SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Re-accredited by NAAC with A+ grade, Accredited by NBA(CSE, IT, ECE, EEE & Mechanical)
Approvedy by AICTE, New Delhi, Recognized by UGC, Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER APPLICATIONS

COURSE

23CAE717
Cloud Computing

UNIT III

Cloud Infrastructure

TOPIC

Resource Provisioning

Semester

II Semester / I MCA

❑ A global technology company offer software services to the multiple business domain has a private cloud. They require reliability engineering services for supporting in the following areas

- Cloud automation

- Infrastructure support

- Application monitoring

- Increased productivity, efficiency etc..
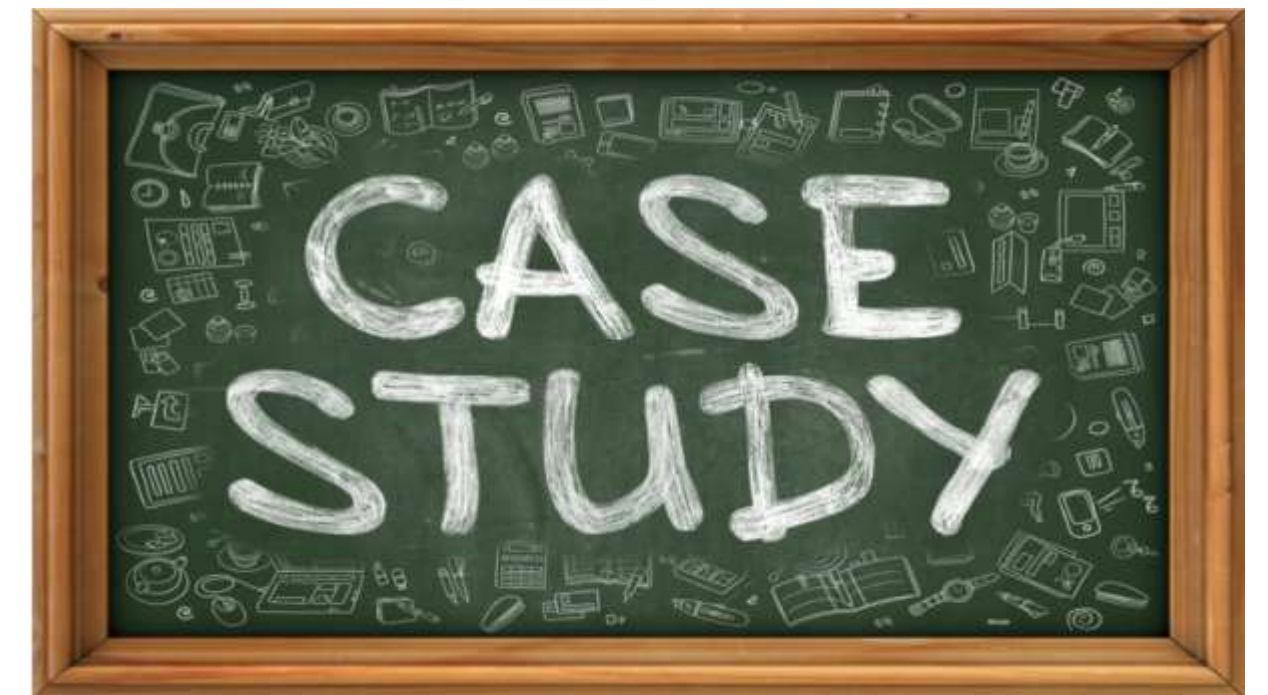
Resource Provisioning

❑ Cloud creates illusion of pool of resources

❑ An activity of Allocation of cloud provider's resources to a customer

❑ When a cloud provider accepts Client's request, it must create number of VMs and allocate resources to support them

❑ It takes Service Level Agreement (SLA) into consideration for providing service to the cloud users

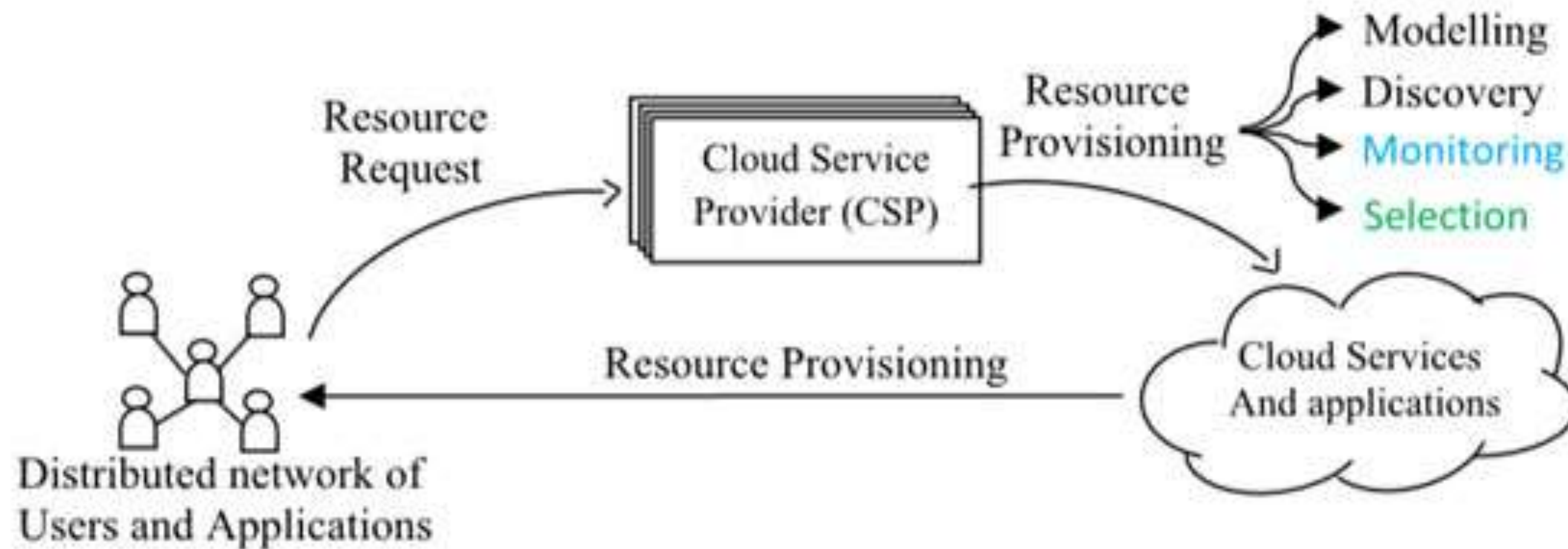❑ Decision on VM size (fixed/dynamic) and VM Placement

Resource Provisioning

❑ Resource provisioning means the

- Selection

- deployment, and

- run-time management of S/W & H/W resources for ensuring guaranteed performance for applications

❑ Resource provisioning phases

- Reservation phase – reserve resources

- Expending phase - utilize resources

- On-demand phase–provision of more resources

Fig 1: Resource Provisioning System in Cloud Computing (RPS)

Based on the Application, it may be

❑ Static Provisioning –unchanging demands

❑ Dynamic Provisioning - demands may change or vary

❑ Self Provisioning – customer purchase computing resource from provider

Resource provisioning techniques are used to improve the following

QoS parameters

- ❑ Response time

- ❑ Minimizing cost

- ❑ Maximizing Revenue

- ❑ Fault tolerant

- ❑ Reduced SLA violation

- ❑ Reduced power consumption

Demand driven

❑ Add/remove instances based on current utilization level of the allocated resources

❑ when a resource has surpassed a threshold for a certain amount of time, the scheme increases that resource based on demand.

❑ When a resource is below a threshold for a certain amount of time

Event driven

❑   Add/remove instances based on specific time event

❑   Anticipates peak traffic before it happens

Popularity driven

❑   Internet searches create popularity demand

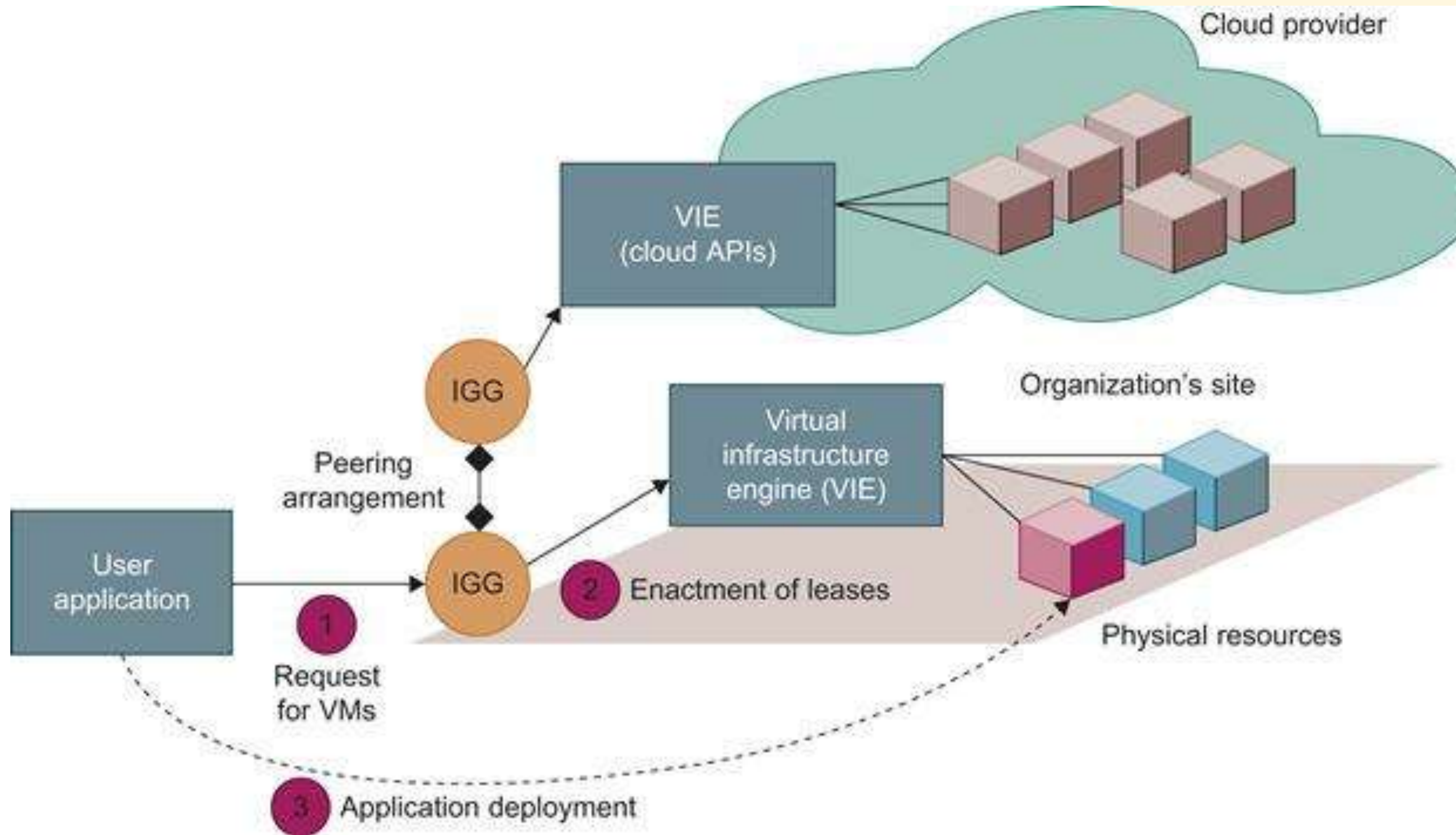❑   scheme anticipates increased traffic with popularity

Dynamic Resource Deployment

❑ To achieve scalability in performance

❑ InterGrid-managed infrastructure was developed by a Melbourne University

❑ A Java- software allows to create execution cloud environments on top of all participating grid resources

❑ Intergrid gateway (IGG) allocates resources from a local cluster to deploy applications in three steps:

    (1) Requesting the VMs

    (2) enacting the VMs

    (3) deploying the VMs as requested

Dynamic Resource Deployment

Cloud Architecture /23CAE717- Cloud Computing & Virtualization Tech./ Dr.N.Nandhini/ MCA/ SNSCT

Dynamic Resource Deployment

- ❑ IGG interacts with another IGG that can allocate resources from provider under peak demand

- ❑ A grid has predefined peering arrangements with other grids, which the IGG manages.

- ❑ Through multiple IGGs, the system coordinates the use of InterGrid resources.

- ❑ An IGG is aware of the peering terms with other grids, selects suitable grids that can provide the required resources, and replies to requests from other IGGs

Dynamic Resource Deployment

❑ InterGrid allocates and provides a distributed virtual environment (DVE). This is a virtual cluster of VMs that runs isolated from other virtual clusters.

❑ DVE manager component performs resource allocation and management on behalf of specific user applications

❑ Scheduler, core component for implementing provisioning policies and peering with other gateways

❑ communication component provides an asynchronous message-passing mechanism.

❑ Received messages are handled in parallel by a thread pool

Dynamic Resource Deployment

❑ Distributed file system for storing large-scale data

❑ Another form of data storage is (Key, Value) pairs

❑ GFS (Google File System): implemented to meet rapidly growing demands of Google's data processing needs

❑ HDFS(Hadoop Distributed File System): implemented for the purpose of running MapReduce applications

❑ Amazon S3 and EBS: S3 used to store/retrieve data from remote servers whereas EBS built on top of S3 for using virtual disks in running EC2 instances

❑ Traditional database system

❑ Provisioning for structured and semi structured data management

Storage Resource Provisioning

❑ Data storage layer is built on top of the physical or virtual servers

❑ A distributed file system is ideal for storing large-scale data

❑ Databases are built with stored data files or in the form of (Key, Value) pairs,

❑ Examples are

➢ GFS: Google FileSystem

➢ HDFS: Hadoop Distributed File System

➢ Amazon S3 and EBS

Storage Resource Provisioning

❑ Cloud databases stores data in structural or semi-structural ways.

  ▪ BigTable from Google

  ▪ SimpleDB from Amazon

  ▪ SQL service from Microsoft Azure

❑ Scaling of such a database might be quite large for processing huge amounts of data

❑ Cloud creates illusion of pool of resources

❑ An activity of Allocation of cloud provider's resources to a customer

❑ When a cloud provider accepts Client's request, it must create number of VMs and allocate resources to support them

❑ It takes Service Level Agreement (SLA) into consideration for providing service to the cloud users