Using dotSeech on Amazon Web Services with EC2 & EBS

By Mike Y Walker

Using dotSEEK - Win

Articles
Using dotSeech on Amazon Web Services with EC2 & EBS
Introduction
Project Prerequisites
Create an EC2 Instance
Create an Elastic IP Address
Create an EBS Volume
Using the VPC Wizard to Set up a NAT
Storage Configuration
Program Startup, Check for the Engine
Handling Query Updates From The Engine
The Index.Vol
Calling the .Net
Deploying to AWS
Seeing the Results
Wrapping Up
Licenses
Share
About the Author
Comments and Discussions

Download dotSeekOnAmazonAws.zip - 202.5 KB

Introduction

This article is the Amazon Web Services version of my previous C# Search Client. Here, we have set the stage to run dotSeek on Amazon's EC2 instance and store the index on Amazon's EBS volume.

Project Prerequisites

Before we can begin, we need to set up an Amazon Web Services account. Amazon has a free tier that will work for most of what I do in this article with the one little catch and that is the t2 instance type. Now, I found that the t2.micro instance type is not powerful enough to run dotSeek. So, I used the t2.small instance type which is not free. Also, I used the Windows Server 2012 R2 Base AMI.

Create an EC2 Instance

Log into your AWS account and click on the EC2 link.

Click on the Launch Instance button.

Click on the Windows tab and then click on the Microsoft Windows Server 2012 R2 Base AMI.

Choose the t2.small instance type. Then click on the Configure Instance Details button.

In the next step, we will change the Shutdown behavior to Terminate. Then click on the Add Storage button.

Click on the Add New Volume button.

Change the Volume Type to Provisioned IOPS. Then change the IOPS to 100. Finally, change the Delete on Termination to Yes. Then click on the Tag Instance button.

Click on the Add Tag button. Then enter Name for the Key and dotSeek for the Value. Then click on the Configure Security Group button.

Select an existing security group and choose the default security group. Then click on the Review and Launch button.

Review the settings and then click on the Launch button.

Select an existing key pair or create a new key pair. Then click on the Launch Instances button.

Click on the View Instances button to navigate to the EC2 Dashboard/Instances Screen. The instance should be running.

Create an Elastic IP Address

Next, we need to create an Elastic IP address. Click on the Elastic IPs link in the navigation pane. Then click on the Allocate New Address button.

Click on the Allocate button.

Click on the Close button.

Select the Elastic IP address that you just created and click on the Associate Address button.

Select the instance that you created earlier. Then click on the Associate button.

Create an EBS Volume

Next, we need to create the EBS volume that we will be using for the dotSeek index. Click on the Volumes link in the navigation pane. Then click on the Create Volume button.

Change the Volume Type to Provisioned IOPS. Then change the IOPS to 100. Finally, change the Size to 4 GiB. Then click on the Create button.

Click on the Close button.

Select the volume that you just created and click on the Attach Volume button.

Select the instance that you created earlier. Then click on the Attach button.

Using the VPC Wizard to Set up a NAT

Go back to the EC2 Dashboard and click on the VPC link.

Click on the VPC Wizard button.

Select the VPC with a Single Public Subnet and NAT option. Then click on the Select button.

Click on the Create VPC button.

Click on the OK button.

Click on the Subnets link in the navigation pane. Then select the public subnet that you just created. Then click on the Route Table tab. Then click on the route table that is associated with the public subnet.

Click on the Edit button.

Change the Destination to 0.0.0.0/0. Then change the Target to the NAT instance that you created earlier. Then click on the Save button.

Click on the Close button.

Storage Configuration

Now that we have the EBS volume attached to the EC2 instance, we need to configure it. Click on the Instances link in the navigation pane. Then select the instance that you created earlier. Then click on the Connect button.

Click on the Download Remote Desktop File button.

Click on the Get Password button.

Click on the Choose File button. Then select the private key file that you created earlier. Then click on the Decrypt Password button.

Click on the Download Remote Desktop File button.

Click on the Cancel button.

Click on the Open button.

Click on the Connect button.

Click on the Yes button.

Click on the Server Manager icon.

Click on the File and Storage Services link.

Click on the Disks link. Then right click on the disk that is offline. Then click on the Bring Online menu item.

Click on the Yes button.

Right click on the disk again. Then click on the New Volume menu item.

Click on the Next button.

Select the disk that you want to configure. Then click on the Next button.

Enter I for the Volume label. Then click on the Next button.

Click on the Next button.

Click on the Create button.

Click on the Close button.

Now, we can see that the I drive is available in the Windows Explorer.

Program Startup, Check for the Engine

Now let's look at the modified .NET code going to look at the updated processing logic. Here is the new Program.cs file. Let's look at the new code.

```
using System.Configuration;

namespace dotSeekOnAmazonAws
{
    class Program
    {
        static void Main(string[] args)
        {
            string indexPath = ConfigurationManager.AppSettings["IndexPath"];
            string enginePath = ConfigurationManager.AppSettings["EnginePath"];
            bool isEngineRunning = false;
            Engine.CheckForEngine(enginePath, indexPath, ref isEngineRunning);
            Engine.Run(indexPath, isEngineRunning);
        }
    }
}
```

The Program.cs file has not changed much. The only change is that we are getting the enginePath from the app.config file. Then we are passing the enginePath to the CheckForEngine() method.

```
public static void CheckForEngine(string enginePath, string indexPath, ref bool isEngineRunning)
{
    string engineExe = enginePath + "\\dotSeek.Engine.exe";
    if (!File.Exists(engineExe))
        throw new Exception("Engine not found at " + enginePath);
    Process[] processes = Process.GetProcesses();
    foreach (Process process in processes)
    {
        if (process.ProcessName == "dotSeek.Engine")
        {
            isEngineRunning = true;
            break;
        }
    }
    if (!isEngineRunning)
    {
        Process.Start(engineExe, indexPath);
        isEngineRunning = true;
    }
}
```

The new CheckForEngine() method is the same as the one in the dotSeek project that we ported to Amazon Web Services. The only change is that we are passing the enginePath to the Process.Start() method.

Handling Query Updates From The Engine

Now, let's look at the new Engine.cs file.

```
public static void Run(string indexPath, bool isEngineRunning)
{
    string queryPath = indexPath + "\\Query.txt";
    string resultsPath = indexPath + "\\Results.txt";
    string lockPath = indexPath + "\\lock";
    string query = "";
    DateTime lastRead = DateTime.MinValue;
    while (true)
    {
        Thread.Sleep(100);
        if (File.Exists(lockPath))
            continue;
        DateTime lastWrite = File.GetLastWriteTime(queryPath);
        if (lastWrite != lastRead)
        {
            lastRead = lastWrite;
            query = File.ReadAllText(queryPath);
        }
        if (query == "")
            continue;
        if (query == "!!exit!!")
            break;
        Console.WriteLine("Query: " + query);
        Searcher searcher = new Searcher(indexPath);
        Hits hits = searcher.Search(query);
        Console.WriteLine("Hits: " + hits.Length());
        using (StreamWriter writer = new StreamWriter(resultsPath))
        {
            writer.WriteLine(hits.Length());
            for (int i = 0; i < hits.Length(); i++)
            {
                writer.WriteLine(hits.Doc(i).Get("path"));
            }
        }
        query = "";
    }
    if (!isEngineRunning)
        Process.GetCurrentProcess().Kill();
}
```

The Run() method is the same as the one in the dotSeek project that we ported to Amazon Web Services. The only change is that we are passing the isEngineRunning to the Run() method. Then we are passing the isEngineRunning to the Process.GetCurrentProcess().Kill() method. This is to make sure that the dotSeek.Engine.exe process is killed when the dotSeek.exe process is killed.

The Index.Vol

In the dotSeek.exe.config file, we have the following:

```
<add key="IndexPath" value="I:\Index.Vol" />
```

This is the location of the index on the EBS volume. Now, let's look at the new Searcher.cs file.

```
public Searcher(string indexPath)
{
    _directory = FSDirectory.Open(new DirectoryInfo(indexPath));
    _searcher = new IndexSearcher(_directory, true);
    _analyzer = new StandardAnalyzer(Version.LUCENE_30);
}
```

Here, we are passing the indexPath to the FSDirectory.Open() method.

Calling the .Net

Finally, let's look at the C:\Start.bat file.

```
I:\dotSeek.exe I:\Index.Vol
```

The Start.bat file is launched when the instance is started. Then it calls the dotSeek.exe and passes the I:\Index.Vol path to it. Now, let's look at the new Form1.cs file.

```
public Form1()
{
    InitializeComponent();
    _indexPath = ConfigurationManager.AppSettings["IndexPath"];
    _queryPath = _indexPath + "\\Query.txt";
    _resultsPath = _indexPath + "\\Results.txt";
    _lockPath = _indexPath + "\\lock";
    _searcher = new Searcher(_indexPath);
}
```

Here, we are getting the indexPath from the app.config file. Then we are using the indexPath to create the _queryPath, _resultsPath and _lockPath. Then we are passing the indexPath to the _searcher.

```
private void SearchTextBox_TextChanged(object sender, EventArgs e)
{
    string query = SearchTextBox.Text;
    if (query.Length < 3)
        return;
    File.WriteAllText(_lockPath, "lock");
    File.WriteAllText(_queryPath, query);
    File.Delete(_lockPath);
    Thread.Sleep(100);
    string[] results = File.ReadAllLines(_resultsPath);
    int hits = int.Parse(results[0]);
    ResultsListBox.Items.Clear();
    for (int i = 1; i <= hits; i++)
    {
        ResultsListBox.Items.Add(results[i]);
    }
}
```

Here, we are creating the _lockPath file. Then we are writing the query to the _queryPath file. Then we are deleting the _lockPath file. Then we are reading the results from the _resultsPath file.

Deploying to AWS

The easiest way to get the dotSeek.exe and dotSeek.Engine.exe deployed is to use Amazon's Elastic Beanstalk. First, I'm creating a set of batch files to publish the dotSeek.exe and dotSeek.Engine.exe to the C:\Publish folder.

```
dotnet publish -c Release -o C:\Publish
```

Here is the Publish.bat file. It is in the dotSeek.exe project folder.

```
dotnet publish -c Release -o C:\Publish
```

Here is the Publish.bat file. It is in the dotSeek.Engine.exe project folder.

Next, I'm using the Elastic Beanstalk Command Line Interface (EB CLI) to deploy the dotSeek.exe and dotSeek.Engine.exe to AWS. Here is the documentation for the EB CLI. Here is the documentation for installing the EB CLI.

Now, I'm going to create a new folder called C:\Deploy. Then I'm going to copy the C:\Publish folder to the C:\Deploy folder. Then I'm going to rename the C:\Deploy\Publish folder to C:\Deploy\dotSeek. Then I'm going to create a new file called C:\Deploy\dotSeek\Start.bat. Here is the Start.bat file.

```
I:\dotSeek.exe I:\Index.Vol
```

Now, I'm going to create a new file called C:\Deploy\dotSeek\.ebignore. Here is the .ebignore file.

```
dotSeek.Engine.exe
```

Now, I'm going to create a new file called C:\Deploy\dotSeek\.ebextensions\dotSeek.config. Here is the dotSeek.config file.

```
commands:
  01_kill_engine:
    command: taskkill /F /IM dotSeek.Engine.exe
    ignoreErrors: true
```

Now, I'm going to create a new file called C:\Deploy\dotSeek\.ebextensions\dotSeek.Engine.config. Here is the dotSeek.Engine