



擁抱大數據與人工智能

新世代雲服務平台的挑戰與 建置案例

Patrick Fu

2018/01/26



Agenda

- 雙子星 公司簡介
- 智能科學家 的工作範疇
- 新世代 雲服務管理平台
- 管理功能介紹
- 總結

關於雙子星雲端



- 雲端軟體原廠

來自工研院雲端中心，至今7年雲端系統軟體開發資歷團隊，為國內唯一100%自主開發OpenStack IaaS+PaaS雲端系統Turnkey Solution軟體原廠

- 厚實投資團隊

工研院、創新工業(工研院創投)、國際知名創投(漢鼎、萬豐)、中鋼集團、行政院國發基金



工業技術研究院
Industrial Technology
Research Institute



ITIC
Industrial Technology
Investment Corporation



CSC GROUP

中鋼集團



行政院國家發展基金管理會
National Development Fund, Executive Yuan



- 大型成功案例

私有雲—工研院工程雲(Matlab Cloud), 台南教網大數據雲
公有雲—是方電訊應用雲(Chief App Cloud)

- 國際市場輸出

日本、歐洲、東南亞，都有潛在客戶與合作夥伴洽談中

Gemini Open Cloud Inc.



雙子星 GOC – 以簡馭繁，直上雲端

一套完整的解決方案 (HW+IaaS+PaaS+SaaS)

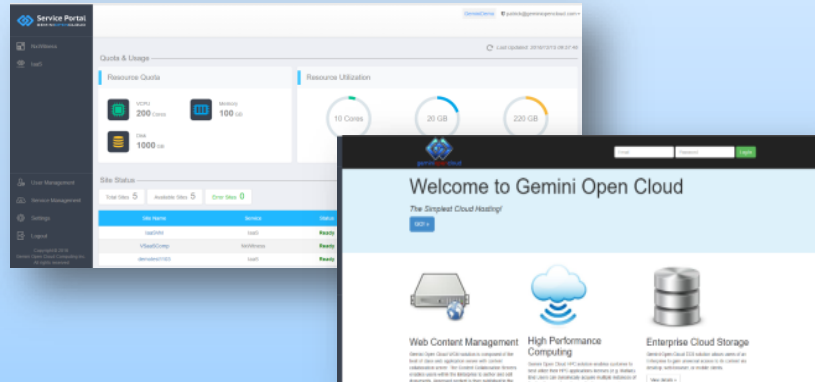
IaaS

Mainstream Open Source



PaaS

Cloud services life management



SaaS

any time, any where, any device

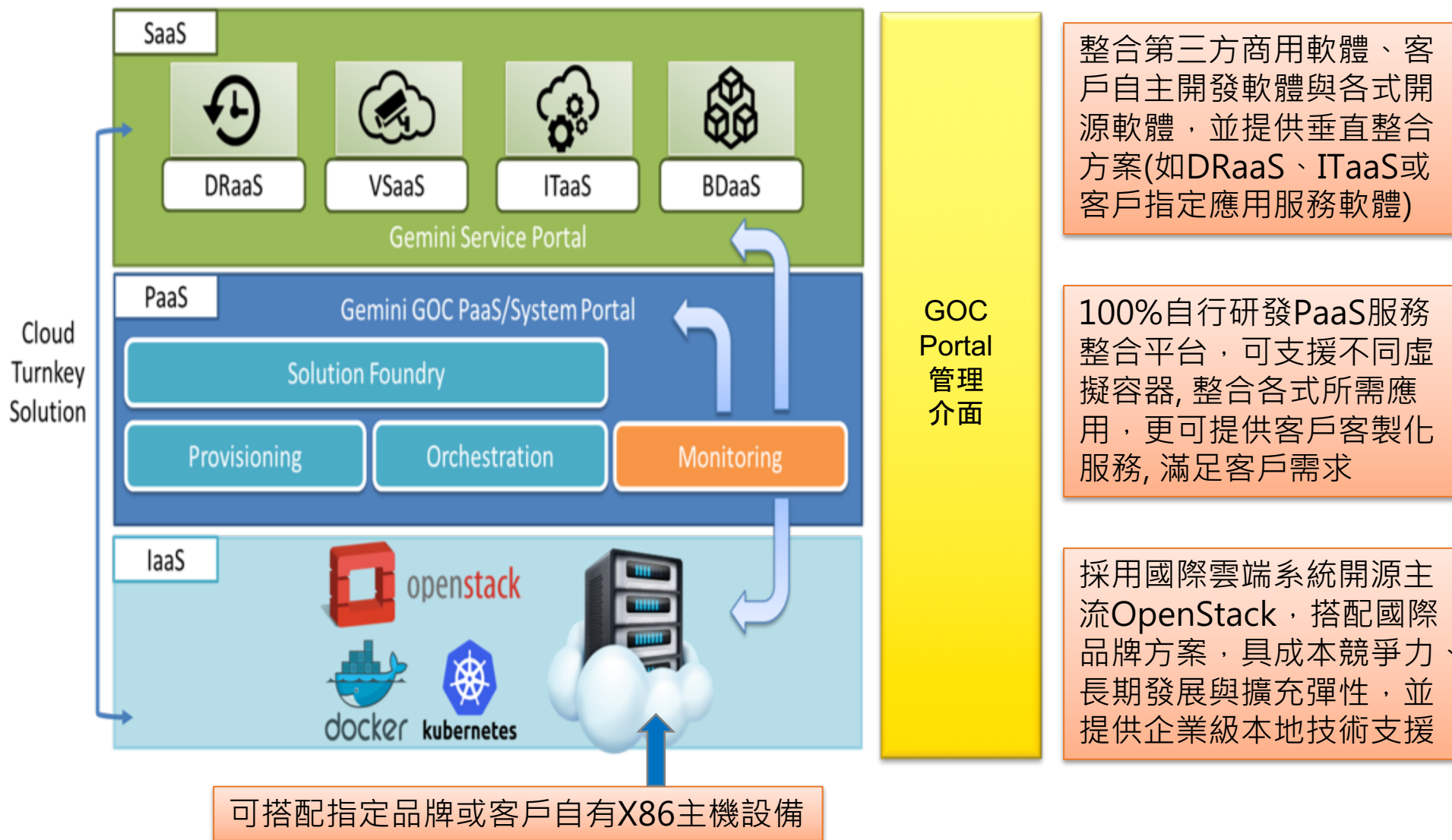


Gemini Open Cloud Inc.

HW

- Rack
- Hyper-converged

雙子星雲端解決方案之架構

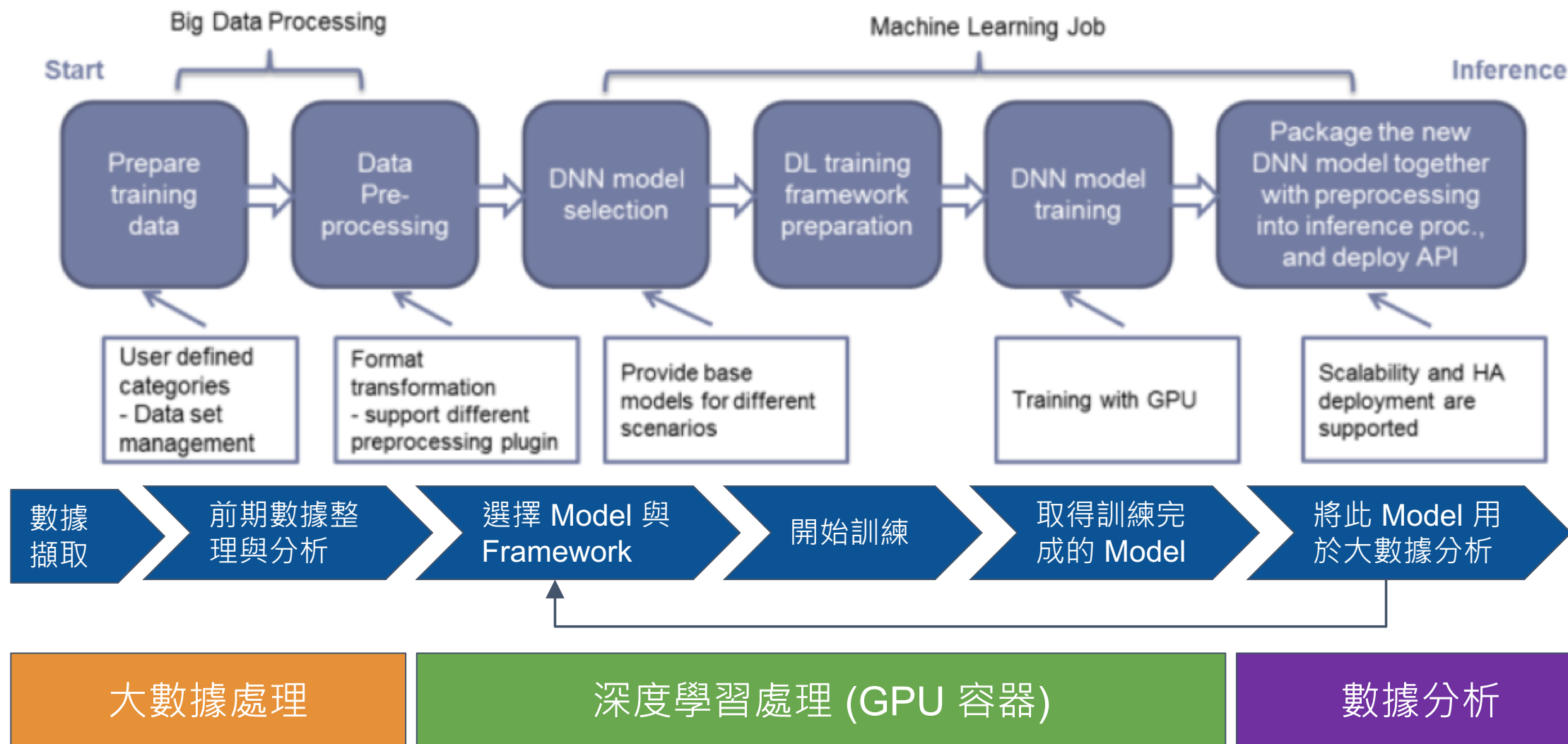




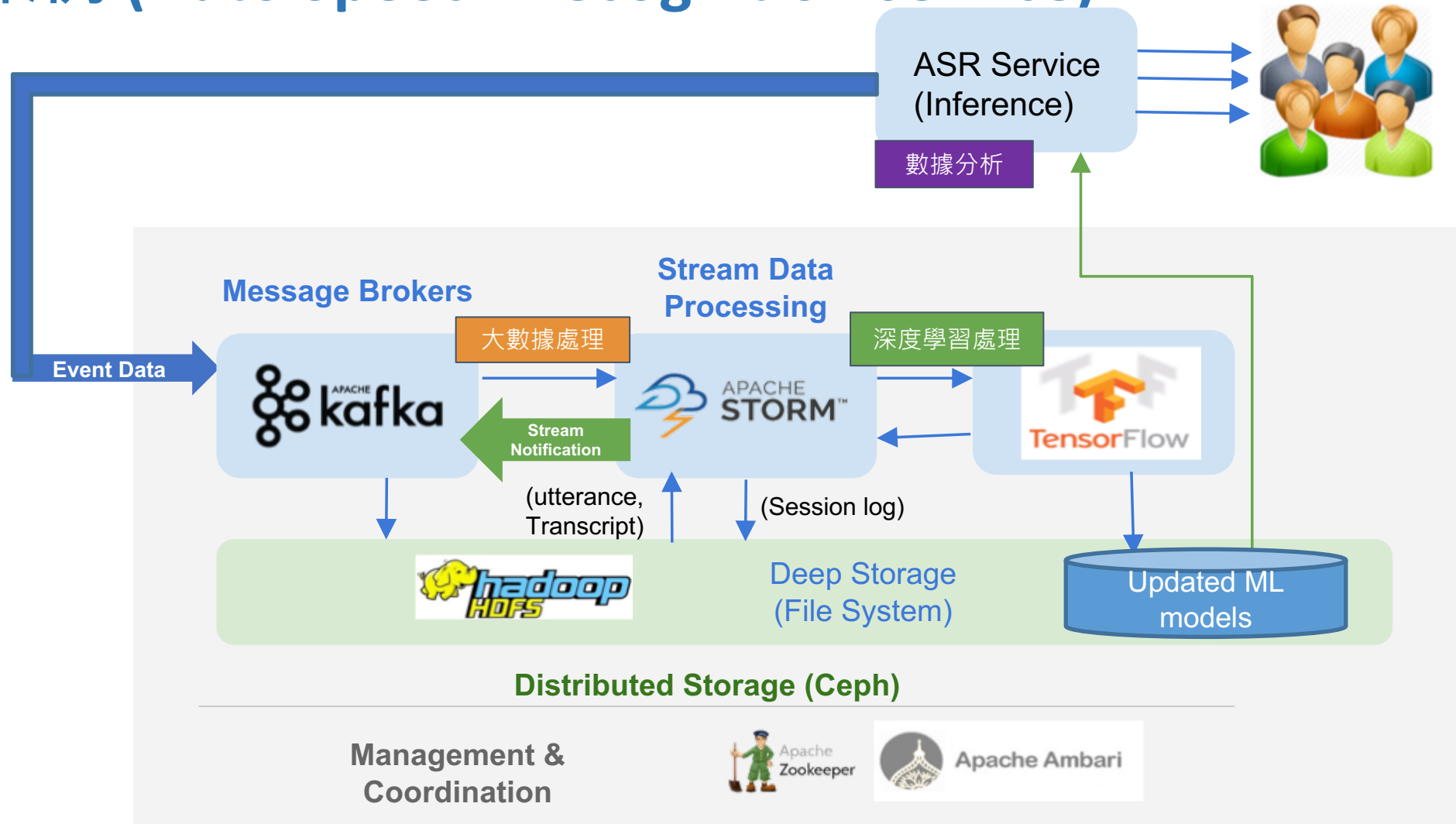
Next Generation Cloud Service Platform



智能科學家的工作範疇



應用案例 (Auto Speech Recognition service)



人工智慧運算平台的挑戰



- 硬體 + 架構成本
 - CPU, Storage, GPU
- 系統延伸性 (Scalable clusters)
 - 實驗數據 與 實際 機器學習數據叢集差異
- 龐大複雜 軟體工具組合
 - Hadoop – Data Extraction, Map Reduce, Scheduler, Streaming data,
 - AI – models, Machine Learning IDE, Back Propagation tools, Model Reduction
 - Data Analytics Visualization
- 多元性數據處理
 - Text, Multi-media, Streaming Video/Audio
 - 跨叢集/平台間數據分享
- 異質平台 之 兼容性

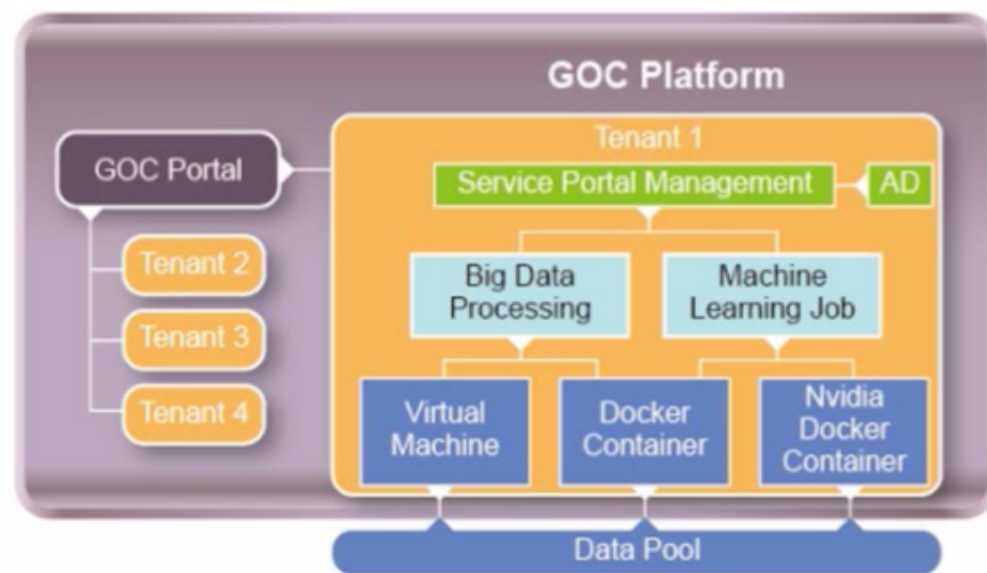
Gemini Open Cloud Inc.



多元性之雲端服務平台



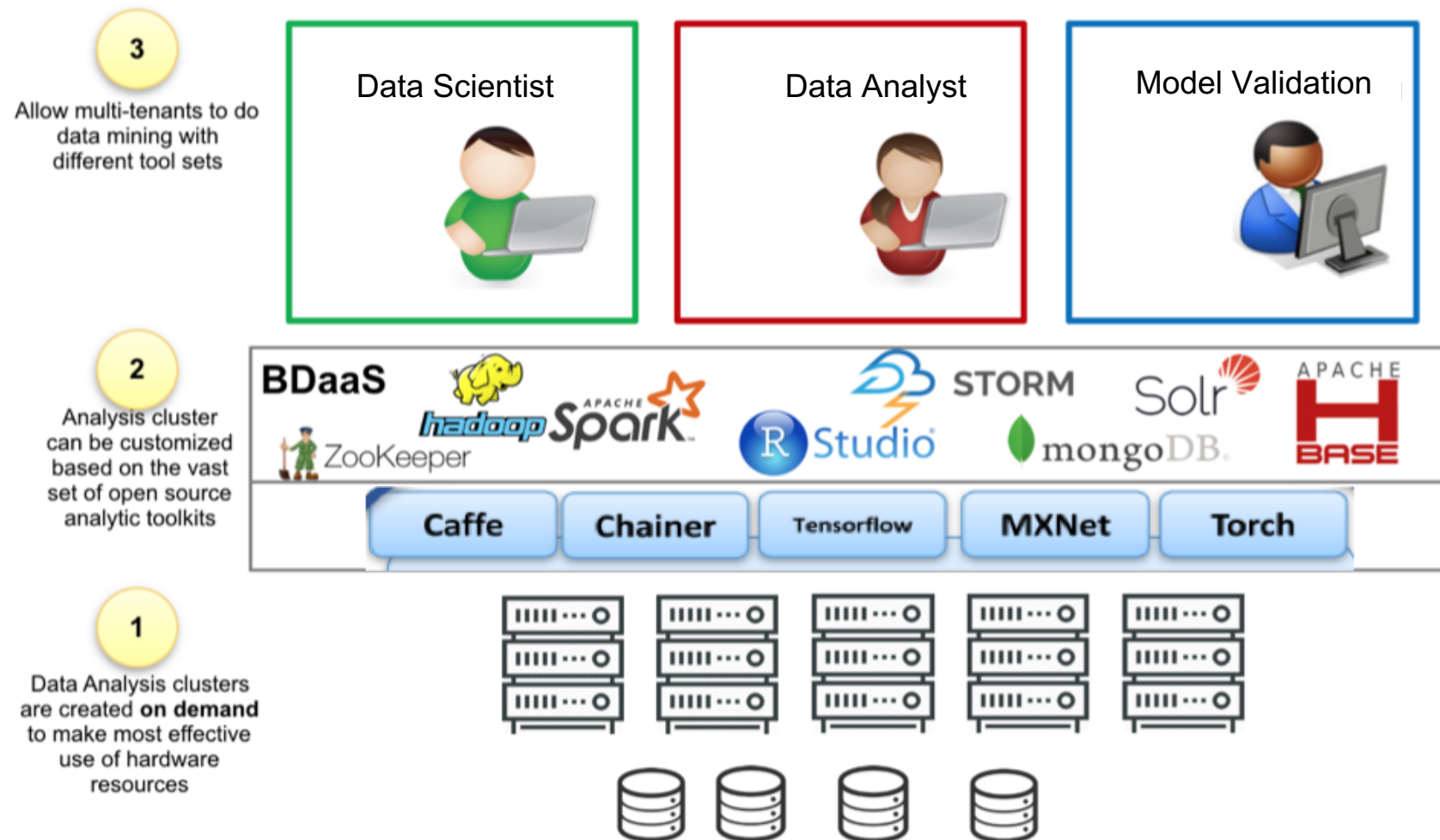
- 每個雲端服務均有自己的 Service Portal
- 使用者可以自我開啟自己的運算叢集 (Big Data, DNN Machine Learning, etc)
- 分享數據資源池 (Share Data pool)
- 系統自動建置 異質資源之雲端服務 (VM's, Dockers, and Physical servers)
- 雲服務生命週期管理





Big Data as a Service (BDaaS)

- ✓ Provide the most popular analytic toolkits on one platform
- ✓ Allow Data Scientists to customize their toolkits
- ✓ Self-service provisioning allows Enterprise to do data mining without worry about infrastructure resource
- ✓ More effective sharing of hardware - scale out cluster on demand
- ✓ Provide Sandbox function for testing or demo scenario
- ✓ Always-up-to-date software version



Cloud Service Provider Object Hierarchy



Solutions

☐ Solution
☐ DrupalDocker
☐ Redmine
☐ Git
☐ CCS
☐ MDCS
☐ IVA
☐ ICLM
☐ HadoopJob
☐ HadoopCluste
☐ OpenEMM

Showing 1 to 10 of 21

Tenants

☐ Tenant
☐ CUS_GKB
☐ CUS_GreatLink
☐ CUS_ICLM
☐ CUS_LR
☐ GEMINI
☐ GeminiDemo
☐ geminimdc
☐ GEMINI_DEMC
☐ IONetworks
☐ NavioWeb

Showing 1 to 10 of 12

Users

☐ User
☐ gkb@
☐ jinha
☐ wordj
☐ iclm_
☐ great
☐ nxdei
☐ test@
☐ jc@n
☐ cc@r
☐ tw@v

Showing 1 to 10 of 12

Volumes

☐ Name
☐ WCMcontent
☐ HakanTest
☐ Extension
☐ vol01
☐ vol01restore

Showing 1 to 5 of 5

Images

☐ Name	Status	Public	Create Time
☐ Ubuntu14.04-2015.3.0	ACTIVE	True	2015-12-31T16:00:00Z
☐ CentOS6.6	ACTIVE	True	2015-12-31T16:00:00Z
☐ Drupal	ACTIVE	True	2015-12-31T16:00:00Z
☐ Alfresco	ACTIVE	True	2015-12-31T16:00:00Z
☐ Wordpress	ACTIVE	True	2015-12-31T16:00:00Z
☐ Windows7-2015.3.0	ACTIVE	True	2015-12-31T16:00:00Z
☐ MatlabLicenseWin7	ACTIVE	True	2015-12-31T16:00:00Z
☐ Matlab2014aWin7All50G	ACTIVE	True	2015-12-31T16:00:00Z
☐ Matlab2014aWin7All50GSplashtop	ACTIVE	True	2015-12-31T16:00:00Z
☐ Matlab2014aWin7OnlineInstall50G	ACTIVE	True	2015-12-31T16:00:00Z

Showing 1 to 10 of 66 rows 10 records per page

Roles & Responsibilities



Features & Capacity	system-admin	tenant-admin (Service Provider)	Service Users (Service Portal)
Create tenant/Quota	O	X	X
user-tenant management	O	X	X
user-role management(*2)	O	O	X
enable / disable solution	O	X	X
create / delete tenant network	O	X	X
site Management	X	O	O
VM operation (start, stop, etc.)	X	O	O
VM resource Mgmt (Image, Volume, Snapshots)	O	O	X
generate report	O	O	O
Manage policies (Load Balancing, Auto-scaling, Firewall)	O	O	X
Solution Foundry (Create & Delete Solutions from GSP)	O	O	

雲端服務方案鑄造 (Gemini Solution Foundry)



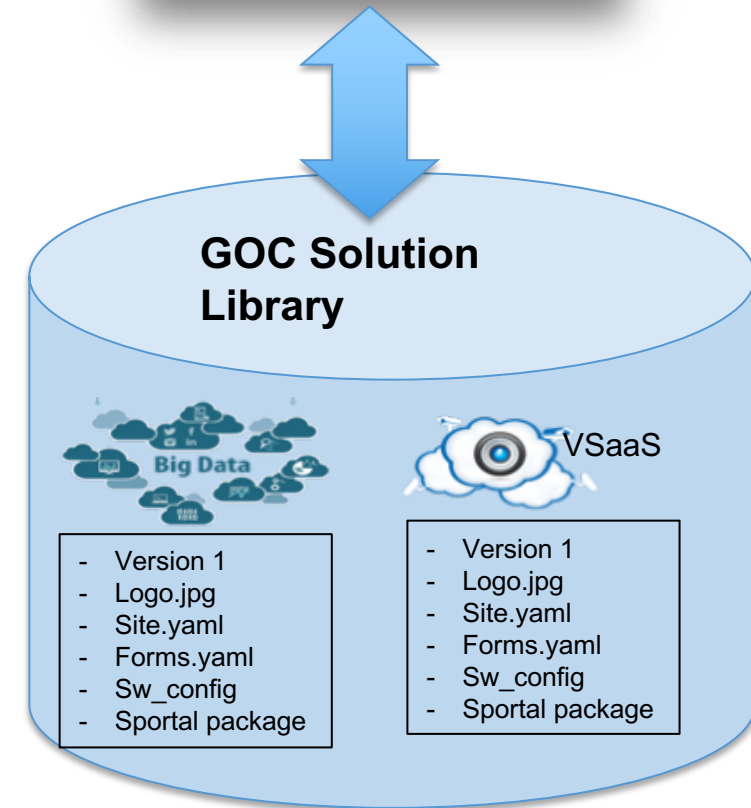
- *Cloud Service Definition can be written in templates under GOC.*

BDaaS GSP (version 2) - Contains Spark & Tensor Flow Apps



A new version of BDaaS solution is being uploaded to GOC. Note a GSP is not a cloud service until it's assigned to a tenant & instantiated.

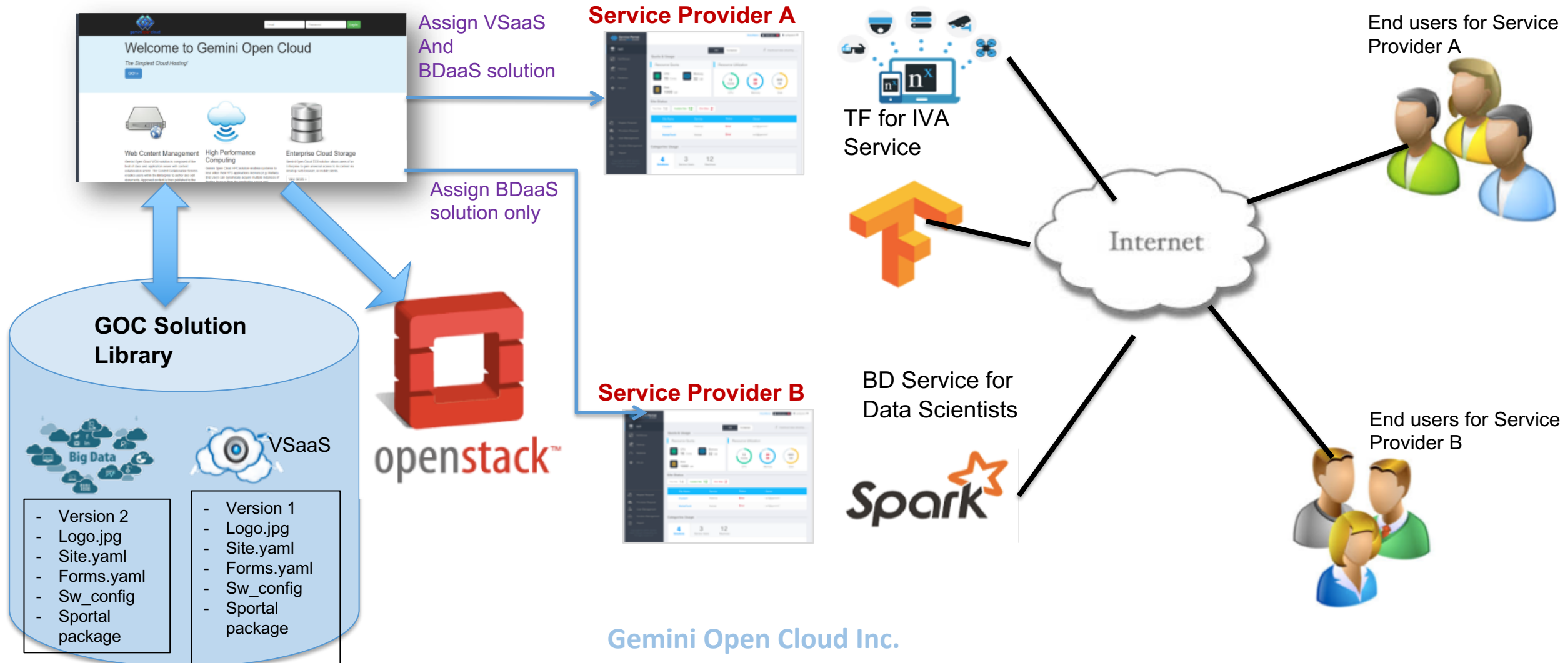
Gemini Open Cloud Inc.



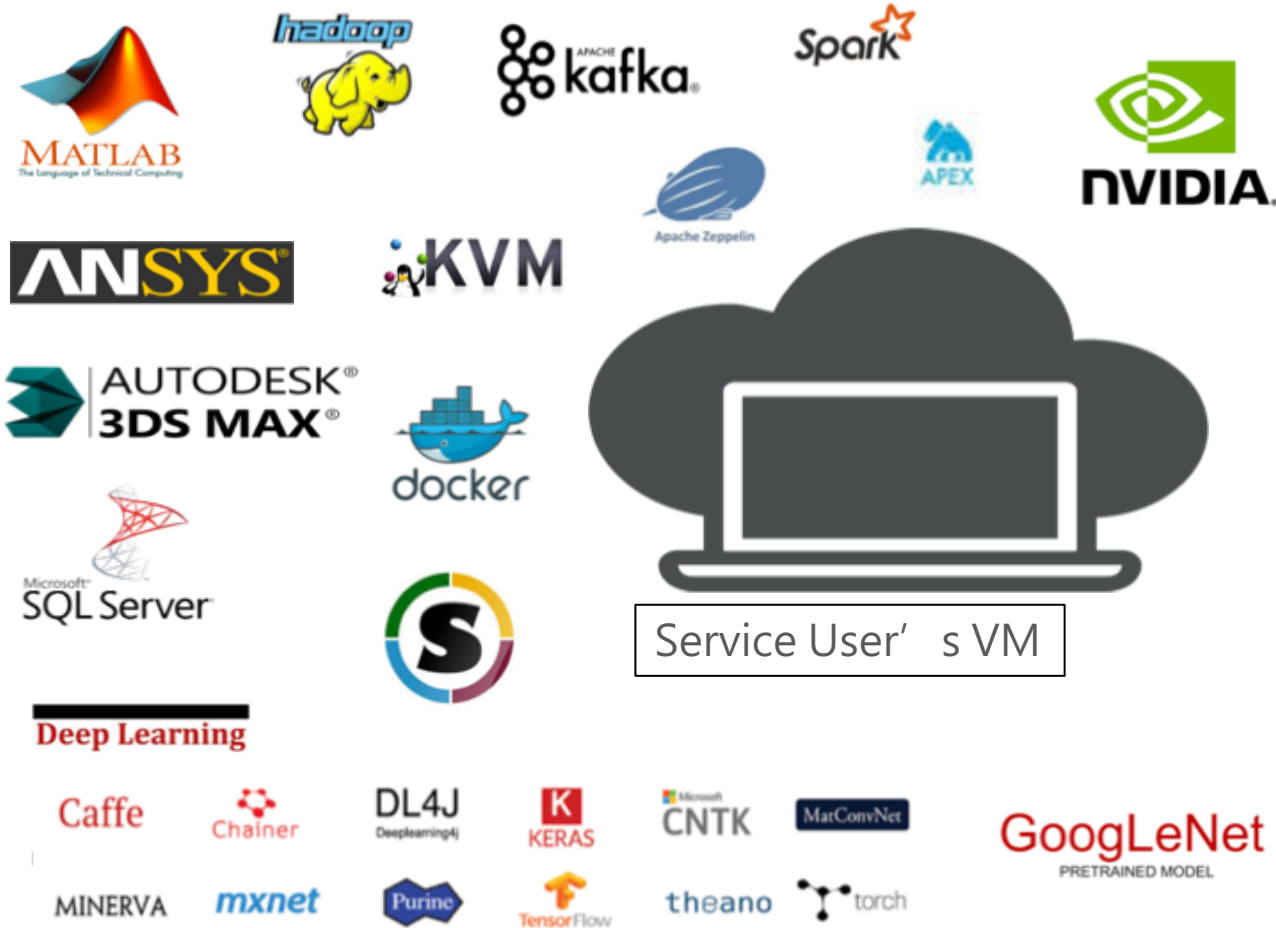
Multi-Tenancy (多租戶服務)



- *GSP solutions can be assigned to multiple tenants to be operated on the same Cloud Instance.*



使用者按需要自建雲端服務



On-demand Cloud Services

Service Usage Configurations

Application Installation Files
Application Silent Installation Scripts
Application License Files



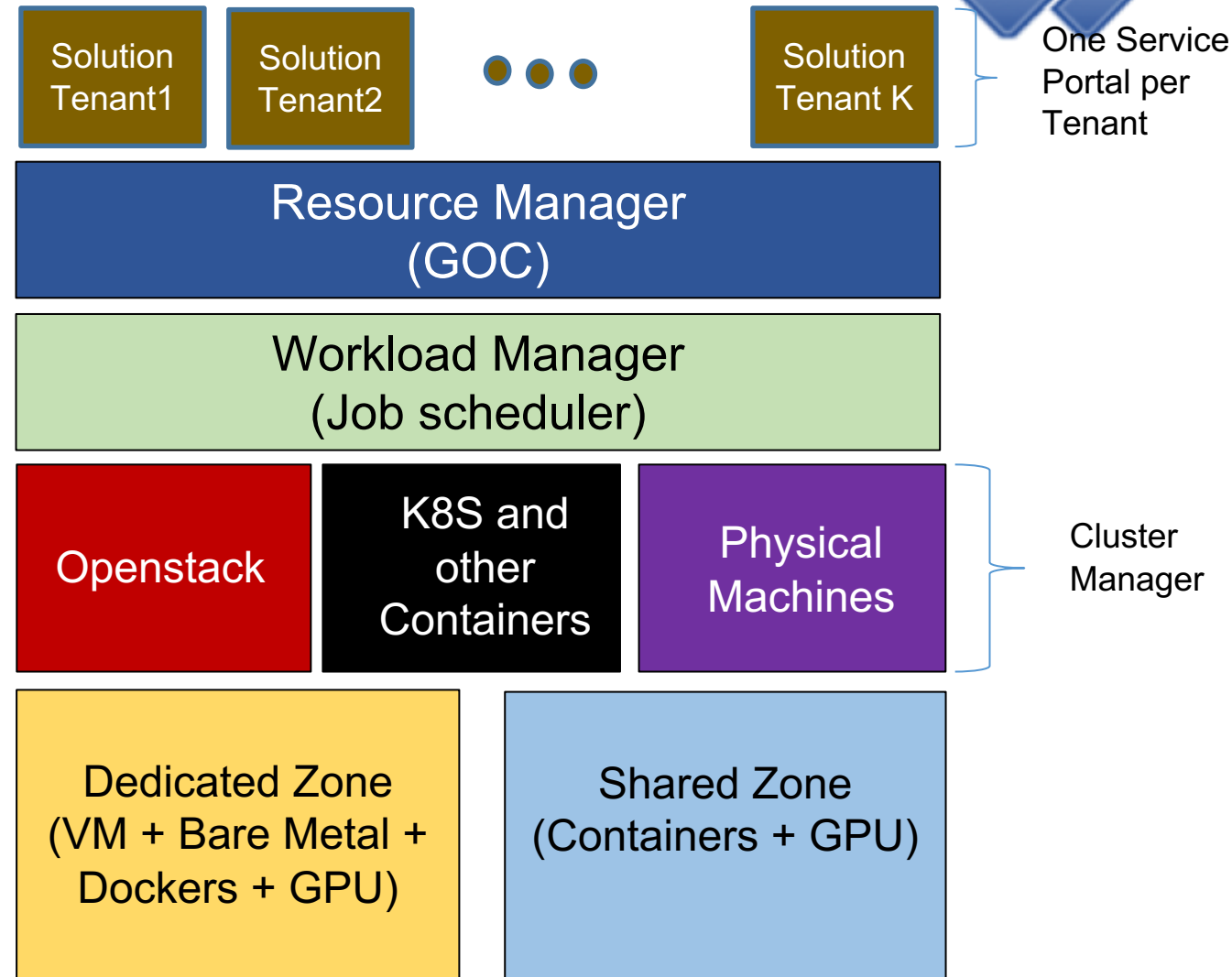
Deployment Management Tools

- A new approach to bring up cloud service with flexibility and manageability
Gemini Open Cloud Inc.

雲服務平台之模組



- Hardware resources are provisioned into zones to accommodate different tenants
- Solutions are uploaded via Solution Foundry and can optionally have Service portal
- Resource Manager manages multiple tenants & determines where the resource is available and dispatches the job
- Workload Manager schedules a job for solutions (e.g. Tensorflow) and optionally specifies the resources it needs (not tenant aware)
- Cluster manages a group of homogeneous resources and track the status of a job
 - Openstack manages VM's & Physical machines provisioned by Ironic
 - K8S manages Dockers
 - Other forms of containers

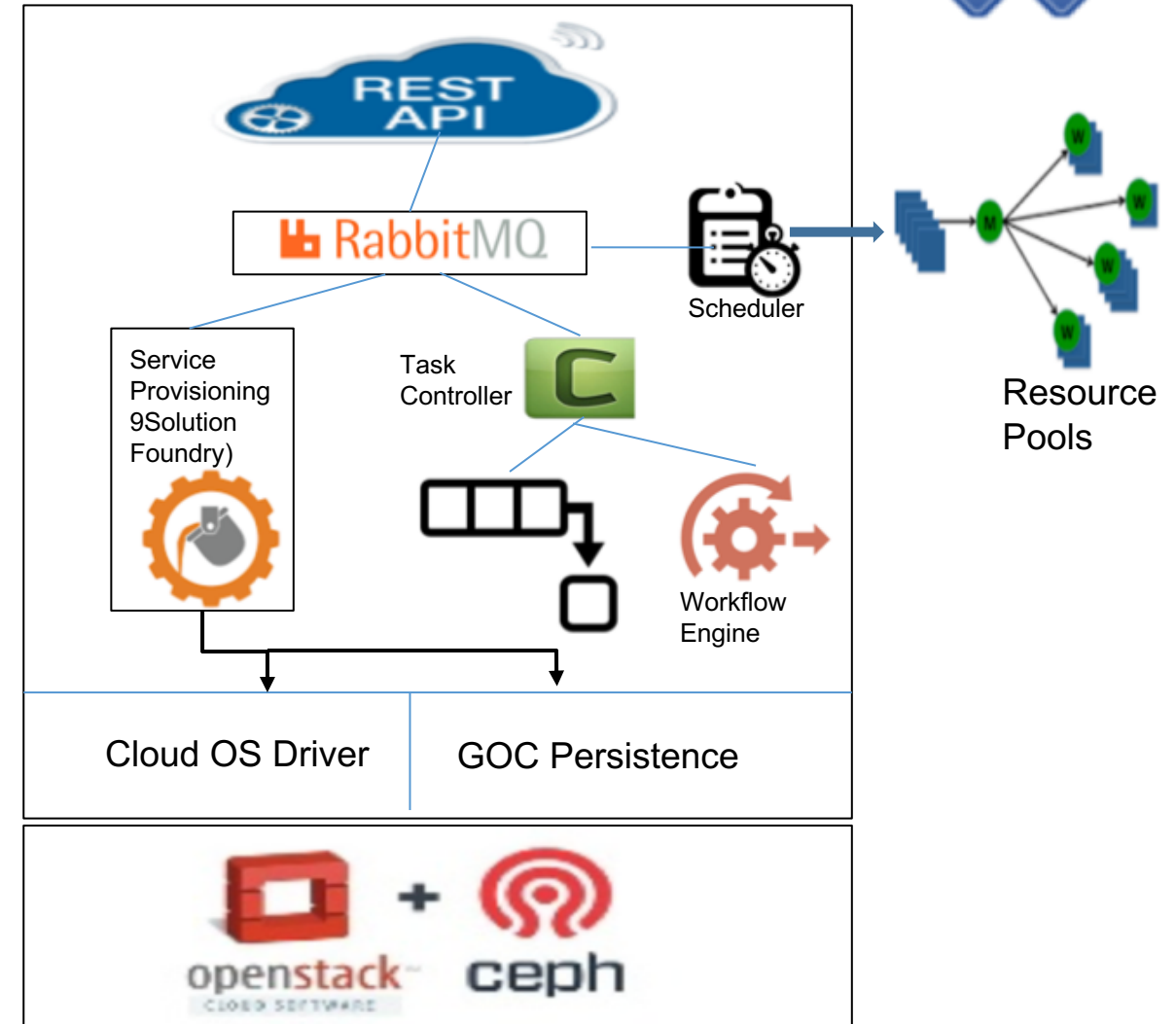


Resource Manager (資源管理)



- Responsible for managing a system of multiple virtual resource types
 - Each virtual resource type can have its own virtual cluster manager
- Features
 - Provisioning Manager (Solution Foundry)
 - API Server
 - Message Queues
 - Queueing Manager
 - Task Controller
 - Workflow
 - Scheduler
 - Quota by Tenant
- Resources can be separated into multiple **resource zones** (typical based on resource types and its tenant owners)

GOC Resource Manager

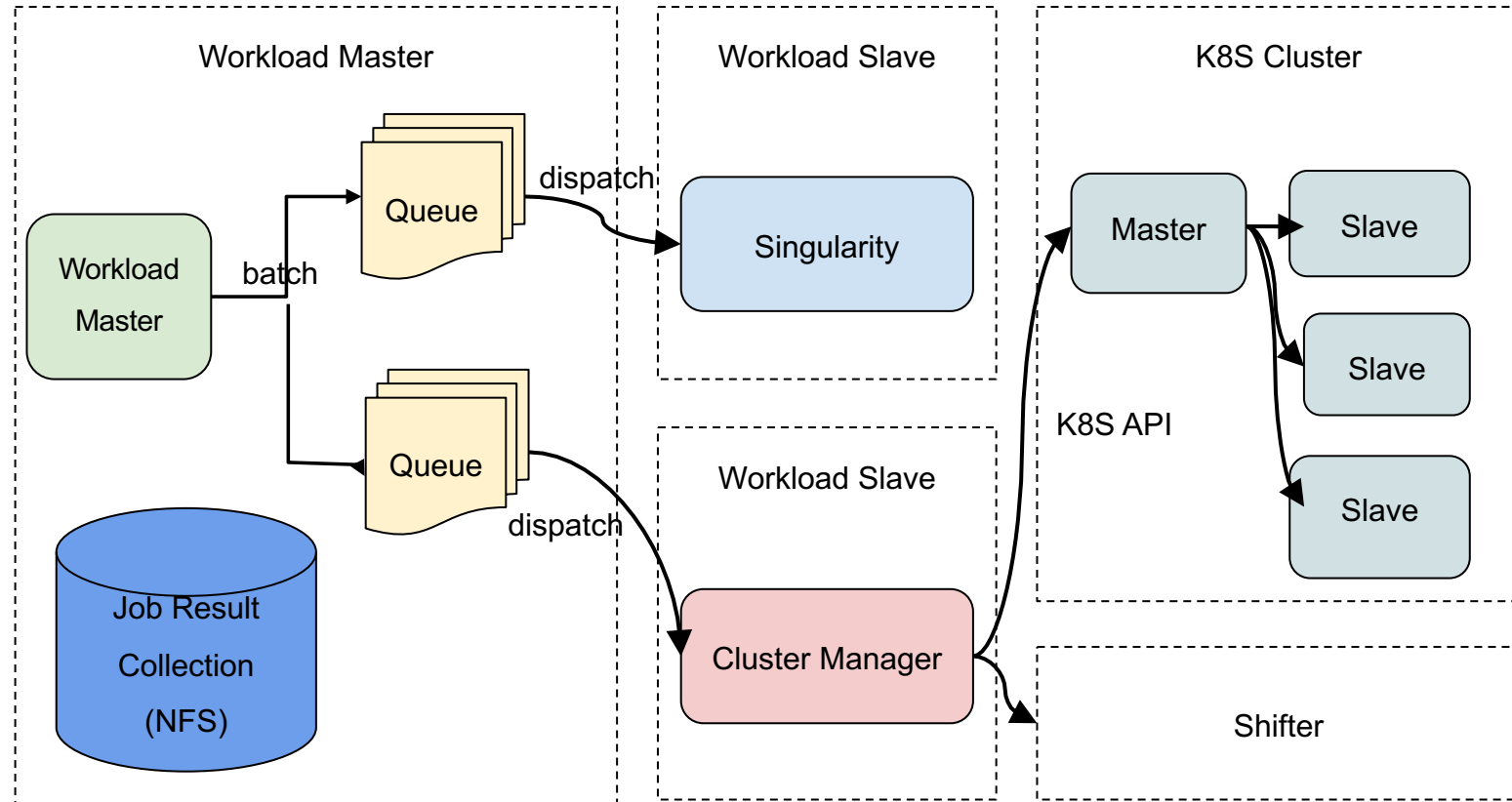




Workload Manager (排程管理)

- Responsible for Queuing & Scheduling jobs to multiple systems
- Independent of the resource manager on each system
- Can have its own portal or CLI

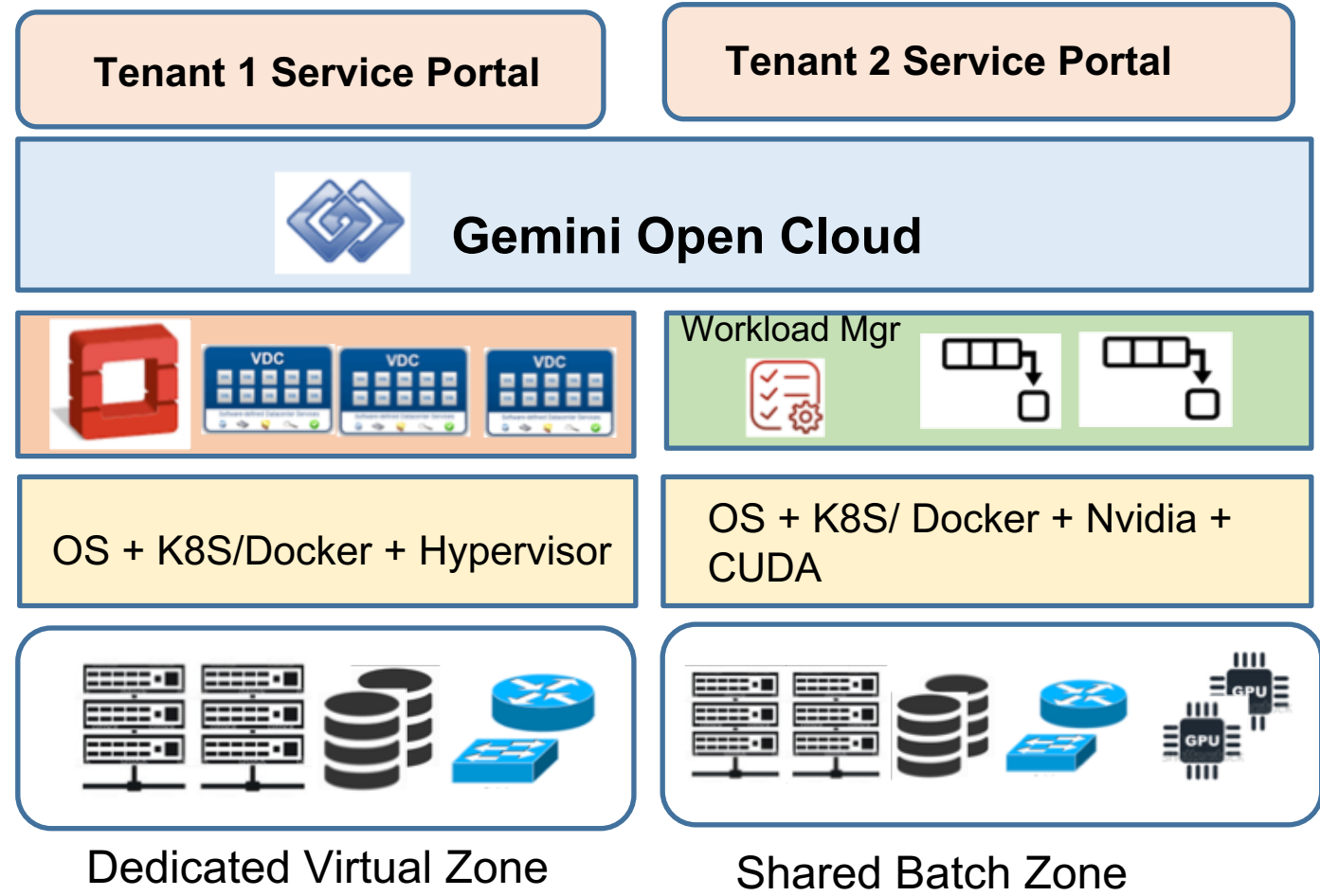
e.g. LSF (Load Sharing Facility), UGE (Univa Grid Engine), Slurm



多層資源池 (專屬區 或 共享區)



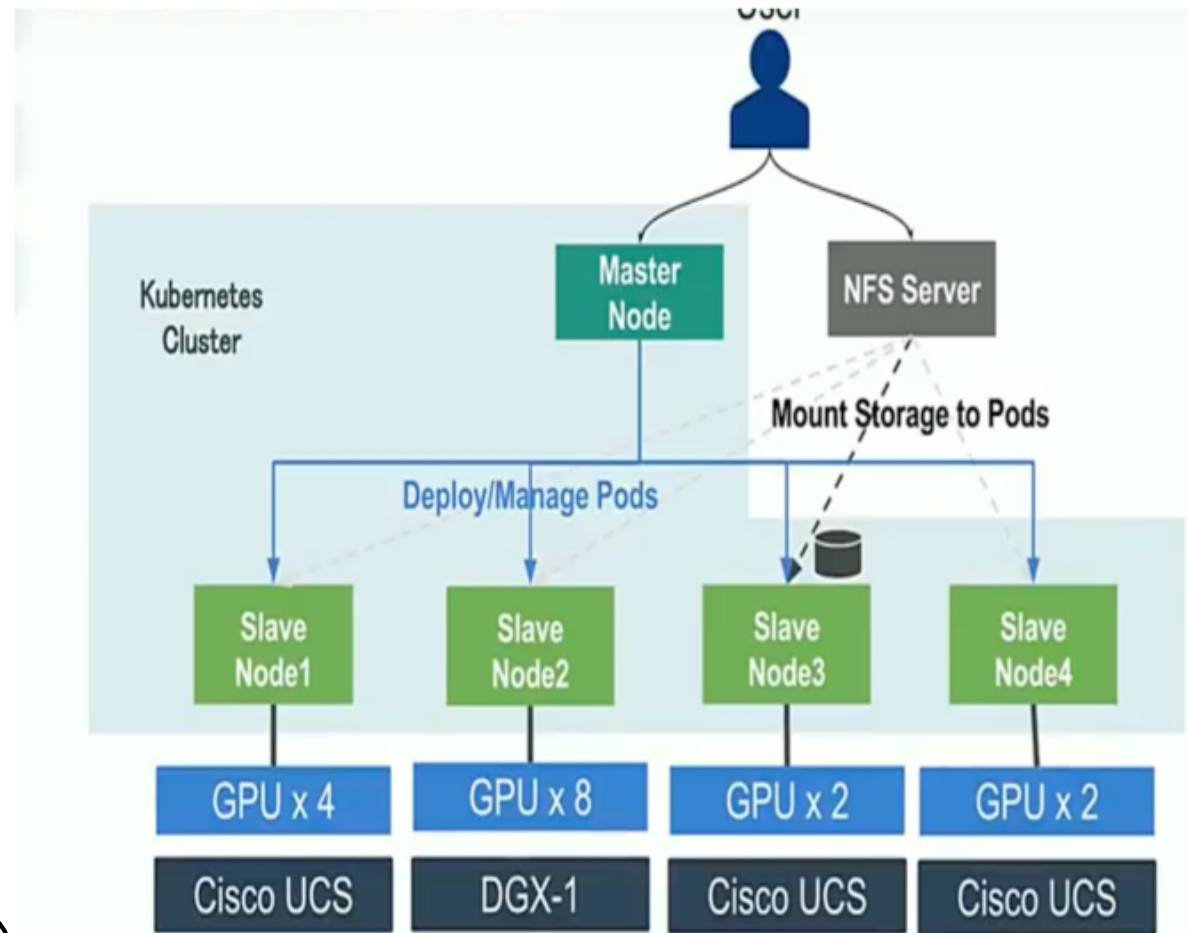
- Virtual Data Center's are 'Rent by Capacity'
- GPU's can be more efficiently utilized in a job queue environment
- GOC workload manager allow users to share a GPU cluster based on 'Rent by Utility'
- Workload manager queues job to Kubernetes (for dockers) & Singularity containers



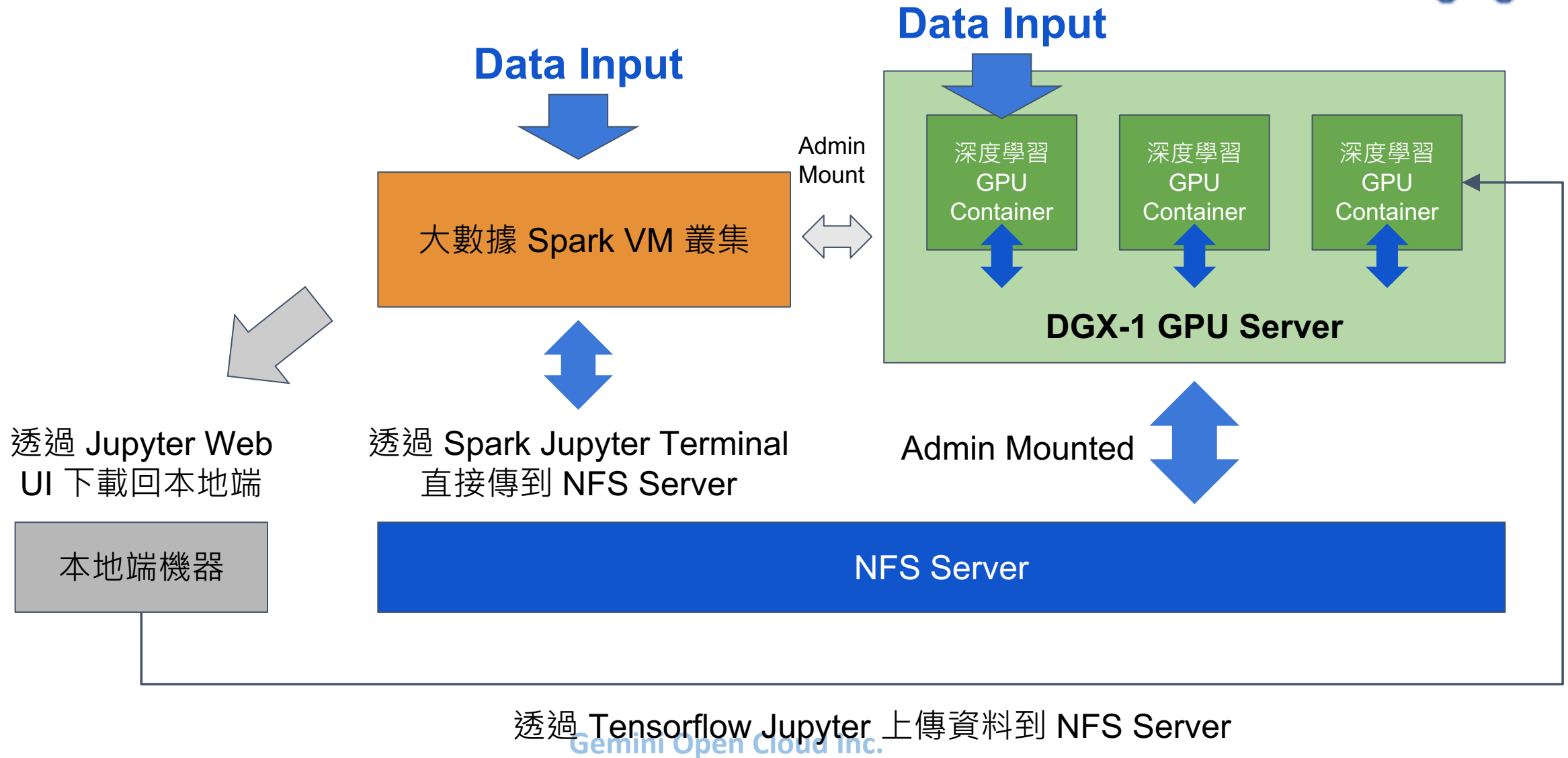
提供 GPU 雲端分享服務



- 用戶能指定用不同種類 的 **GPU 需求**
- 管理者有權力訂定使用者最多可以使用多少資源 (**quota limits**)
- **GPU isolation:**
 - 用戶只看到自己的 GPU 資源
 - 用戶不可同時分享 GPU
- 用戶可提交批次工作
- 批次工作結束後, 資源自動回歸到資源池
- 用戶可監控其 GPU 使用資源狀況

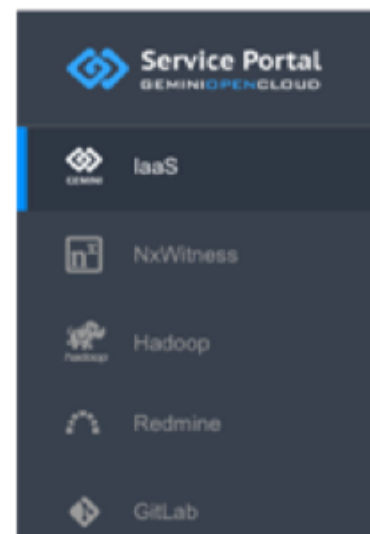


不同運算叢集可分享數據 (Data Sharing)



雲端服務入口(Service Portal)

Item	Description
Self-Registration	使用者自行註冊帳號
Self-Provisioning	使用者自行申請服務，並由管理者核准後自動 Provisioning
Service Life Cycle	自行管理服務的生命週期
Resource Monitoring	監控資源池使用狀態，以及目前使用資源的分布狀態
Solution Management	自訂服務申請表單與顯示資訊
User Management	使用者與權限管理
Service Report	雲端資源與服務使用報表
Service Customized UI	客製雲端服務使用者操作介面
Automatic Notification	雲端服務與 Portal 狀態通知 (含 Portal 通知與 Email)



Cluster Management Interface

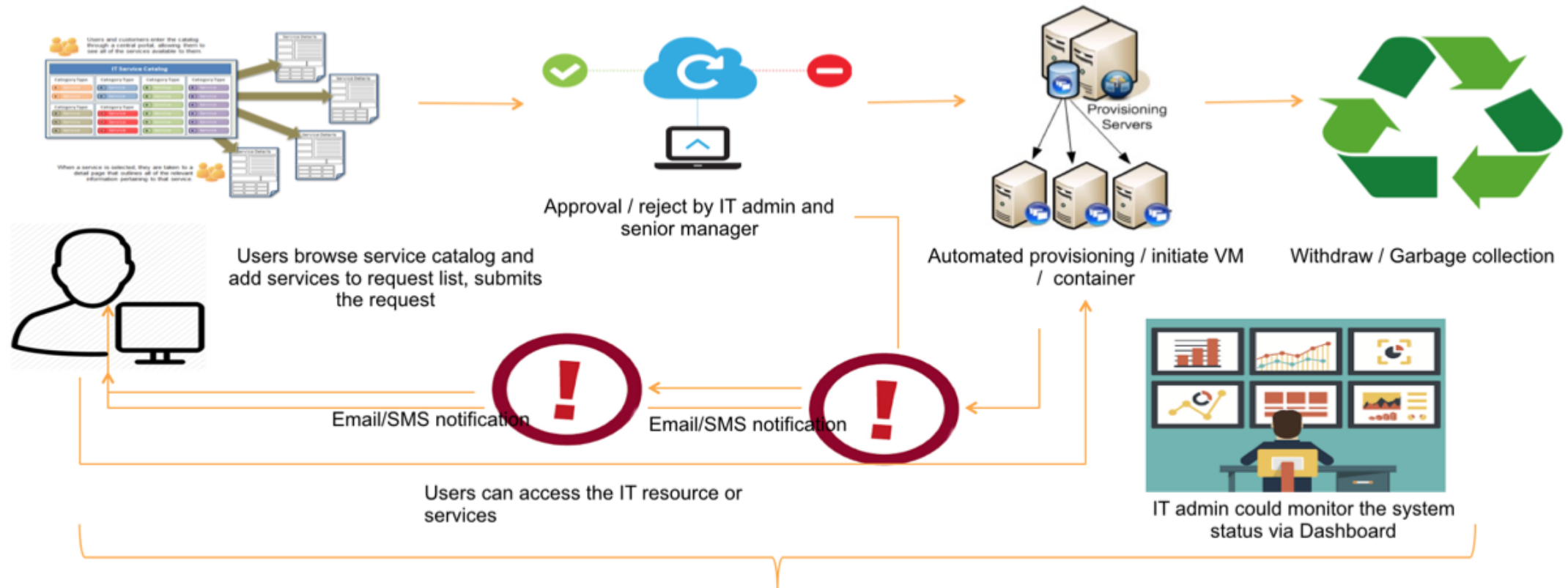
Buttons: Delete Cluster, Start Cluster, Stop Cluster, Search..., + Launch Cluster

Cluster Name	Status	Type	Public IP	Created By	Create Time	Share
service01	Ready	Core Hadoop	10.113.1.66	sp@gemini.com	2017/01/09 18:24:19	Private
service02	Ready	Spark	10.113.1.19	sp@gemini.com	2017/01/09 18:24:19	Private
service03	Ready	MongoDB	10.113.1.66	sp@gemini.com	2017/01/09 18:24:19	Private
service04	Ready	HBase	10.113.1.19	sp@gemini.com	2017/01/09 18:24:19	Private

Showing 1 to 1 of 4 entries

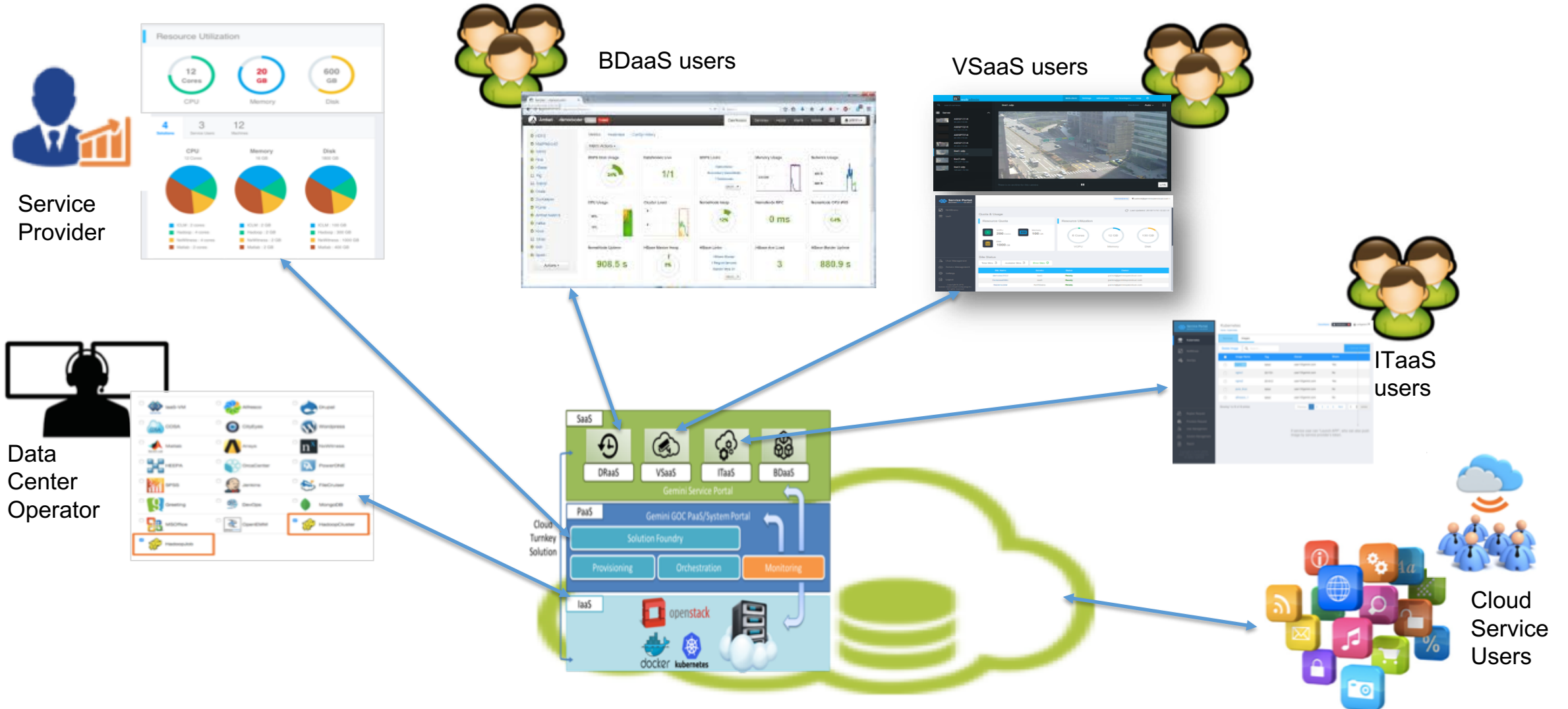
Navigation: Previous, 1, 2, 3, 4, 5, Next, 5 entries

客戶自助服務流程



- GOC self-service portal integrates the predefined services and advanced automation, allowing users to request and receive IT resources without intervention and then monitor, manage, and control those resources

高度客制化界面



The screenshot displays the Service Portal Gemini Open Cloud interface. On the left is a dark sidebar with navigation links: HPC, Kubernetes, Volume Management, Service Request, User Management, Solution Management, and Report. The main content area is titled 'HPC' and contains a 'Create Job' button. Below this is a navigation bar with 'Job', 'Storage', 'Console', and 'Cluster' tabs, where 'Console' is selected. A 'Login Password' field and a 'Connect' button are present. The top right shows user information: 'TenantName', 'Notification 99', and 'sp@usermail.com'. The main content area displays system information as of Fri Nov 10 07:45:36 UTC 2017, including system load, processes, memory usage, and IP address. It also provides links for graphing data and getting cloud support, and mentions available updates and a new release. The bottom of the screen shows a footer with copyright information and a 'Volume' section.

GOC – All for your Cloud Needs



Meet Industrial Standard: **STABILITY + RELIABILITY + HIGH AVAILABILITY**

STABILITY ✓

OpenStack IaaS, the main stream open source Cloud OS

RELIABILITY ✓

Gemini Open Cloud professional service @ 7/24 level

HIGH AVAILABILITY ✓

SLA 99.995%
Full HA Architecture

Unique Advantages: **SIMPLICITY + INTEGRATED SOLUTION + SELF-PROVISIONING**

SIMPLICITY ✓

Turning legacy IT into cloud service easily with simplified UI

INTEGRATED SOLUTION ✓

Integrated Solutions like Hadoop, VSaaS, DR and ITaaS

SELF-PROVISIONING ✓

Service Portal
Solution Foundry
System Portal



Thank you