

Task Description

The goal of this experiment is to detect and resolve a performance problem in a given Palladio model. For detection, we provide a detailed model description and describe the steps to identify the performance problem. For resolving, your task is to apply Architectural Templates and to show based on simulations that you resolved the performance problem.

0 Background Questions

1. For the next steps, use this print for filling it out.

2. Please enter your name

3. Please enter your e-mail address. We might use this to contact you with regard to your answers to the open questions.

4. How would you rate your knowledge of Architectural Templates?

none	low	medium	high	expert

5. How many month of experience do you have with Architectural Templates?

6. How would you rate your knowledge of Eclipse?

none	low	medium	high	expert

7. How many month of experience do you have with Eclipse?

1 Performance Problem Detection

Please execute the following tasks:

1. Install Palladio (latest release), Architectural Templates (latest release), and configure a new workspace as described in the attached “ATInstallationGuide.pdf”. Afterwards, you will have the CloudStore model in your workspace. If you already installed Palladio and Architectural Templates, you have to >>Check for Updates<< and only have to Import the CloudStore model as described into a clean workspace.

Reminder:

You can find the Wiki describing available Architectural Templates here:

<http://wiki.cloudscale-project.eu/index.php/Wiki/>

2. Please note the starting time.

_____ :hh _____ :mm

3. Read the “CloudStoreDescription.pdf” as attached.

4. Please answer the following questions:

1. How many Assembly Contexts are described in the “CloudStore Description”?

2. How many Assembly Contexts are specified in the system of the CloudStore model?

3. How many Resource Containers are described in the “CloudStore Description”?

4. How many Resource Containers are specified in the resource environment of the CloudStore model?

5. How many actions are modeled for the “getHome” Operation of the “Homepage” Component of the CloudStore model?

6. What is the purpose of the “getWorker” and the “returnWorker” Actions in the “getHome” Operation of the CloudStore model?

7. Why is the “returnWorker” Action not executed as last?

5. Run the “AT-CloudStore.launch” run configuration (right-click → Run As.. → AT-CloudStore) that is part of the imported CloudStore project and investigate the analysis result in the Palladio perspective.

6. Please answer the following questions:

1. When is the first time that the response time for the usage scenario is above two seconds?

2. How many CloudStore customers are within the system at that point in time? (note: inspect the DLIM model (see Palladio workshop) for determining an estimate.)

3. Which system operation call(s) have response times over two seconds?

4. What is the bottleneck resource causing response time over two seconds?

5. Given that response times should stay below two seconds, which options do you see to resolve the situation?

7. Please note the current time when you finished all the tasks above.

_____ :hh _____ :mm

2 Resolving the Performance Problem

We have the following requirements defined:

- **Performance:** Response times shall stay below two seconds. Violations are allowed in the limits given in the “Elasticity” requirement.

- **Elasticity:** When violations of the performance requirement are detected, CloudStore shall return to a stable state within 20 seconds.
- **Cost-Efficiency:** The operation costs for operating CloudStore shall be minimized.

Next, your task is to resolve the detected performance problem to meet above requirements. You have to investigate two options for resolving: vertically scaling the CPU of the database and horizontally scaling the database.

2.1 Vertical Scaling

In vertical scaling, a fixed server can dynamically speed-up its processing resources over time. Figure 1 illustrates such a scaling for a server.

In your case, you have to scale-up the CPU of CloudStore's database server. For this task, execute the following steps:

1. Please note the starting time.

_____ :hh _____ :mm

2. Apply an appropriate Architectural Template for vertical scaling:

- Note that the monitor repository (see Palladio workshop) of CloudStore includes a monitor (named: “Response Times – Browsing Mix”) that calculates the mean response times for the overall usage scenario in an interval of 5.0 seconds. This calculated value is used by default to determine whether scaling needs to be triggered. Use the values of the following table during your modeling tasks, if feasible:

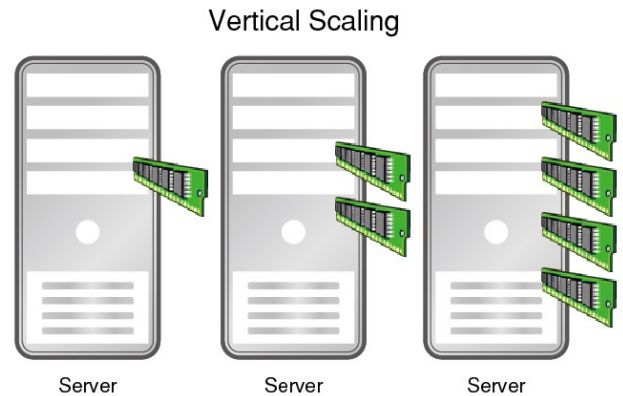


Figure 1: A server scales its processing resources up (from left to right)

Scale-Down Threshold	0, 1
Scale-Up Threshold	0, 8
Step Size	13.350.000.000
Min Rate	13.350.000.000
Max Rate	133.500.000.000

Here, the “Scale-Down Threshold” specifies that the current CPU processing rate should be reduced if the calculated mean response times are lower than 0.1 seconds. Likewise, the “Scale-Up Threshold” specifies that the current CPU processing rate should be increased if the calculated mean response times are higher than 0.8 seconds. The “Step Size” is the rate by which scaling adapts the processing rate. “Min Rate” and “Max Rate” specify lower and upper bounds for this rate, respectively.

- You have a maximum of **1 hour** (starting from the time you noted in step 1) to apply the AT. If you do not manage to provide model with a suitable AT applied in the given time, skip to step 5.

3. Note the current time.

_____ :hh _____ :mm

4. Run a CloudStore simulation with your applied AT and inspect the results.

5. Make a screenshot of the “XY Plot” of the “Response Times – Browsing Mix” and store it for later use.

6. Note the current time.

_____ :hh _____ :mm

7. Please answer the following questions:

1. Did you manage to apply a suitable AT? (If yes, name it here; if no, skip to question 5.)

2. Is the performance requirement always met?

3. In case of a violation of the performance requirement, does the system return to a stable state within 20 seconds?

4. Note the current time.

_____ :hh _____ :mm

5. Please list any issues during your task:

2.2 Horizontal Scaling

In horizontal scaling, a given server is dynamically replicated. Each replica is added to an according loadbalancer that distributes workload among them. Figure 2 illustrates such a horizontally-scaled server with three replica.

In your case, you have to scale-out CloudStore's database server. This represents an alternative to vertical scaling and allows you to assess whether it is more cost-efficient. For this task, execute the following steps:

1. Please note the starting time.

_____ :hh _____ :mm

2. Apply an appropriate Architectural Template for vertical scaling:

- Note that the monitor repository (see Palladio workshop) of CloudStore includes a monitor (named: "Response Times – Browsing Mix") that calculates the mean response times for the overall usage scenario in an interval of 5.0 seconds. This calculated value is used by default to determine whether scaling needs to be triggered. Use the values of the following table during your modeling tasks, if feasible:

Scale-In Threshold	0, 1
Scale-Out Threshold	0, 4
Number of Initial Replica	1

Here, the "Scale-In Threshold" specifies that the current number of replica should be reduced by 1 if the calculated mean response times are lower than 0.1 seconds. Likewise, the "Scale-Out Threshold" specifies that the number of replica should be increased by 1 if the calculated mean response times are higher than 0.4 seconds. The "Number of Initial Replica" gives the number of database servers at simulation start.

- The loadbalancer shall work as follows: It forwards workload to a given replica with a probability of $1/(\text{number of current replica})$.
 - You have a maximum of **2 hours** (starting from the time you noted in step 1) to apply the AT. If you do not manage to provide model with a suitable AT applied in the given time, skip to step 5.
3. Note the current time.

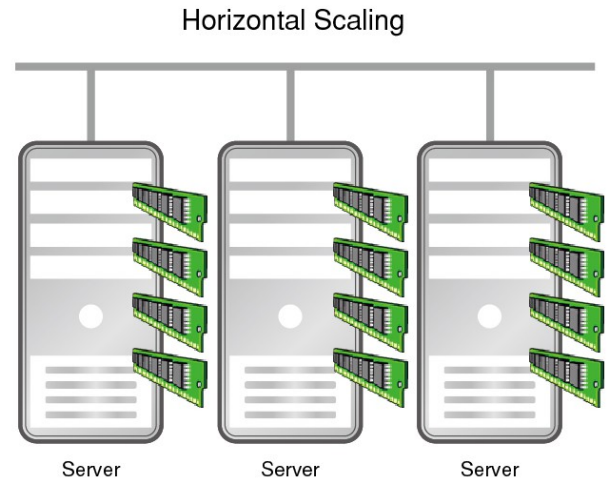


Figure 2: A server is replicated and load-balanced (scale-out)

_____ :hh _____ :mm

4. Run a CloudStore simulation with your reconfiguration rule and inspect the results.
5. Make a screenshot of the “XY Plot” of the “Response Times – Browsing Mix” and store it for later use.
6. Note the current time.

_____ :hh _____ :mm

7. Please answer the following questions:

1. Did you manage to apply a suitable AT? (If yes, name it here; if no, skip to question 5.)

2. Is the performance requirement always met?

3. In case of a violation of the performance requirement, does the system return to a stable state within 20 seconds?

4. Note the current time.

_____ :hh _____ :mm

5. Please list any issues during your task:

3 Finalization

In order to finalize this experiment, you have to execute these last tasks:

- Name your above taken screenshots (forename_lastname_vertical.png and forename_lastname_horizontal.png)
- If you have any remarks, feel free to add them here:

-
-
-
-
-
- Hand-in this paper and send the screenshots to Christoph Nützel
<christoph.nuetzel@s2011.tu-chemnitz.de>. Add Sebastian Lehrig
<sebastian.lehrig@informatik.tu-chemnitz.de> as CC in your e-mail.

THANK YOU! :)