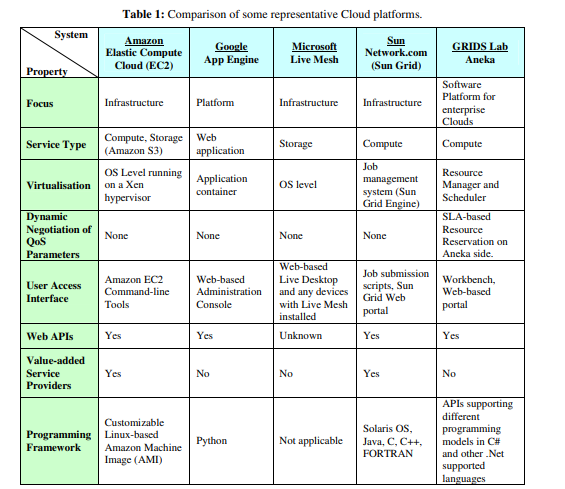
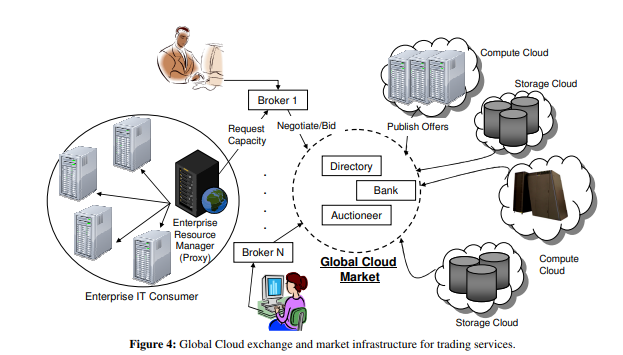
|  |
| --- |
| **UNIT – 4** |
| **Market Based Management of Cloud** |
| **Market Based Management of Clouds**  As consumers rely on Cloud providers to supply all their computing needs, they will require specific Quality of Service (QoS) to be maintained by their providers in order to meet their objectives and sustain their operations. Cloud providers will need to consider and meet different QoS parameters of each individual consumer as negotiated in specific  SLAs. To achieve this, Cloud providers can no longer continue to deploy traditional system-centric resource management architecture that does not provide incentives for them to share their resources and still regard all service requests to be of equal importance. Instead, market- oriented resource management is necessary to regulate the supply and demand of Cloud resources at market equilibrium, provide feedback in terms of economic incentives for both Cloud consumers and providers, and promote QoS-based resource allocation mechanisms that differentiate service requests based on their utility. Figure shows the high-level architecture for supporting market-oriented resource allocation in Data Centers and Clouds.  awd  There are basically four main entities involved:  •Users/Brokers: Users or brokers acting on their behalf submit service requests from anywhere in the world to the Data Center and Cloud to be processed.  •SLA Resource Allocator: The SLA Resource Allocator acts as the interface between the |

|  |
| --- |
| Data Center/Cloud service provider and external users/brokers. It requires the interaction of the following mechanisms to support SLA-oriented resource management:   * Service Request Examiner and Admission Control: When a service request is first submitted, the Service Request Examiner and Admission Control mechanism interprets the submitted request for QoS requirements before determining whether to accept or reject the request. Thus, it ensures that there is no overloading of resources whereby many service requests cannot be fulfilled successfully due to limited resources available. It also needs the latest status information regarding resource availability (from VM Monitor mechanism) and workload processing (from Service Request Monitor mechanism) to make resource allocation decisions effectively. Then, it assigns requests to VMs and determines resource entitlements for allocated VMs. * Pricing: The Pricing mechanism decides how service requests are charged. For instance, requests can be charged based on submission time (peak/off-peak), pricing rates (fixed/changing) or availability of resources (supply/demand). Pricing serves as a basis for managing the supply and demand of computing resources within the Data Center and facilitates prioritizing resource allocations effectively. * Accounting: The Accounting mechanism maintains the actual usage of resources by requests so that the final cost can be computed and charged to the users. In addition, the maintained historical usage information can be utilized by the Service Request Examiner and Admission Control mechanism to improve resource allocation decisions. * VM Monitor: The VM Monitor mechanism keeps track of the availability of VMs and their resource entitlements. * Dispatcher: The Dispatcher mechanism starts the execution of accepted service requests on allocated VMs. * Service Request Monitor: The Service Request Monitor mechanism keeps track of the execution progress of service requests.   •VMs: Multiple VMs can be started and stopped dynamically on a single physical machine to meet accepted service requests, hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests. In addition, multiple VMs can concurrently run applications based on different operating system environments on a single physical machine since every VM is completely isolated from one another on the same physical machine.  •Physical Machines: The Data Center comprises multiple computing servers that provide resources to meet service demands.  Commercial offerings of market-oriented Clouds must be able to:   * support customer-driven service management based on customer profiles and requested service requirements, * define computational risk management tactics to identify, assess, and manage risks involved in the execution of applications with regards to service requirements and customer needs, * derive appropriate market-based resource management strategies that encompass both * customer-driven service management and computational risk management to sustain SLA-oriented resource allocation, |

+

* incorporate autonomic resource management models that effectively self-manage changes in service requirements to satisfy both new service demands and existing service obligations, and
* leverage VM technology to dynamically assign resource shares according to service requirements.





References:

R. Buyya, C. S. Yeo and S. Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities," 2008 10th IEEE International Conference on High Performance Computing and Communications, 2008, pp. 5-13, doi: 10.1109/HPCC.2008.172.