

Cloud and Big Data Infrastructure

Tuyet-Trinh Vu, Binh-Minh Nguyen, Viet-Trung Tran
Hanoi University of Science and Technology

contact: trinhvt@soict.hust.edu.vn

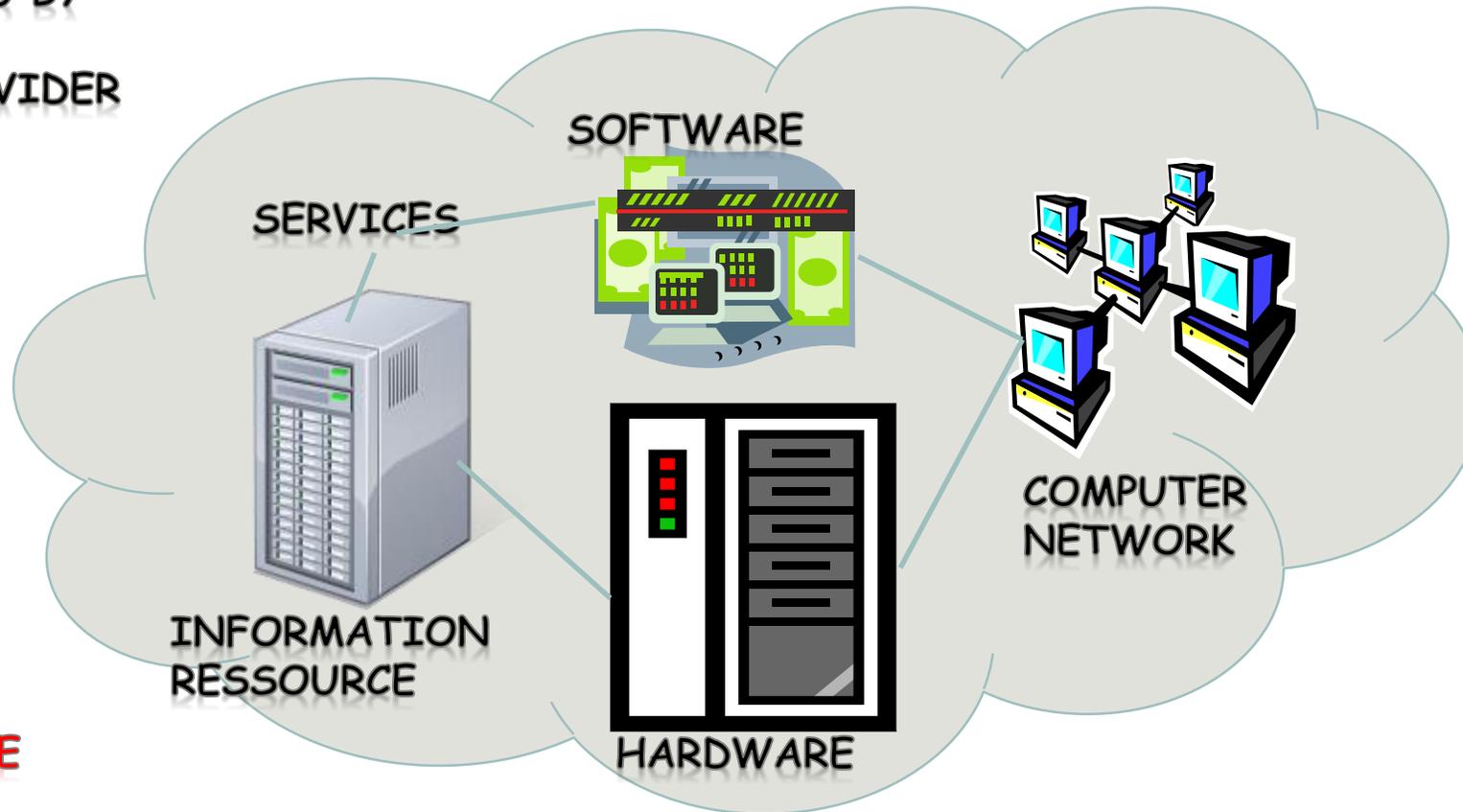


Outline

- Cloud computing
- Cloud service models ... Interoperability
- Infrastructure for big data
- Our approach & work
- Concluding remarks

Cloud computing

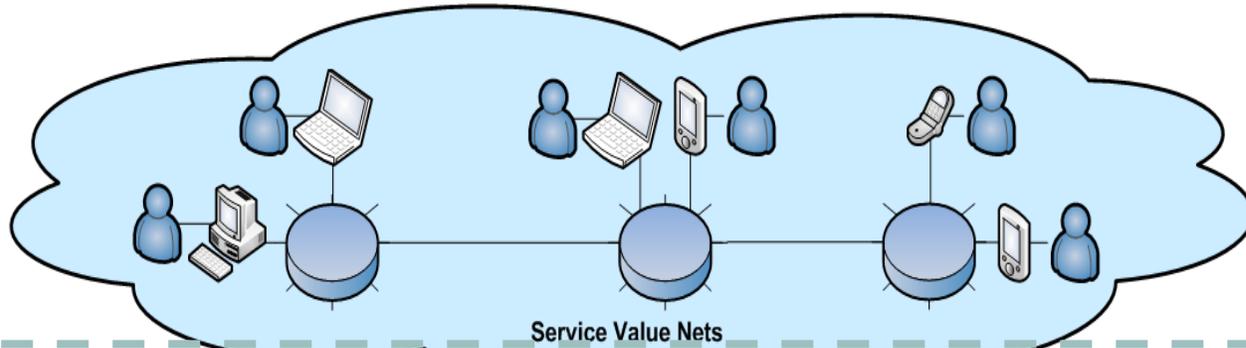
PROVISIONED BY
THE
SERVICE PROVIDER



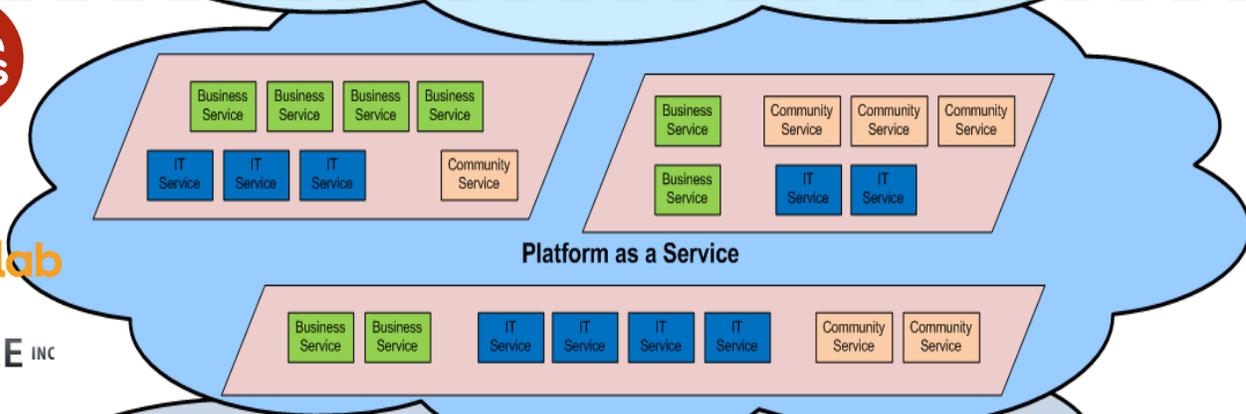
ON DEMAND
NETWORK
ACCESS

SHARED POOL
OF
CONFIGURABLE
COMPUTING
RESOURCES

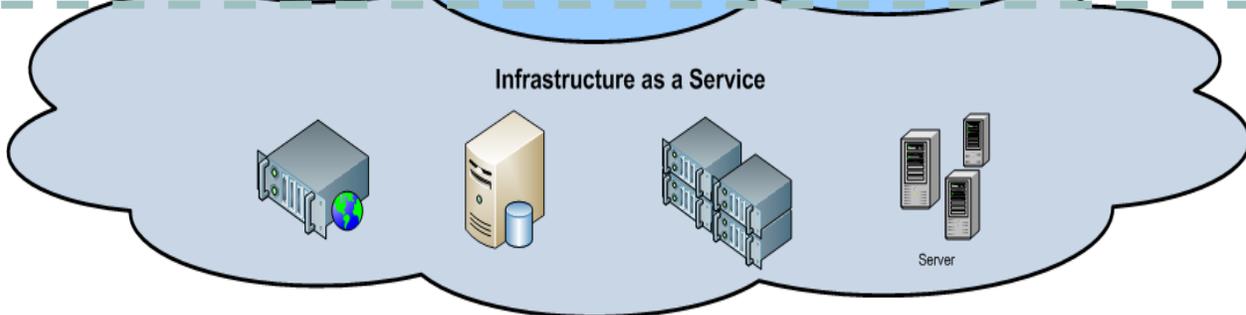
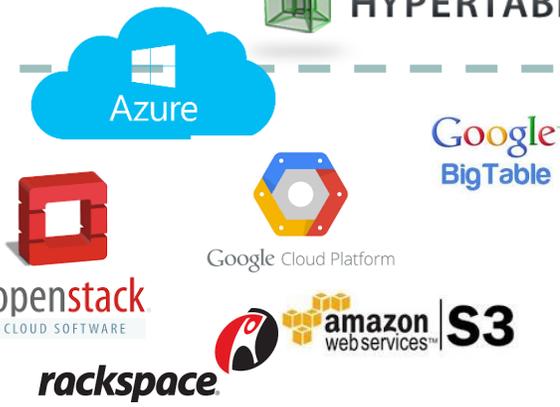
Cloud service models



Use ready application/service



Develop, maintain, manage application



Rent processing, storage, network, capacity and computing resource

From Cloud-onomics

- There are hundreds of cloud vendors ...

- Due to
 - Business competition
 - Lack of official standards



Each vendor imposes its own stack of technologies

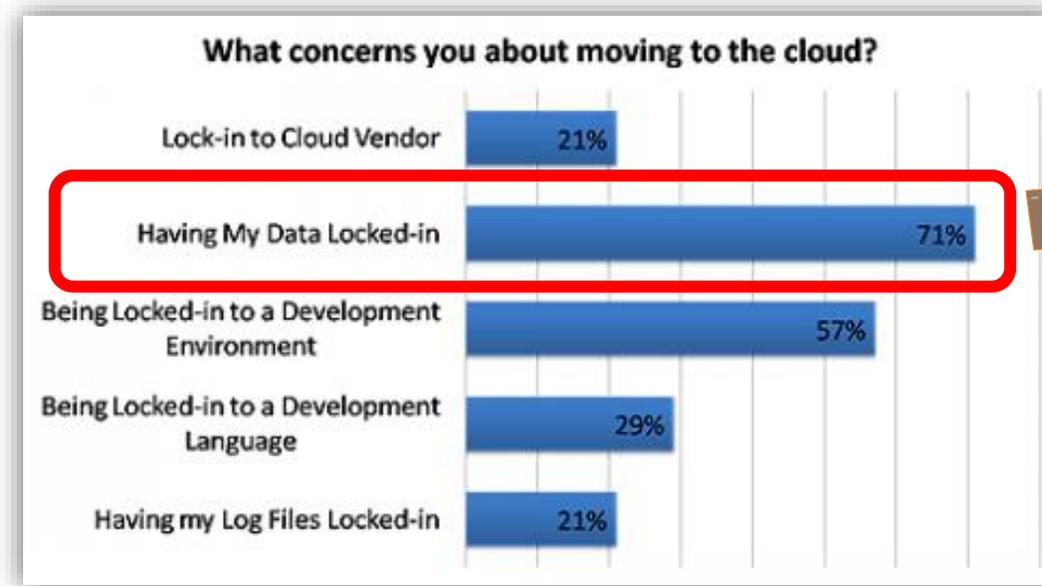
- Differences among the stacks: hypervisor, networking infrastructure, data storage facilities, management means, ...
- Vendor lock-in issue:
 - Lock cloud users into services provided by only one vendor!
 - Can you transfer data and applications to and from the clouds at the same time?

Some critics, such as Richard Stallman*, have called it “a trap aimed at forcing more people to buy into locked, proprietary systems that will cost them more and more over time”

*Richard Stallman is founder of GNU Project and Free Software Foundation



Vendor Lock-in Figures



Source: RighScale blog

Need of innovative solutions => appear keywords in context of cloud computing

Interoperability
Portability

Avoid Vendor Lock-in => More Service Choices => Lower Cost



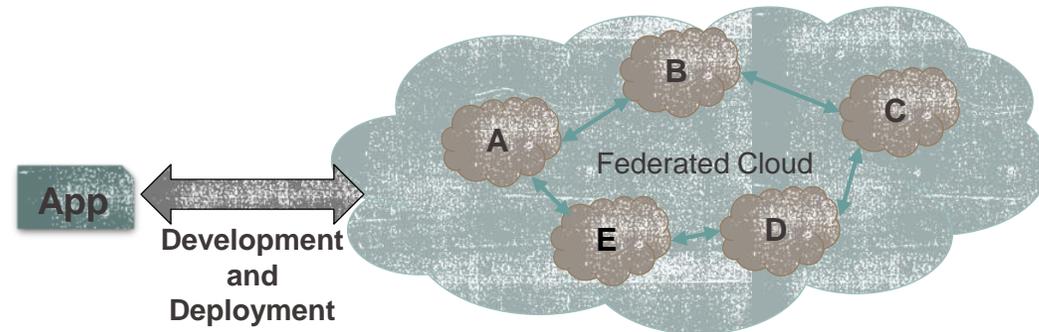
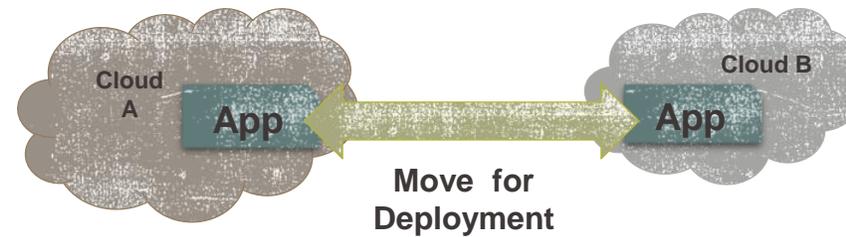
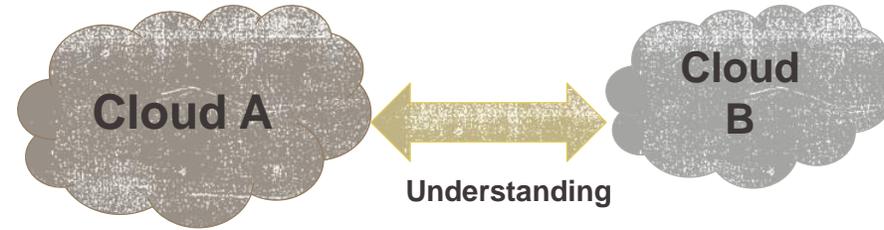
Why Interoperability?

- Avoid vendor lock-in
- Take full advantage of the cloud as itself and not as a vendor infrastructure, platform or service
- Develop applications/services once, deploy anywhere
- Open research directions:
 - Enable hybrid clouds
 - **Brokering cloud services**
 - Cloud service marketplace



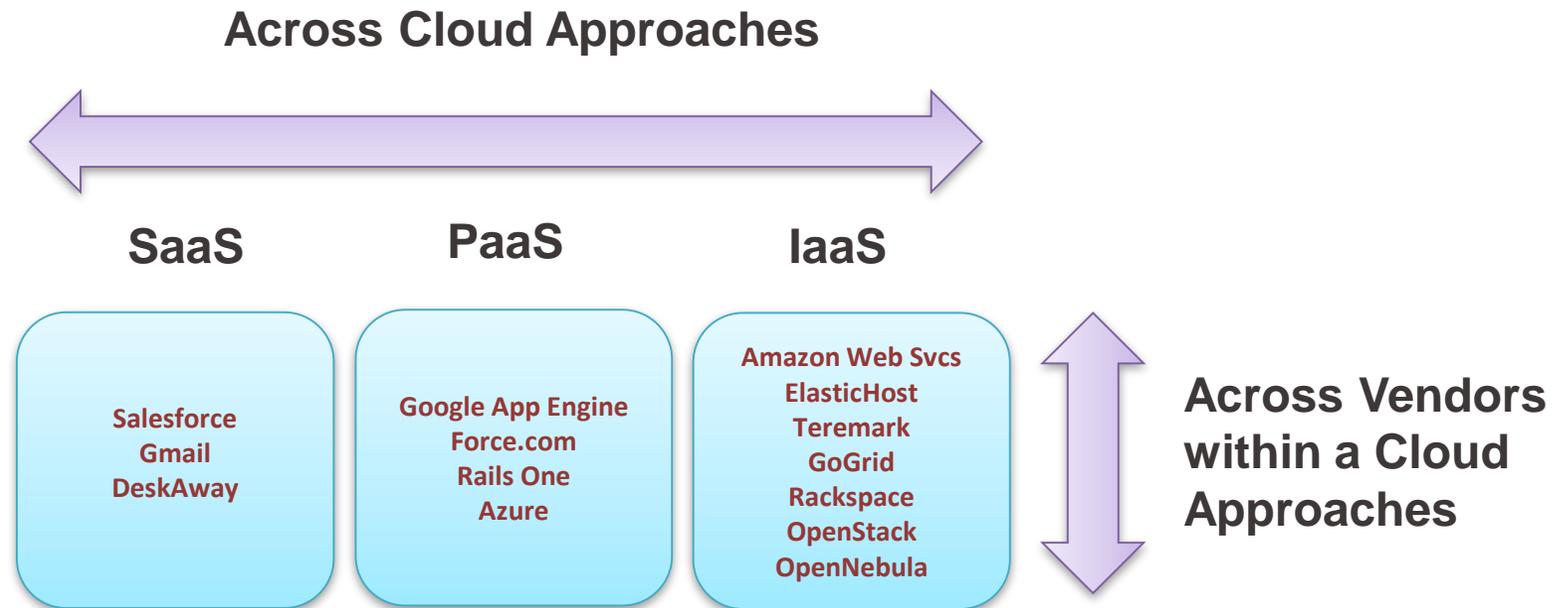
Concepts

- **Interoperability:** Ability for different cloud to talk to each other
- **Portability:** Ability to move application, data, tools from one cloud to another
- **Federation:** Ability to bring together services from various cloud vendors to provide a solution



Interoperability between Clouds?

- Ability to use the cloud services provided by multiple vendors
 - Across vendors within cloud approach
 - Across cloud approach
- Ability to move data and code from one cloud to another or back to the enterprise(portability)



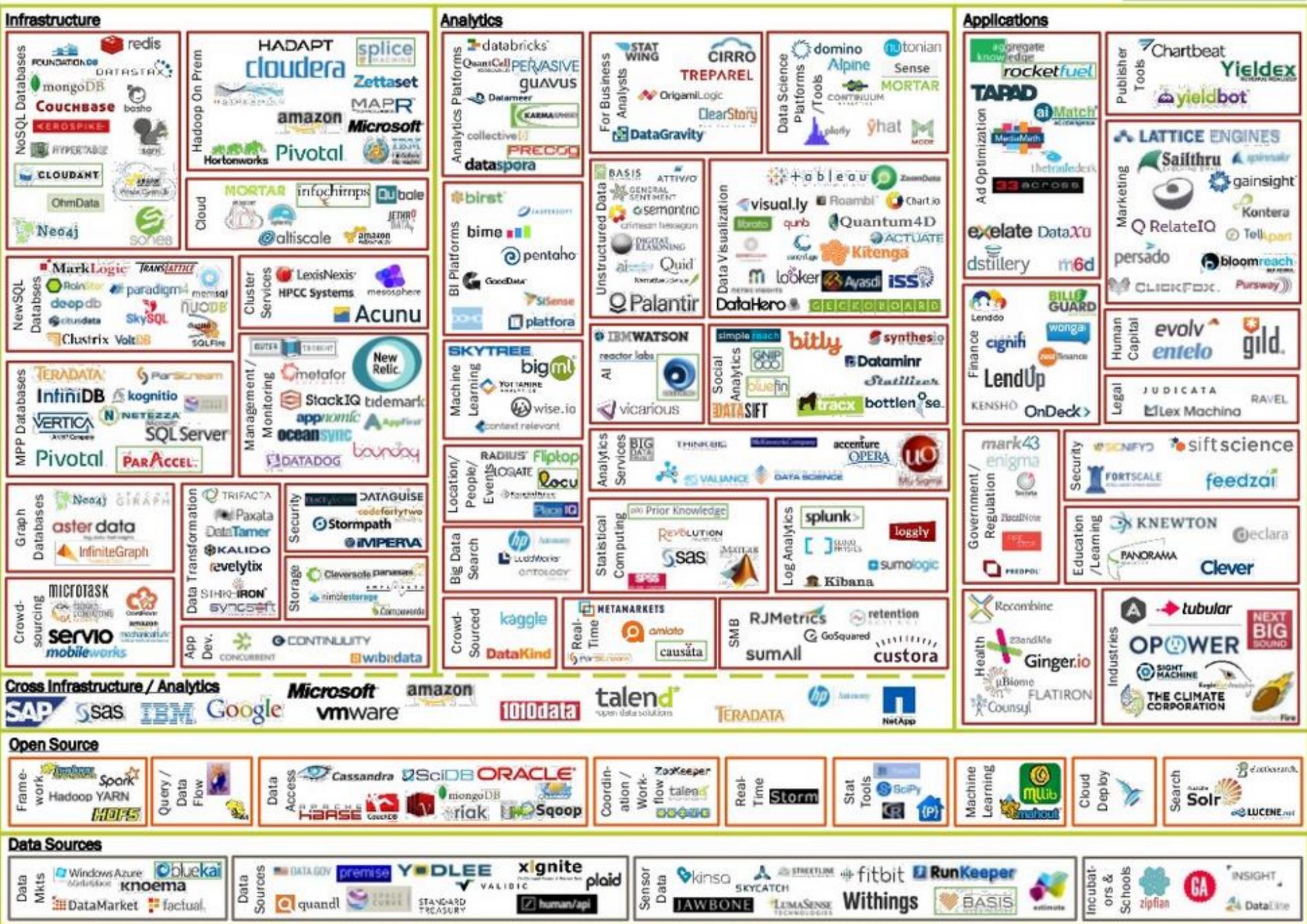
A Cloud **Standardization**? A **Solution** does **not depend** on Cloud providers?
Or **both**?



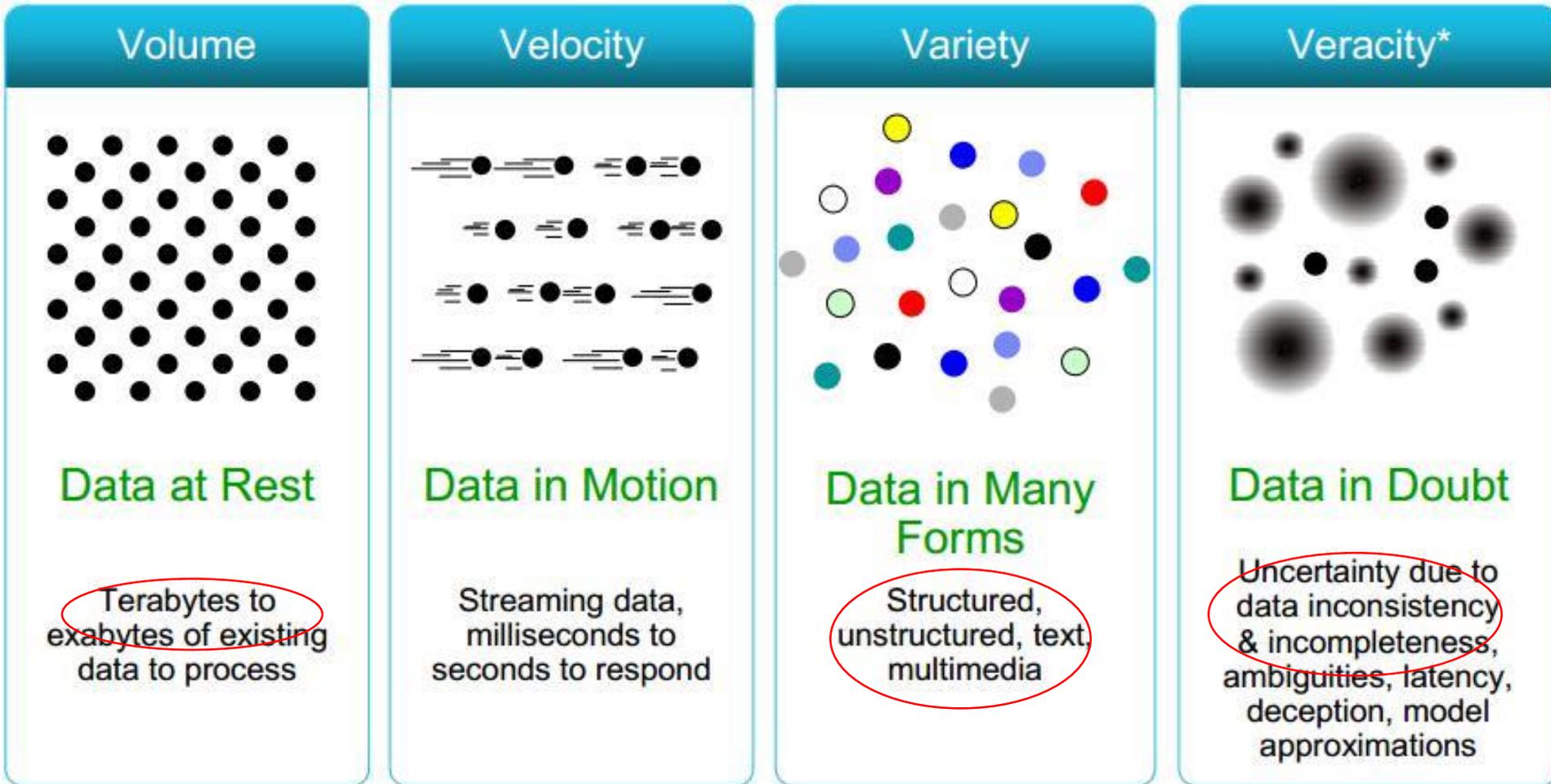


BIG DATA LANDSCAPE, VERSION 3.0

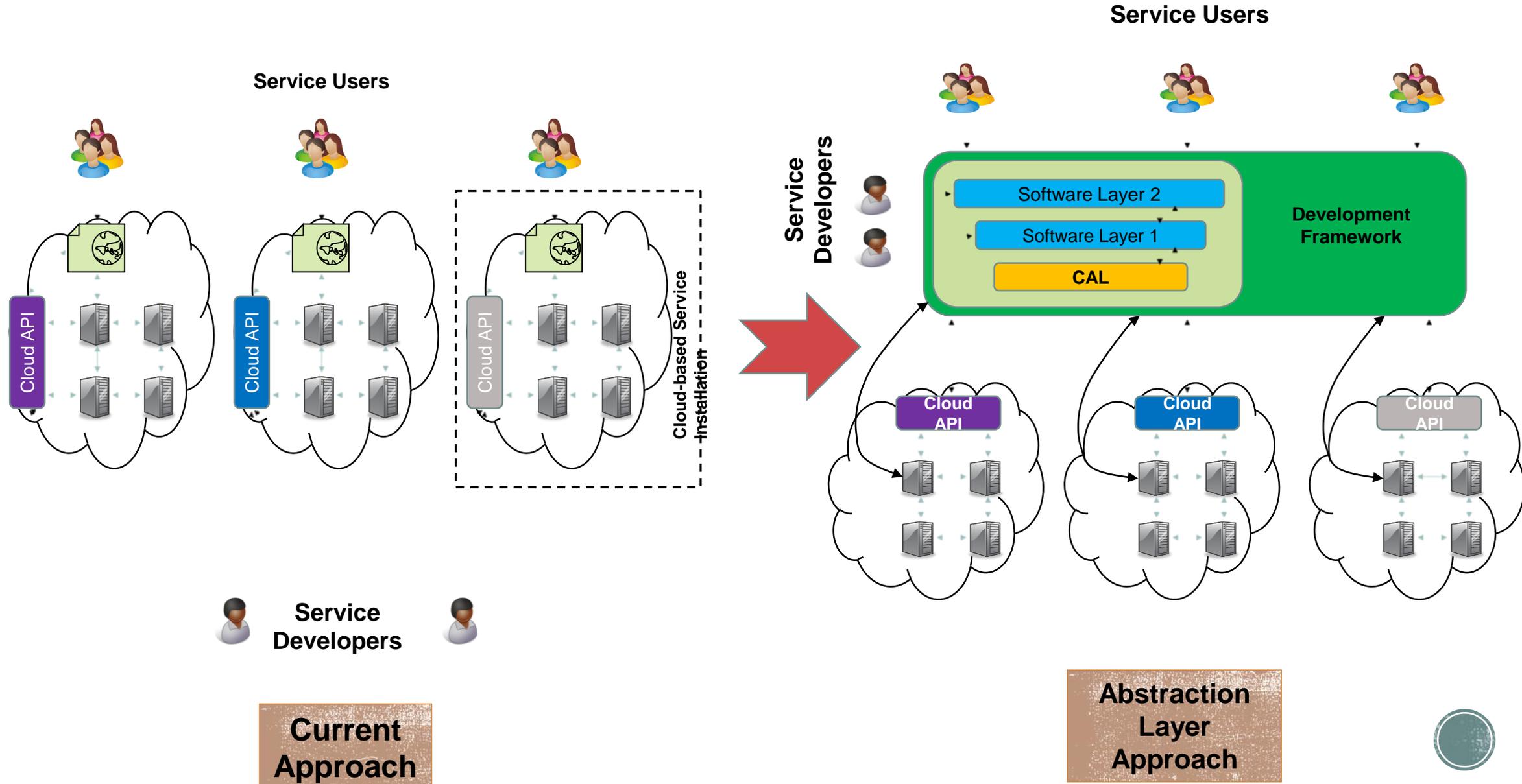
Exited: Acquisition or IPO



Characteristics: 4V

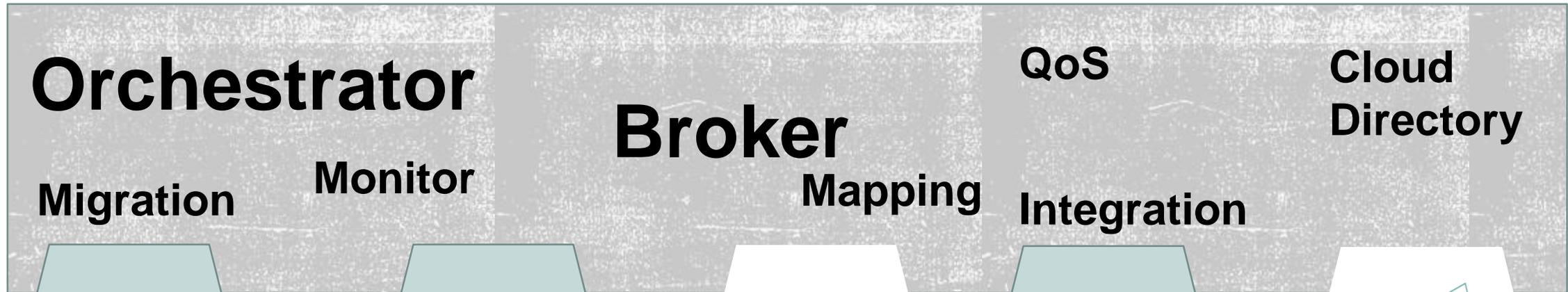


Abstraction approach in Cloud Environment



Our approach

Interoperability,
Extensibility,
Openness,
Adaptability



Cloud de-facto Standards:
OCCI, CDMI, EC2 API, S3 API,

Connector

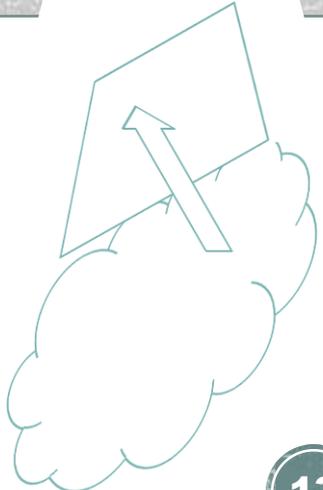
Windows Azure
The Future Made Familiar

Connector

amazon web services

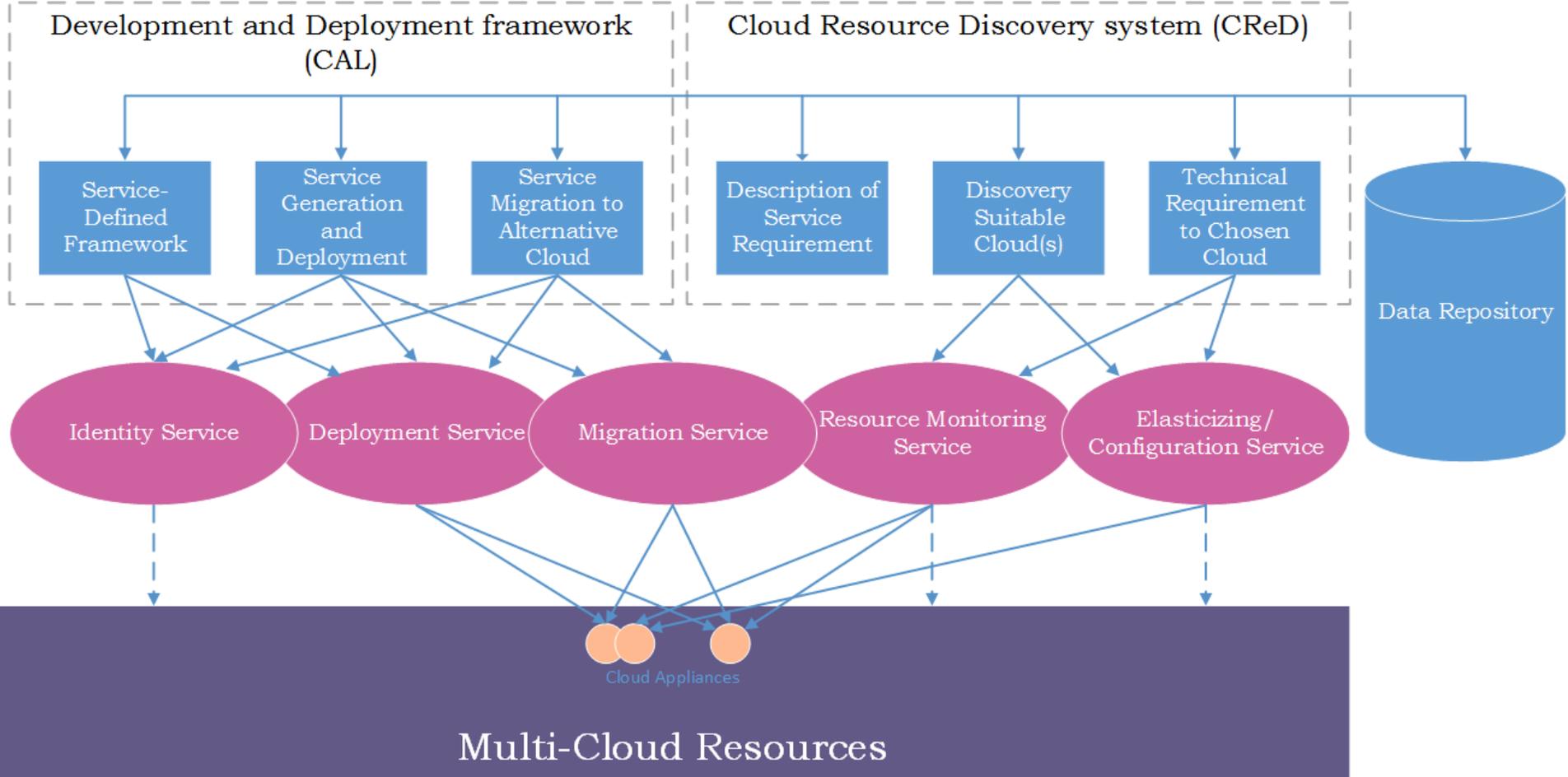
Connector

Google App Engine

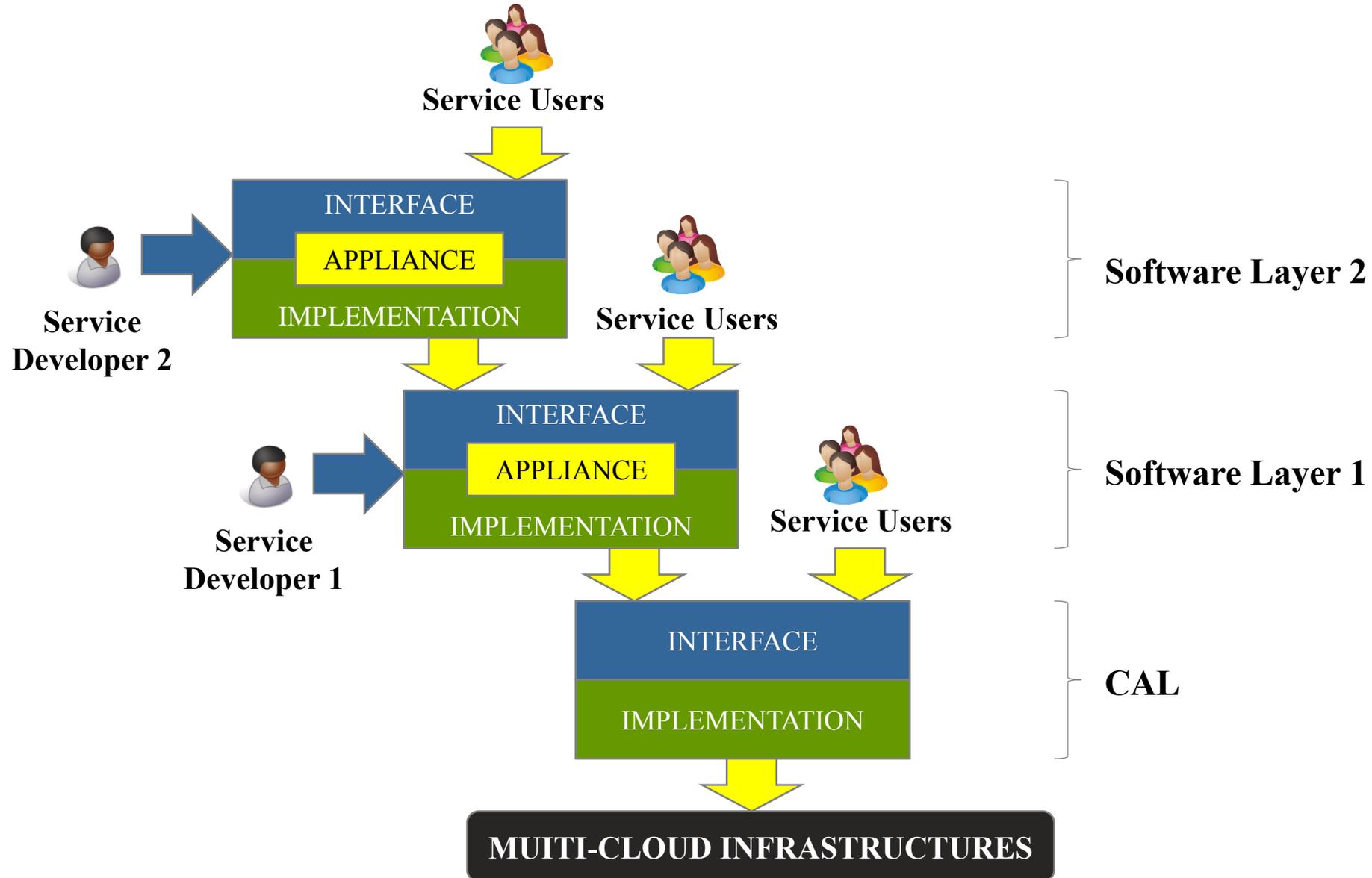


- CAL: Cloud Abstraction Layer for Development and Deployment of Cloud Applications.
- Semantic Cloud Resource Model
- Description of developed application requirements (use fully-distributed intelligent classification algorithm (light-weight, online learning, and fast responses) that is accurate and scalable for large systems (multiple clouds))
- Resource provision: Add intermediation state for VM (MIDDLE state beyond ON and OFF) improve QoS
- Scalable Technique: Resource Consumption Prediction: ANFIS (neutron + Fuzzy): improve QoS

Cloud Resource Broker



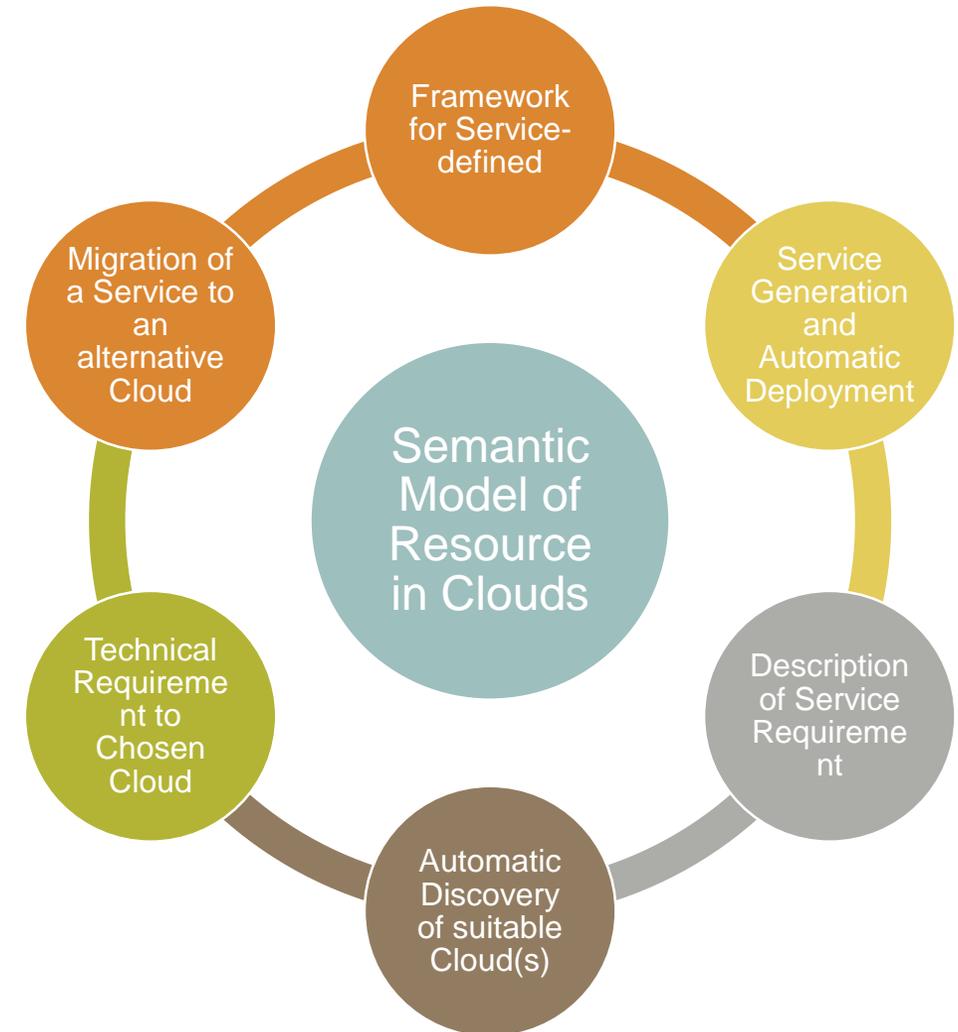
CAL – Cloud Abstraction Layer [1]



Semantic Cloud Model [5]

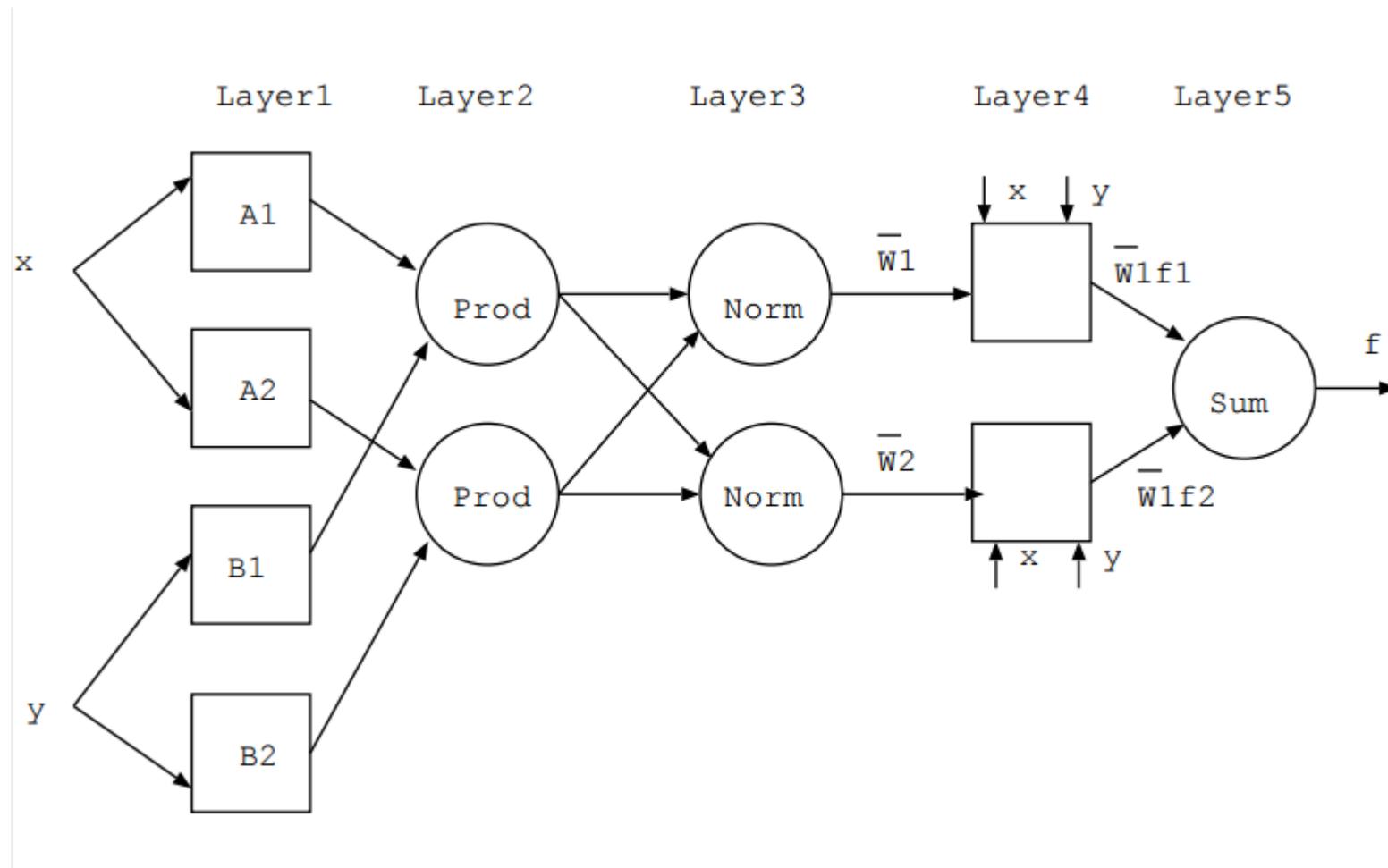
1. Definition and description of services
2. Automatic creation (generation, composition)
3. Description of developed application requirements.
4. Automatic search of resources.
5. Technical adjustment according to application requirements
6. Automatic migration/deployment

Semantic Model provides a platform with unique strengths – interoperability, scalability and understanding of services and data

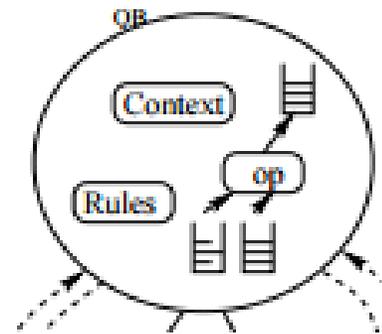
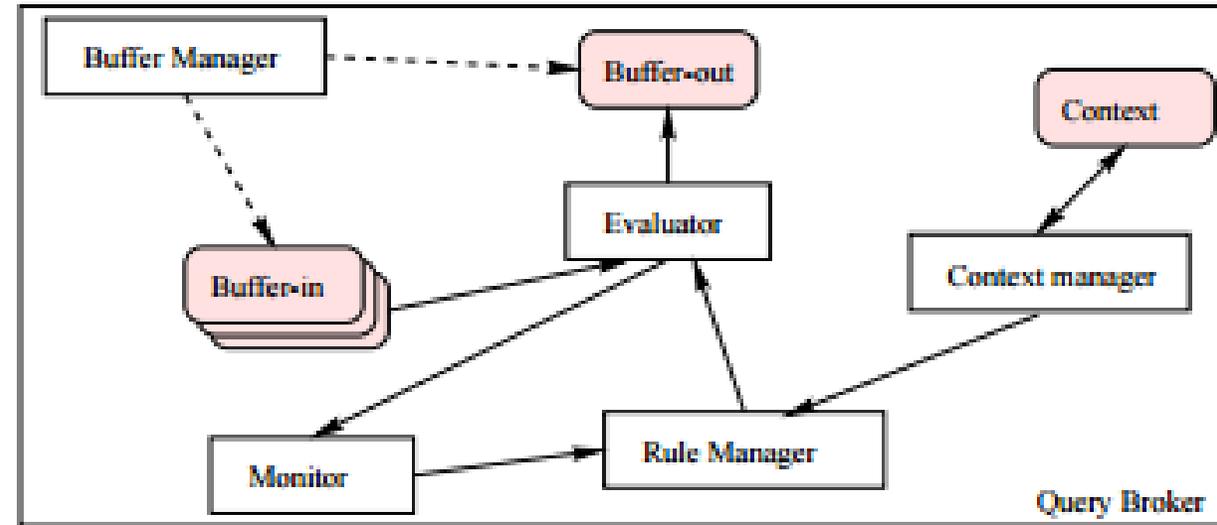
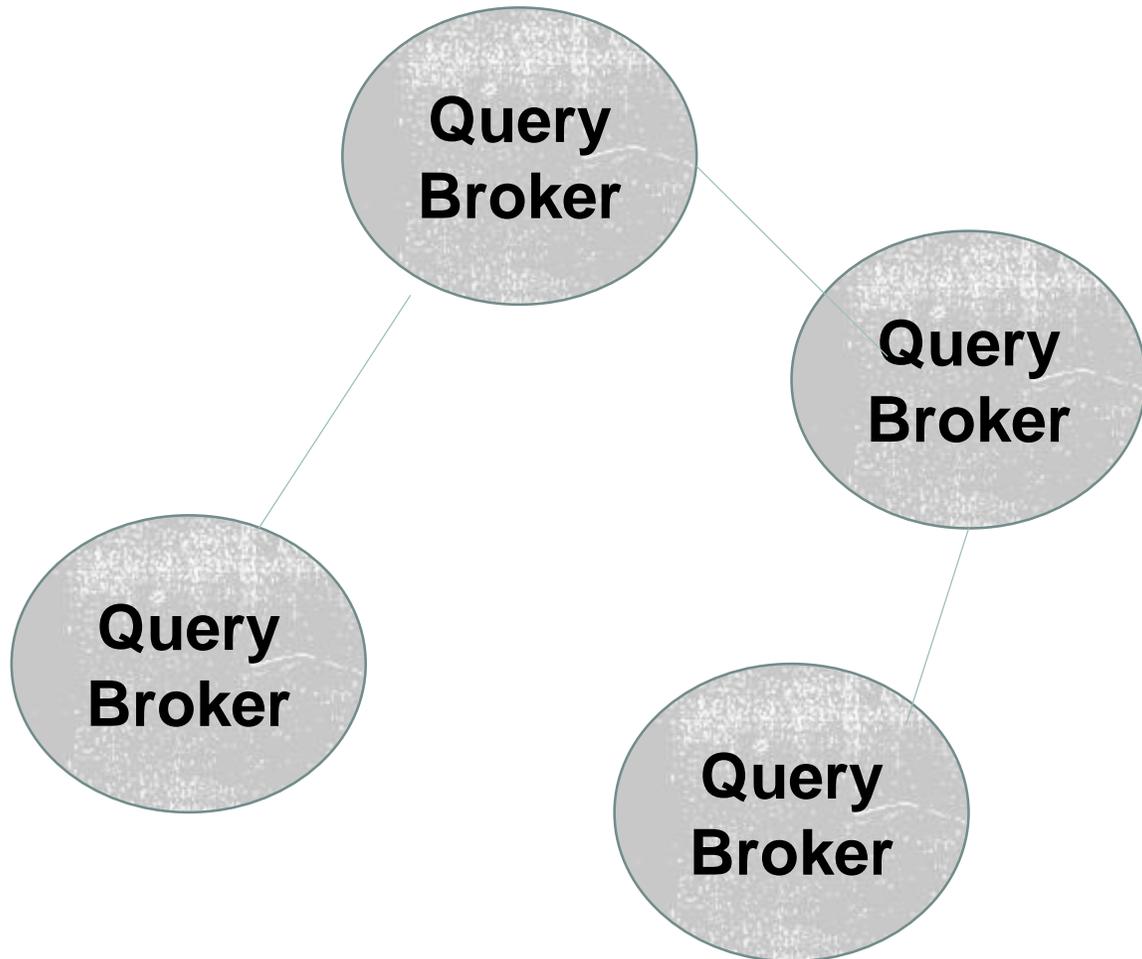


Usage Resource Prediction Model

- ANFIS architecture



Uncertain and incomplete data management [3]



Adaptable,
incomplete and
uncertain

Concluding remarks

- The cloud is inevitable choice for hosting Big Data EcoSystem
- Needs of openness, extensibility, flexibility, adaptability, interoperability
- Toward a brokering framework for cloud and big data services
 - Cloud abstraction layer
 - Uncertain and incomplete data management: integration and querying
 - different optimization aspects (overhead cost and adaptability)
 -privacy ...

References

1. Binh Minh Nguyen, Dang Tran, Quynh Nguyen. A Strategy for Server Management to Improve Cloud Service QoS. DS-RT 2015.
2. Cuong Dao-Duc, Hua Xiaohui, Olivier Morere. Maritime Vessel Images Classification Using Deep Convolutional Meural Network. Submitter to SolCT 2015
3. Quang-Duong Ha, Tuyet-Trinh Vu. Using cache to improve top-k query processing in P2P system. ISIPS 2015
4. Debabrota Basu, Qian Lin, Weidong Chen, Hoang Tam Vo, Zihong Yuan, Pierre Senellart, Stéphane Bressan. Cost-Model Oblivious Database Tuning with Reinforcement Learning. DEXA 2015
5. Binh Minh Nguyen, Quang Minh Dao. Towards a Semantic Model of Resource in Cloud Environment. SolCT 2014.

