Task Force 4: Cloud Federations

Wojciech Burakowski, Attila Kertesz, Steven Latre, Patrick Gwydion Poullie and Maciej Sosnowski
Partners:

• Warsaw University of Technology, Poland
• University of Szeged and MTA SZTAKI, Hungary
• University of Antwerp-iMINDS, Belgium
• University of Zurich, Switzerland
and (maybe)
• Alexandru Ioan Cuza University of Iasi, Romania
The objective of TF4
Architectures for CFs
Planned studied scenarios
Models for CFs
  - On task flow level
  - On resource level
  - Model of resources
Research topics
Experiments possibilities
Conclusions
The objectives:

- To recognize important research topics corresponding to Cloud Federations
- To specify adequate models (for analysis, simulations and experiments) and traffic scenarios
- To present solutions
- To test solutions, if possible
- In addition
  - To enforce international cooperation (e.g. common papers, to establish European projects?)
Architecture:

• Try to define a reference architecture we will assume for our studies
European Commission’s definition of Clouds²

A 'cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality.
Architectural view by the EC

Private Cloud
SP → IP

Public Cloud
SP → IP

Hybrid Cloud
SP → IP
IP1
IP2

Community Cloud
SP → IP
IP1

Inter-Cloud overview

Benefits of an Inter-Cloud

For users:
• Diverse geographical locations
  – Legislation compliant services: decide where data is stored
• Better application resilience
  – Avoid cloud service outages, multiple data centers for fault tolerance
• Avoidance of vendor lock in
  – Distribute workload, price-sensitive usage, easy migration

For providers:
• Expand on demand
  – Offload resource utilization to other providers
• Better SLA to customers
  – Support worst-case scenarios, survive outages
Inter-Cloud developments

a) Centralised Inter-Cloud Federation. Clouds use a central entity to facilitate resource sharing.

b) Peer-to-Peer Inter-Cloud Federation. Clouds collaborate directly with each other.

c) Multi-Cloud Service. Clients access multiple clouds through a service.

d) Multi-Cloud Library. Clients develop their own brokers by using a unified cloud API in the form of a library.
### Inter-Cloud survey

<table>
<thead>
<tr>
<th>Project</th>
<th>Type, Organization</th>
<th>Architecture</th>
<th>Brokering Approach</th>
<th>Application Type</th>
<th>Awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>InterCloud</td>
<td>Research project, University of Melbourne</td>
<td>Centralised federation</td>
<td>SLA based and Directly managed</td>
<td>Singular Jobs</td>
<td>Geo-location, Pricing</td>
</tr>
<tr>
<td>Contrail</td>
<td>Private and public European research organisations Funded by EU</td>
<td>Centralised federation</td>
<td>SLA based</td>
<td>Singular Jobs</td>
<td>Pricing</td>
</tr>
<tr>
<td>Dynamic Cloud Collaboration (DCC)</td>
<td>Academic research project supported by South Korean research funds.</td>
<td>Centralised federation</td>
<td>SLA based</td>
<td>Singular Jobs</td>
<td>Pricing</td>
</tr>
<tr>
<td>Federated Cloud Management (FCM)</td>
<td>Academic research project supported by EU funds.</td>
<td>Centralised federation</td>
<td>SLA based</td>
<td>Singular Jobs</td>
<td>Pricing</td>
</tr>
<tr>
<td>RESERVOIR</td>
<td>Private and public European research organisations Funded by EU</td>
<td>Peer-to-peer federation</td>
<td>SLA based and Trigger-Action</td>
<td>Singular Jobs</td>
<td>Pricing</td>
</tr>
<tr>
<td>Open Cirrus</td>
<td>Research testbed by academic and industry partners. Partially funded by US NSF</td>
<td>Peer-to-peer federation</td>
<td>Directly managed</td>
<td>Singular Jobs</td>
<td>Data location</td>
</tr>
<tr>
<td>OPTIMIS</td>
<td>Private and public European research organisations Funded by EU</td>
<td>Peer-to-peer federation and Independent Service</td>
<td>SLA based</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Pricing</td>
</tr>
<tr>
<td><strong>Arjuna Agility</strong></td>
<td>Commercially owned</td>
<td>Peer-to-peer federation</td>
<td>Trigger-Action</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Local resources</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td><strong>Global Inter-Cloud by Bernstein et al.</strong></td>
<td>Publications are by people from miscellaneous companies - CISCO, Huawei Technologies, EMC Corporation</td>
<td>Peer-to-peer federation</td>
<td>SLA based</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Data location, Local resources</td>
</tr>
<tr>
<td><strong>mOSAIC</strong></td>
<td>Private and public European research organisations Funded by EU</td>
<td>Independent service</td>
<td>SLA based</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Pricing</td>
</tr>
<tr>
<td><strong>STRATOS</strong></td>
<td>York University. Supported by Canada’s NSERC funds, Amazon and CA Inc.</td>
<td>Independent service</td>
<td>SLA based</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Geo-location, Pricing, Legislation/ Policy and Local resources</td>
</tr>
<tr>
<td><strong>Commercial Cloud Management Systems (RightScale, EnStratus, Scala, Kaavo)</strong></td>
<td>Commercially owned</td>
<td>Independent service</td>
<td>Trigger-Action</td>
<td>Singular Jobs</td>
<td>Geo-location, Data location, Pricing, Legislation/ Policy and Local resources</td>
</tr>
<tr>
<td><strong>Libraries (JClouds, LibCloud, DeltaCloud, SimpleCloud, Apache Nuvem)</strong></td>
<td>Open source projects</td>
<td>Multi-Cloud libraries</td>
<td>Directly managed</td>
<td>Singular Jobs, Periodical Jobs, Compute and Data Intensive Interactive application</td>
<td>Geo-location, Data location, Pricing, Legislation/ Policy and Local resources</td>
</tr>
</tbody>
</table>
Generalized FCM (covering all presented scenarios)

Two levels of brokering:
- at the federated level
- at single cloud level
Planned scenarios for using Cloud Federations:

- **Scenario 1**: each Cloud belonging to Cloud Federation has its own customers, which regard it as **Home Cloud**. So, such Cloud at the first wants to provide service to these customers using its own resources.
  - If this cloud is saturated then it looks for resources from other clouds from the Federation.
  - If this cloud has spare resources, then it wants to rent them

- **Scenario 2**: The request for service is generated by a customer to Cloud Federation. Cloud federation treats each customer in the same way. It means that such Cloud Federation manages allocation of resources to a given request in a “fairness” way taking into account all Clouds
Models on task flow level:

- Two types of requests for services:
  - Type#1: Requests on demand for existing services (on-line)
  - Type#2: Requests for provisioning new services (off-line)
Requests of type#1 (on-line): for offered services

- Sequence of tasks (a bit simplified description)
- Request rate per task sequence
- Description of each task: sequence number, type of resource the task can be executed
Requests of type #2: for provisioning new services (offline)

- Requests for virtual machine configurations
- Application can be decomposed into multiple subservices

Each subservice will consume resources

Connections between subservices also have an impact on the resources

Subservices can be re-used by other applications
TF4: Cloud Federations

Models of cloud resources:

• Type of resources:
  – Compute resources
Characterization of resources offered by Clouds

Compute resources are packed in racks and organized in clusters

• Rely heavily on parallelization of resources
  Need high-bandwidth network access
  – Gigabit Ethernet
  – Infiniband

• Two main aspects
  – Network topology
  – QoS

Metrics
• Throughput (Mbps)
• Latency
• Packet loss
Characterization of resources offered by Clouds

Compute resources are often connected to shared storage infrastructure

• Type of database system
  – ACID with limited scale-out (SQL)
  – BASE with unlimited scale-out (NoSQL)

• Has an impact on
  – Database size
  – Elasticity
  – Consistency

Metrics
• Consistency speed
• Scale-out factor
• Size
Characterization of resources offered by Clouds

Data centers consume energy

• Almost 50% of costs relate somehow to energy

• Prices may vary over time
• Prices may vary across locations

• Often less explicitly defined.

Metrics
• KWh
• Share of renewable energy
TF4: Cloud Federations

Research topics:

• Task allocation
• Resource sharing
• Firness
• Virtual Network Infrastructure for Cloud Federations
One of the essential problems in Cloud Federations is to make optimal task allocation to the resources in order to meet some requirements for service quality and resource usage.

We need to specify the rules for task allocation to resources.
Planned research topics: Task allocation (when we provision service)

Dynamic environment which is constantly changing
• New applications coming
• Existing applications being removed
• Existing applications change
Task allocation (when we provision service): previous work

Maintaining a tree of peers that manage the application placement

Task allocation in federations

Substrate resources on different sites
Different sites have
- Important differences in capabilities
- Important differences in reliability
- Important differences in connectivity

Application requests, analysis (KPN), monitoring and workload prediction
Elastic migration, redundancy, multi-hopping multi-path, monitoring and networking prediction
Node and link failure, heterogeneity
Resource sharing:

- Interdependency of host resources, e.g., a VM utilizes RAM slower, when it receives less CPU.
- Which dependencies exist and how to describe them mathematically?
- Hitherto no consideration for dependencies or overly simplified assumptions e.g., Leontief utility functions.
- Investigation by executing workloads on different VM configurations.
- Combined with different stress tests.
- Monitoring resource consumption over time.
Fairness:

• Straight forward definition for single resource: Max-Min fairness Clouds
• Heterogenous resources on two levels
  – Different types of physical machines
  – Different types of host resources, e.g., CPU, RAM
• Heterogenous resources make share values subjective
• No access to utility functions
• Envy-freeness (economic concept) not applicable
• How to define fairness in clouds?
• Relevant for private clouds, where resource allocation is not prescribed by SLAs
Virtual Network Infrastructure for Cloud Federations:

- What traffic is carried by the network
- Which Classes of Services are needed, if needed
- To find algorithm to network dimension and adaptation to current needs
Testing capabilities:

- CloudSim for simulations
- European test-beds for experiments
  - Fed4Fire (more detailed information during next meeting)
  - OPEN CIRRUS - a federated testbed of distributed clusters for systems and applications research
Conclusion:

- The TF4 just started
- We plan to finish „white paper” (next week) and to upload it on the server
- We still discuss research topics
- We are open for new partners and co-operation with other TFs: e.g. with TF3 for specification of task sequences and rules for tasks allocation at the provisioning and service execution levels.

Any suggestions are welcome