Gateway Portal Load Balancing

Pre-requisites

We advise you to start by reading our Load Balancing overview.

Generated Clients and Web Access

There are two ways to connect to a Load Balanced cluster:

- Using a **Generated Client** (executable program, created by the Connection Client Generator)
- Using a **Gateway Portal Web Access** (web page in the browser, created by the Webmaster Toolkit)

Please note that the Gateway Portal is only enabled in ServerLink Corporate edition.

**This page describes how to use the Load Balancing Manager for use with a Gateway Portal.**

Gateway and Application Server Roles

The Load Balancing Manager tool **must be used on the server having the Gateway role**. Using it on another server of the load balanced cluster will have no effect on the load balancing.
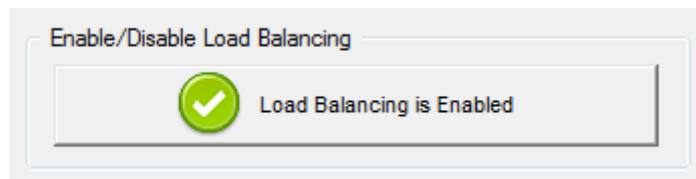
Quick Start

**Activating Load Balancing is really easy. It only takes three steps:**

- 1) Generate a Web Access page with the option "Generate a Gateway Portal enabled Web Page" checked (using the Webmaster Toolkit):

SERVERLINK
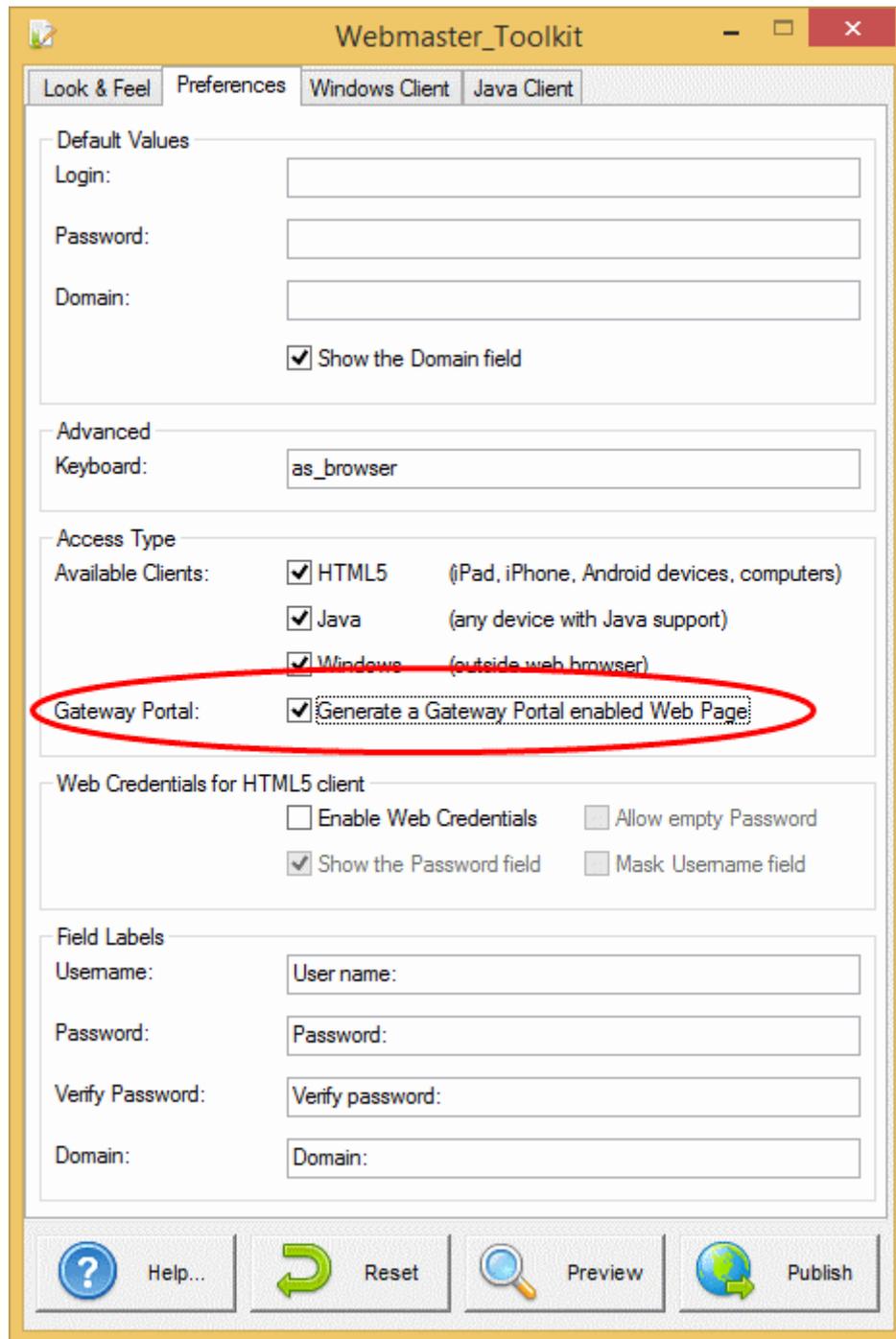
- 2) Enable Load Balancing by clicking on the "Enable/Disable Load Balancing" button (in Load Balancing Manager):

- 3) Open a browser and navigate to the Web Access page you generated in step 1 (by default: http://localhost/index.html). After typing a login, you will see that **the Web Access page choose the less loaded server of your farm**:
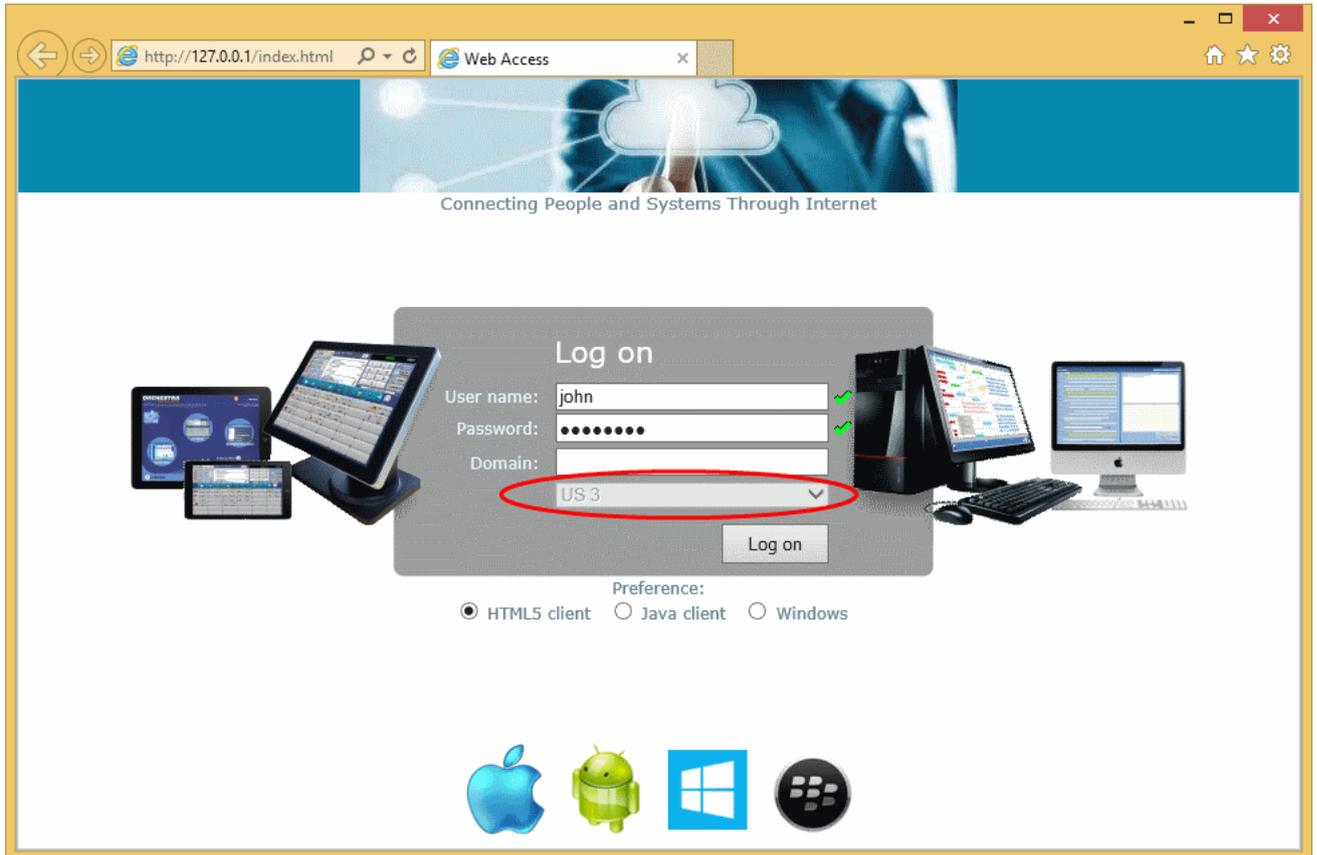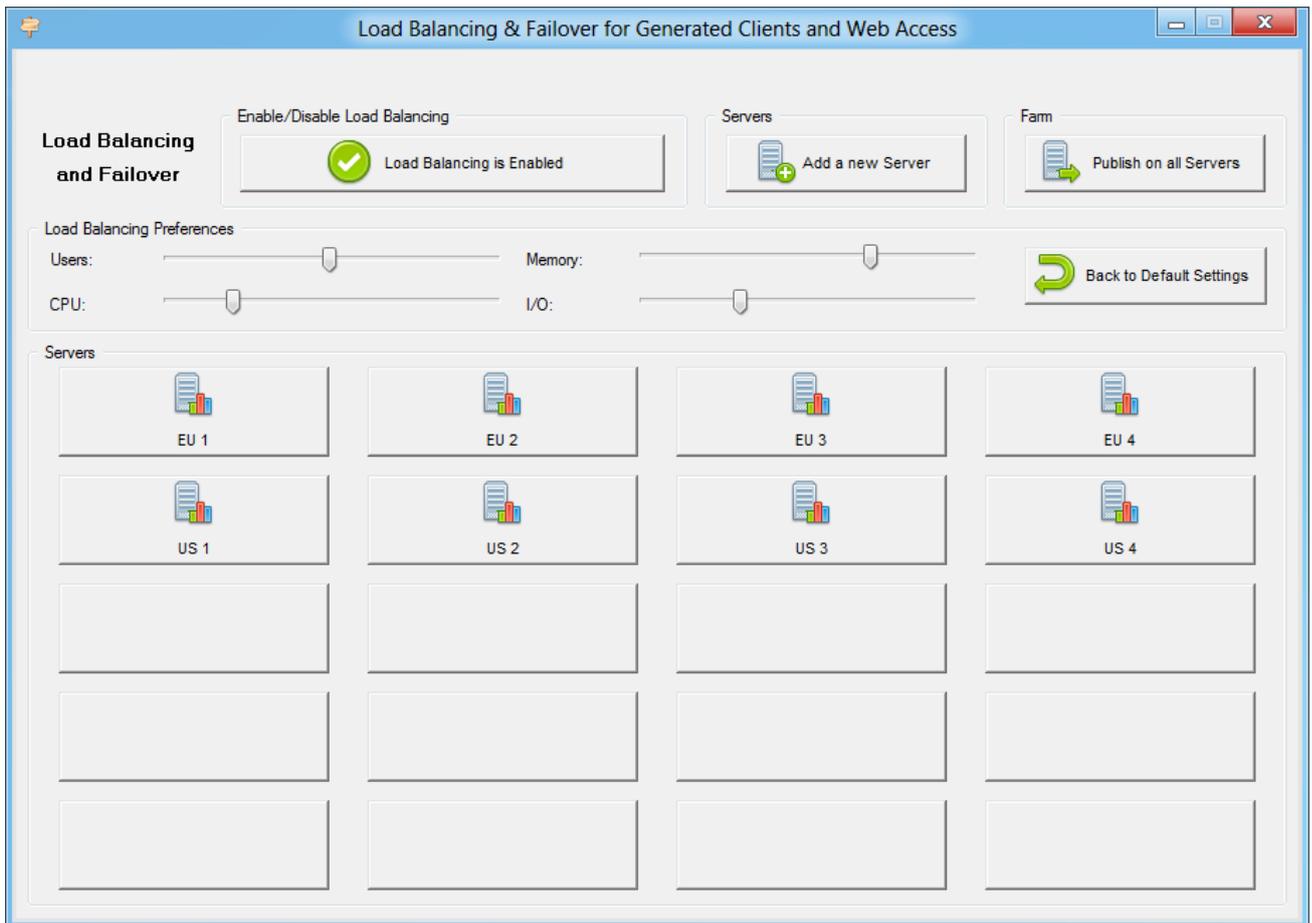
Load Balancing Main Window

The main window displays a lot of data and allows you to configure your Gateway Portal Load Balancing.

It lists all the servers in your Load Balanced server farm, and allows you to add a new server (by clicking on the "Add a new Server" button) or displays an existing server (by clicking on it).

It also allows you to enable, disable and configure the Load Balancing (more on this below).

Enabling/Disabling Load Balancing

On top of the Gateway Portal Load Balancing window, you will see a big button:

- This button display the current state of the Load Balancing.
- If you click on it, it will enable or disable the Load Balancing depending on its current state.

Here is the button when the Load Balancing is **Disabled**. A click on it will activate it, and disable the Users/Servers Assignment.
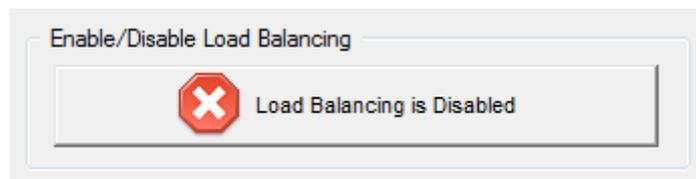
Here is the button when the Load Balancing is **Enabled**. A click on it will deactivate it, and enable the Users/Servers Assignment.
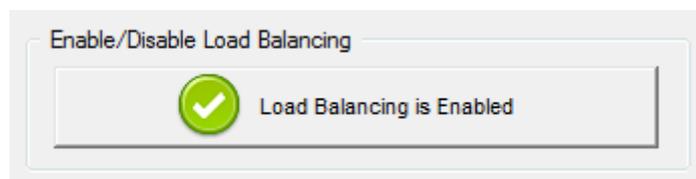
## How does Load Balancing choose a server?

When Load Balancing is enabled:

- **if a user already has an opened Windows session on one of the application servers, he will be sent to this server**;
- Otherwise, the user will be sent to the less loaded server at the time of his connection.

Configuring the Computation of Servers' Loads

To determinate which is the less loaded server, the load of every server is computed using a weighted average between several performance indicators:

- number of connected users
- processor usage
- memory usage
- disk usage

You can modify the weight (importance) of these indicators by using the following sliders:
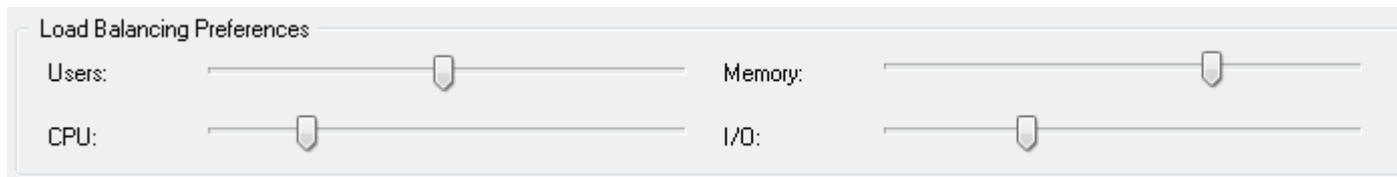
Using these sliders, you will be able to fine-tune the Load Balancing, and to optimize it to your own needs. For instance if your users launch a business application with big memory requirements, it might be a good idea to increase the impact of the RAM in the load computation using the "Memory" slider.

You can also reset these values by clicking on the "Back to Default Settings" button.

How is Server's Load computed?

The load of each server is computed when needed in order to decide on which server the user must be sent.

For this computation, we use a **weighted average** between 4 hardware metrics.

The 4 sliders allow you to give more (or less) weight on each of these metrics, which are:

- Users : number of connected users
- CPU : percentage of non-idle processor time
- Memory : percentage of used memory
- I/O : percentage of non-idle disk time

For example, if you put the "Memory" slider on its right side and all the other sliders on their left side, the load of each server will be computed using almost only the percentage of used memory - and your users will be sent to the server having the most available memory.

- When a slider is completely on the left, the weight used in load computation will be 1.
- When a slider is completely on the left, the weight used in load computation will be 100.

Moreover, the load of a given server depends on the ratio [used resources / total resources] ; so if a server A is twice as powerful as a server B, twice more users should be sent to server A than to server B (all other things being equals).

Adding a new Server

To add a new server, simply click on the "Add a new Server" button. The following window will be displayed:
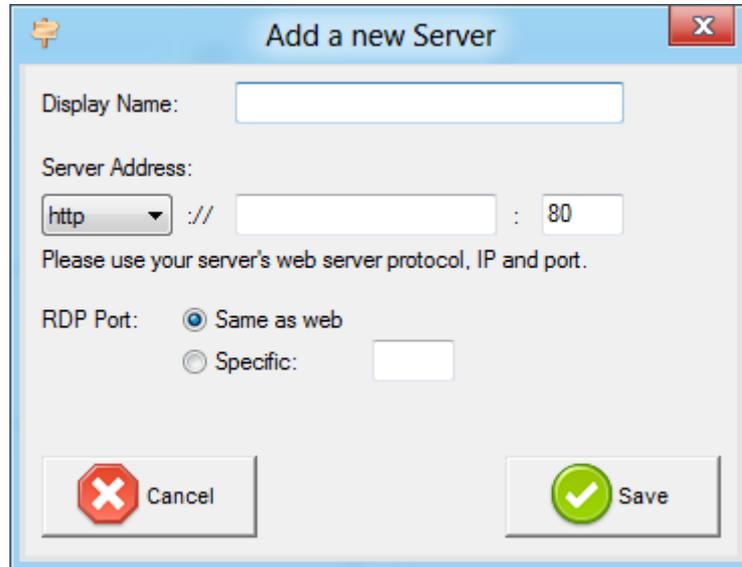
The "Display Name" is the title that will be shown to your users on the Web Access HTML page. It is supposed to be more user-friendly than a technical value (such as an IP), for instance "US Server" or "Blue Zone".

When using the Gateway Portal Load Balancing, the "RDP Port" will not be used. It is only used when connecting using a Generated Client. We advise you to keep the default setting ("Same as web").

Editing an existing Server

To modify an existing server, simply click on the "Edit" button of the server you want to update. The following window will be displayed:
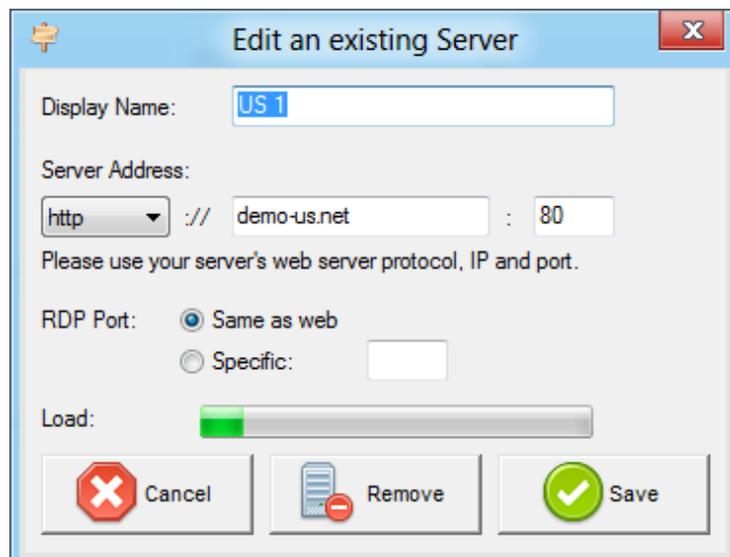
The progress bar shows the current load of the server. It also confirms you that the server is well configured and can be contacted by the Gateway.

About the "Publish on all Servers" Button

This button has to be used only when you are connecting to your load balanced farm of servers with a Generated Client.

In present case, when using the Gateway Portal Web Access (HTML page), you don't have to use this button, because only the gateway must know about your farm configuration.