piXlogic

The Visual Search Company

piXserve Elastic Cloud

Standard, Professional and Enterprise Editions

piXserve-ECs/ECp/ECe

Getting Started Guide

piXlogic, Inc.
4984 El Camino Real
Suite 205
Los Altos, CA 94022

T. 650-967-4067
http://www.piXlogic.com

Third-Party Information

This document contains links to third-party websites that are not under the control of piXlogic, Inc. ("piXlogic") and piXlogic is not responsible for the content on any linked site. If you access a third-party website mentioned in this document, then you do so at your own risk. piXlogic provides these links only as a convenience, and the inclusion of any link does not imply that piXlogic endorses or accepts any responsibility for the content on third-party sites.

Trademarks

piXlogic, piXserve, and piXsearch and related logos are trademarks of piXlogic, Inc., and are registered in the United States and internationally.

Other product names, logos, designs, titles, words, or phrases mentioned may be trademarks, service marks or trade names of other entities and may be registered in certain jurisdictions including internationally.

Third Party Copyright Notices

The names and marks of third-party companies, products, and organizations mentioned in this document are the property of their respective owners.

.

Introduction

Amazon Elastic Cloud Computing (Amazon EC2) is a web service from Amazon that provides flexible, easy to provision computing resources. piXserve is a scalable search solution, developed by piXlogic, that is focused on images and videos.   piXserve is available in a range of editions (Workgroup Edition, Standard Edition, Professional Edition, Enterprise Edition, Hadoop Edition). piXserve-Elastic Cloud Standard Edition (piXserve-ECs)   and   piXserve-Elastic   Cloud Professional Edition (piXserve-ECp) and piXserve-Elastic Cloud Enterprise Edition (piXserve-ECe) package respectively the Standard, the Professional and Enterprise versions of the software in a virtual machine suitable for deployment on Amazon EC2. This document describes how to install and configure piXserve-ECs/piXserve-ECp/piXserve-ECe on Amazon EC2.

This document assumes that you have already signed up for an Amazon EC2 account, and that you have downloaded the tools necessary to set up your Amazon EC2 environment and have read the accompanying documentation. You will need access to the EC2 command line tools and the AWS Management Console. You can obtain more information about Amazon EC2 and download tools and documentation by going to:

http://www.amazon.com/ec2

http://console.aws.amazon.com

This document also assumes some familiarity with the piXserve product line. You should contact piXlogic technical support for a free walkthrough of the software.   The piXserve operating manual is available online once you have logged into an instance of piXserve-ECs/ECp/ECe.

piXserve-ECs, piXserve-ECp, and piXserve-ECe are each deployed as an Amazon Machine Image (AMI) in Amazon EC2. Each AMI comprises one instance of piXserve-Standard Edition or piXserve-Professional Edition, or piXserve-Enterprise Edition pre-packaged in a Amazon Linux 2 virtual machine, which is styled after RHEL/Centos-7 You can deploy one or more AMIs and configure them to work independently or as a group.   You only be charged for what you use. piXserve-ECs/ECp/ECe are billed on an hourly fee basis. Amazon CPU, bandwidth, storage, etc. charges will also accrue to your account. The billing processed is managed by Amazon and you will be billed directly from them. Please be sure to read and adhere to the terms of use that are described at the time you sign up for using piXserve-ECs/ECp/ECe. These terms are also provided in the online documentation of the product.

To sign up and start using piXserve-ECs/ECp/ECe, or to learn more about the different versions of piXserve that are available, please visit the piXlogic website:

https://pixlogic.com/solutions

Starting a piXserve-ECs, ECp, or ECe Instance

After signing up for piXserve-ECs at Amazon EC2 you will need to get the current piXlogic AMI ID. You can get these IDs from the piXlogic website: https://www.pixlogic.com/solutions/elastic-cloud-support. The ID references shown on the website are used to start an instance of piXserve-ECs or piXserve-ECp or piXserve-ECe.

When you launch an instance of these AMI you need to be mindful of the region designator that you are using for that deployment. Amazon EC2 currently supports 24 regions and 77 availability zones. As part of the process you will attach storage and other resources to your instance, and all need to reside in the same region.

The piXserve AMIs comprise an Amazon Linux2 virtual machine that has been preconfigured with piXserve and with supporting software (JDK 13, PostgreSQL 12). Included for convenience is also an httpd server configured for streaming videos. A more detailed description of the environment is provided in the section titled "Instance Details" later in this document.

You can manage your AMI environment either through Amazon's command line tools, or through the Amazon AWS Management console. In this guide we will use the AWS Management Console for simplicity. Documentation on the command line tools can be found here:
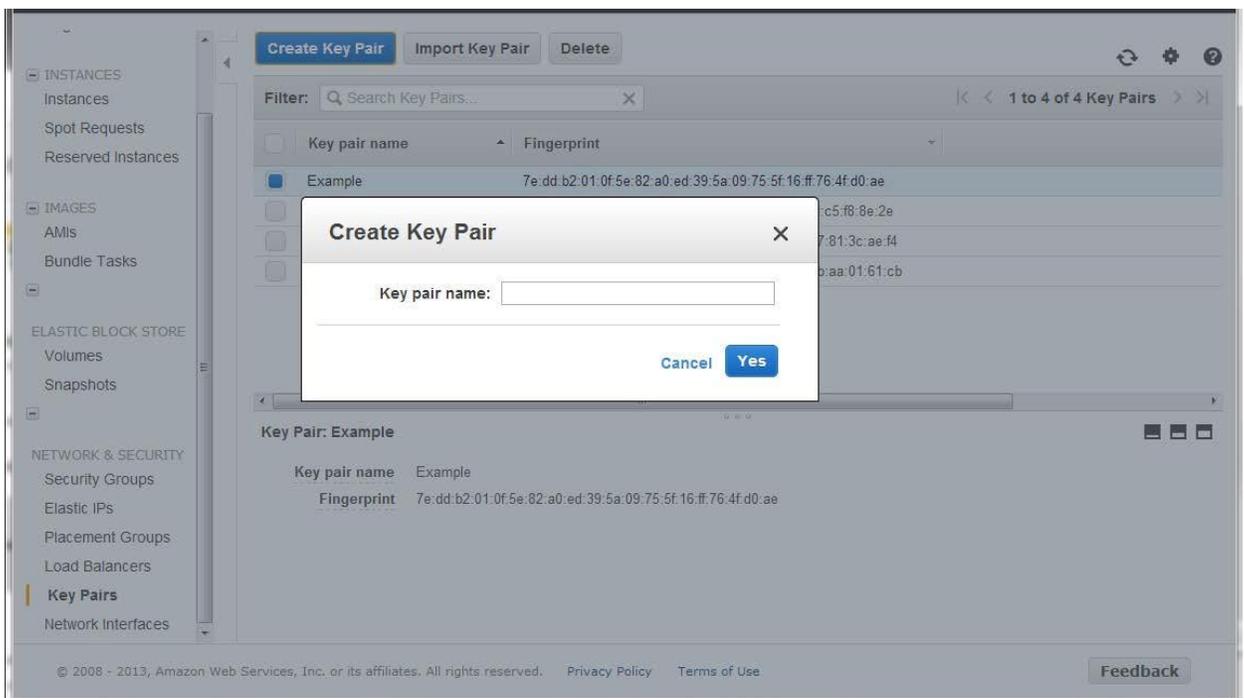
http://docs.amazonwebservices.com/AWSEC2/latest/GettingStartedGuide/

To start an instance, proceed as follows.

1. Using your web browser, sign in to Amazon Web Services and navigate to the AWS Management console: https://console.aws.amazon.com/ec2/home
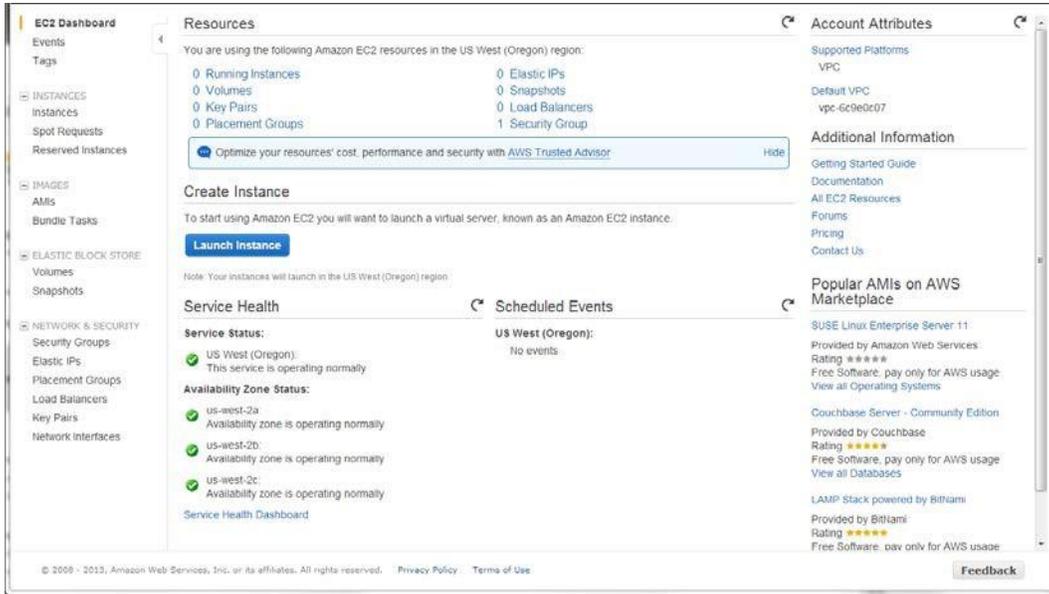
2. Navigate to the "Key Pairs" menu, and create a new private key:

3. The keypair will be created and saved, and a copy downloaded to your computer. The contents of the file will contain something like this:
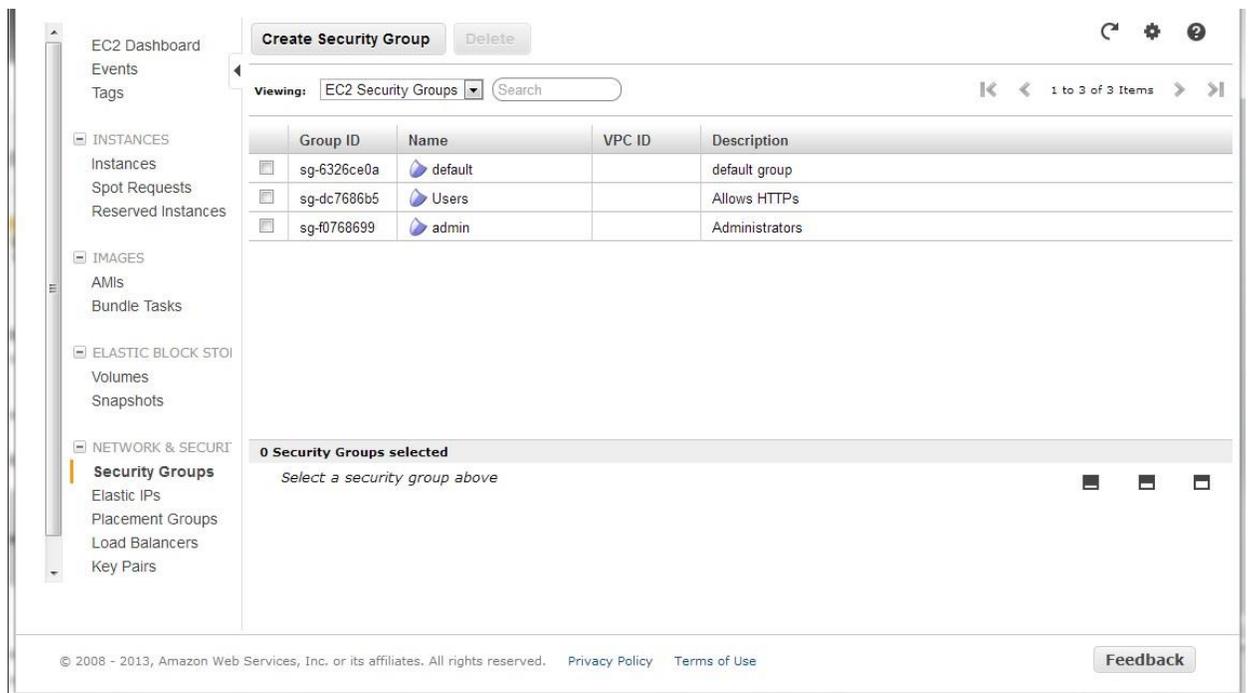
-----BEGIN RSA PRIVATE KEY-----

      ......
      ......
       ..

-----END RSA PRIVATE KEY-----

You should save the key in a safe location and you will see it listed on your key-

pair page. You will use this key-pair later, when you connect to piXserve-ECs, ECp,

or ECe.

4. Next, you will adjust security groups as follows. Go to the "Security Groups" tab of the AWS Management Console, and create an "admin" and a "users" group.


The admin group should be allowed connections on ports:

a. 80, 443, 8080, 8443 – connections to the web

server

b. 82, 554 – connections to the local httpd streaming server (optional), and to RTSP

sources (optional)

c. 5432 – connections to the Postgres database that is running on the piXserve instance (piXserve uses Postgres 12). This port is used if you are configuring multiple piXserve instances on the network to work together, or if you intend to connect to Postgres from an external computer using a PgAdmin client.

d. 22 – SSH and SCP port to connect to the ami instance

The users group should be allowed connections on ports:

a. 80, 443, 8080, 8443 – connections to the web server

b. 82, 554 – connections to the local streaming server (optional)

5. As a next step, navigate to the "Elastic IPs" section on the AWS Management Console. Add one or more addresses as you need to reach your piXserve-ECs or ECp instances. Note that there is a monthly charge from Amazon for each address/block of addresses that you reserve.

6. You are now ready to create/attach storage to your instance(s). The piXserve AMI is running on a virtual hard disk (an EBS volume (Elastic Block Storage)). You can specify the size at the time you configure the AMI. The minimum recommended size is 100GB. At that level, a little over half of this space is available to store piXserve created databases, cached thumbnails, and log files. piXserve created databases are stored by default in the local PostgreSQL instance. As described in the piXserve documentation, you may choose to store these data on other machines running PostgreSQL.

Depending on the version of piXserve you are deploying, your use cases may involve indexing files stored on hard disks, indexing live video sources, indexing/crawling web sites. In the case of files, these need to be located on a storage device that is visible and mounted on the AMI. Therefore, you will probably be attaching EBS disks to your running piXserve AMIs, and these disks will contain the images/videos that you want your users to work with. EBS volumes are persistent (e.g. they do not disappear if an AMI instance is turned off). Amazon also makes available VPN connections so that you may create a network that ties together your AMIs and your local network resources (including network shares). Files stored in S3 are usually directly accessible to piXserve via https. There are storage and/or data transfer charges that apply from Amazon. Create an EBS volume that is big enough to hold your image/video repository. Each EBS can be a maximum of roughly 17.6 TB in size, but you can have multiple EBS units. These EBS units should be NTFS formatted, and you can attach them with their own drive letter (F, G, H, etc.).
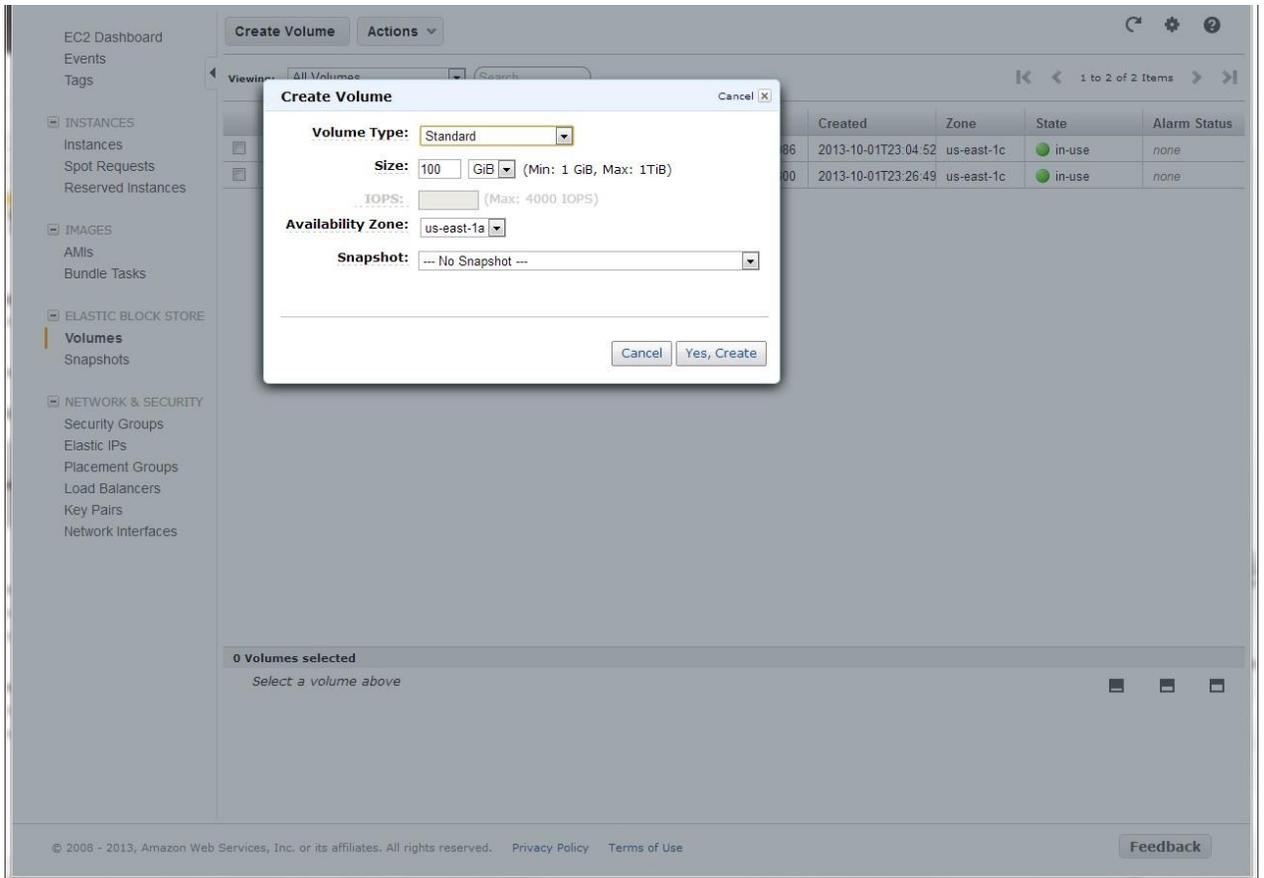
Make sure that the Availability Zone for the EBS volume matches the availability zone of the AMI that you will be attaching it to. You can attach multiple volumes to an instance of piXserve-ECs/ECp/ECe. Keep in mind that the drive letters are auto-assigned in sequence, starting with the letter "D". If the order in which you attach volumes changes, then your data will reside on a drive path that may be different than what piXserve used at the time it indexed your data. To avoid this problem, always attach your volumes in the same sequence.

7. Next, you are now ready to start an instance of piXserve-ECs, piXserve-ECp or piXserve-ECe. Go to the piXlogic cloud support page (https://www.pixlogic.com/solutions/elastic-cloud-support) and get the ID of the AMI you want. Go to the "AMIs" navigation tab on the AWS Management Console, select "All Images" from the dropdown, and scroll until you find the piXserve AMI you wanted or search for it using the search box:
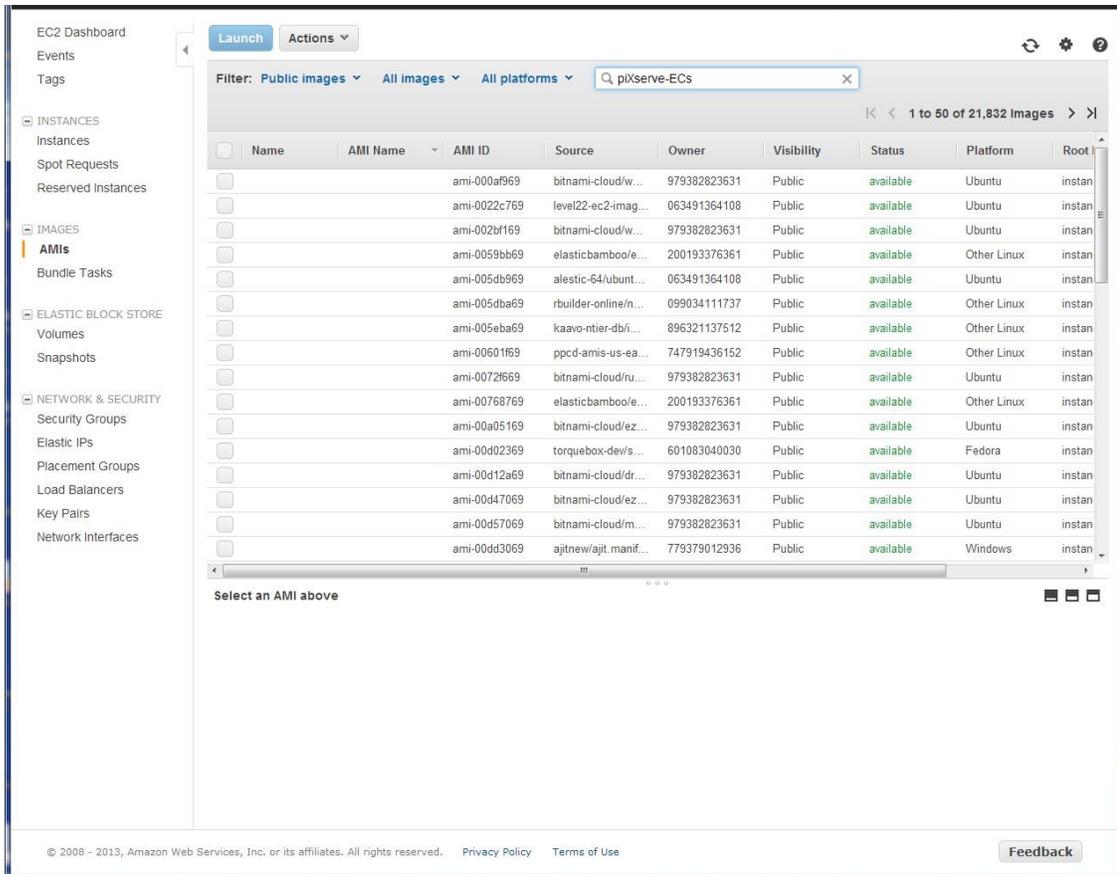
8. You are now ready to start the instance. Right mouse click on the AMI, and choose "Launch" from the drop-down menu. Before launching, you will need to select the instance type and provide some configuration details:

The first decision you will make is the type of instance you want to launch. Depending on your needs and budget, you will most likely choose either a "General Purpose" image or a "Compute Optimized" one. piXserve will take advantage of as many CPUs/Cores/Computing Unit resources available. Contact piXlogic for additional information about sizing piXserve for your specific application. Note that piXserve can also take advantage of NVIDIA GPU accelerators and indexing will go much faster when a GPU enabled AMI is used. Therefore, GPU instances are generally recommended. The amount of GPU RAM available and the number of CPU cores available are key factors to the configuration of piXserve (contact piXlogic for additional information about configuration parameters.

From the left of the screen, as an example, clicking on "General Purpose", yields this screen:
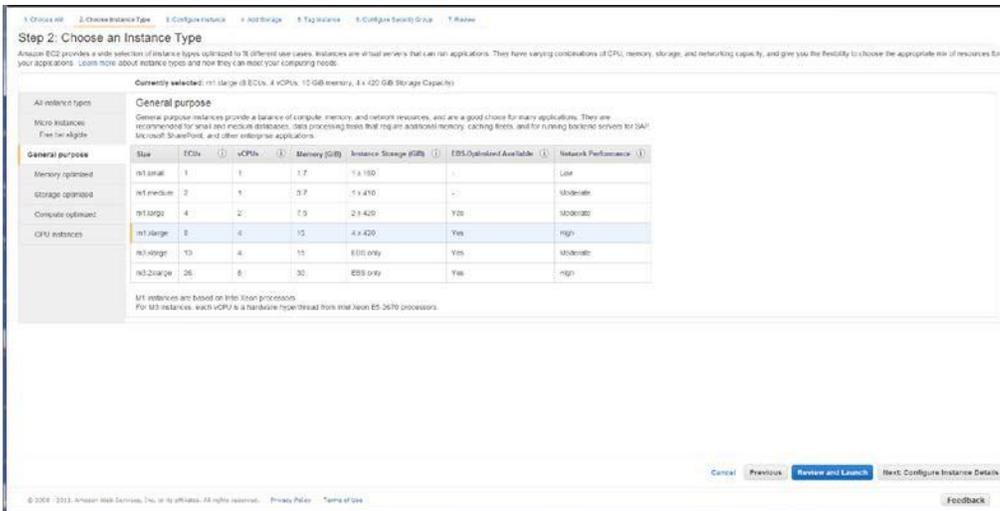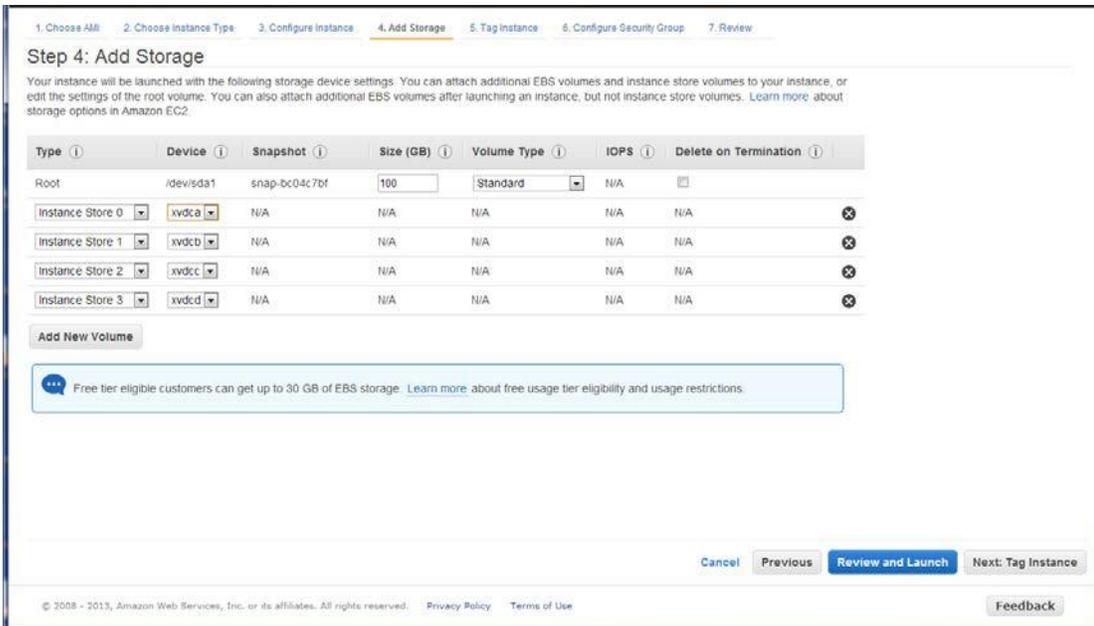
From here we can choose, for example, a g2.2xlarge (with GPU) or a c3.2xlarge (without GPU) instance type, which will give our AMI access to 8 virtual CPUs and 15Gb of RAM.

Next we can specify how many instances to launch and some additional preferences:
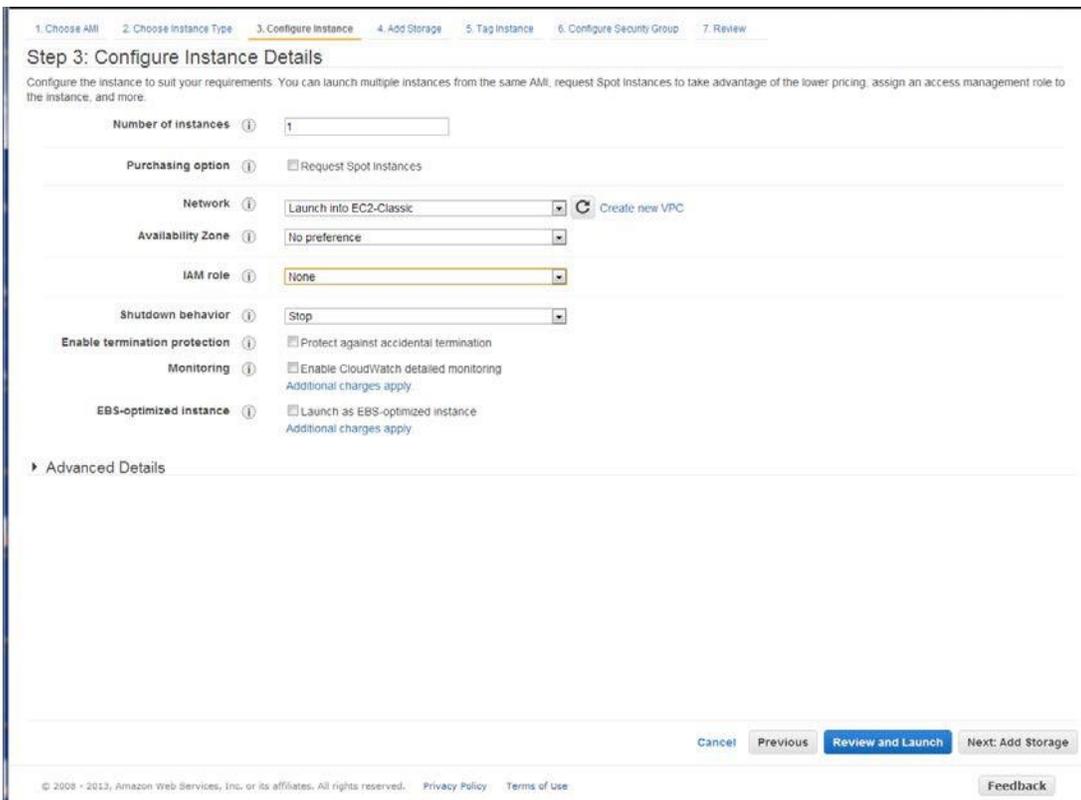
After making your choices, click next to attach storage devices:

You can attach one of more EBS devices (up to four in this particular instance type). Make sure that your device name(s) are sequential. The last character in the device name corresponds to the drive letter that Windows will assign to that device (ex: "xvcdf" will map to "F:\"). These paths will be important later on when you run piXserve.

Clicking next let you create an optional "tag" for the instance:
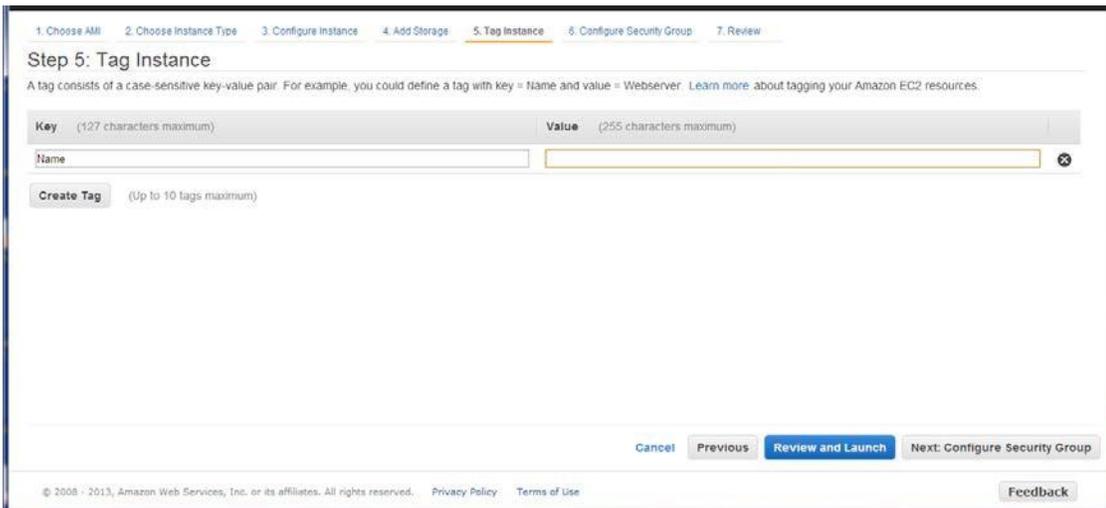
Clicking "next" again lets you select the security groups that will apply. If you created the "users" and "admin" groups described earlier, you should see them on this screen, and you should select them both. If you select "admin", piXserve administrator access will be enabled. Similarly, if you select "users", only users will be able to login.
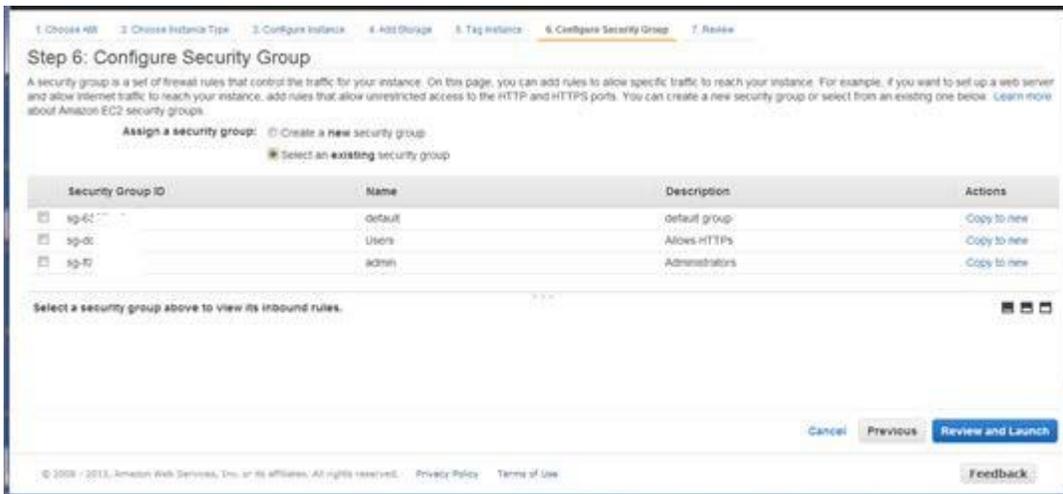
At this point you can review your configuration, make changes if needed, and launch your instance(s). It will take a little time for the instance to launch. You can follow the progress and status on the "Instances" tab of the AWS Management Console. When it's ready you will see the green "running" icon. If you click on the name of the instance, you can see more detailed info in the bottom part of the window.

9. To complete the preparation of your AMI(s) you may also wish to assign an IP address to the instance. You can go to the "Elastic IPs" section of the Management Console, select the IP address you want to assign, right mouse click, and choose to associate it with your desired piXserve-ECs/p/e instance.

10. Once your AMI has been launched, piXserve is automatically running on that instance[1]. To verify this, open your browser and point to the piXserve login page:

http://xxx.xxx.xxx.xxx/piXserve/loginpage.htm

where the IP address is the one that you assigned from your Elastic IP pool. You should see the login page come up.

---

[1] Depending on the type of AMI instance you have chosen, and the resources allocated to that instance, it may take a while before the Windows operating system is up and running and the piXserve service has launched. This process could take a couple of minutes or longer. Once the instance is fully up, accessing piXserve is a fast process.
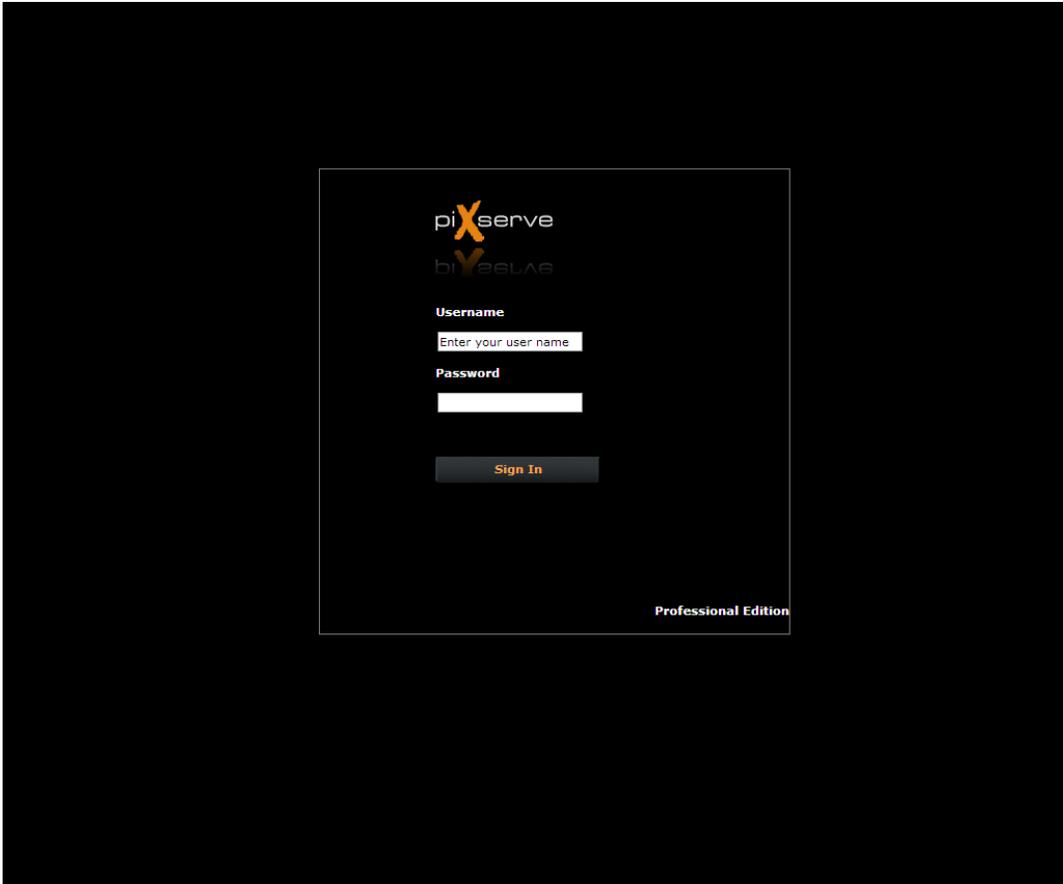
For your initial session, the credentials are "admin" and a password that is derived from the id of the AMI instance that you have launched. So, for example, if your Instance ID is "i-123a456b" then your initial admin password will be "i-123a456b" (without the quotes). You should change the default value as soon as you are logged in the first time. When you log in as "admin" you have access to both the Administrator Workspace and the User Workspace in piXserve (logging in as a regular user only gives access to the "User Workspace"). You will see this screen:
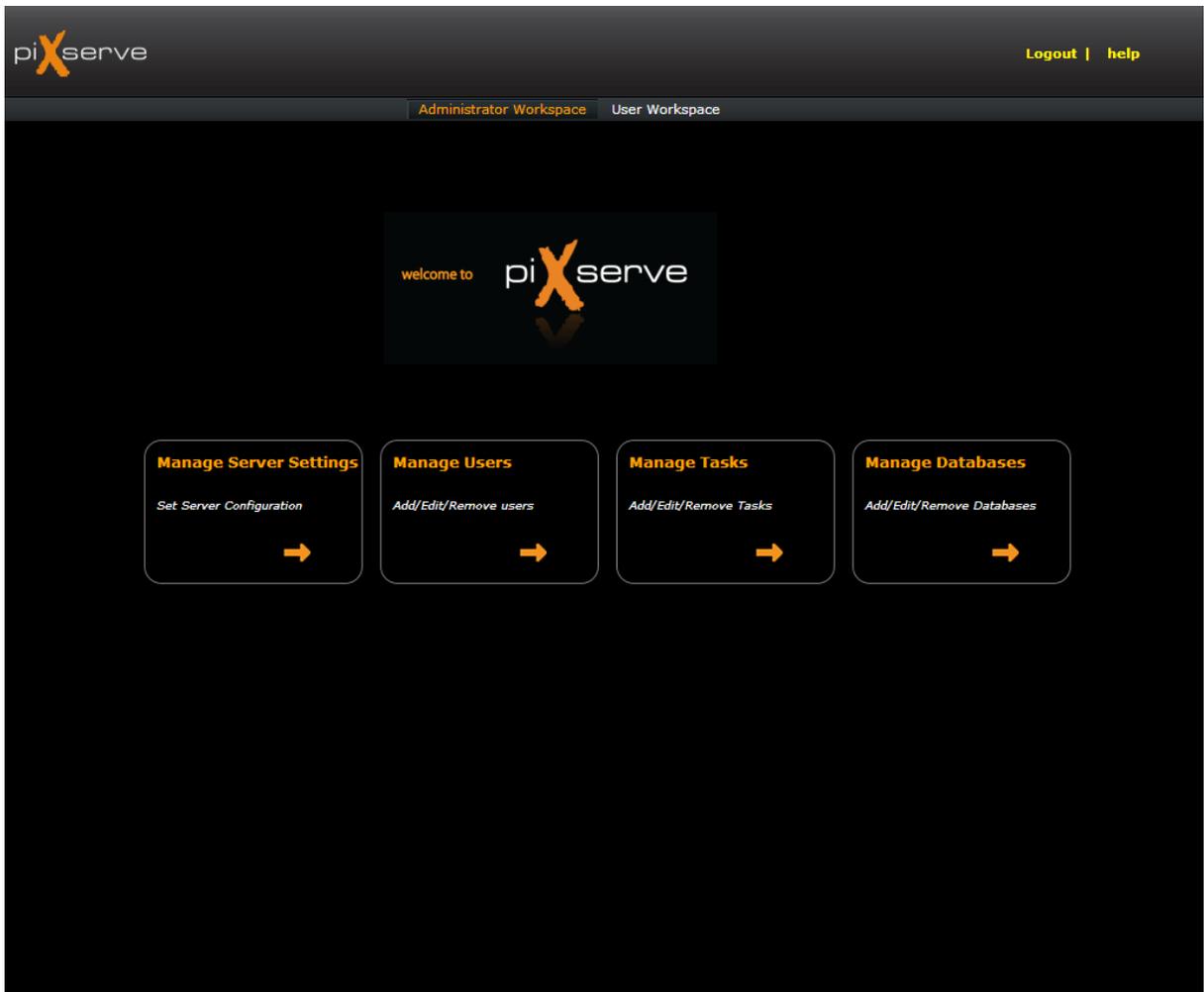
Click on "Manage Users" and you will come to a page where you can change the admin password as well as add/manage users.

Context sensitive help as well as general help is available online on each page of the piXserve application (click the "help" button on the top right of the page). As part of your initial set-up process, piXlogic will walk you through some of the key aspects of the application over the telephone, so call ahead to schedule a session (Tel. +1-650-967-4067, Monday-Friday, 9am-5pm Pacific Standard Time, except regular US holidays).

11. The piXserve AMIs are set up so that you can carry out most administrative functions through the piXserve user interface. The on-line documentation provides information about the various configuration options, and you are encouraged to call piXlogic for an informal telephone training session and to ask more detailed questions or receive specific help. The default http interface is through port 80 and for https through port 443. To enable the https connection, you will need to provide an authentication certificate. Please contact piXlogic technical support for information on how to make this change. Generally speaking,

you will need to SCP your certificate to the "/opt/piXlogic/conf/SSL" directory, and adjust the "/opt/piXlogic/conf/server.xml" file to reflect the name and location of your certificate.

12. The piXserve AMIs can operate both on CPU-only instances and GPU-assisted instances. GPU assisted instances are highly recommended as piXserve can use GPU resources to improve calculation speeds. piXserve takes advantage of as many CPU cores as are available on the AMI, and each core in turns takes advantage of GPU resources, if available, to carry out its work. Typically, the amount of on-board GPU RAM is the limiting factor with respect to how many CPU cores can make use of the underlying GPU card. Amazon makes available G3 and G4 instances powered by NVIDIA cards, and these are suitable for the purposes of piXserve.

    The number of cores that piXserve is allocated to use on the local machine, and the relationship to GPU memory are parameters that can be defined by the Administrator in the piXserve "Server Settings" administrator UI page. Please contact piXlogic for suggestions on parameter choices, and refer to the piXserve documentation for additional information.

13. If you choose to run piXserve in a GPU environment, make sure that the latest NVIDIA drivers are installed on your server.   The choice of drivers   depends on the instance type.   Instructions are provided here:

    https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/install-nvidia-driver.html#preinstalled-nvidia-driver

    Here is an example of how to do it:

yum erase nvidia cuda	This is to remove any prior installation
yum update -y	This is to make sure that we are working with an up to date system
yum install -y gcc kernel-devel-$(uname -r)	This is needed by the NVIDIA installer to update kernel configuration data
./NVIDIA-Linux-x86_64-450.80.02.run	This is an NVIDIA installer downloaded from the NVIDIA support site. It contains the latest linux drivers as of the time of this writing. Your version number may vary.
nvidia-smi	Run this command to check that the drivers are properly accessible. This is a sample return:

Note that an incorrect configuration or the absence of the driver will cause piXserve to abort if piXserve was configured to use GPU resources. Also, note that if you are planning to use multi-GPU machines, it is convenient to choose ami's that leverage NVLink high bandwidth interconnects so that they can be addressed as a single GPU block by piXserve.

Additional Resources

piXlogic Support page:

https://www.pixlogic.com/elastic-cloud-support

Amazon Web Services - Resource Center:

http://aws.amazon.com/articles/

Amazon Web Services: Discussion Forums:

https://forums.aws.amazon.com/index.jspa?#

Amazon EC2 Getting Started Guide:

http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/gsg-aws-intro.html

Amazon EC2 Management Console:

https://console.aws.amazon.com/console/home?#