

Beyond the Clouds, The Discovery Initiative



How Should Next Generation Utility Computing Infrastructures Be
Designed to Solve Sustainability & Efficiency Challenges ?

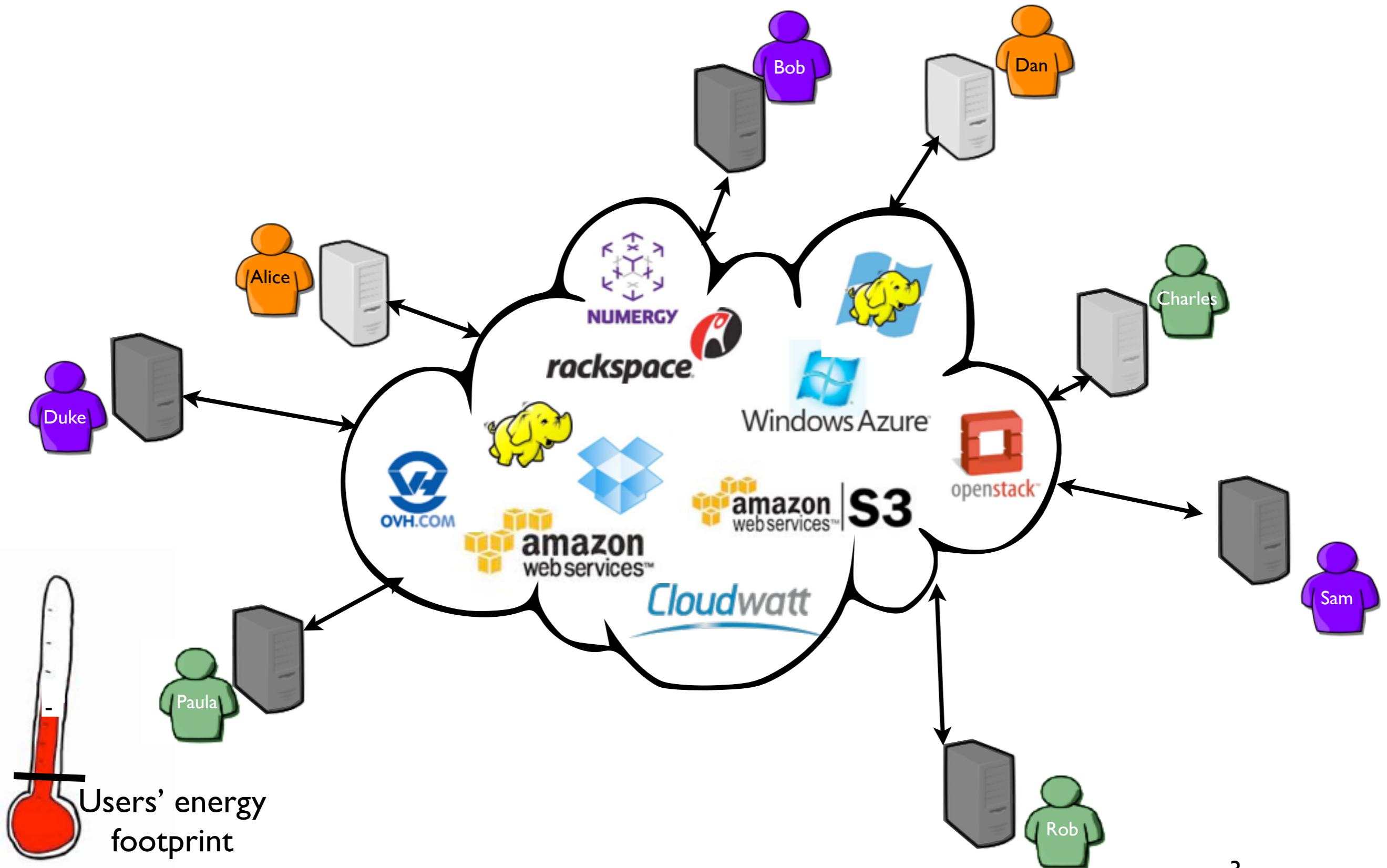


Adrien Lebre
Journée SUCCES - Nov 2015

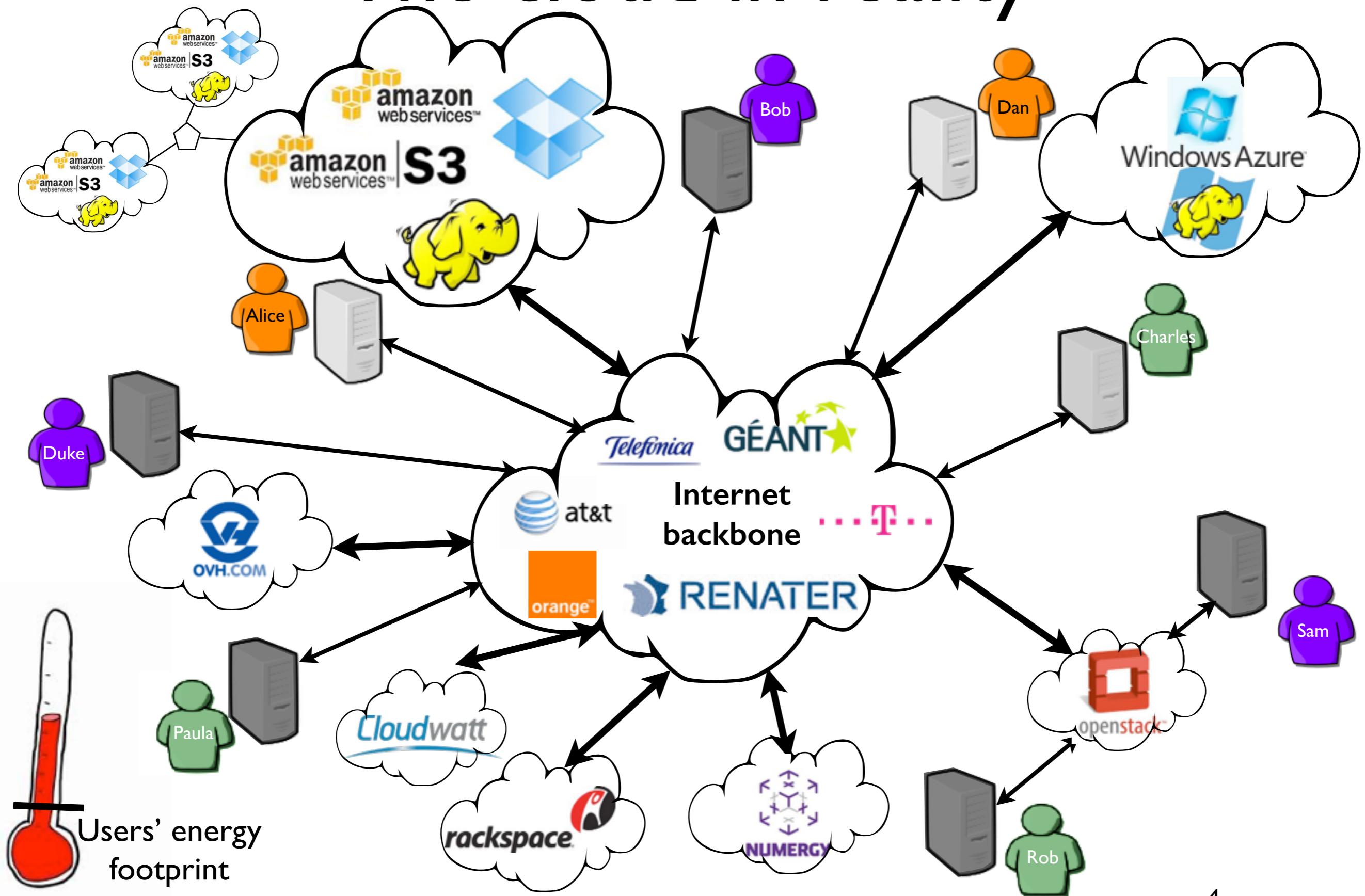
Localization is a key element to deliver
efficient as well as *sustainable* Utility
Computing solutions

A simple Idea
Bring Clouds back to the cloud

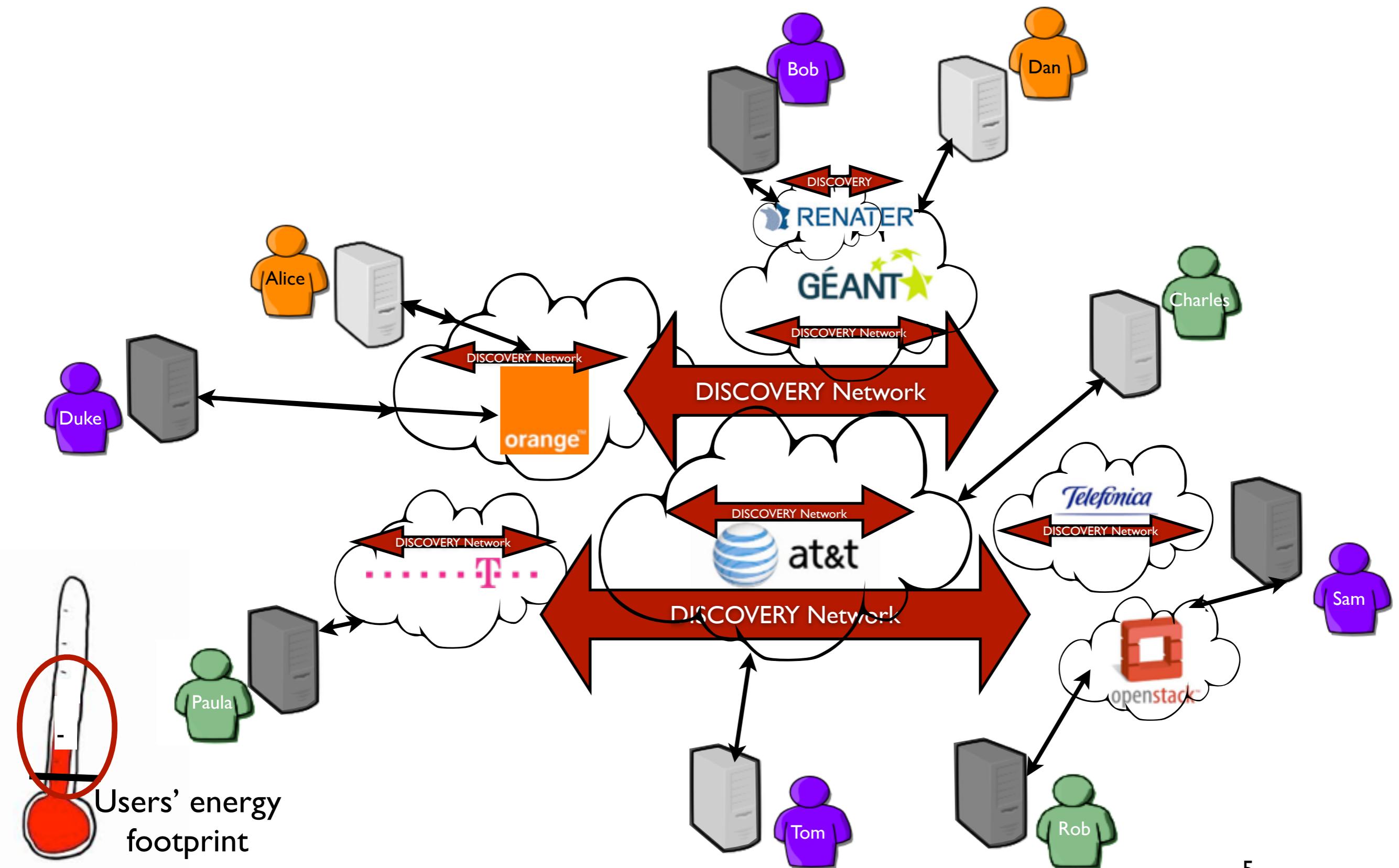
The cloud from end-users



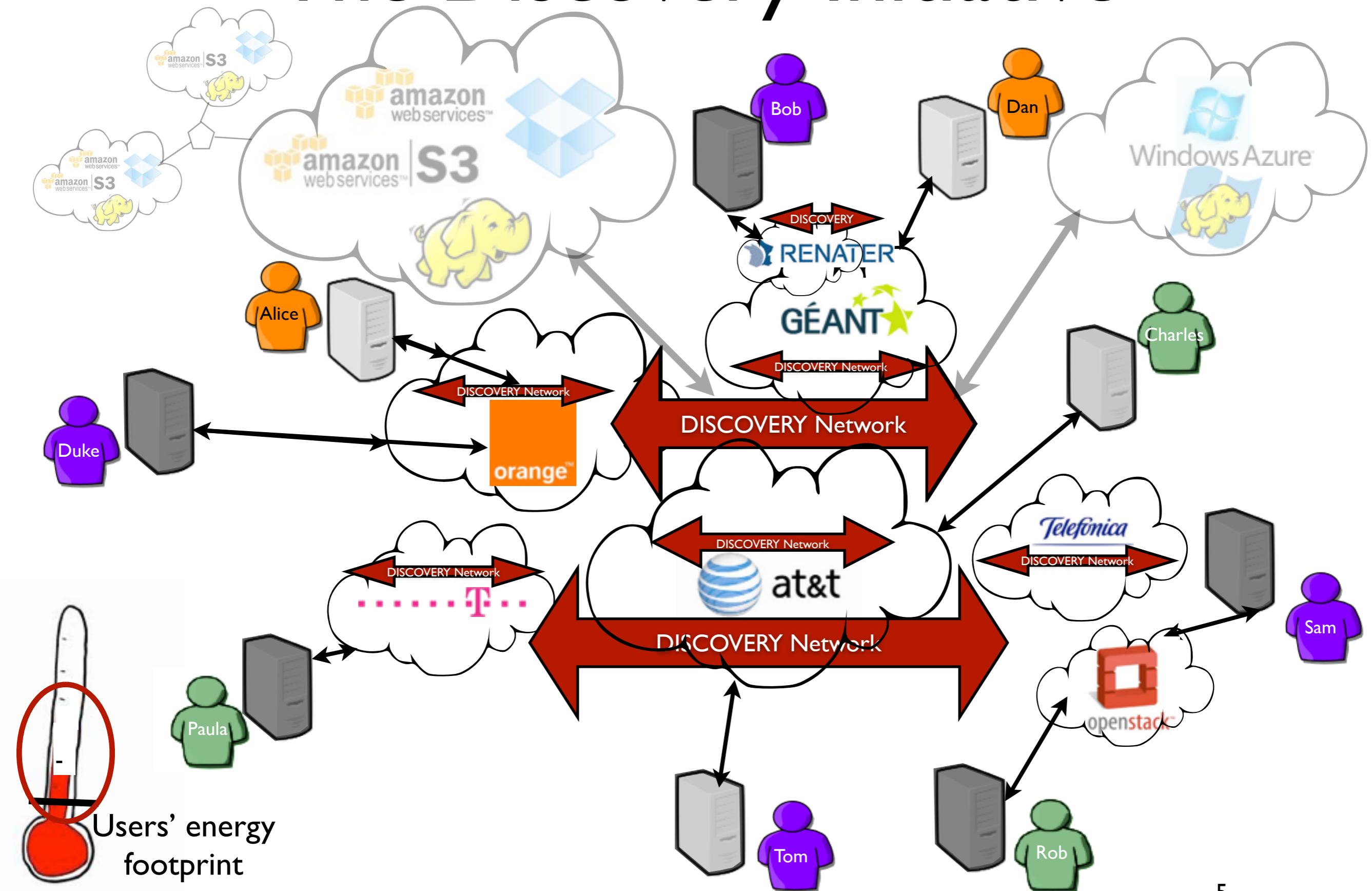
The cloud in reality



The Discovery Initiative



The Discovery Initiative



Why ?

Let's give a look to
the current situation

The Current Trend: Large off shore DCs

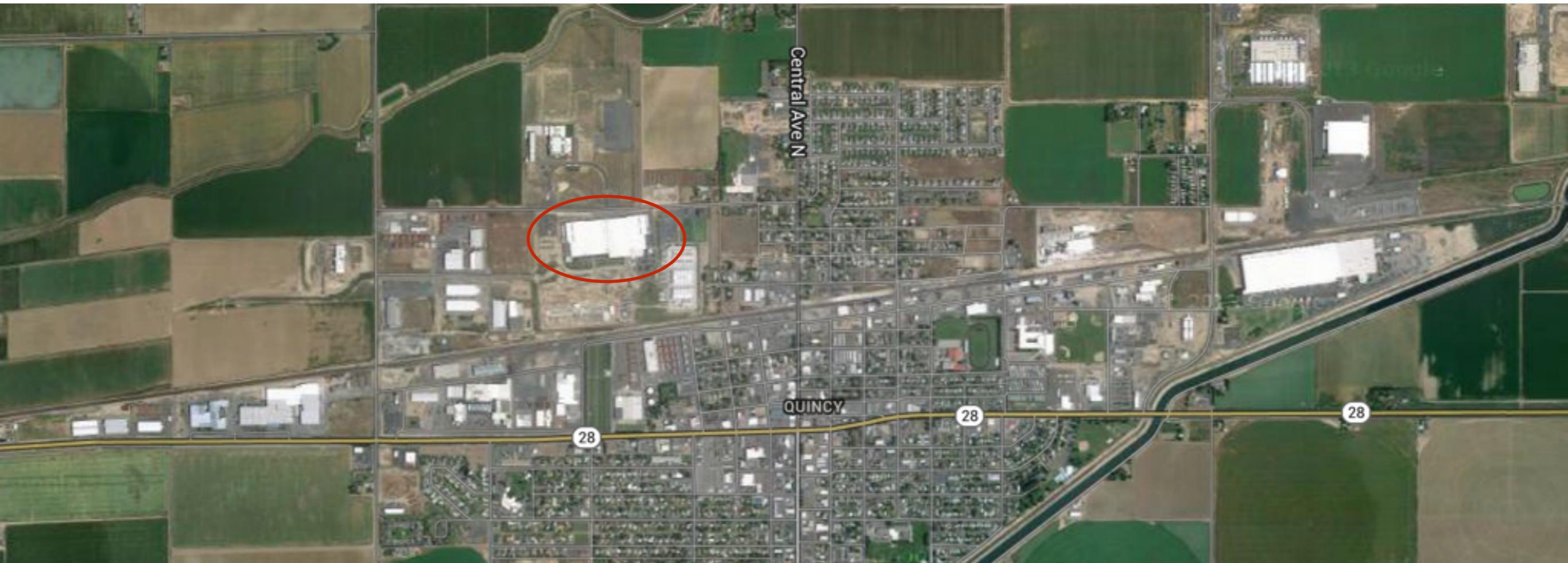
- To cope with the increasing UC demand while handling energy concerns but...



credits: datacentertalk.com - Microsoft DC, Quincy, WA state

The Current Trend: Large off shore DCs

- To cope with the increasing UC demand while handling energy concerns but...



credits: google map - Quincy

The Current Trend: Large off shore DCs

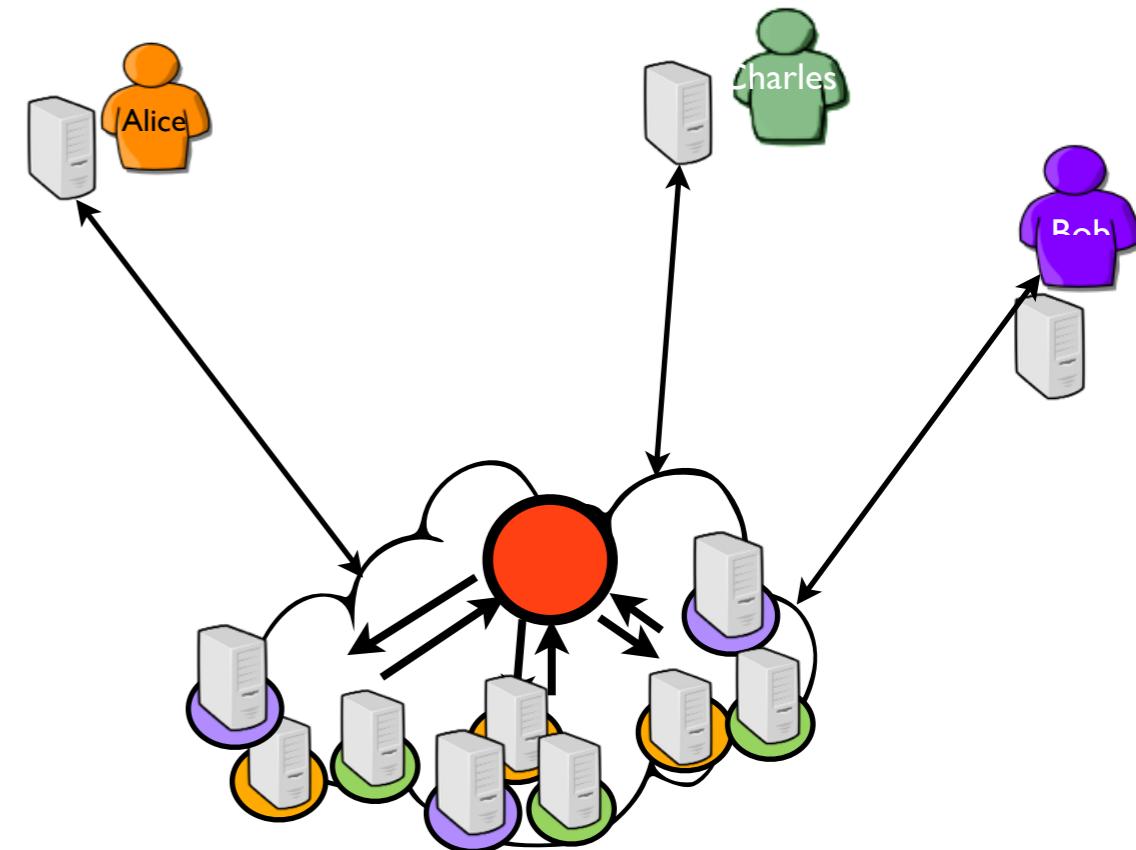
- To cope with the increasing UC demand while handling energy concerns but...



credits: coloandcloud.com

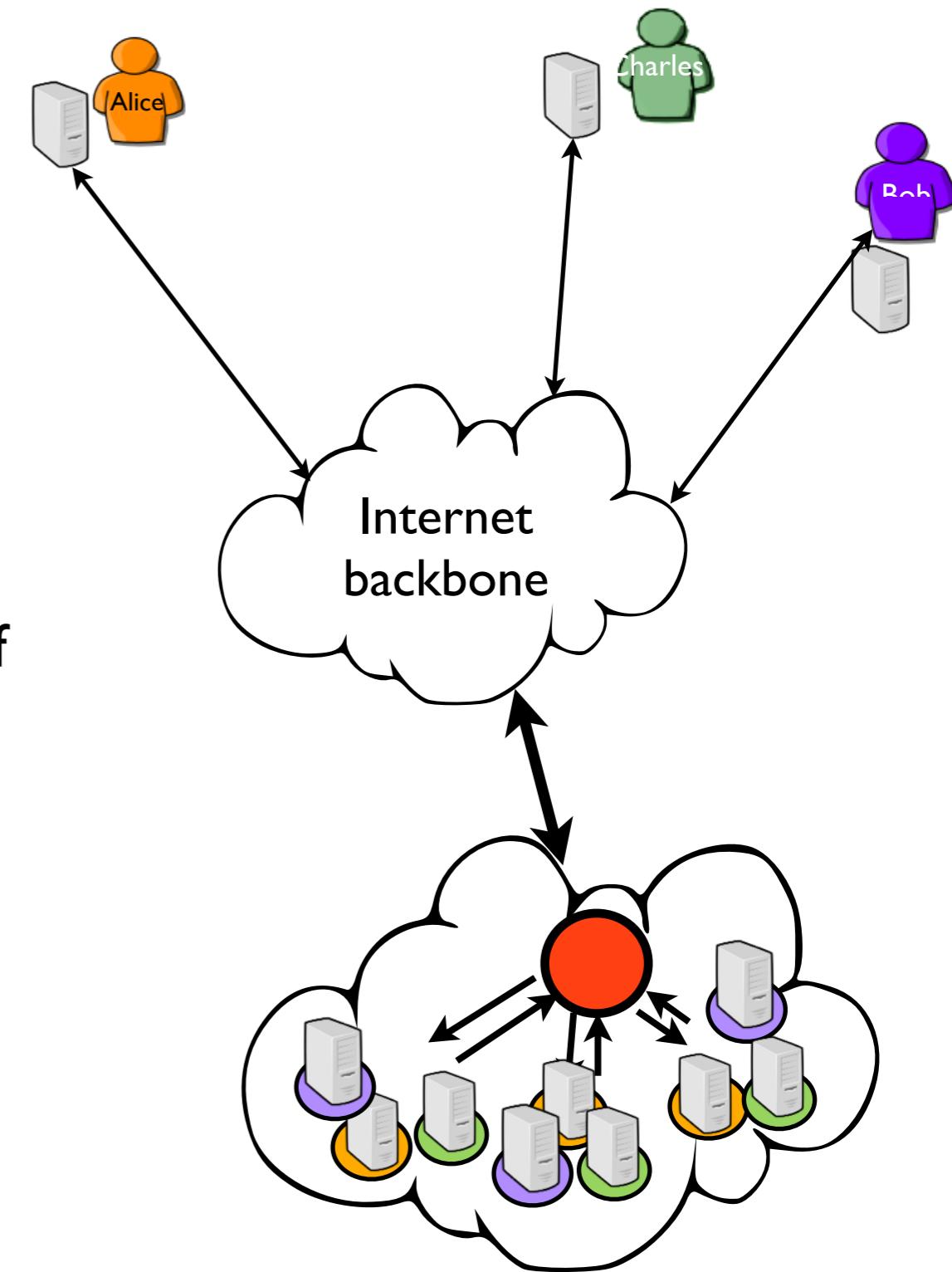
Inherent limitations of current solutions

- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 - I. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)



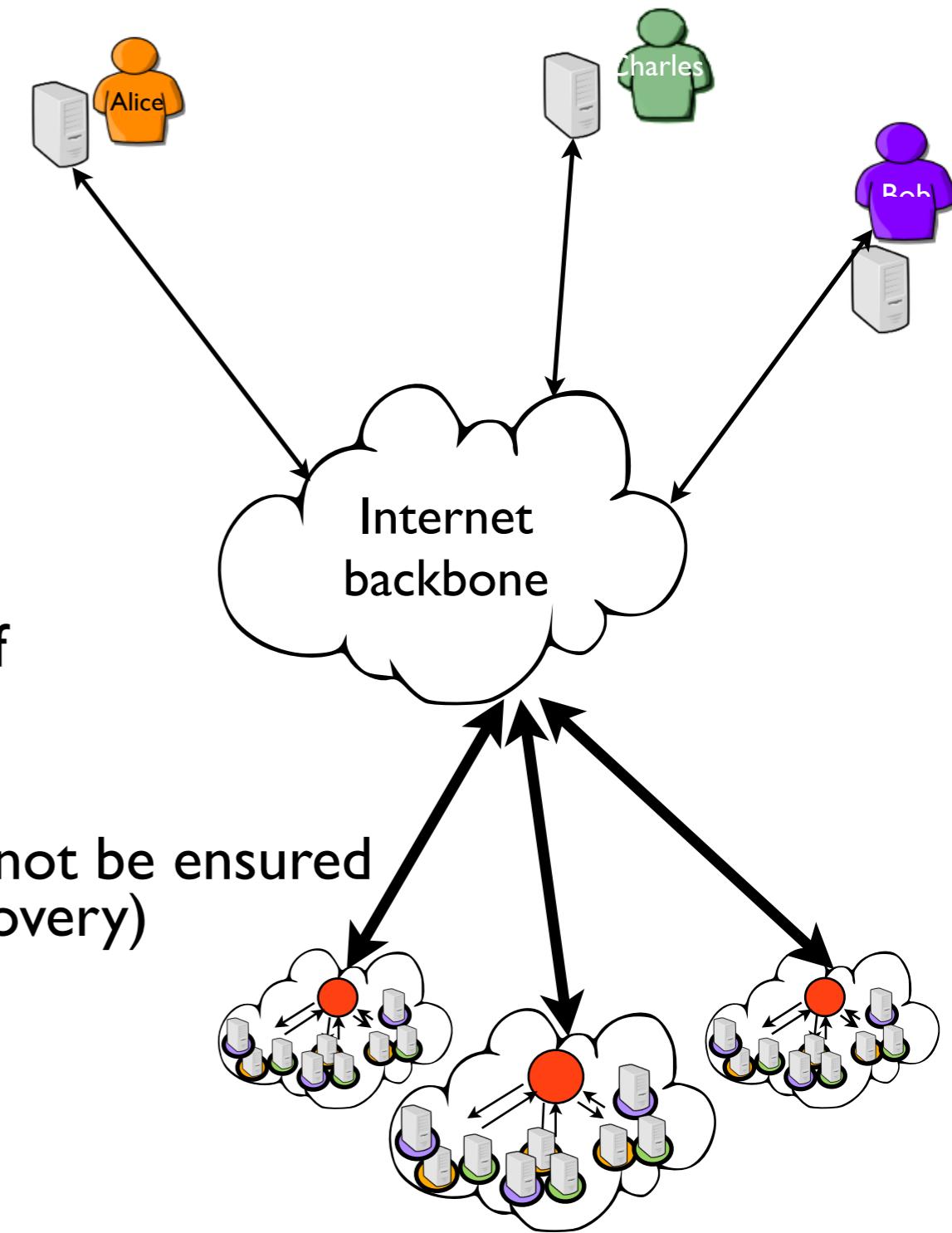
Inherent limitations of current solutions

- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 - I. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)
 2. Overhead implied by the unavoidable use of the Internet to reach distant platforms



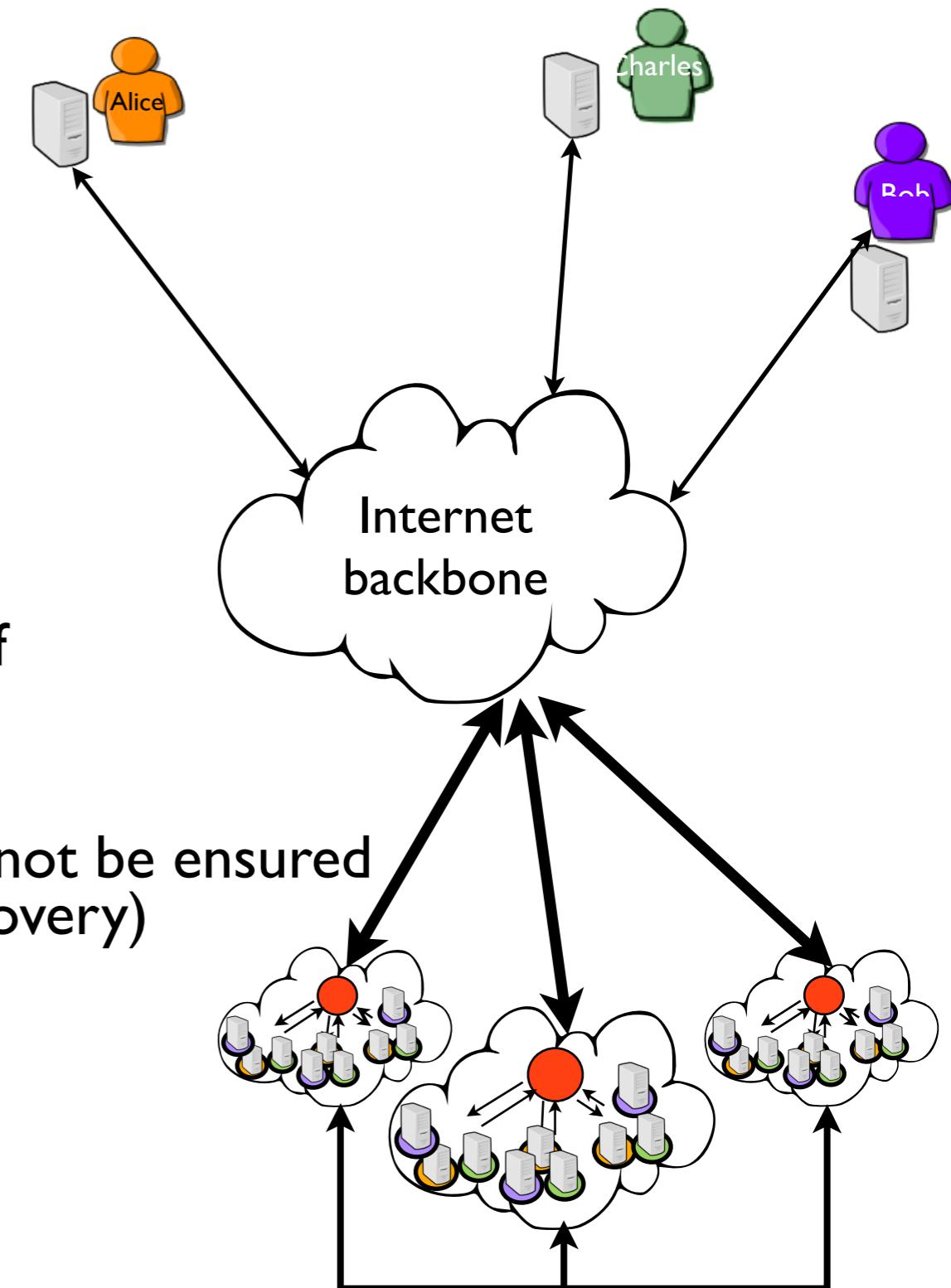
Inherent limitations of current solutions

- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 1. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)
 2. Overhead implied by the unavoidable use of the Internet to reach distant platforms
 3. The connectivity to the application/data cannot be ensured by centralized dedicated centers (disaster recovery)



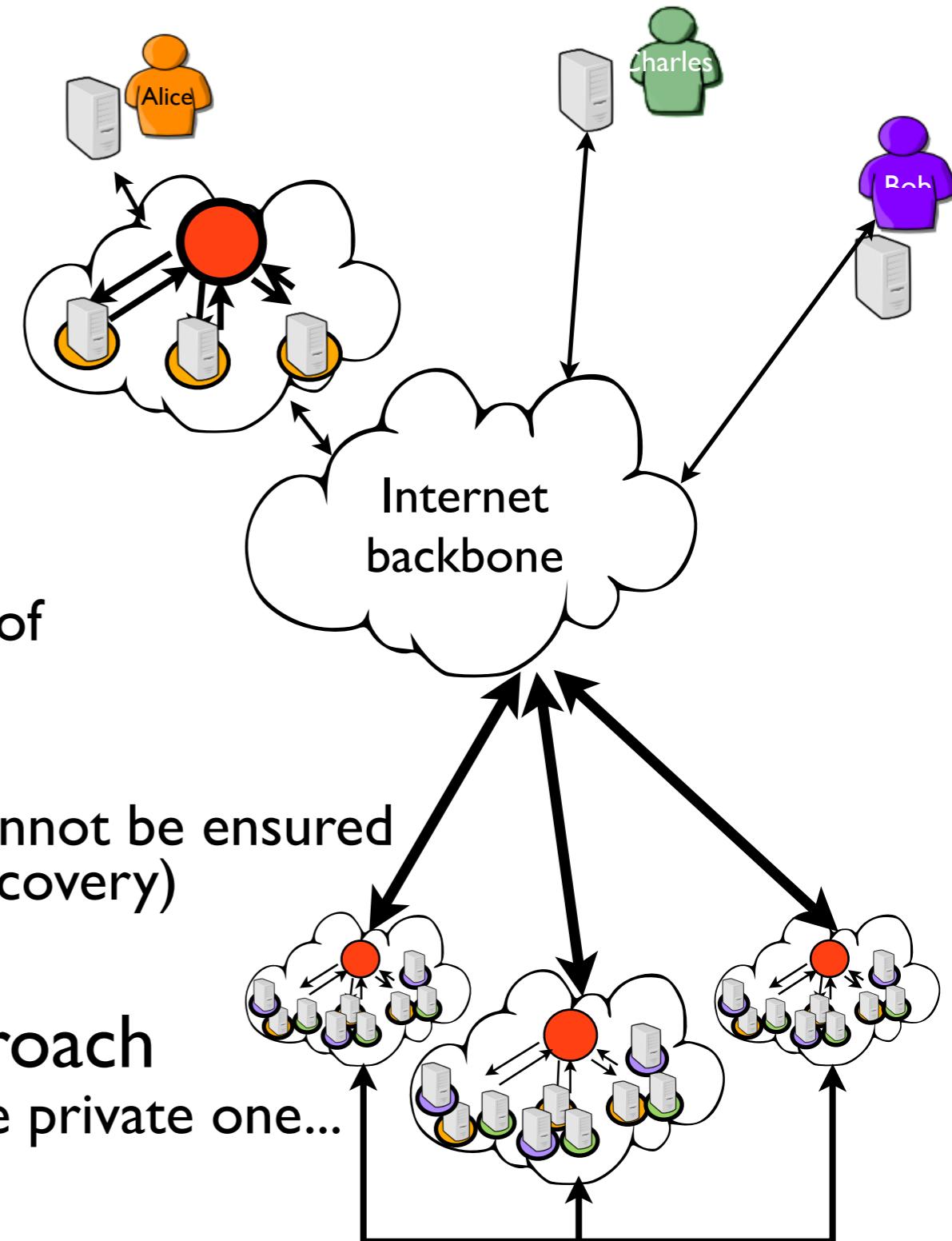
Inherent limitations of current solutions

- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 1. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)
 2. Overhead implied by the unavoidable use of the Internet to reach distant platforms
 3. The connectivity to the application/data cannot be ensured by centralized dedicated centers (disaster recovery)



Inherent limitations of current solutions

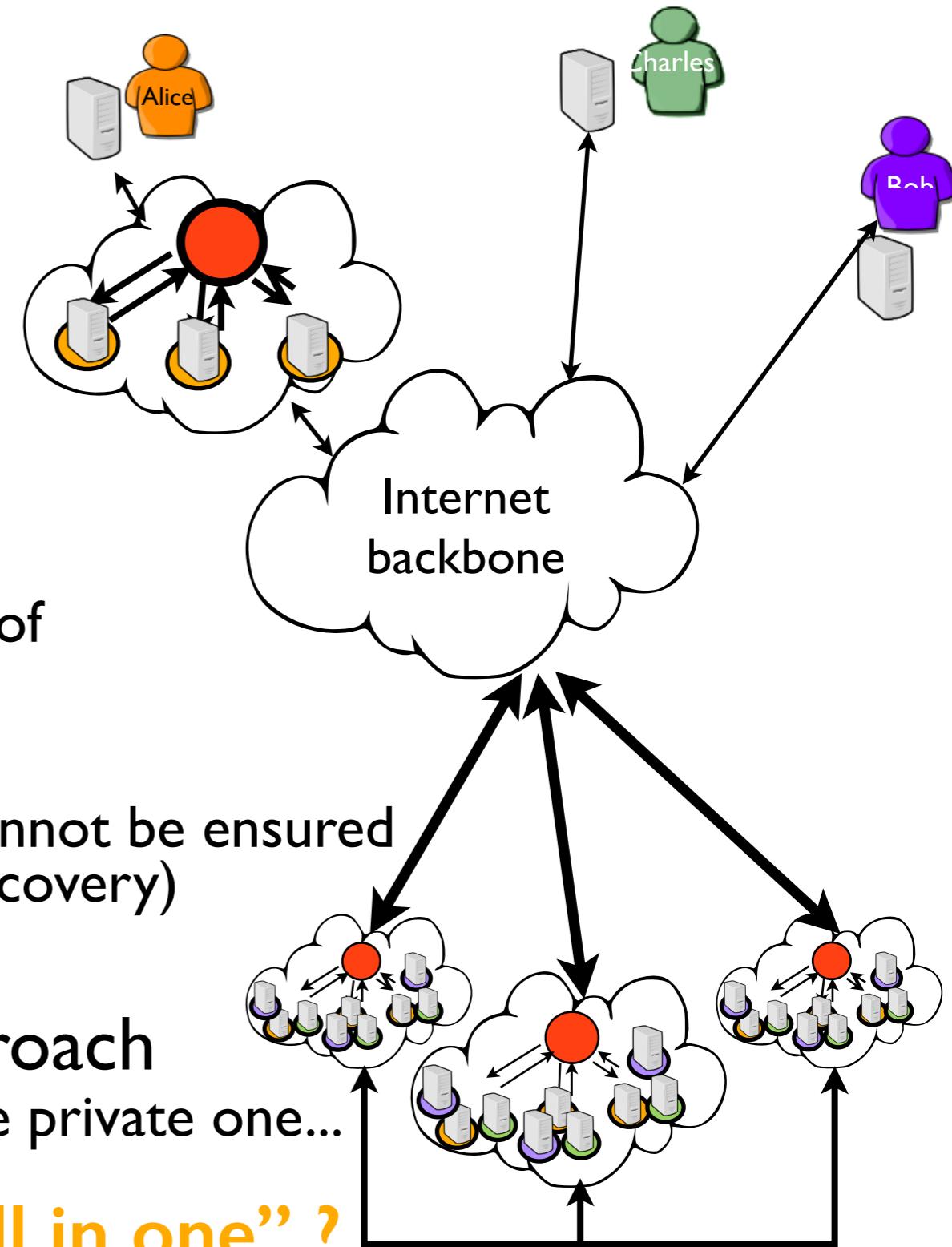
- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 1. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)
 2. Overhead implied by the unavoidable use of the Internet to reach distant platforms
 3. The connectivity to the application/data cannot be ensured by centralized dedicated centers (disaster recovery)
- Hybrid platforms: a promising approach
It depends how you are going to extend the private one...



Inherent limitations of current solutions

- Large off shore DCs to cope with the increasing UC demand while handling energy concerns but...
 1. Externalization of private applications/data (jurisdiction concerns, PRISM NSA scandal, Patriot Act)
 2. Overhead implied by the unavoidable use of the Internet to reach distant platforms
 3. The connectivity to the application/data cannot be ensured by centralized dedicated centers (disaster recovery)
- Hybrid platforms: a promising approach
It depends how you are going to extend the private one...

Can we address these concerns “all in one” ?
μ/nDC concept



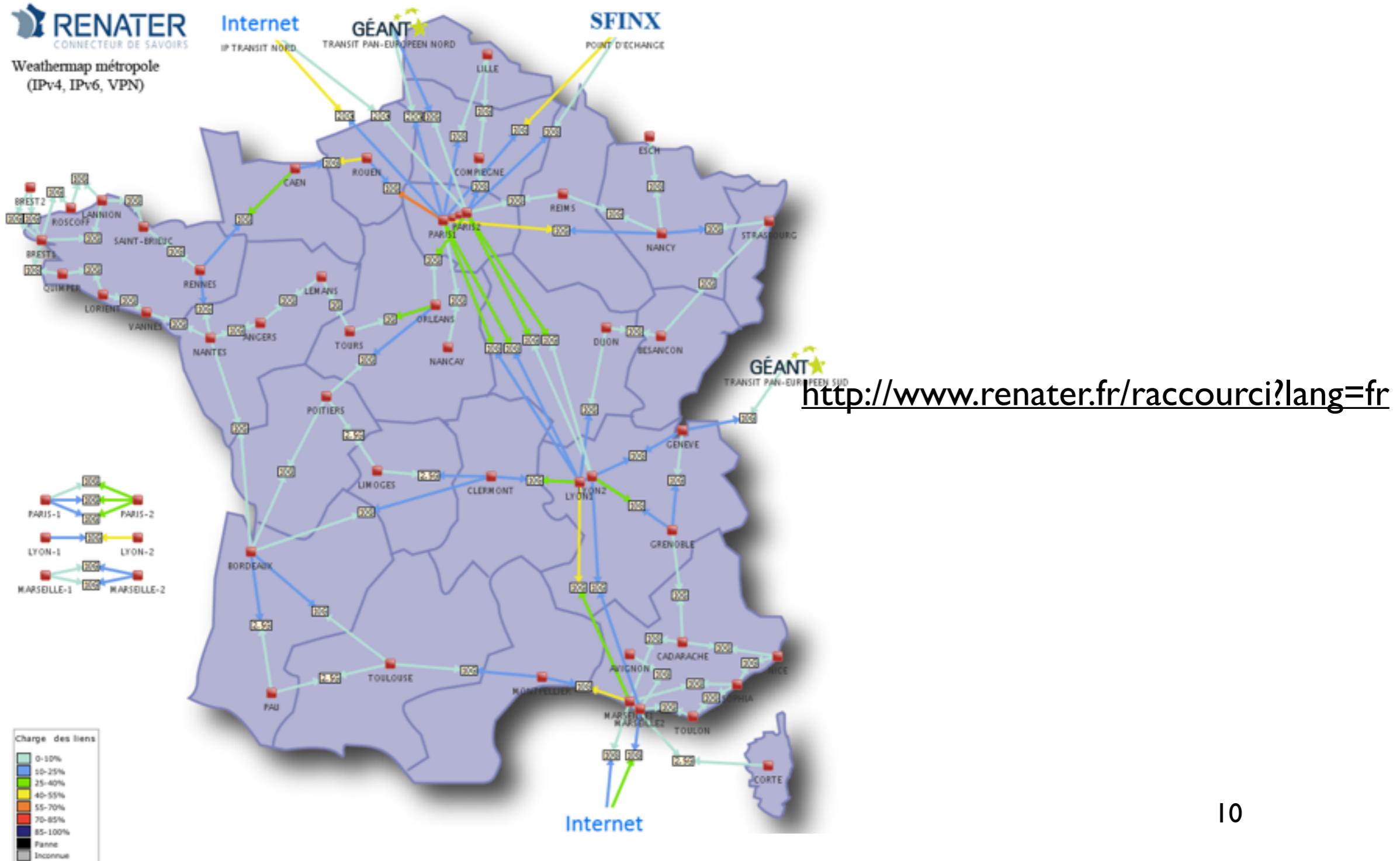
How and where the μ DC concept can be deployed ?

Locality Based Utility Computing Toward LUC Infrastructures

Beyond the Clouds, the DISCOVERY Initiative

• Locality-based UC infrastructures

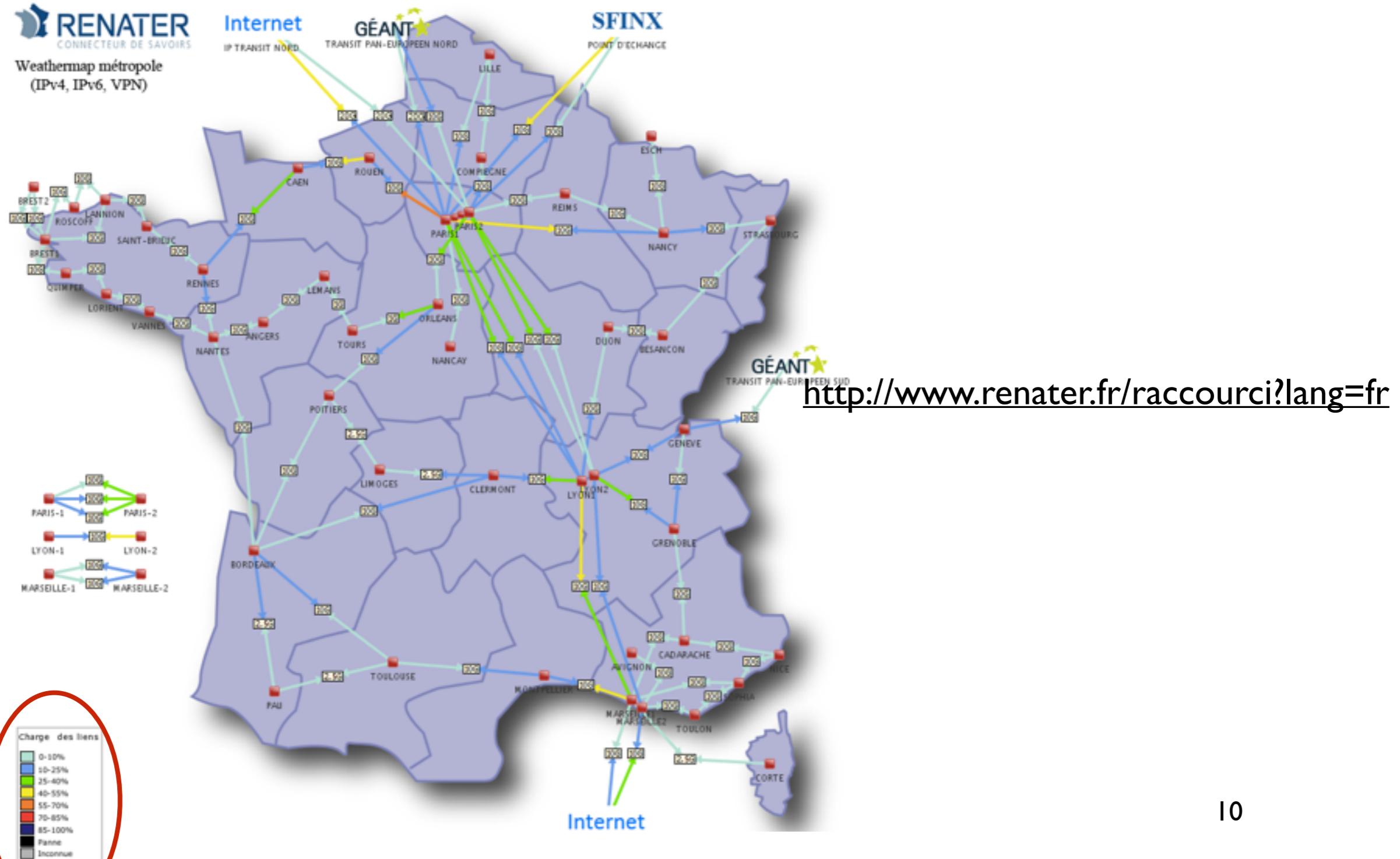
A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.



Beyond the Clouds, the DISCOVERY Initiative

• Locality-based UC infrastructures

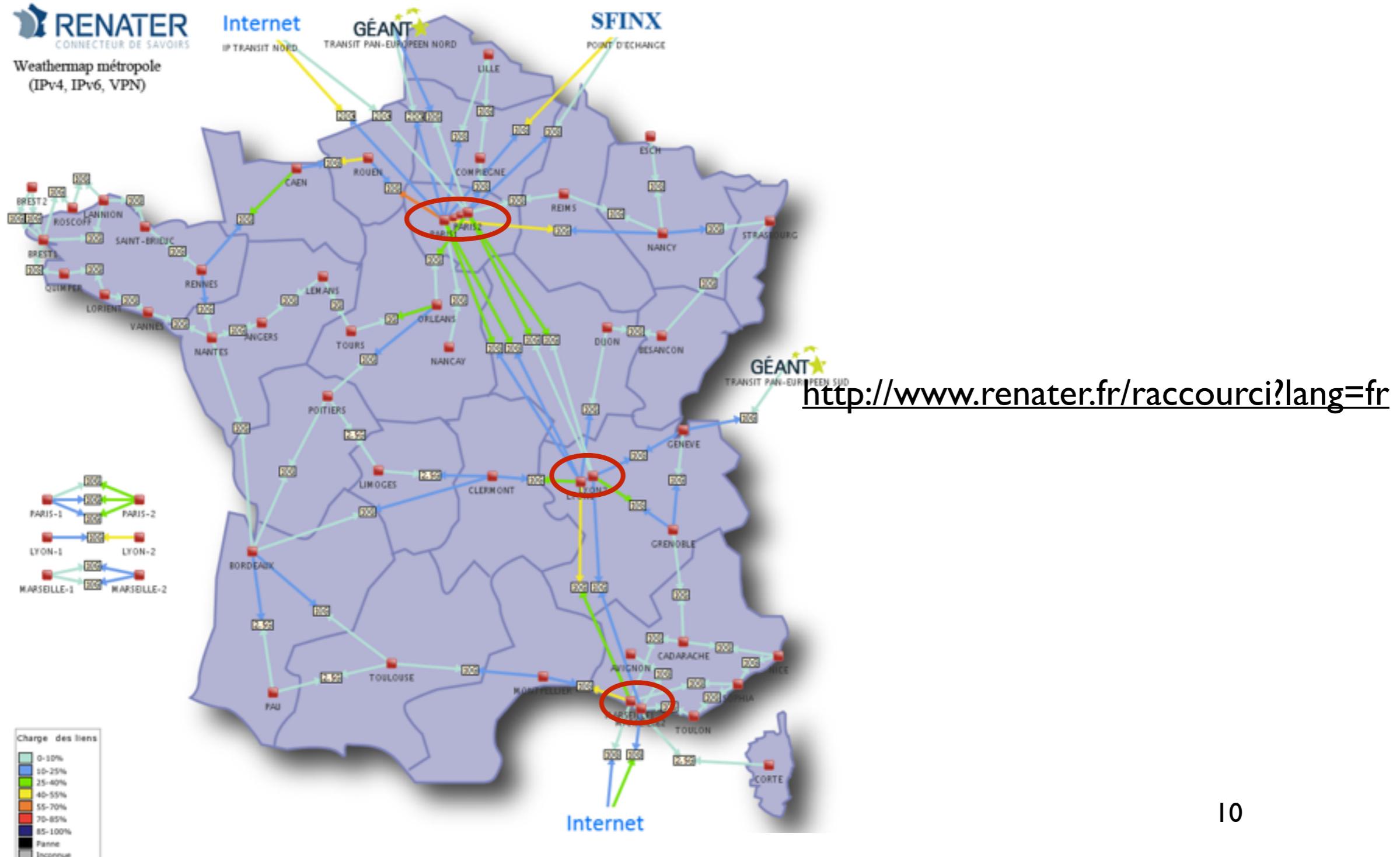
A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.



Beyond the Clouds, the DISCOVERY Initiative

• Locality-based UC infrastructures

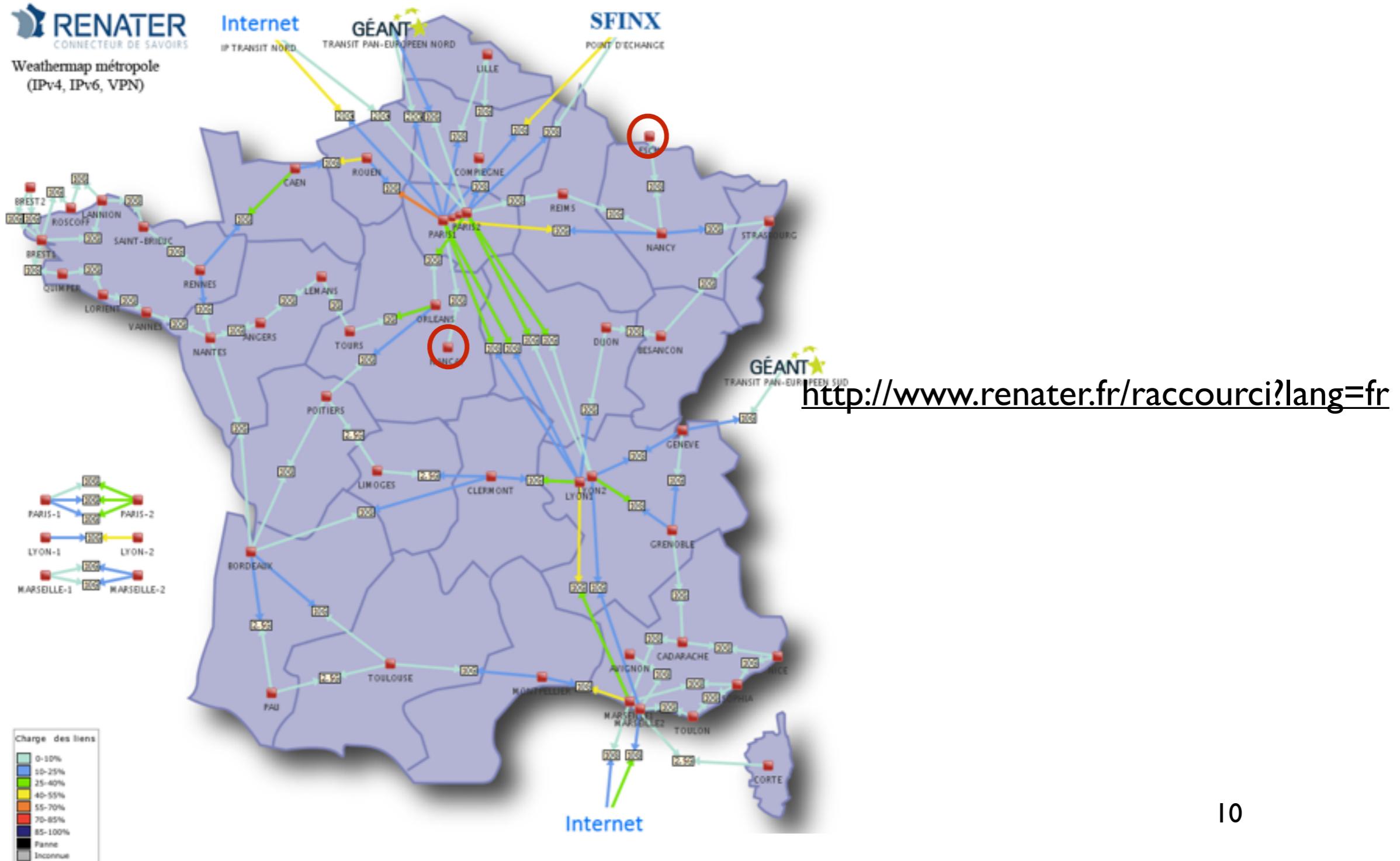
A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.



Beyond the Clouds, the DISCOVERY Initiative

• Locality-based UC infrastructures

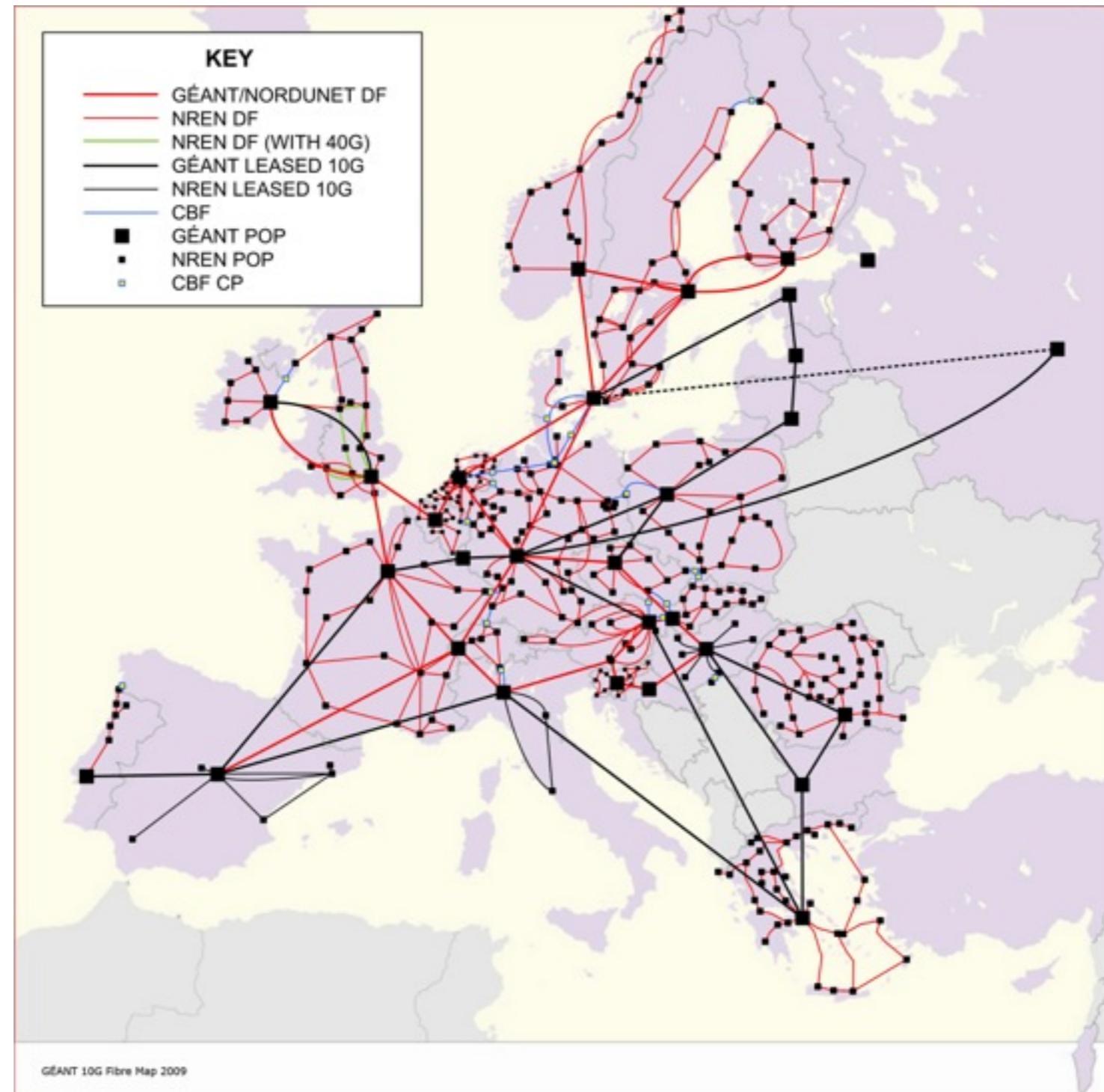
A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.



Beyond the Clouds, the DISCOVERY Initiative

- Locality-based UC infrastructures

A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.



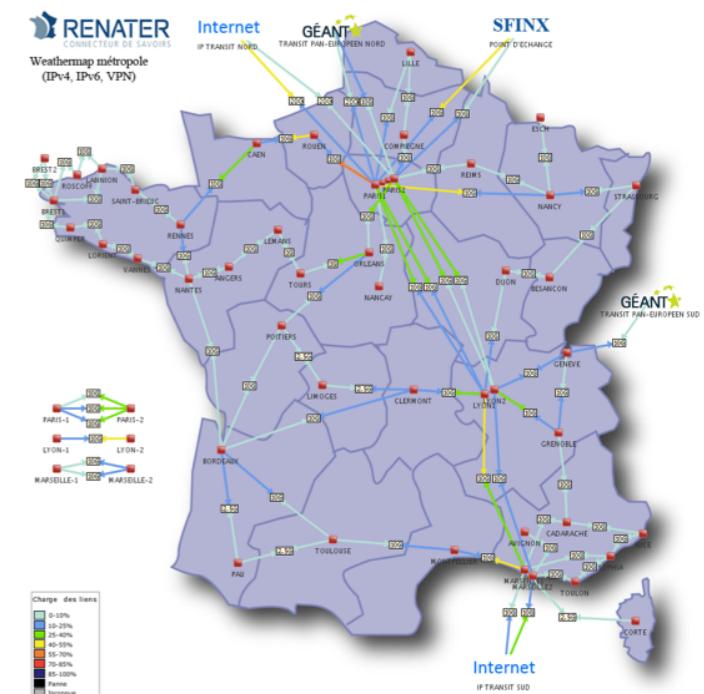
Beyond the Cloud, the DISCOVERY Initiative

- Locality-based UC infrastructures

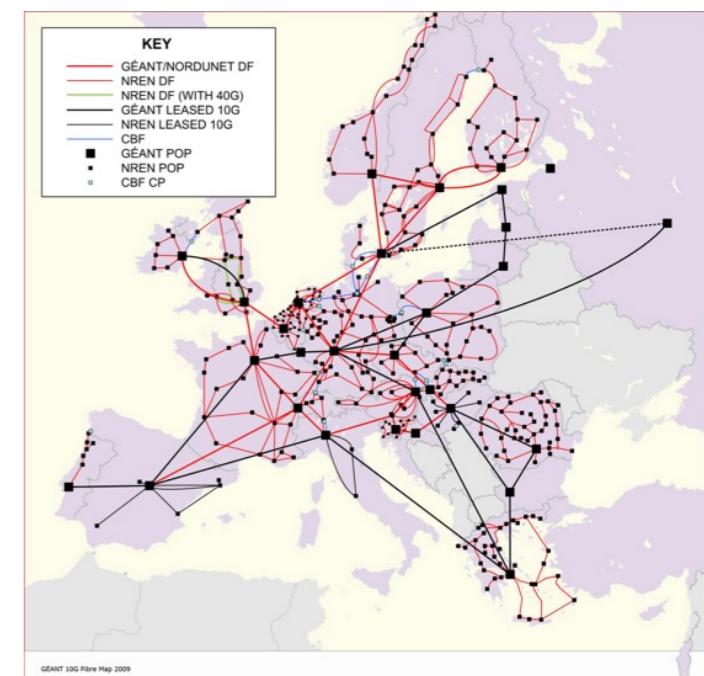
A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users.

- Leveraging network backbones

Extend any point of presence of network backbones with UC servers (from network hubs up to major DSLAMs that are operated by telecom companies and network institutions).

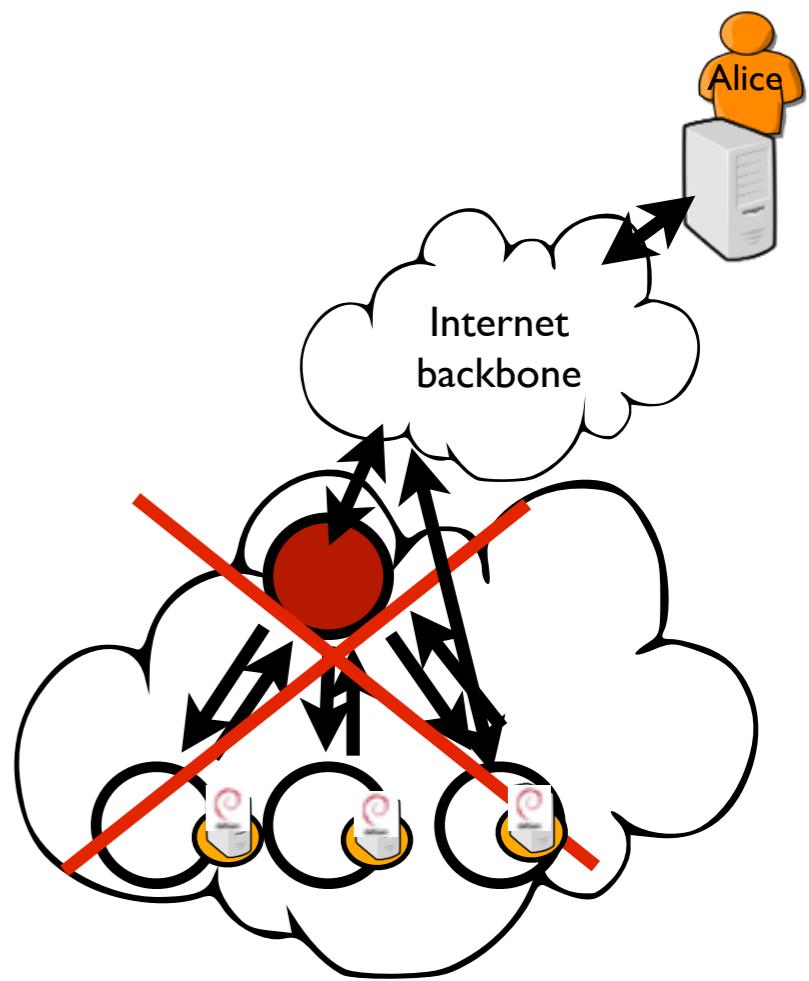


⇒ Operating such widely distributed resources requires the definition of a fully distributed system



The DISCOVERY Proposal

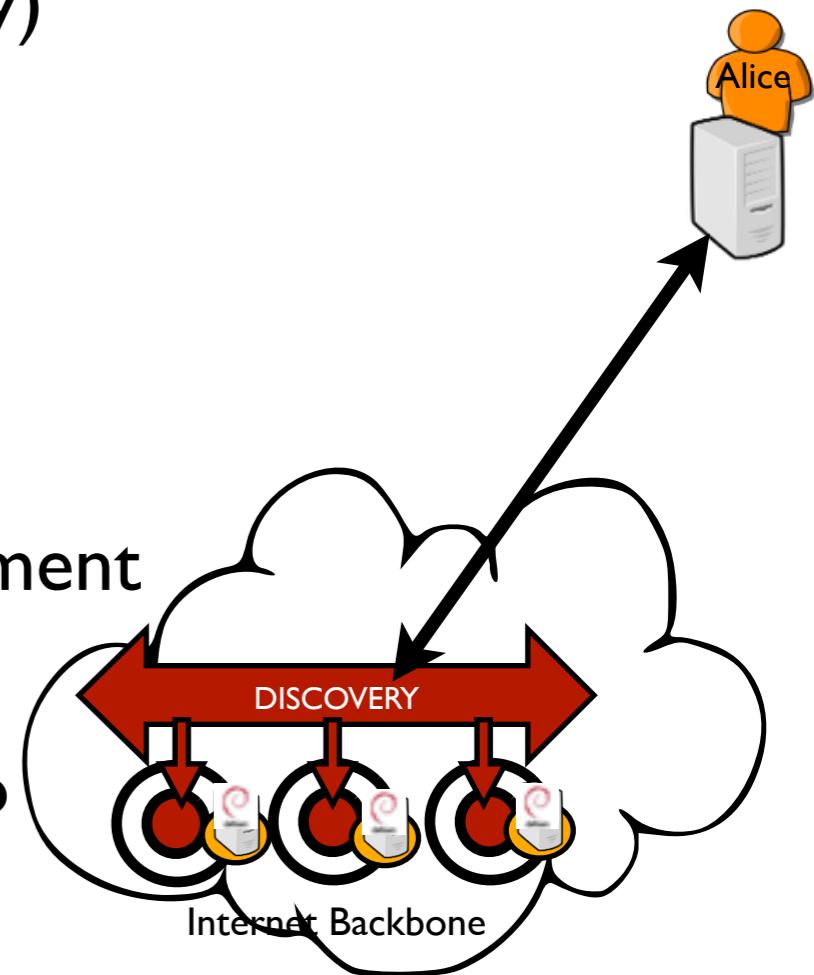
- DIStributed and COoperative framework to manage Virtual EnviRonsments autonomously



The DISCOVERY Proposal

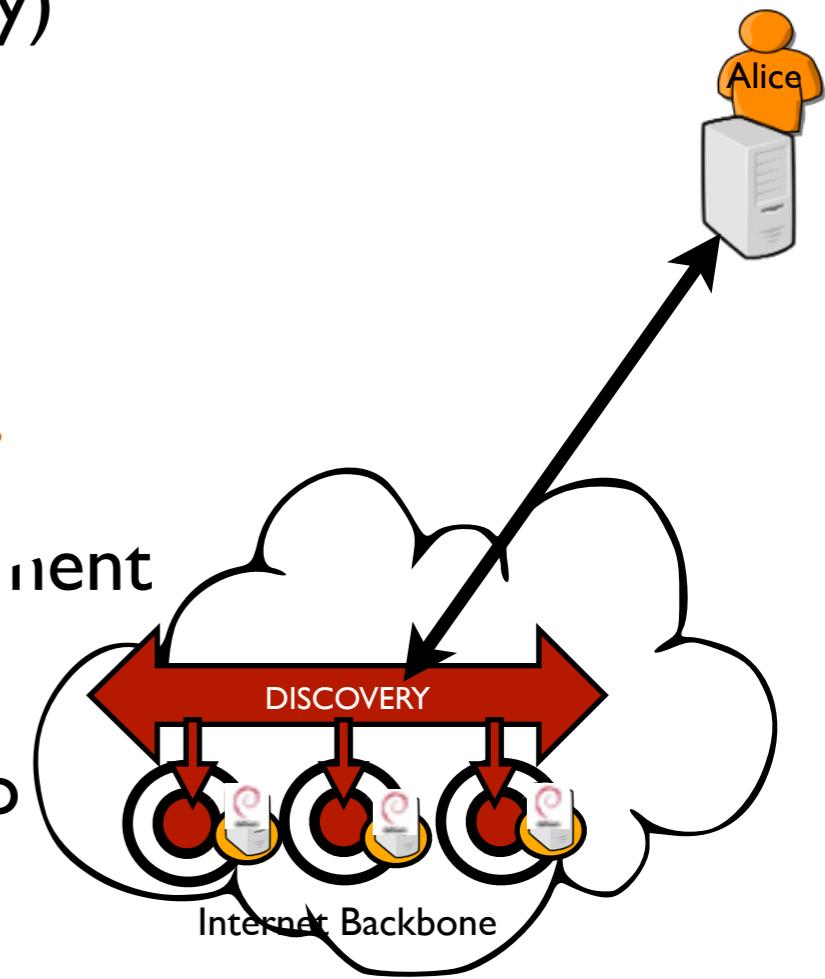
- DIStributed and COoperative framework to manage Virtual EnviRonsments autonomously
- The LUC OS
 - A fully distributed IaaS system and not a distributed system of IaaS systemS. We want to/must go further than high level cloud APIs (cross-cutting concerns such as energy/security)
 - Leverage P2P algorithms and self-* approaches
- lots of scientific/technical challenges

Cost of the network !? partial view of the system !? Impact on the others VMs !?, management of VM images !? Which software abstractions to make the development easier and more reliable (distributed event programming)? How to take into account locality aspects? ...



The DISCOVERY Proposal

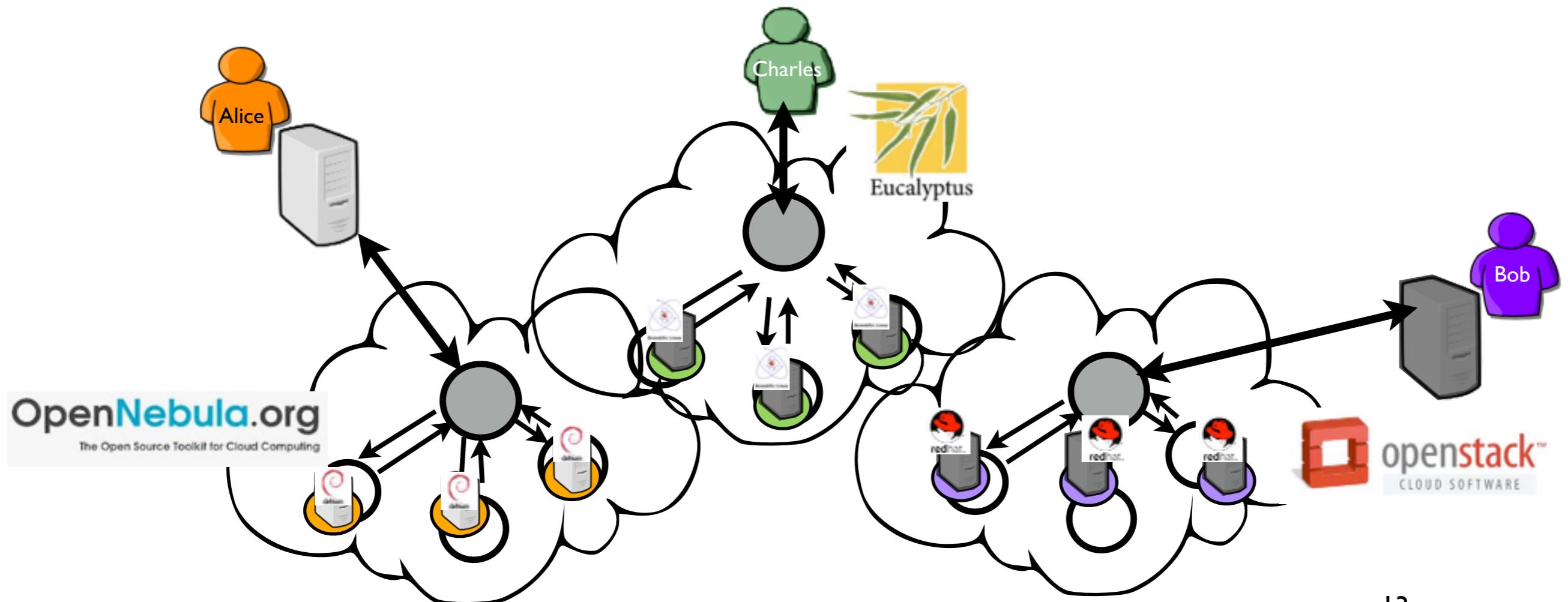
- DIStributed and COoperative framework to manage Virtual EnviRonsments autonomously
- The LUC OS
 - A fully distributed IaaS system and not a distributed system of IaaS systemS. We want to/must go further than high level cloud APIs (cross-cutting concerns such as energy/security)
 - Leverage P2P algorithms and self *
- lots of scientific/technological issues
 - Cost
 - ?? A distributed version of the EGI Core that directly manipulates resources
 - <http://www.egi.eu/infrastructure/cloud> ??
 - more reliable ...ing)? How to take into account ...?
 - ...



Why not a broker ?

- “federation of clouds” (sky computing)

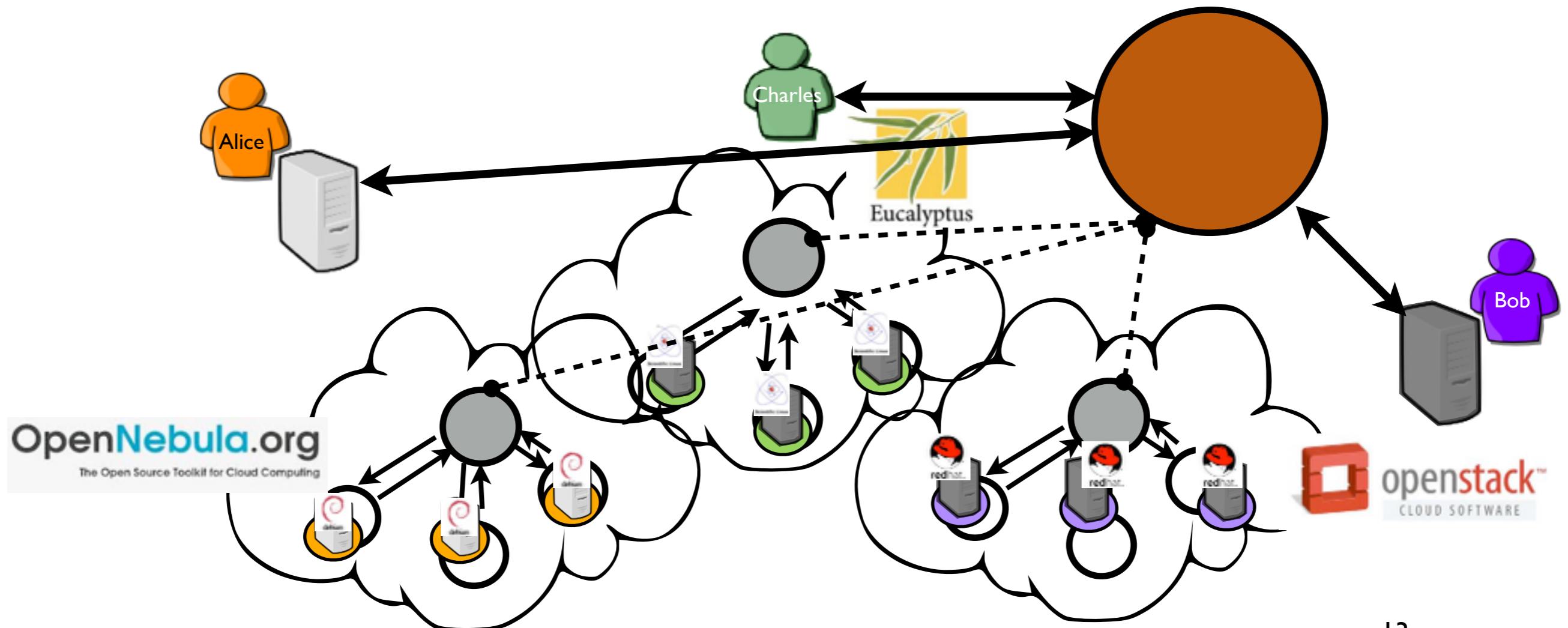
Sporadic (hybrid computing/cloud bursting) almost ready for production
While standards are coming (OCCI, OVF,), current brokers are rather limited



Why not a broker ?

- “federation of clouds” (sky computing)

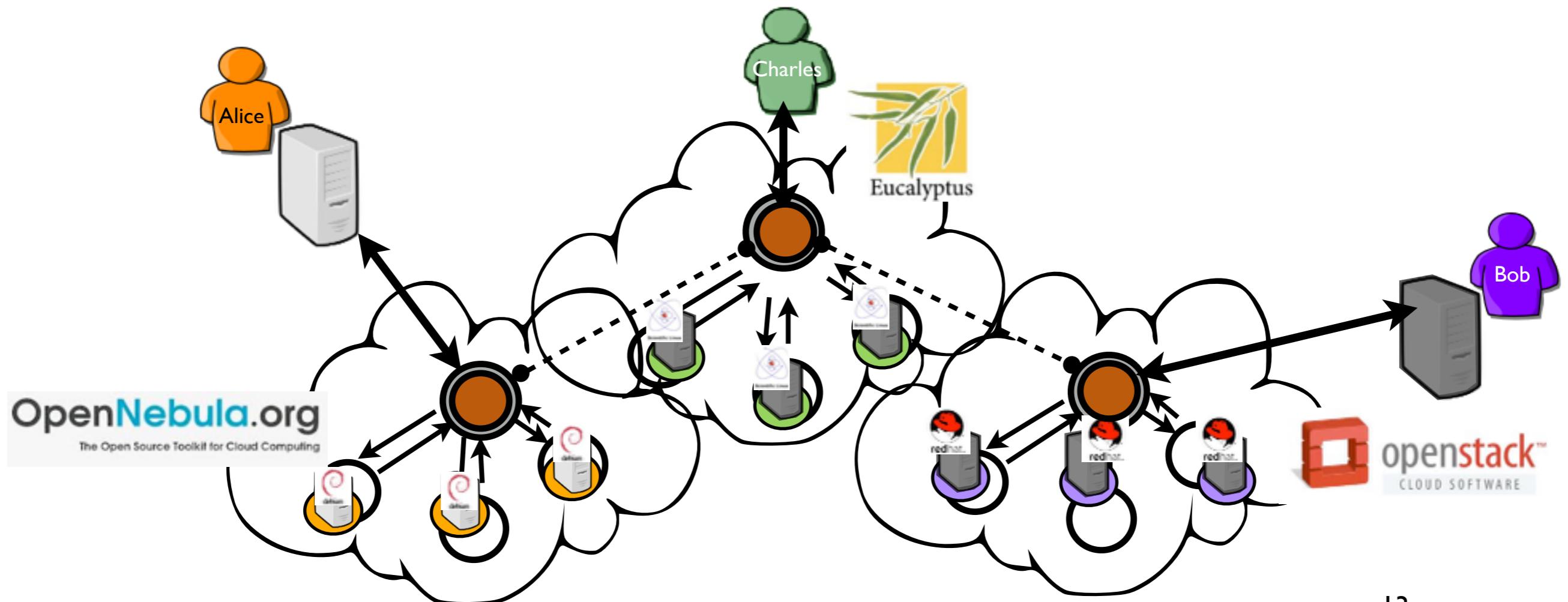
Sporadic (hybrid computing/cloud bursting) almost ready for production
While standards are coming (OCCI, OVF,), current brokers are rather limited



Why not a broker ?

- “federation of clouds” (sky computing)

Sporadic (hybrid computing/cloud bursting) almost ready for production
While standards are coming (OCCI, OVF,), current brokers are rather limited

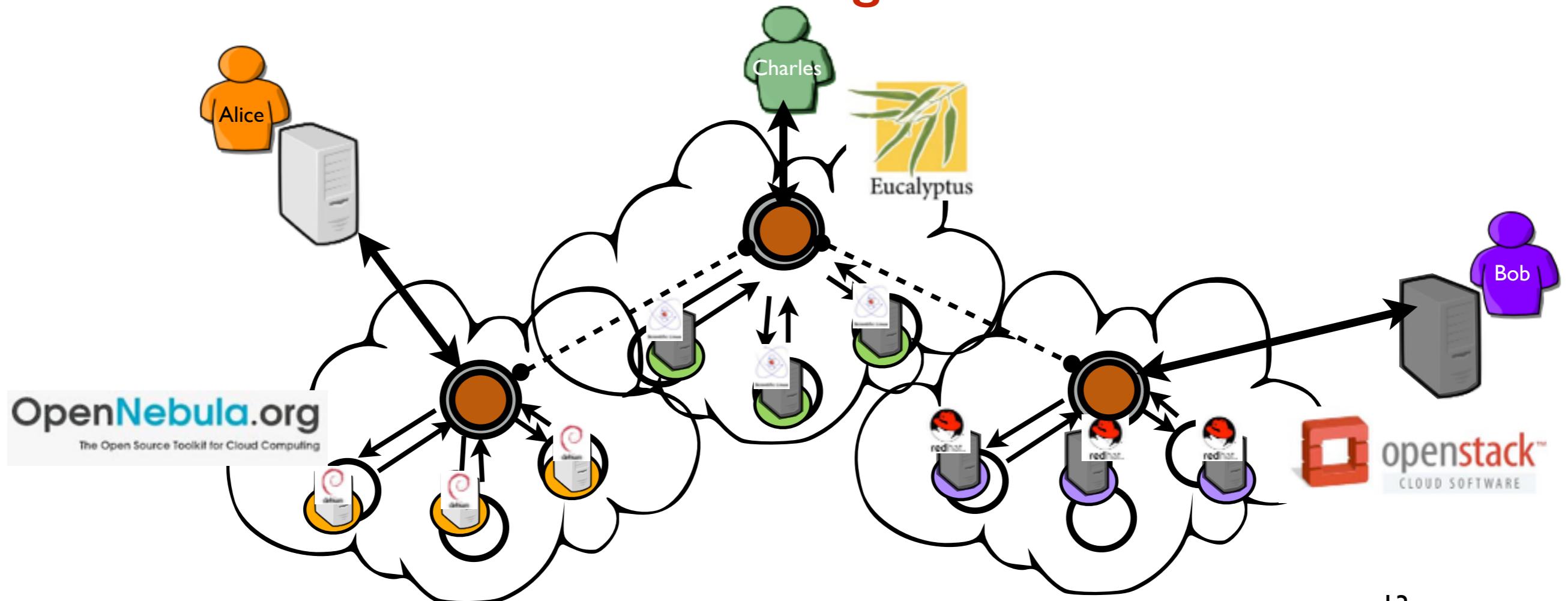


Why not a broker ?

- “federation of clouds” (sky computing)

Sporadic (hybrid computing/cloud bursting) almost ready for production
While standards are coming (OCCI, OVF,), current brokers are rather limited

Advanced brokers must reimplement standard IaaS mechanisms while facing the API limitation



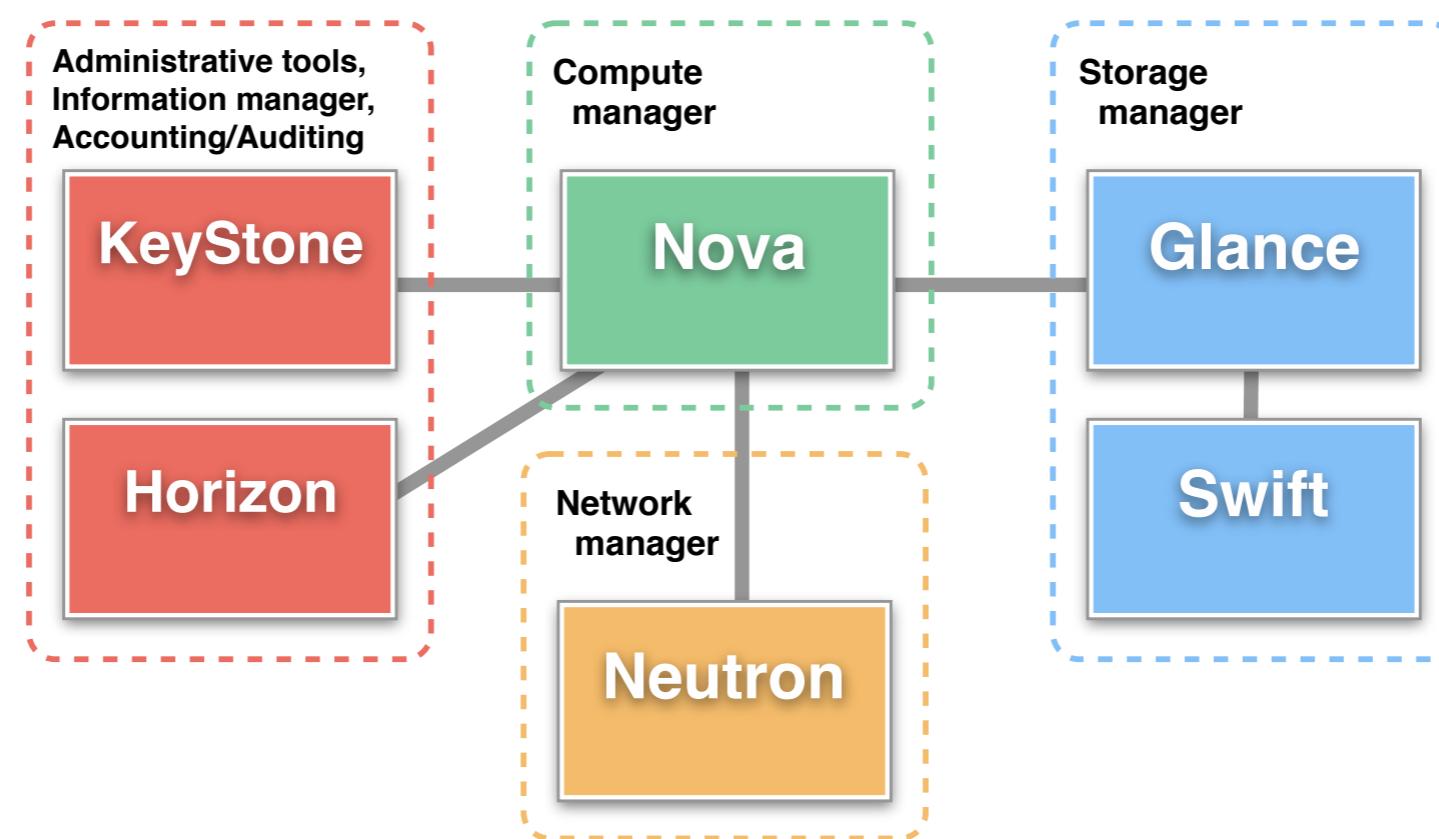
Would OpenStack be the solution?

- Do not reinvent the wheel ...it is too late
- OpenStack



Open source IaaS manager with a large community

Composed of several services dedicated to each aspect of a cloud



Would OpenStack be the solution?

- Do not reinvent the wheel ...it is too late
- OpenStack



Open source IaaS manager with a large community
Composed of several services dedicated to each aspect of a cloud

The screenshot shows a web browser displaying the OpenStack Architecture Design Guide. The header includes the OpenStack logo, a search bar, and navigation links for 'CONTENTS', 'SEARCH', 'SIDEBAR', 'PREV | UP | NEXT', and help icons. The main content area is titled 'Technical considerations' and contains sections on 'Infrastructure segregation', 'Host aggregates', 'Availability zones', and 'Segregation example'. A detailed paragraph discusses repurposing an environment for scalability. Below this, a section on 'Infrastructure segregation' notes the challenge of scaling database management systems and message queues. At the bottom, a note about traditional clustering techniques is present.

Technical considerations

[Infrastructure segregation](#)
[Host aggregates](#)
[Availability zones](#)
[Segregation example](#)

Repurposing an existing OpenStack environment to be massively scalable is a formidable task. When building a massively scalable environment from the ground up, ensure you build the initial deployment with the same principles and choices that apply as the environment grows. For example, a good approach is to deploy the first site as a multi-site environment. This enables you to use the same deployment and segregation methods as the environment grows to separate locations across dedicated links or wide area networks. In a hyperscale cloud, scale trumps redundancy. Modify applications with this in mind, relying on the scale and homogeneity of the environment to provide reliability rather than redundant infrastructure provided by non-commodity hardware solutions.

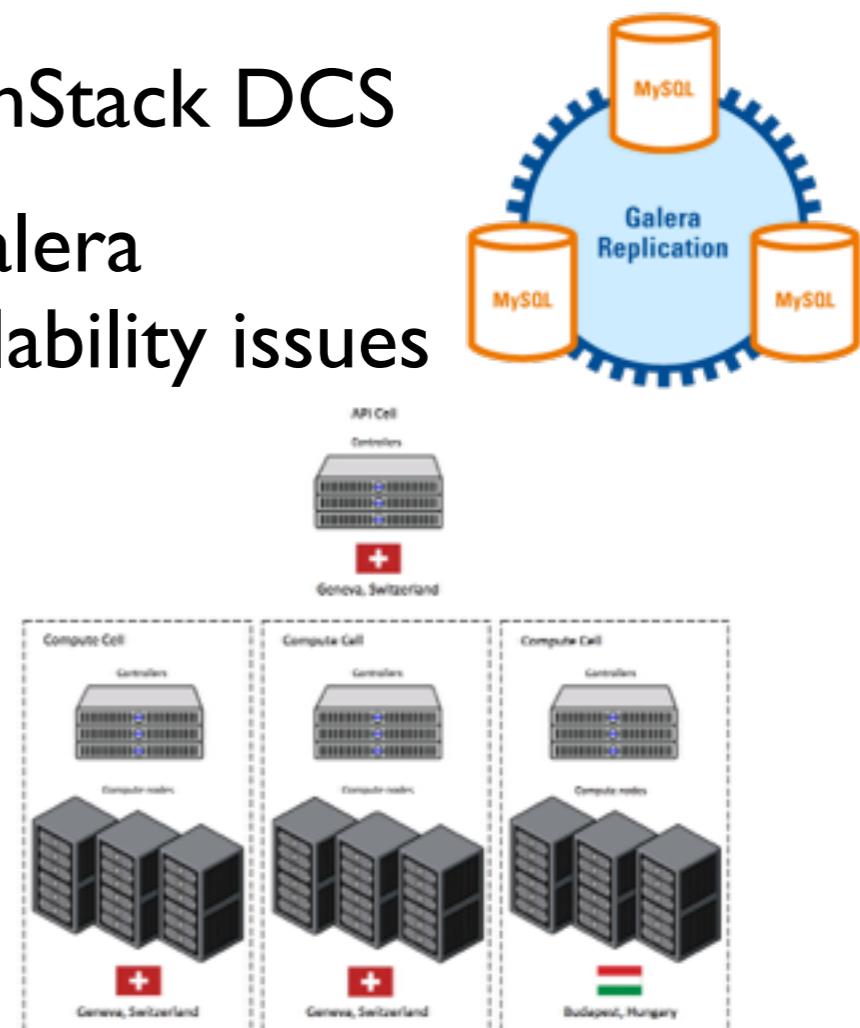
Infrastructure segregation

OpenStack services support massive horizontal scale. Be aware that this is not the case for the entire supporting infrastructure. This is particularly a problem for the database management systems and message queues that OpenStack services use for data storage and remote procedure call communications.

Traditional clustering techniques typically provide high availability and some additional scale for these environments. In the quest for massive scale, however, you must take additional steps to relieve the performance pressure on these components in order to prevent them from negatively impacting the overall performance of the environment. Ensure that all the components are in balance so that if the massively scalable environment fails, all the components are near maximum capacity and a single component is not causing the failure.

Distributing OpenStack

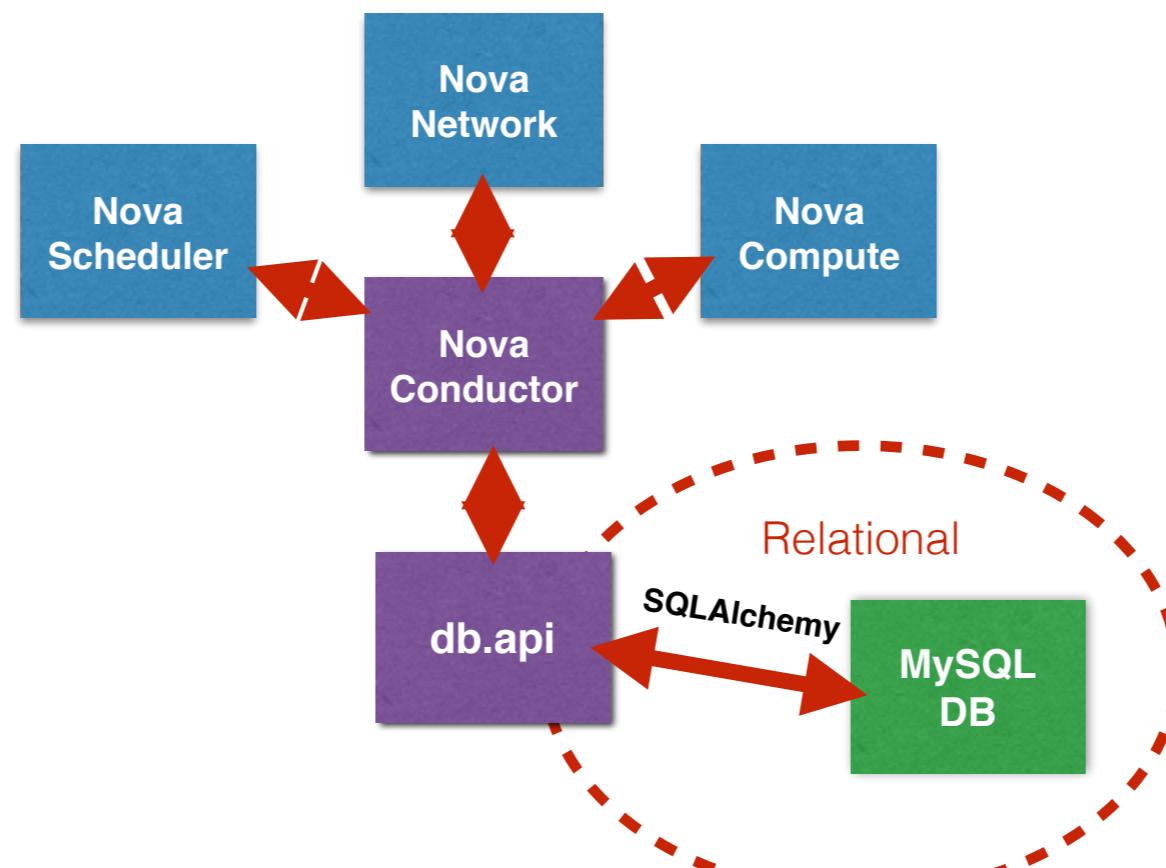
- Services collaborate through
 - A messaging queue 
 - A SQL database 
- Few proposals to federate/operate distinct OpenStack DCS
 - ‘Flat’ approach: leveraging HaProxy and Galera (Active replication) ⇒ Complexity and scalability issues
 - Hierarchical approaches:
 - Cells based (CERN: 3 Sites / 50K cores)
⇒ the top cell is a SPOF
 - Cascading OpenStack.
- You know others!? please mail us!



<http://beyondtheclouds.github.io/dcc.html>

Leveraging a key/value store DB

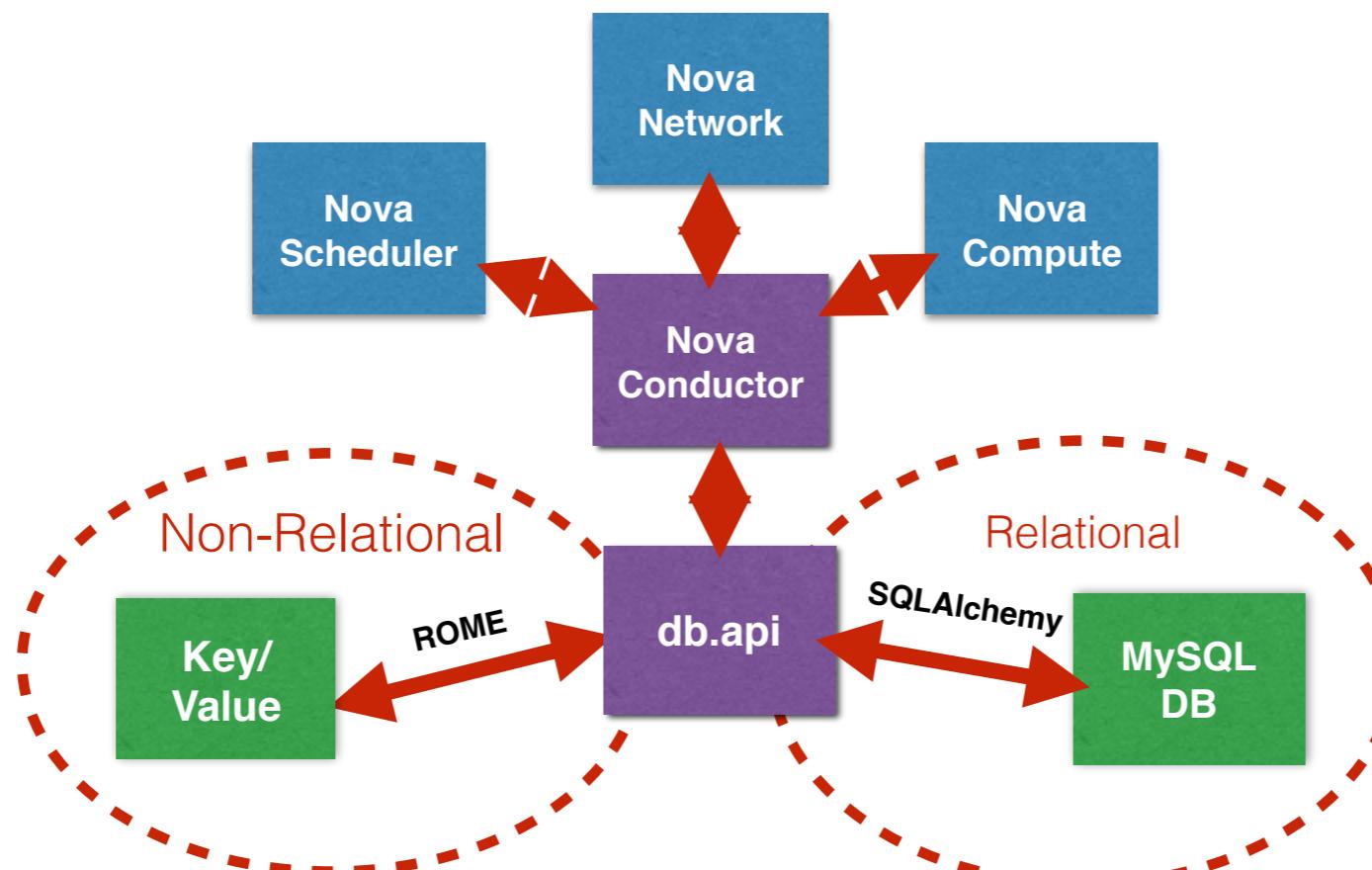
- Alternate solutions exists for storing states over a highly distributed infrastructure ⇒ NoSQL DB
- How can we switch between a SQL solution and a NoSQL system for storing inner states of OpenStack ?



Nova (compute service) - software architecture

Leveraging a key/value store DB

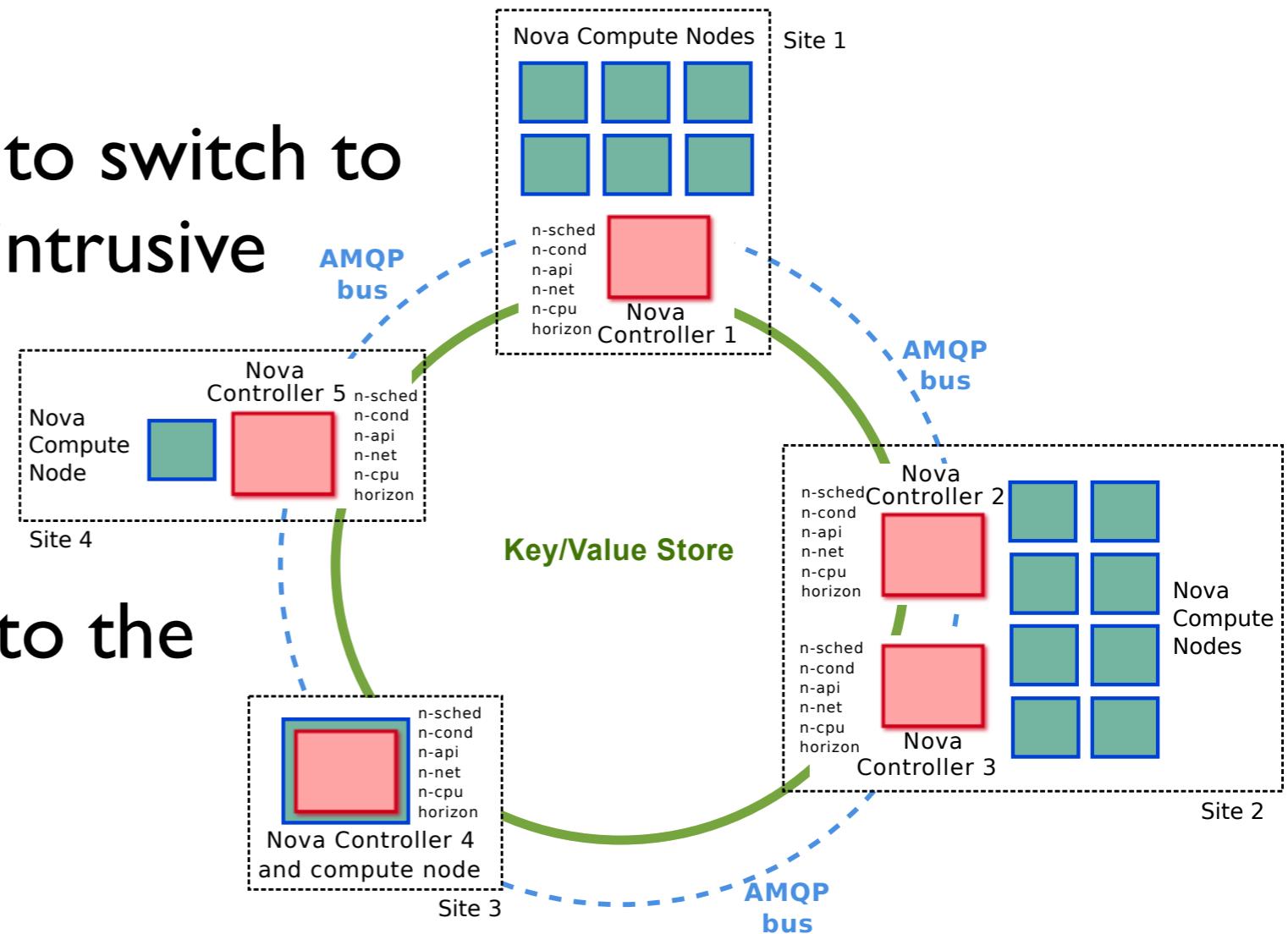
- Alternate solutions exists for storing states over a highly distributed infrastructure ⇒ NoSQL DB
- How can we switch between a SQL solution and a NoSQL system for storing inner states of OpenStack ?



Nova (compute service) - software architecture

ROME

- Relational Object Mapping Extension for key/value stores
Jonathan Pastor's Phd - <https://github.com/badock/rome>
- Enables the query of KVS (key/value store) with the same interface as SQLAlchemy
- Enables Nova OpenStack to switch to a KVS without being too intrusive
- The Key/value store is clustered on controller
- Compute nodes connect to the Key/value cluster



The Discovery Initiative Pros/Cons

- Pros

- Locality (jurisdiction concerns, latency-aware apps, minimize network overhead)

- Reliability/redundancy (no critical point/location/center)

- The infrastructure is naturally distributed throughout multiple areas

- Lead time to delivery

- Leverage current PoPs and extend them according to UC demands

- Energy footprint (on-going investigations with RENATER)

- Bring back part of the revenue to NRENs/Telcos*

- Cons

- Security concerns (in terms of who can access to the PoPs)

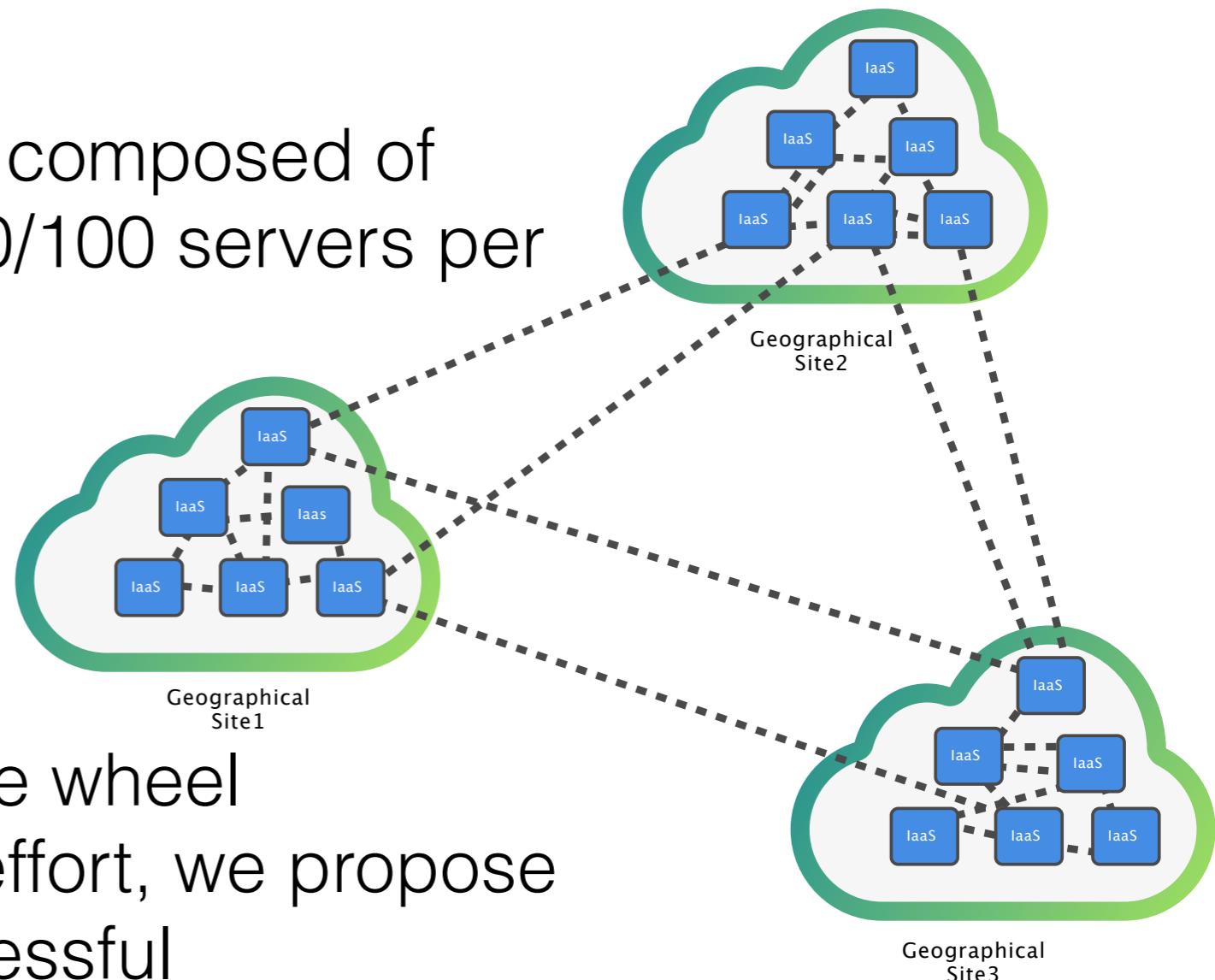
- Operate a fully IaaS in a unified but distributed manner at WAN level

- Not suited for all kinds of applications : Large tightly coupled HPC workloads
50 nodes/1000 cores, 200 nodes / 4000 cores (5 racks),
so 1000 nodes in one PoP does not look realistic ...

- Peering agreement / economic model between network operators

In short

- Building an infrastructure composed of 100/1000 remote sites (10/100 servers per site).
- Target collaboration without any broker.
- But without reinventing the wheel to limit the development effort, we propose to leverage existing successful mechanisms: **OpenStack**.



Conclusion

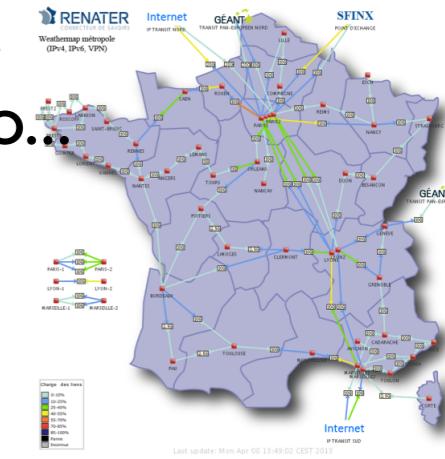
- Cloud Computing technology is changing every day
 - New features, new requirements (IaaS ++ services)
 - One more challenge will be to ensure that such new features/mechanisms can run in a distributed manner.
- Distributed Cloud Computing is happening !
 - Dist. CC workshop (UCC 2013, SIGCOMM 2014/2015)
 - FOG Computing workshop (collocated with IEEE ICC 2013)
 - IEEE CloudNet ...
 - More and more academic papers

One major challenge of the next H2020 call related to Cloud Computing

Beyond Discovery !

- From sustainable data centers to a new source of energy

A promising way to deliver highly efficient and sustainable UC services is to provide UC platforms as close as possible to the end-users and to...



- Leverage “green” energy (solar, wind turbines...)

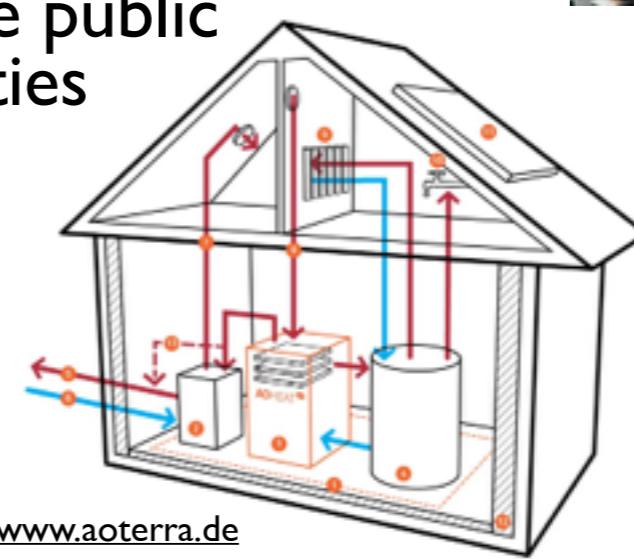
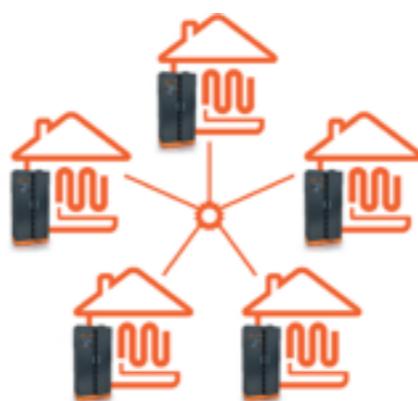
Transfer the green micro/nano DCs concept to the network PoP
Take the advantage of the geographical distribution



<http://parasol.cs.rutgers.edu>

- Leveraging the data furnaces concept

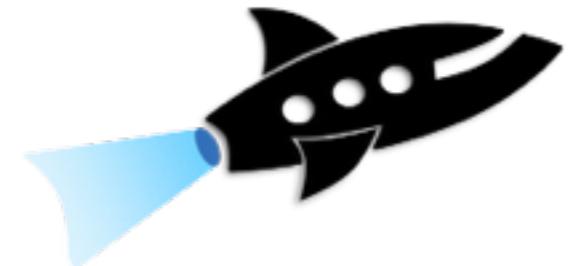
Deploy UC servers in medium and large institutions and use them as sources of heat inside public buildings such as hospitals or universities



<https://www.aoterra.de>

The DISCOVERY Initiative

- Thank you / Questions ?
- Several researchers, engineers, stakeholders of important EU institutions and SMEs have been taking part to numerous brainstorming sessions (BSC, CRS4, Unine, EPFL, PSNC, Interoute, Orange Labs, Peerialism, TBS Group, XLAB, ...)



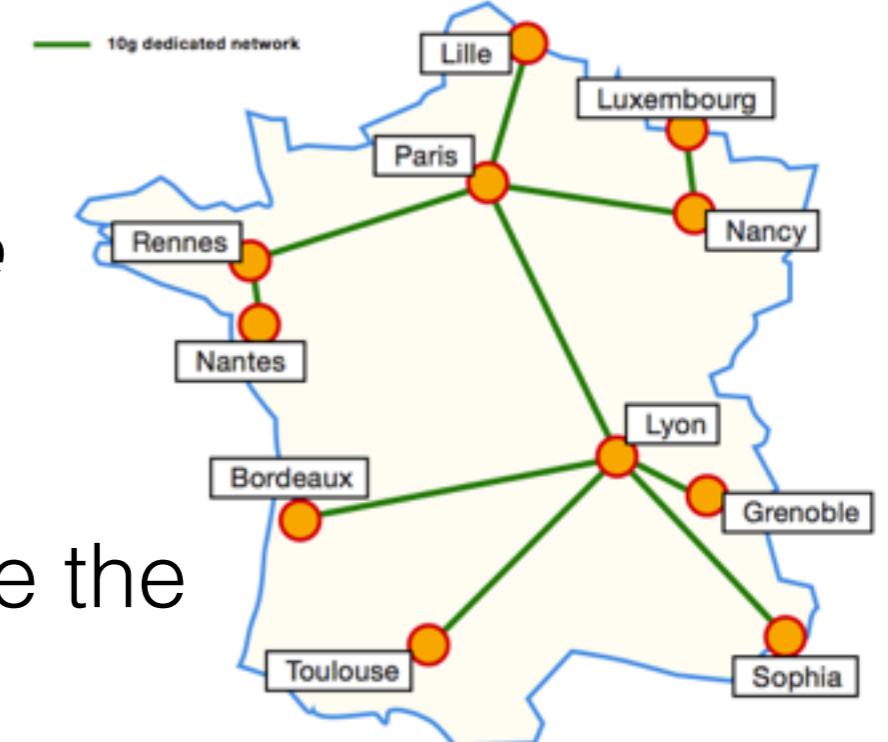
<http://beyondtheclouds.github.io/>



adrien.lebre@inria.fr

Experiments

- Preliminary experiments have been conducted on Grid'5000.
- ***mono-site experiments:*** to evaluate the overhead of using REDIS and the network impact.
- ***multi-site experiments:*** To determine the impact of latency.
- Ask for the creation of 500 VMs, fairly distributed on each controller.



Preliminary results

- **Time measured for creating 500 VMs in parallel.**
- Experiments performed on servers with homogeneous hardware.
- For a fair comparison (routing issues can disturb Galera):
use servers on the same site (Rennes).
- Clusters were simulated by adding latency between nodes with TC.
- **We followed configuration advised by OpenStack multi-site documentation.**

10 ms intersite latency

	Redis	MySQL (no replication)	Galera
1 cluster (no replication)	298	357	-
2 clusters	271	209	2199
3 clusters	280	157	3243
4 clusters	263	139	2011

50 ms intersite latency

	Redis	MySQL (no replication)	Galera
1 cluster (no replication)	298	357	-
2 clusters	723	268	1361
3 clusters	518	210	2202
4 clusters	427	203	1253