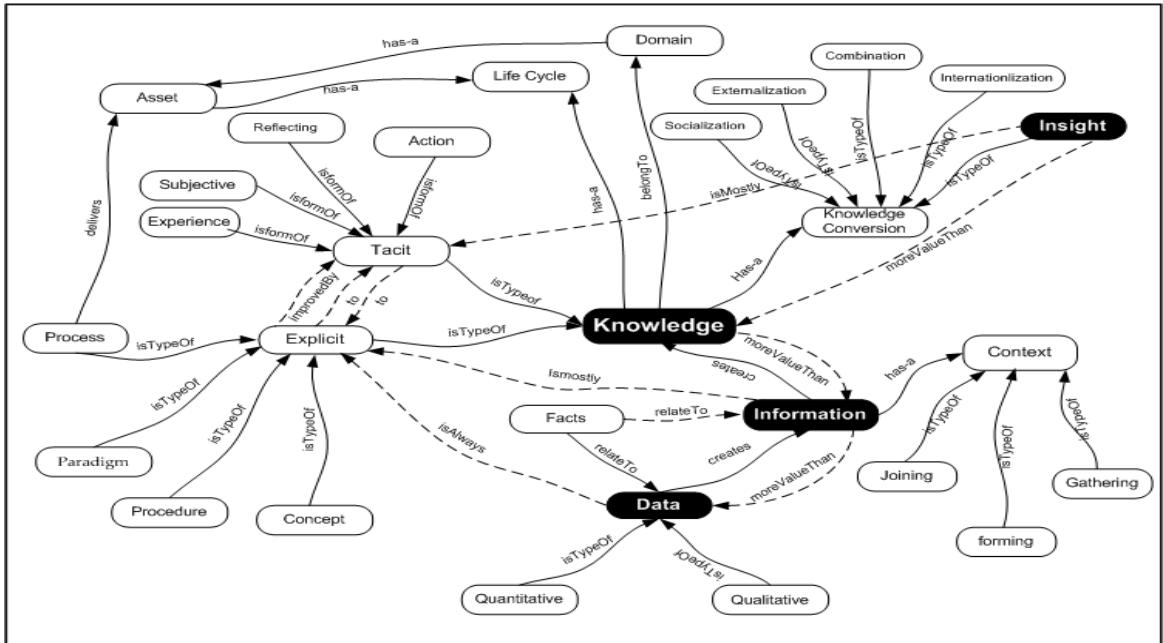
8. ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the sponsors of this study for providing an opportunity to apply Ontology Engineering and KM discovery and mapping theory to practice and collaborating on this educational project.

9. REFERENCES

- Belkin, J.N.(2000) Helping People Find What They Don't Know," Comm. of the ACM, 43(8), pp. 58-61.
- Bellard, F. (2005). Qemu, a fast and portable dynamic translator. In USENIX 2005 Annual Technical Conference, Anaheim, CA, USA.
- Briggs, D. (2000) Maximizing the Knowledge Asset Value within the Enterprise. DM Direct.
<http://www.dmreview.com/dmdirect/2000421/2146-1.html>.
- Das,A., Wu,W. and McGuinness,D. Industrial strength ontology management. In Proceedings of the First Semantic Web Working Symposium. Stanford, USA. 2001.
- Fischer, G., Grudin, J., Lemke, A. C., McCall, R., Ostwald, J., Reeves, B. N., Shipman, F. (1992). Supporting indirect, collaborative design with integrated knowledge-based design environments.
- Fischer, G., Ostwald, J. (2001). Knowledge Management — Problems, Promises, Realities, and Challenges, IEEE Intelligent Systems, January/February, pp. 60-72.
- Goldberg, P. R. (1973). Architecture of virtual machines. In Proceedings of the workshop on virtual computer systems, pages 74-112, Cambridge, Massachusetts, United States. ACM Press.
- Nonaka, I., Takeuchi, H. (1995) The Knowledge Creating Company, Oxford University Press, New York.
- Steenkamp, A.L. and Konda, D. (2003) Information Technology, the Key Enabler for Knowledge Management Information Technology, the Key Enabler for Knowledge Management, a Methodological Approach, International Journal of Knowledge, Culture and Change Management, Vol. 3, Monograph MC03-0070-2003, (Ed.M. Kalantzis and B. Cope),
- Sterling. T. (2001). An Introduction to PC Clusters for High Performance Computing, California Institute of Technology and NASA Jet Propulsion Laboratory, USA.
http://www2.sccc.ru/News/Prezent/final-White_paper.pdf
- VMware, Inc. <http://www.vmware.com>.

APPENDIX 1. ONTOLOGY VIEWPOINT OF KNOWLEDGE MANAGEMENT



APPENDIX 2. EXAMPLE OF CLUSTER AND HOST INSTANCES AND RELATIONSHIPS

