AVOID RESOURCE CONTENTION WITH E4C TECHNOLOGY

A PRIMARY TELCO USE CASE

Overcommitment and Contention

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.

Why these counters are important

Side effects

CPU Ready Time and Ballooned Memory are symptoms of contention on CPU and RAM respectively. These metrics represent, in IT literature, the universally recognized most significant indicators of the fact that the virtual machines are experiencing bad performances.

Acceptable thresholds

The generally accepted industry best practice based on VMware's guidelines is that CPU Ready Time values up to 5% (per vCPU) falls within acceptable parameters.
Memory Ballooning is the first technique the hypervisor uses to reclaim memory. Avoiding ballooning is a sign of good health for a virtual farm.

What Workload Consolidation do

The Workload Consolidation intelligence computes the ideal placement of VMs among the physical hosts in order to decrease both CPU Ready Time and memory ballooning, so to enable higher performances and VMs density.

Test Workflow

WITH VMWARE DRS

Vmware DRS is set in Fully Automated mode and E4C Workload Consolidation is switched off on the target clusters

WITH ECO4CLOUD TECNOLOGY

E4C Workload Consolidation is enabled on the target clusters, while VMware DRS is configured in Partially Automated mode

FINAL RESULTS

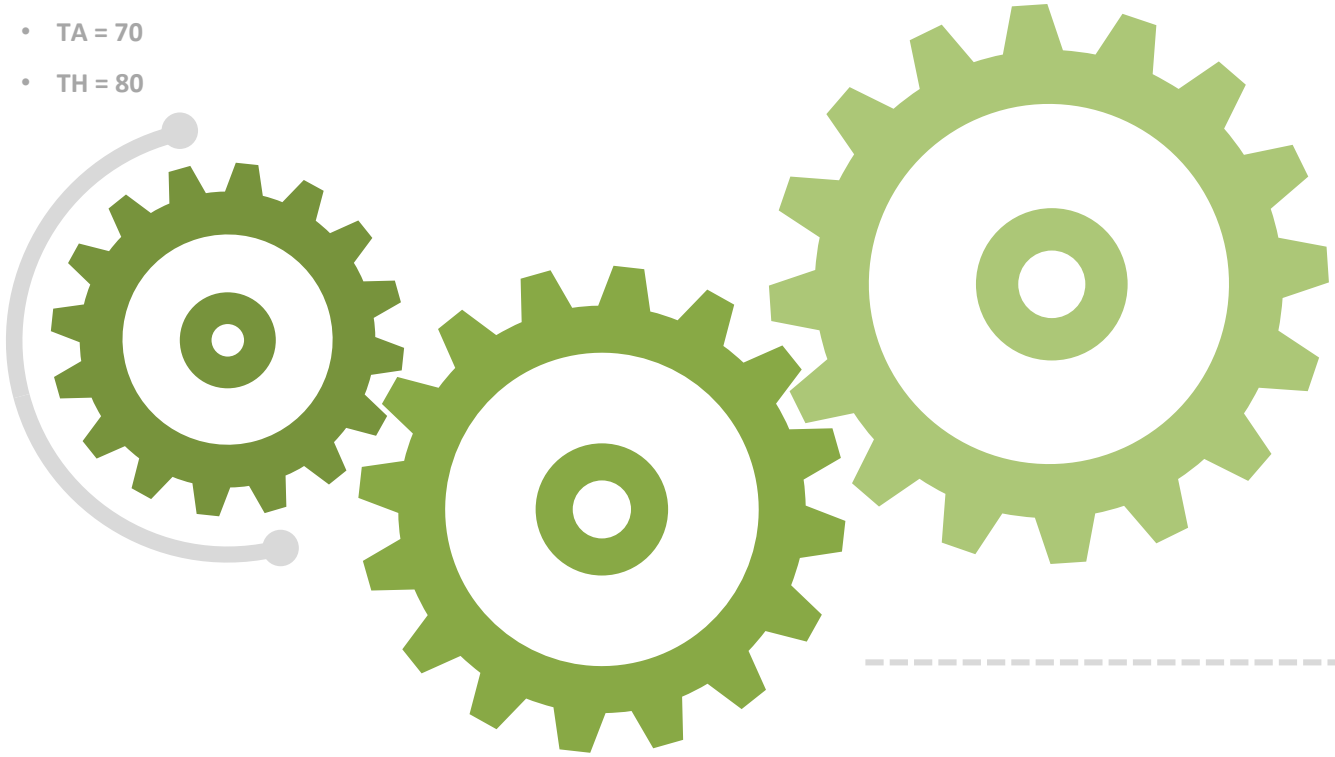
Analysis of the collected data and Report production

Testbed: production site on E4C's customer premises

E4C in Action

PARAMETERS VALUES

- TA = 70
- TH = 80



DRS MODE

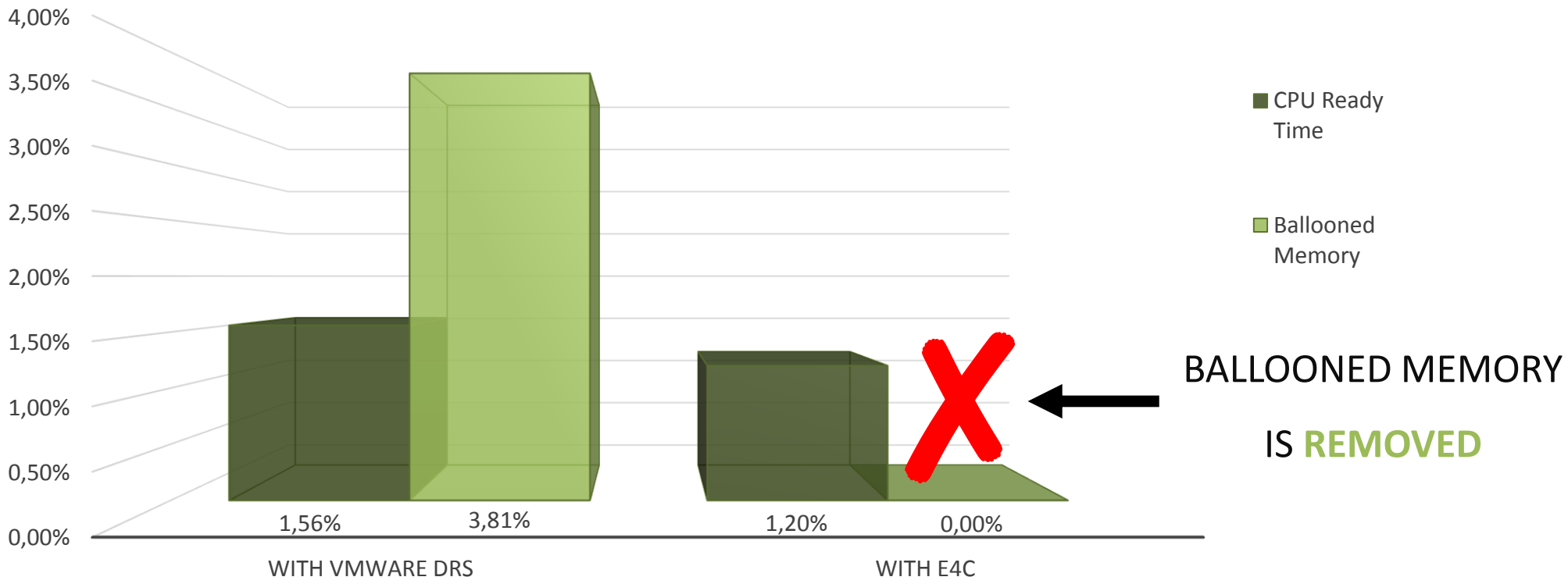
- Partially Automated

HOT VMOTIONS

- Using hot vMotions, E4C is able to choose the best place for each VM in order to minimize CPU and RAM contention

Performance Improvement

Contention decreases using E4C

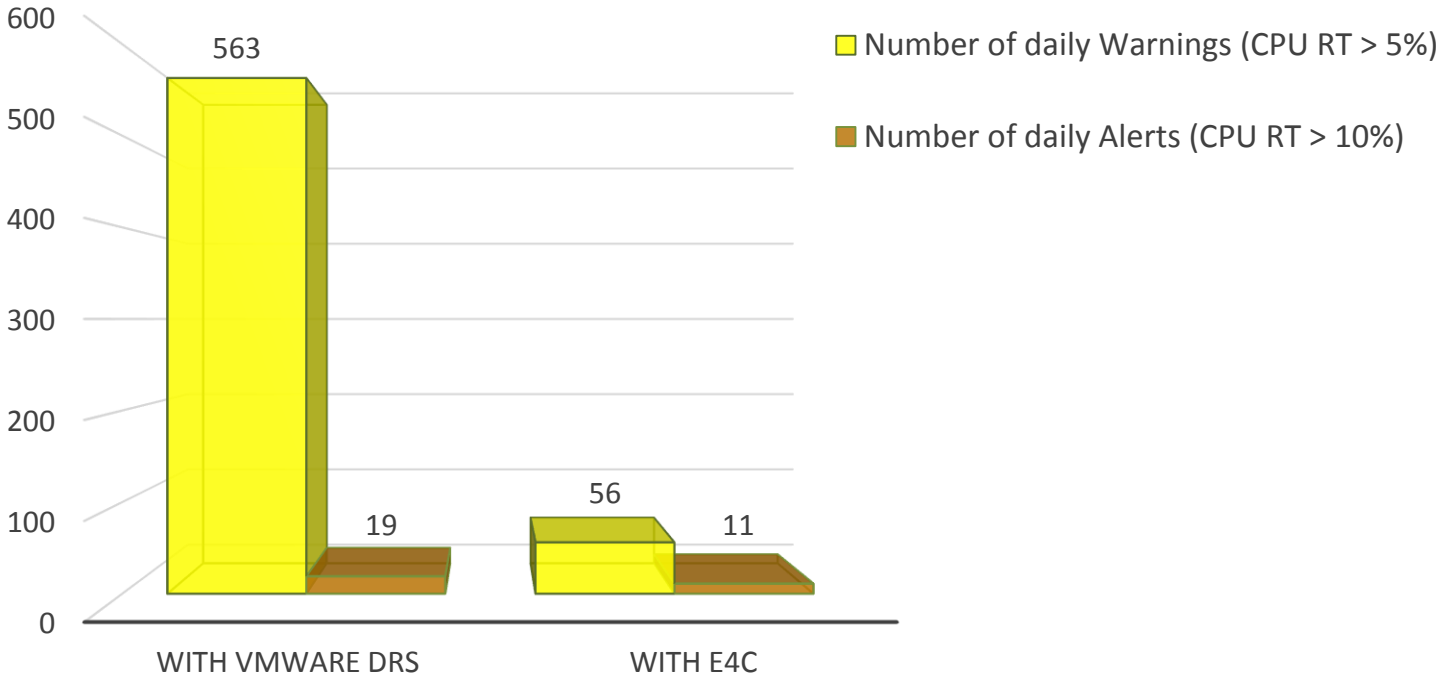


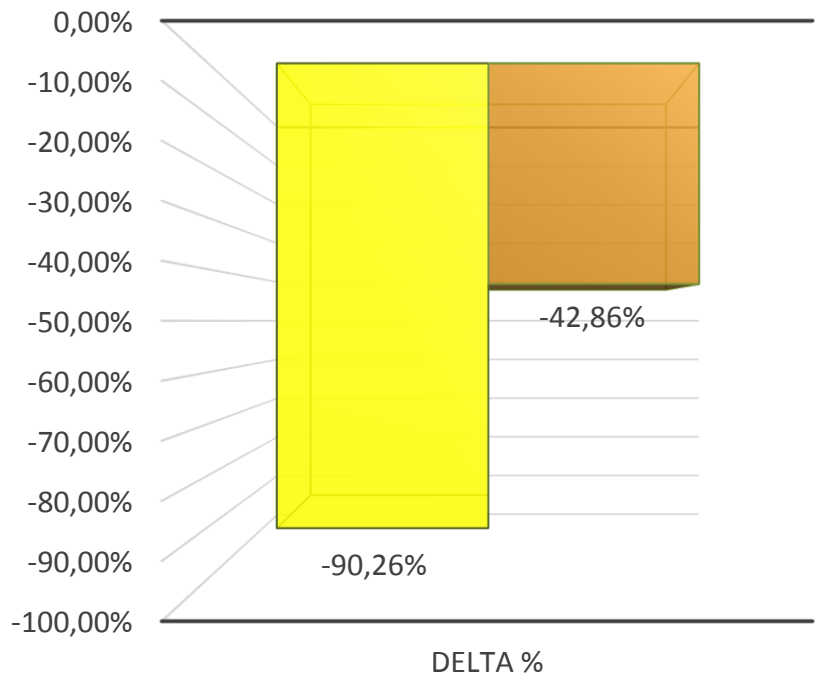
Based on the deployment of Eco4Cloud on a primary Italian Telco, where E4C has been optimizing virtual infrastructure operations since 2013

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 2013

E4C benefits

Less Operations

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

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 2013

Get in Touch

Piazza Vermicelli, Rende (CS), Italy

info@eco4cloud.com

+39 0984 494276