AVOID RESOURCE CONTENTION WITH E4C TECHNOLOGY

A TELECOM ITALIA USE CASE

Things to know

Overcommitment

VMware ESX is a hypervisor that enables impressive memory and CPU consolidation ratios. ESX allows running VMs with total configured resources that exceed the amount available on the physical machine. This is called overcommitment.
Overcommitment raises the consolidation ratio, increases operational efficiency, and lowers total cost of operating virtual machines.

Contention

If out of control, overcommitment leads to Resource Contention, that is the situation of several VMs competing over the same resources, waiting for the VMware scheduler to assign them.
This is the main reason for performance issues in virtualized environment and, as such, it must be seen as the first key performance indicator to monitor in a virtual farm.
Contention is measured via CPU Ready Time and Memory Ballooning.

CPU Ready Time

Definition: CPU Ready time is a metric showing how much time a virtual machine with work to do is waiting to have a physical (or Hyper Threaded) core scheduled by VMware CPU scheduler.
What represents: High CPU Ready time is a symptom of CPU contention.
Effects: In short, the more CPU Ready you see on your VMware Infrastructure, the worse off it is, leading to performance degradation on the virtual guests and bad end user experience.

Memory Ballooning

Definition: VMware ballooning is a memory reclamation technique used when and ESXi host is running low on memory. This allows the physical host system to retrieve unused memory from certain guest virtual machines (VMs) and share it with others.
What represents: Ballooned memory is a symptom of RAM memory contention. If host free memory drops towards the 4% threshold, the hypervisor starts to reclaim memory using ballooning.
Effects: VM memory ballooning can create performance degradation. Ballooning is a CPU intensive process, and can eventually lead to memory swapping, when a balloon driver inflates to the point where the VM no longer has enough memory to run its processes. This will slow down the VM, depending upon the amount of memory to recoup and/or the quality of the storage IOPS delivered to it.

Telecom Italia Use Case

from theory.. to case studies

After discussing the theoretical principles and the practical applications of innovative approaches for consolidation, let's analyze the **case study of Telecom Italia**

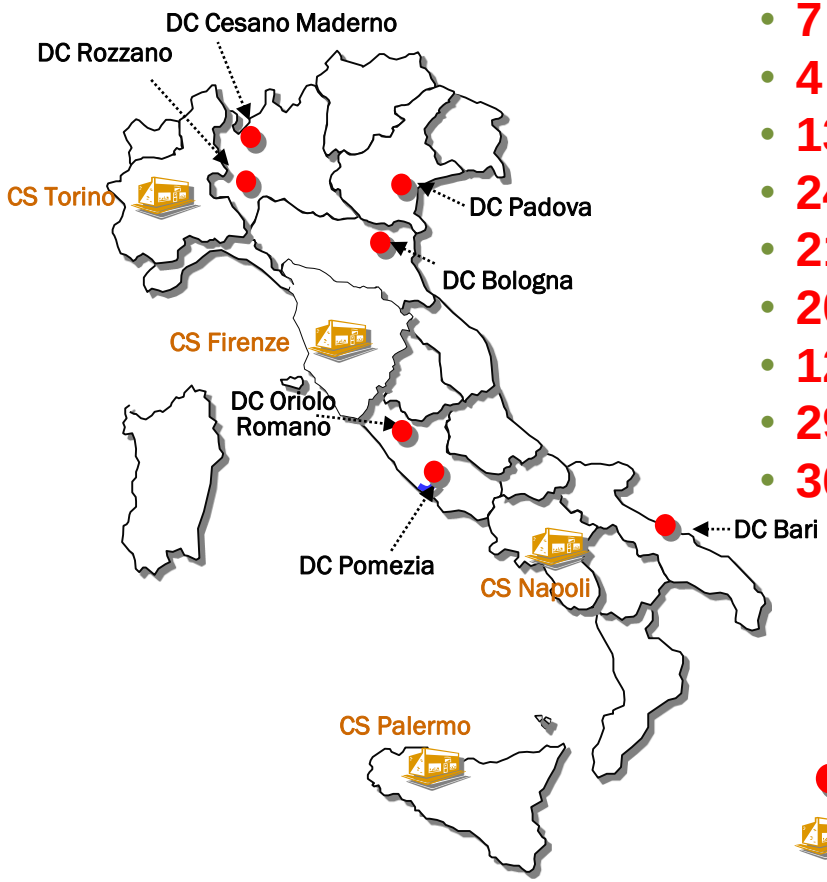
energy saving in Telecom Italia

Workload consolidation is one of the approaches through which Telecom Italia faces the problem of energy saving in data centers.

focus on workload consolidation

We will focus on the **technological and architectural benefits** deriving from the use of **workload consolidation** solutions in Telecom Italia data centers.

Data Centers in Telecom Italia

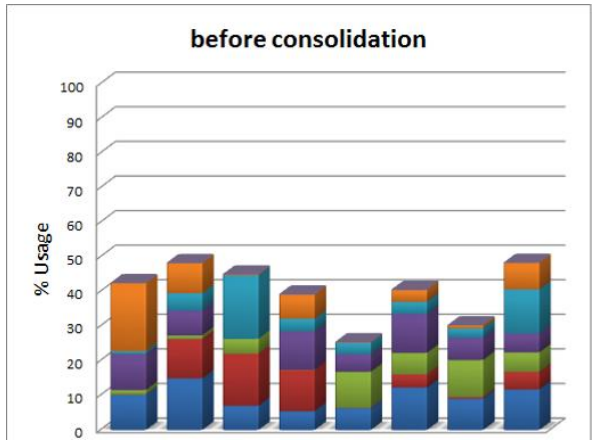


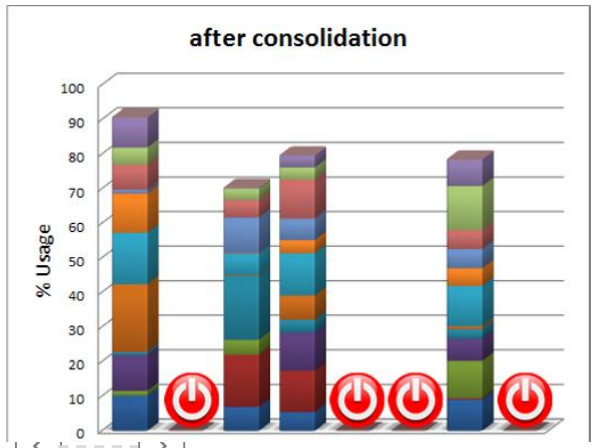
- 7 National Data Centers
- 4 Regional Service Centers
- 135 System Rooms
- 24.000 Servers
- 21.000 TB Storage
- 20.000 TB Backup
- 12.000 TB Backup transmitted/month
- 29 MW Available Power
- 36.000 m^2 Available Floor Area

E4C in Action

Workload Consolidation

Do more with less



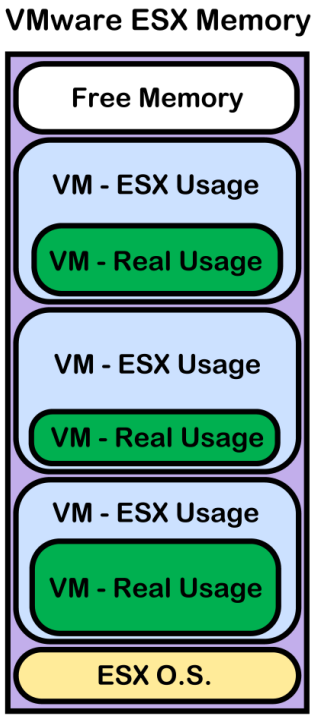


E4C Optimizes VMs placement

Eco4Cloud Workload Consolidation is a Virtual Infrastructure Optimization Solution improving performance and economics of virtualized data centers with an intelligent software platform, which increases efficiency and reduces costs.
It works by dynamically consolidating VMs on the most efficient set of physical resources.

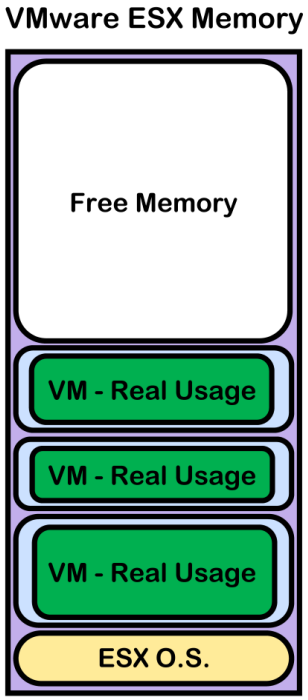
Smart Ballooning

Reclaim RAM Memory



E4C decreases RAM Consumption

Smart Ballooning is a virtual machines memory management software for VMware platforms. Smart Ballooning allows to release memory unused by virtual machines and make it available for vSphere to allocate it to other virtual machines. It works by injecting Memory Ballooning selectively on VMs wasting memory



Smart workload management in Telecom Italia: extensions in progress

The deployment started in Jan 2014 and has progressively extend to all VMware data centers (*about 500 ESX servers*) with the objective of minimizing the number of active servers, **reducing energy consumption** and improving the **overall efficiency**.

Issues

- Due to the complexity of the environment, **monitoring features** were developed on all data centers.
- To solve configuration problems it's important to gain full **commitment** of people in charge of operations.
- **Some obsolete servers** had to be taken out of the perimeter and planned for decommissioning. Their workloads were moved to servers that had been turned off.

Deployment solution

- The software is installed as a ***virtual appliance*** integrated with VMware vCenter
- The appliance suggests vCenter how to **dynamically and optimally consolidate VMs** on physical hosts

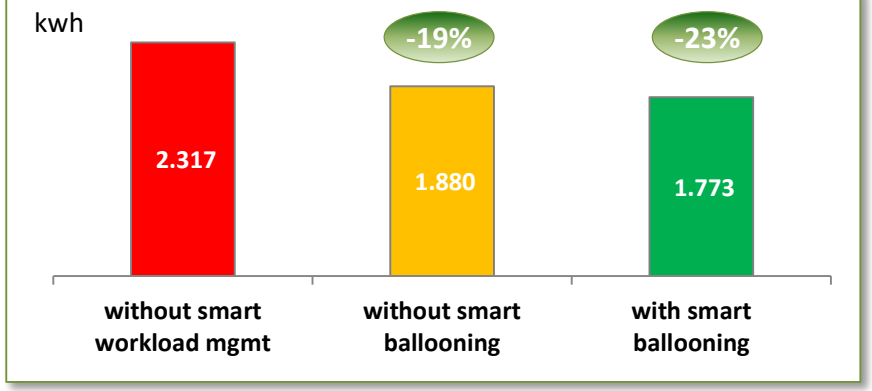
FARM	Total hosts	# Hosts w/E4C Active	# Host in permanent stand-by	% Host in permanent stand-by
Bari Consolidation (Production)	20	8	4	50,0%
Bari Consolidation (Test & Dev)	29	24	7	29,2%
Bari NGDC (Test & Dev)	58	22	13	59,1%
Bari vCloud (Test & Dev)	9	9	1	11,1%
Bologna NGDC (Production)	41	28	4	14,3%
Pomezia Consolidation (Production)	28	26	6	23,1%
Pomezia NGDC (Production)	48	48	8	16,7%
Pomezia NGDC (Test & Dev)	31	31	2	6,5%
Rozzano NGDC (Production)	13	13	1	7,7%
Padova Consolidation (Production)	13	6	1	16,7%
Padova NGDC (Production)	25	25	2	8,0%
Oriolo Consolidation (Production)	43	30	1	3,3%
Oriolo NGDC (Production)	17	17	2	11,8%

Energy savings may **fluctuate,** depending on **the dynamic workload**

Smart workload management in Telecom Italia: results

- So far, the VM consolidation solution has been applied on the physical servers of on-premises Telecom Italia data centers (*about 500 servers*)
- As the utilization of CPU and RAM is variable, **the overall number of servers that can be switched off (and possibly devoted to incremental workload) is estimated to be around 20% of the overall number (*about 100*)**

Energy consumption trends

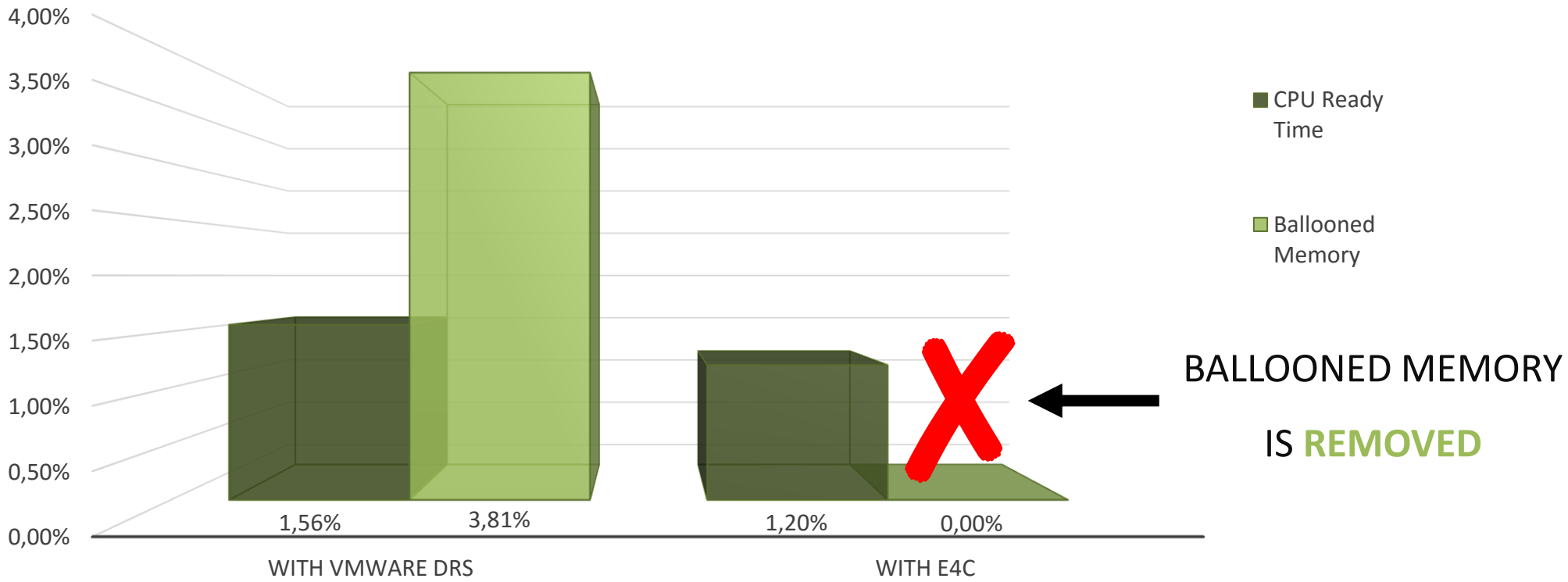


Steady benefits

Overall number of servers	500 ESX
No. of servers that can be hibernated	100 (20% tot)
Energy saving/hour for each hibernated server	450 Wh

Performance Improvement

Contention decreases using E4C

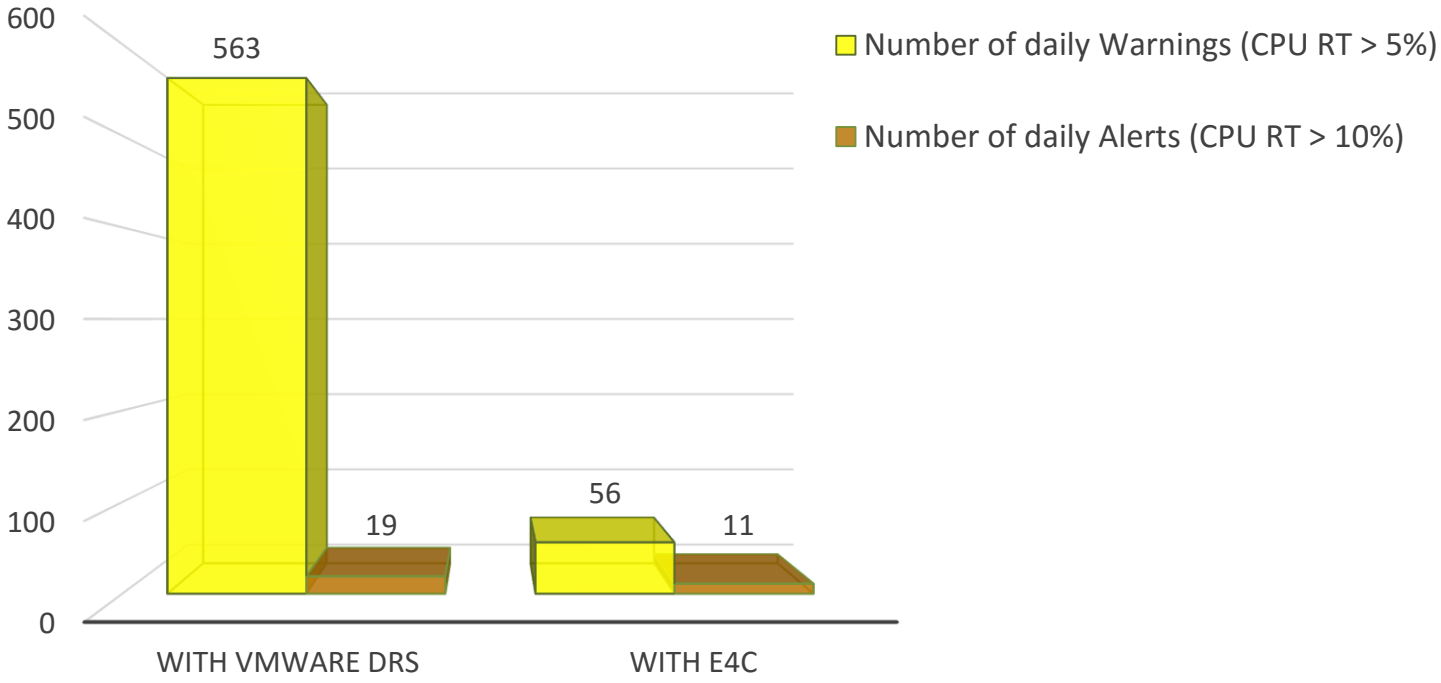


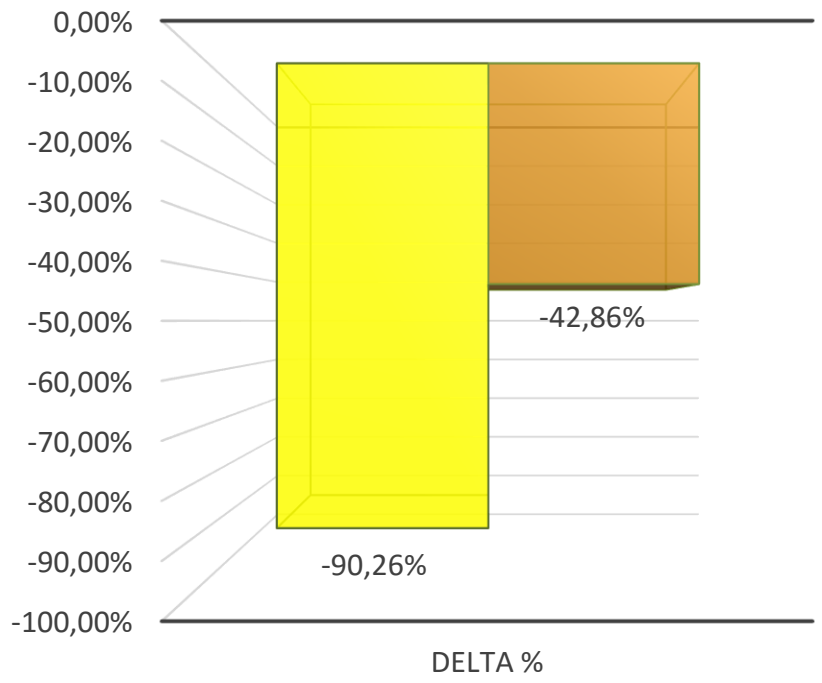
Based on the deployment of Eco4Cloud on a primary Italian Telco, where E4C has been optimizing virtual infrastructure operations since 2014

Performance Improvement

CPU READY TIME

Number of daily warnings and alerts decrease using E4C





Based on the deployment of Eco4Cloud on a primary Italian Telco, where E4C has been optimizing virtual infrastructure operations since 2014

E4C benefits

Less Operations

- 514 less warnings/alerts each day, per cluster
- 3598 less warnings/alerts each week, per cluster

Less hardware

- 100 on 500 server can be switched off
- No need of refresh cycle (buy new server)

Optimize Performances

- - 23% of average CPU Ready
- Ballooned memory is REMOVED

Based on the deployment of Eco4Cloud on a primary Italian Telco, where E4C has been optimizing virtual infrastructure operations since 2014