

# COMMIT/

## E-INFRASTRUCTURE VIRTUALIZATION FOR E-SCIENCE APPLICATIONS/

### TOWARDS A SMART INFRASTRUCTURE

#### BACKGROUND

Complex (computational and data) intensive applications have **strict requirements** regarding the infrastructure:

- reliable, efficient, secure, scalable
- BUT ALSO** easy to use, do not require drastic changes to the application.

COMMIT-P20 is developing methods & tools which when combined create a **smart infrastructure**:

- adapt automatically** to the volume of the data being processed
- exploit the elasticity of the Cloud to provide **reliable** and **secure on-demand** computing resources distributed over multiple Cloud providers.

#### SMART INFRASTRUCTURE

- fully programmable and controllable infrastructure
- Aware of application dynamics
- Automatically adapt to application data traffic/load
- Leverage on existing standards and state of art technologies from Clusters to Clouds

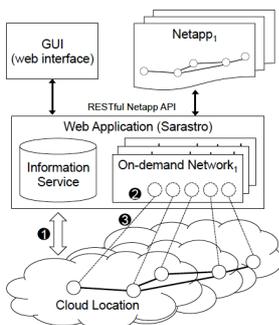


Figure: Architectural overview of the system

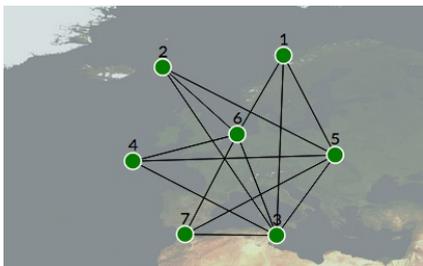


Figure: Active virtual machines multiple cloud locations in Europe. The network is controlled and adapted to the current data flow.

#### IN PRACTICE

The workflow for analysing brain regions using MRI and DTI is:

- Long running application **48 hours** on commodity hardware
- Can be executed in one or more phases
- Data privacy** must be guaranteed
- Repeated for each patient part of the study (~200 patients)

**It will take more than a YEAR to complete the study**

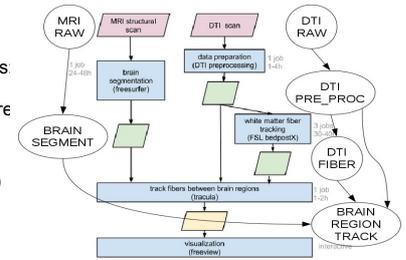


Figure: Workflow used by AMC for tracking brain fibers between regions; superimposed is the data state transition model for the workflow used in Pumpkin.

#### SOLUTION

- No up front investment in a new infrastructure
- Instantaneous on-demand resource provisioning
- Automatically adapt to the applications load
- scaling up and down
- Optimize the usage of available resources
- Resilient to network and resource failure
- No Vendor Locking
- Security
- Secure communication Channels (VPN)
- Secure Computing resources.

**Complete the Analysis of Brain Regions study on 200 subjects in a couple of DAYS**

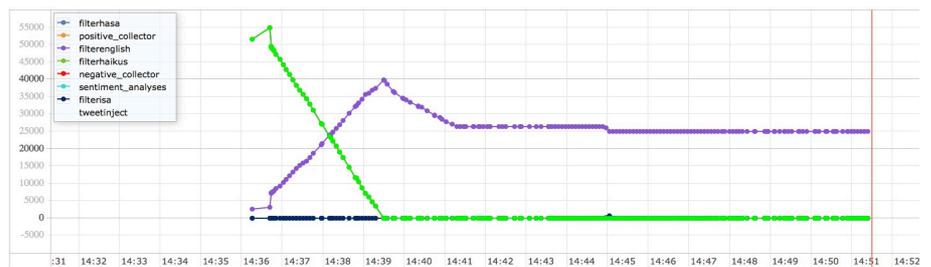


Figure: Live processing and networking statistics from individual processing functions in the virtual network.

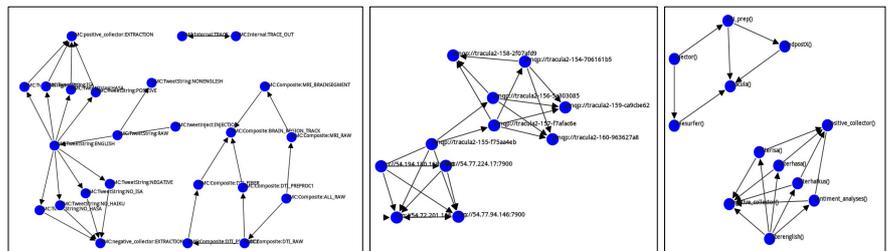


Figure: from left to right: (1) Data state transition diagram capable by the network. The state graph depicts two distinct applications on the network. (2) The workflow as a function network which performs the data transitions; two separate workflows can be seen. (3): The actual cloud network i.e. connections between VMs where communicating functions are hosted.

#### More details about this research can be found:

- Interactive Factories at Super Computing 2013: <https://www.youtube.com/watch?v=DP5TLqW1hW4>
- Intercloud demonstration at Super Computing 2011: <https://www.youtube.com/watch?v=WlnZ6pSEJhs>
- Creating a Software Defined Network on-demand: <https://www.youtube.com/watch?v=GiIMqGCUVA>
- R. Cushing, et al., Exploratory Data Processing Using Non-deterministic Finite Automata, submitted to the BigDataCloud 2014 workshop in conjunction of Euro-Par 2014 Porto, Portugal
- R. Strijkers, et al, Internet Factories: Creating application-specific networks on-demand. Journal of Computer Networks 68 (2014) 187–198

Contributors:  
M.X. Mankers, R. Cushing,  
A.S.Z Belloum, R. Meijer, C.de  
Laat  
Projectleader:  
Prof. Dr. Henri Bal  
H.e.ba@vu.nl

COMMIT  
Science Park 123  
1098 XG Amsterdam  
www.commit-nl.nl  
buro@commit-nl.nl  
+31(0)20 592 4110

This publication was  
supported by the Dutch  
national program COMMIT  
COMMIT is a public-  
private research community

