First Workshop on

Automated Control for Datacenters and Clouds (ACDC09)

June 19th, Barcelona, Spain

Elastic Management of Cluster-based Services in the Cloud

Rafael Moreno-Vozmediano, Ruben S. Montero, Ignacio M. Llorente

dsa-research.org

Distributed Systems Architecture Research Group
Universidad Complutense de Madrid

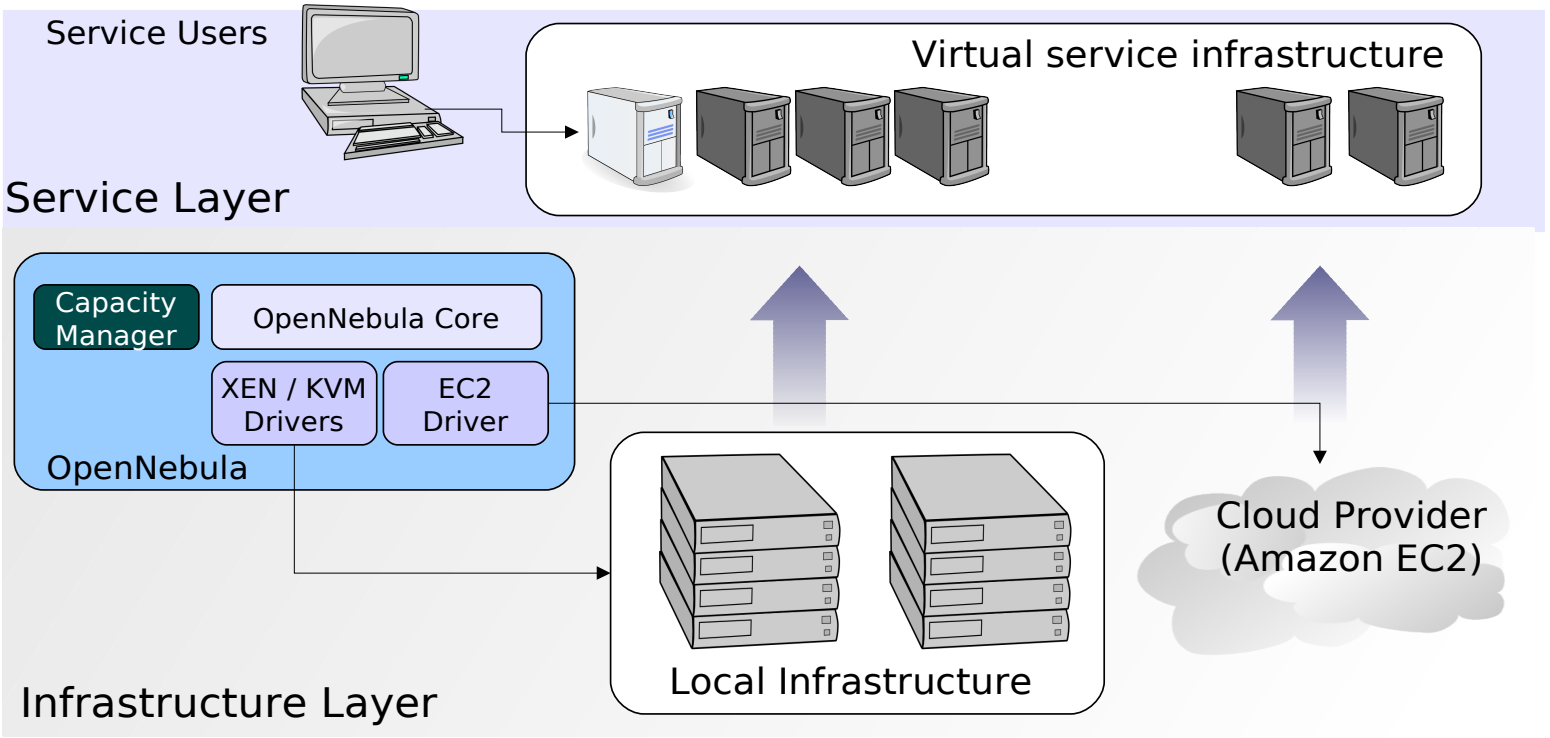
Contents

- Introduction
- Elastic management and implementation of clustered services
- Experimental results
- Conclusions
- Demo

dsa-research.org

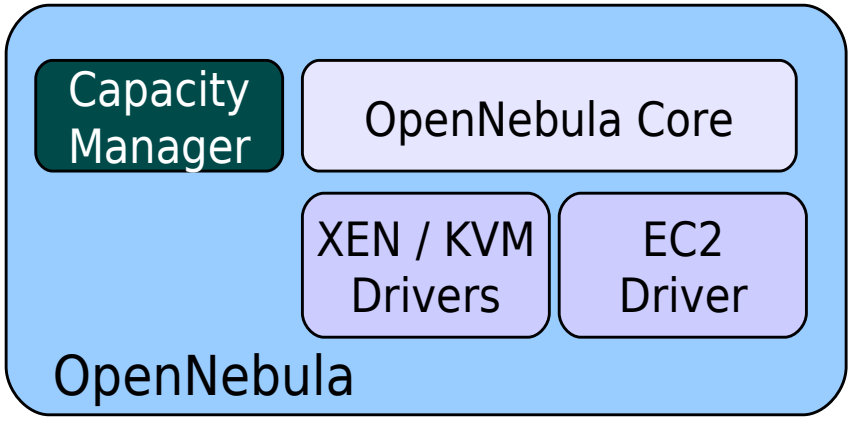
Introduction

- **Goals**
  - Deployment of clustered services on top of a virtualized infrastructure consisting of:
    - Local (in-house) virtual infrastructure
    - Cloud resource provider (Amazon EC2)
    - VM manager (the OpenNebula engine)



- **Motivation of the work**
  - The proposed architecture offers several benefits
    - The virtualization layer allows to separate the infrastructure management (VM provisioning) from the service management
    - Elastic provisioning of virtual resources on demand, to adapt the infrastructure to the service requirements.
    - Fully transparent for the service itself, and independent of the type of service
    - The integration with the cloud provides additional capacity to the virtual infrastructure

- **The VM Manager (OpenNebula)**



http://OpenNebula.org

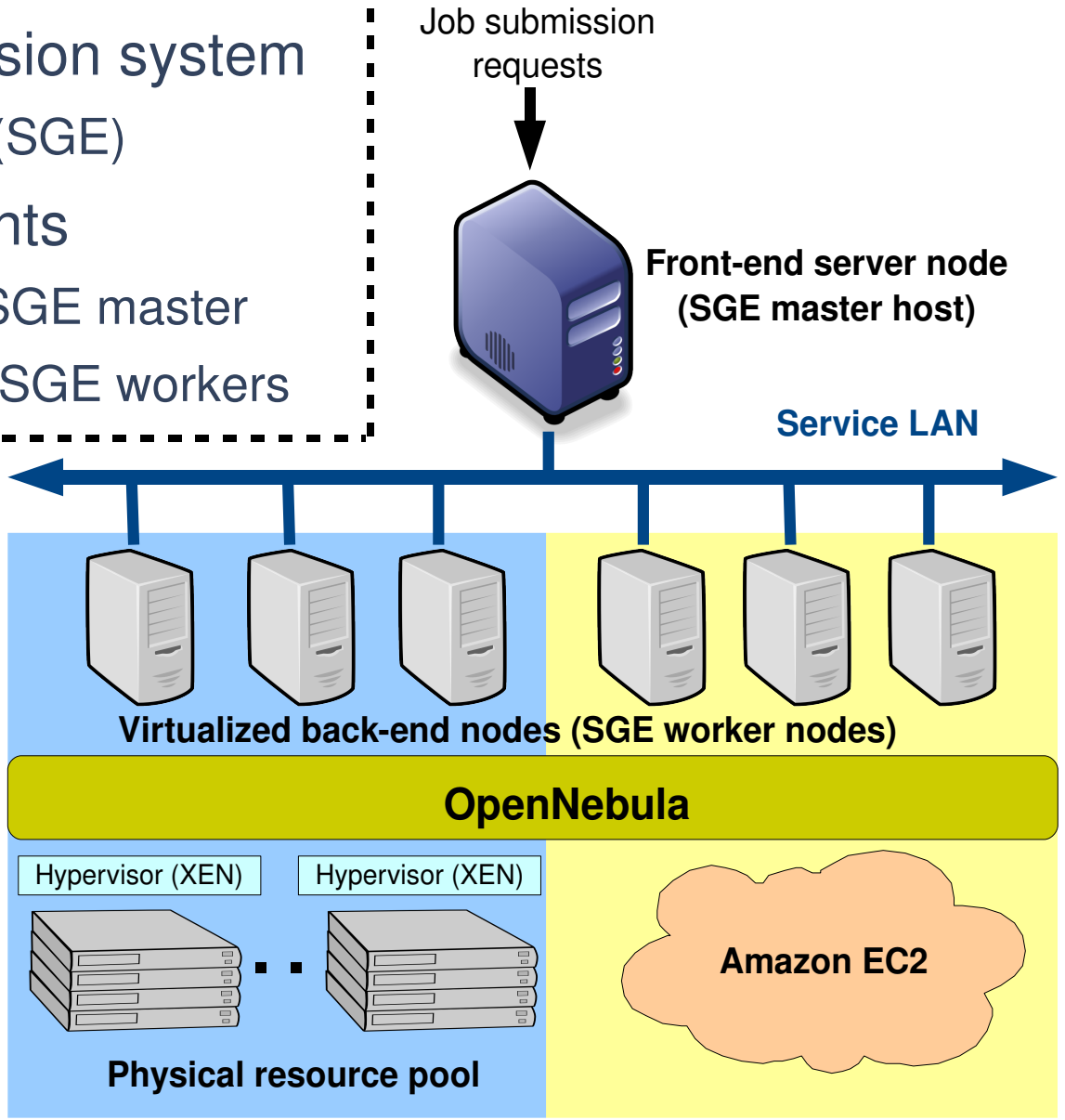
- **OpenNebula Core:** manages the life-cycle of a VM by performing basic VM operations (e.g. deployment, monitoring, or termination)
- **Capacity Manager (scheduler):** Adjusts the placement of VMs based on a set of configurable provisioning policies
- **Virtualizer Access Drivers:** Pluggable drivers that provide an abstraction of the underlying virtualization layer (Xen, KVM, EC2), by exposing the basic functionality of the hypervisor or cloud interface

Contents

- **Benefits**
  - **Elastic cluster capacity**
    - The capacity of the cluster can be modified by deploying (or shutting down) virtual cluster nodes on demand
  - **Cluster partitioning**
    - The available physical and cloud resources can be used to deploy VMs bound to different cluster-based services
    - This approach isolates the different workloads and the performance assigned to each cluster
  - **Heterogeneous configurations**
    - The VMs of a service could have multiple software configurations
    - The different service images can be maintained with a minimal operational cost ("*install once deploy many*" approach)

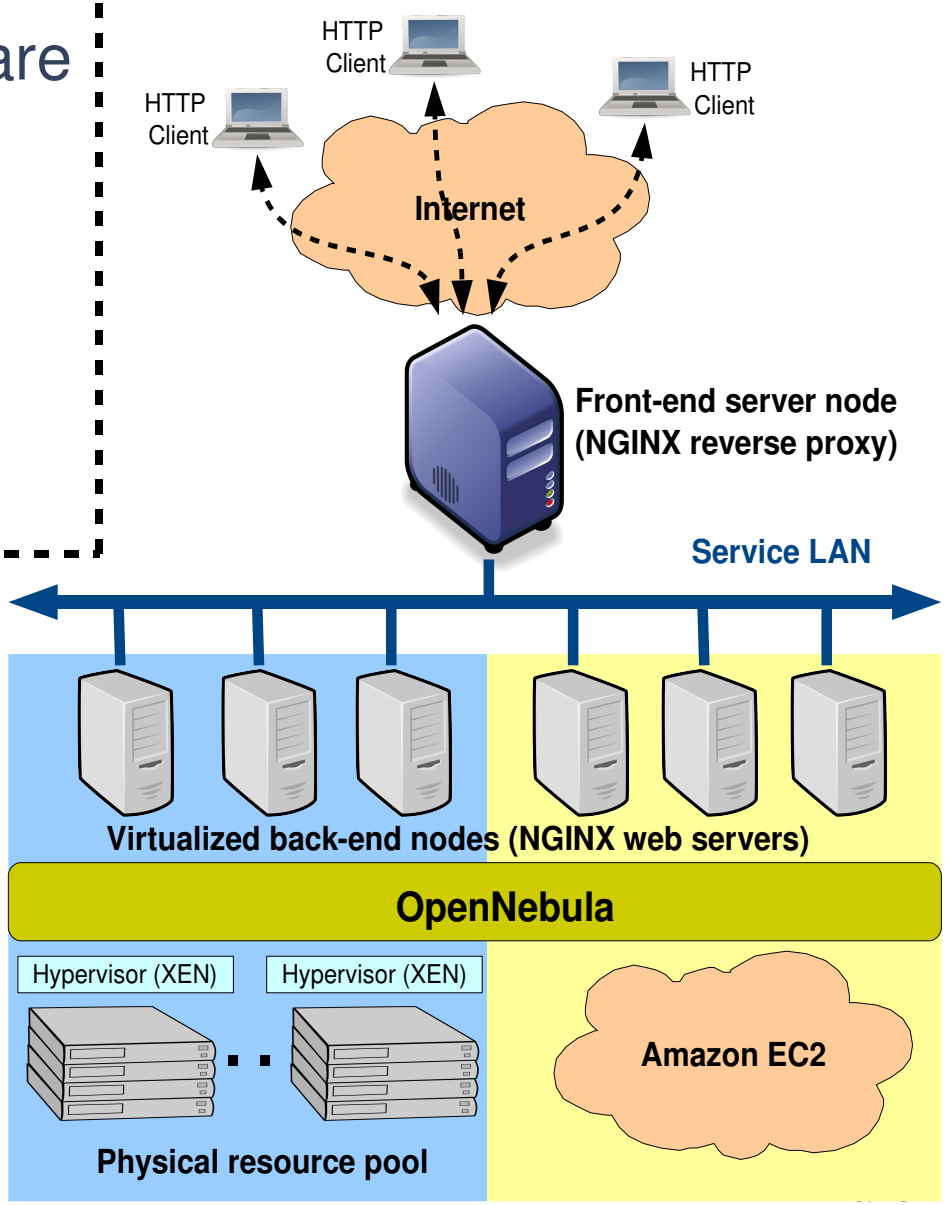
Implementation of clustered services in the cloud

- **Implementation of a computing cluster**
  - Batch job submission system
    - Sun Grid Engine (SGE)
  - Cluster components
    - Front-end node: SGE master
    - Back-end nodes: SGE workers



Implementation of clustered services in the cloud

- **Implementation of a web server cluster**
  - Based on NGINX software
  - Cluster components
    - Front-end node:
      NGINX reverse proxy
    - Back-end nodes:
      NGNIX web servers

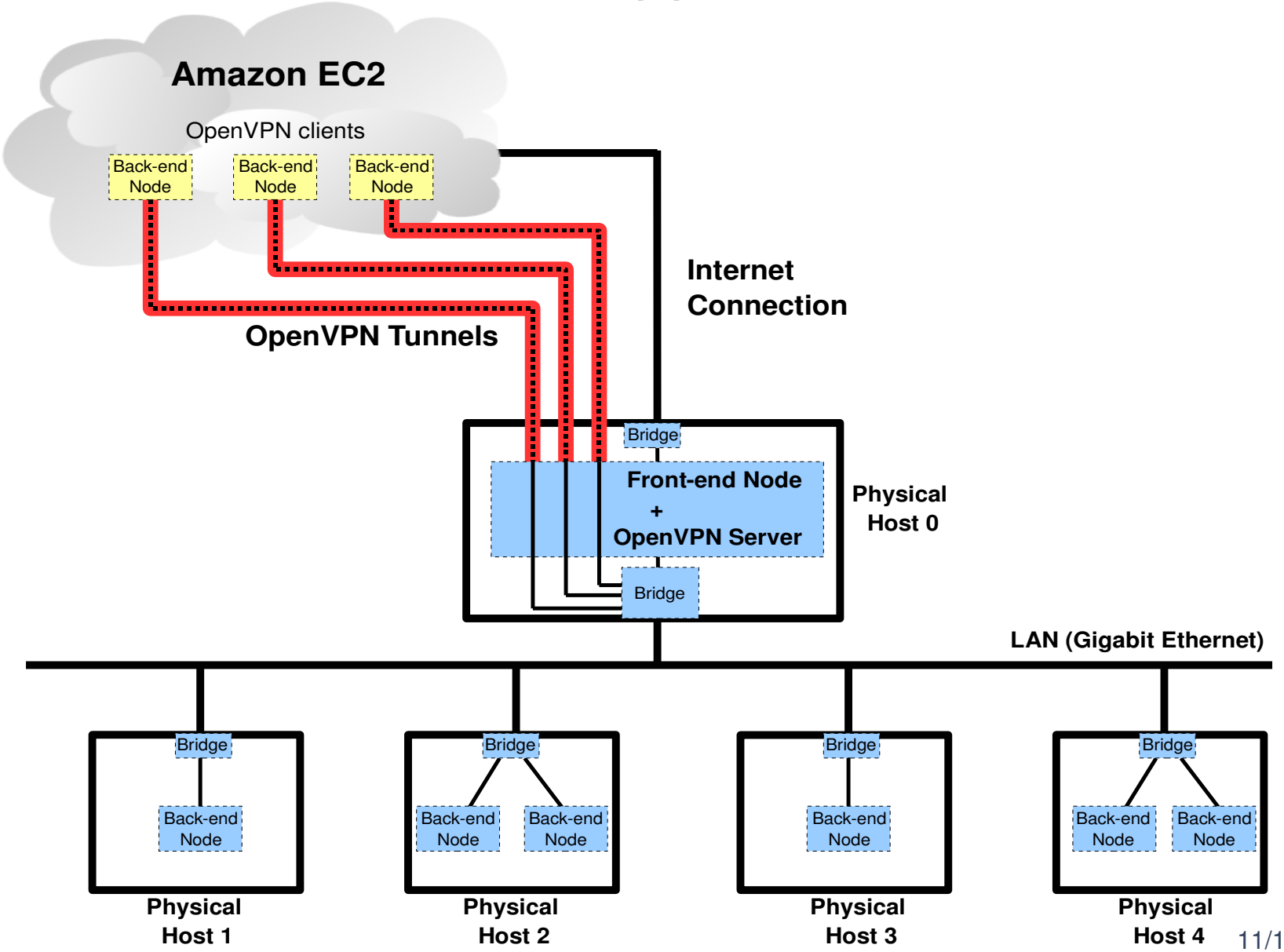


Implementation of clustered services in the cloud

- **The network infrastructure (i)**
  - The service LAN
    - Every virtual node (both local and Amazon EC2 nodes) communicates with the front-end through the service LAN (using private addresses)
  - The local (in-house) virtual nodes
    - Are directly attached to the service LAN by means of a **virtual bridge** configured in every physical host
  - The remote (Amazon EC2) virtual nodes
    - Are connected to the service LAN by means of a **virtual private network (VPN) tunnel**
    - The VPN tunnel is implemented with the OpenVPN software

Implementation of clustered services in the cloud

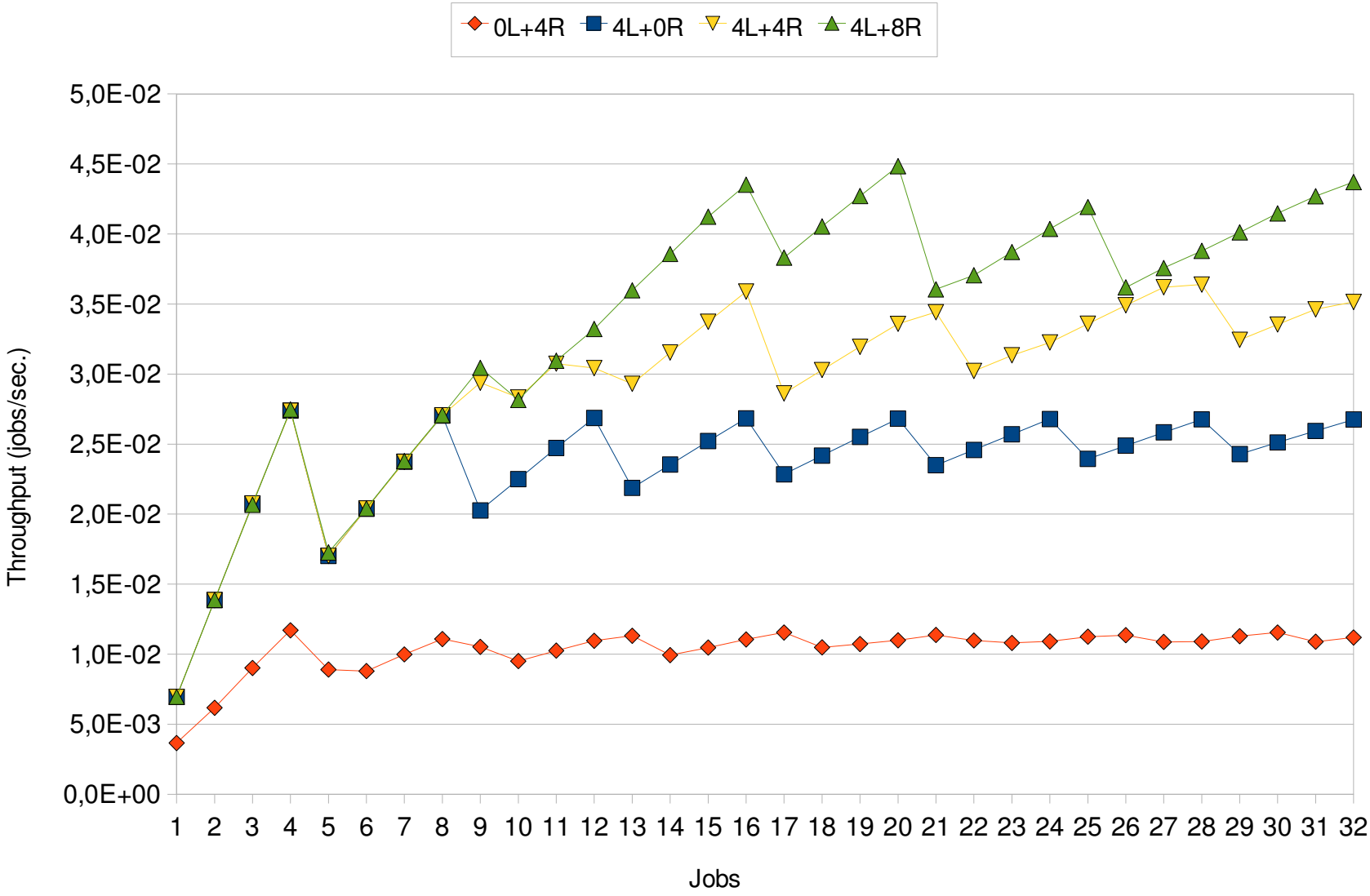
- The network infrastructure (ii)



Contents

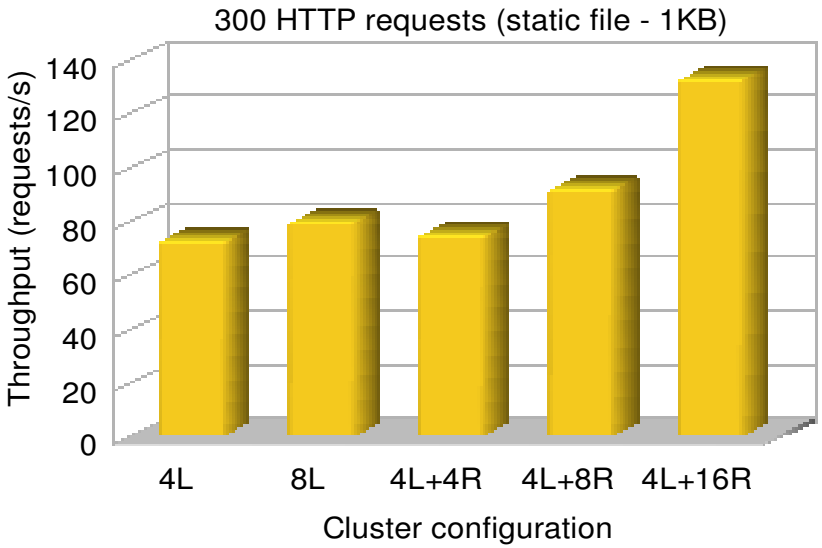
Experimental Results

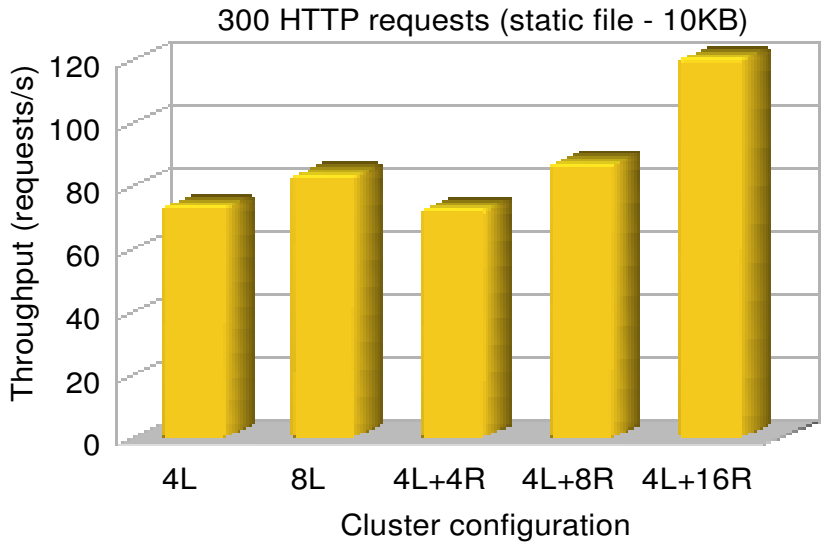
- Computing cluster performance

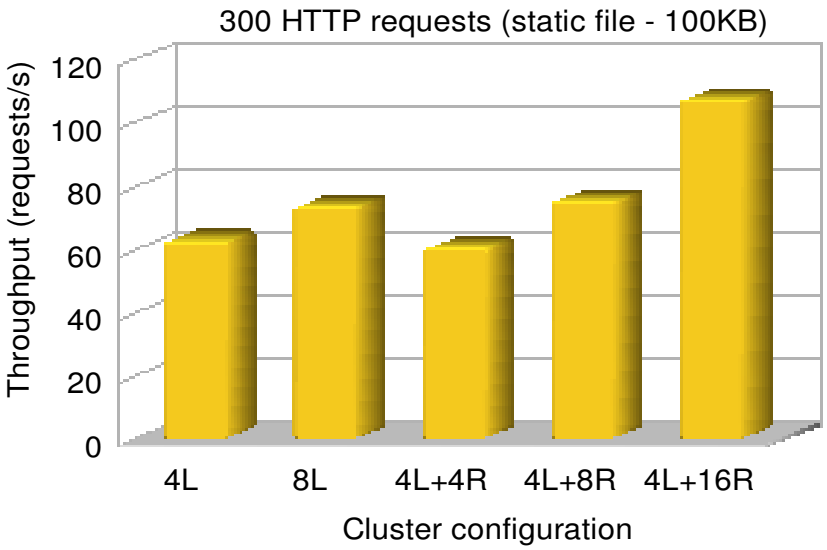


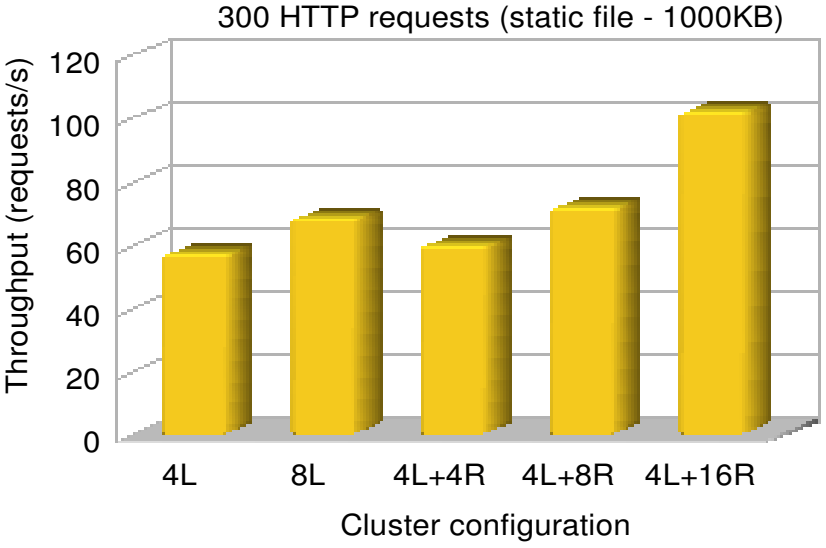
Experimental Results

- **Web server cluster performance**









Contents

- Introduction
- Elastic management and implementation of clustered services
- Experimental results
- Conclusions
- Demo

Conclusions

- We have analyzed the deployment of two cluster-based services on top of a virtualized infrastructure, with the capacity of integrating external cloud resources.
  - Computing cluster
  - Web server cluster
- This approach separates the resource provisioning from the service management, and provides important benefits:
  - Elastic capacity to adapt the cluster to its dynamic workload
  - Cluster partitioning to isolate it from other running services
  - Heterogeneous configurations tailored for each application class
- Performance results show
  - A sustained performance increment when adding a growing number of cloud nodes to the cluster
  - That proves the feasibility of the proposed architecture and provisioning model, and its capacity to support service elasticity.

Contents

THANK YOU FOR YOUR ATTENTION

QUESTIONS?