Implementing VMware vSphere 5 Auto Deploy on the Cisco Unified Computing System

What You Will Learn

This document provides an overview and detailed procedure for installing and configuring the VMware vSphere 5 Auto Deploy feature on a Microsoft Windows server. VMware vSphere is an industry-leading virtualization platform, transforming data centers into dramatically simplified cloud infrastructures and enabling the next generation of flexible, reliable IT services. Cisco Unified Computing System™ (Cisco UCS™) and VMware vSphere together enable rapid application and infrastructure deployment with greater security, availability, and performance leading to increased return on investment (ROI) and reduced total cost of ownership (TCO) for computing environments.

Overview

VMware Auto Deploy, a feature available with VMware vSphere 5.0, combines the features of host profiles, VMware Image Builder, and Preboot Execution Environment (PXE)—a network boot using PXE—to simplify the task of managing VMware ESXi hypervisor installation and upgrades for hundreds of machines. VMware ESXi host images are centrally stored in the VMware Auto Deploy library. New hosts are automatically provisioned, based on rules defined by the user. Rebuilding a server to a clean state is as simple as a reboot. To move between VMware ESXi versions, simply update a rule using the VMware Auto Deploy PowerCLI and perform a test compliance and repair operation.

VMware Auto Deploy customizes VMware ESXi systems using host profiles and other information stored on the managing VMware vCenter Server system. You can set up the environment to use different images and different host profiles for different hosts. VMware Auto Deploy can be installed on a standalone Microsoft Windows host virtual machine or on the VMware vCenter Server itself. It also ships with the VMware vCenter Server appliance. In either case, it is registered with a VMware vCenter Server.

VMware Auto Deploy works in tandem with Dynamic Host Control Protocol (DHCP), Trivial File Transfer Protocol (TFTP), and the VMware vCenter Server. The PXE boot image is downloaded, and the preboot environment runs on the host. It contacts the VMware Auto Deploy server, which checks the rule engine based on which image is offered to the host and which host profile is applied (Figure 1).
The following concepts relate to configuration of the VMware Auto Deploy server:

- **Software depot**: Location where VMware ESXi images, drivers, and other add-ons are stored
- **Image profile**: VMware ESXi images available in the depot
- **Host profile**: Optional configuration template created on VMware vCenter Server
- **Deploy rule**: Pattern-based policy through which the service provider specifies the image profile that hosts that match the pattern should receive
- **Deploy rule set**: Active set of deploy rules

**Integration with Cisco Unified Computing System**

As part of Cisco UCS Manager Release 2.0, the Cisco UCS cluster name, service provider name, and service provider template name are stored in the system management BIOS (SMBIOS) of the blade. This information is now available for the VMware Auto Deploy server to check the rules engine for the host profile to be applied. Note that no configuration needs to be performed to implement this capability on Cisco UCS. As part of the BIOS in Cisco UCS Manager Release 2.0, this information is automatically available to create a rule set based on the information exposed.
Deployment Steps

Step 1: Create Service Profile

Create a service profile template for the VMware ESX cluster that will be deployed using VMware Auto Deploy (Figure 2). The first (and preferably only) entry in the boot order should be **LAN (PXE)**.

*Figure 2.* Creating a Service Profile Template for the VMware ESX Cluster

The Ethernet virtual network interface cards (vNICs) should be created based on requirements for the native VLAN on the PXE vNIC set as the VLAN on which the DHCP server is reachable, either by creating them on the same VLAN or by entering `ip helper-address` on the Layer 3 interface (Figure 3). Create service profiles from the template as required.

*Figure 3.* Configure vNICs

**Note:** The BIOS version on the blades must be from Cisco UCS Release 2.0. Use of a firmware policy as part of the template is highly recommended.

Step 2: Configure DHCP Server

Configure a DHCP server with the relevant scope and boot options (Figure 4). In the following example, the Microsoft Windows DHCP server running on VMware vCenter is used. The following URL provides a step-by-step guide to configuring a Microsoft Windows DHCP server:


The MAC address must be populated from a service profile created in Step 1, which will be used to deploy the first host.
Figure 4. Configuring the DHCP Server

Click Server Options to configure options 66 and 67 (Figure 5). Option 66 will have the TFTP server IP address. The TFTP server could be the same VMware vCenter Server or a different server.

Figure 5. Configuring Server Options

Option 67 will have the value `undionly.kpxe.vmw-hardwired` (Figure 6).
For a Linux-based DHCP server, the configuration shown in Figure 7 applies.

**Figure 7.** Linux-Based DHCP Server Configuration

```plaintext
subnet 10.10.10.0 netmask 255.255.255.0 {
    range 10.10.10.50 10.10.10.60;
    if ((exists user-class) and (option user-class = "gPXE")) {
        filename "https://10.10.10.100:6501/vmware/rbd/tramp;"
    } else {
        filename "undionly.kpxe.vmware-hardwired;"
    }
    option subnet-mask 255.255.255.0;
    option broadcast-address 10.10.10.255;
    option routers 10.10.10.1;
    option next-server 10.10.10.10;
}
```

**Note:** 10.10.10.100 is the IP address of the VMware Auto Deploy server.

To rule out DHCP server configuration problems at this time if the service provider is associated, the blade should receive an IP address defined in the reservation operation, but should fail at the TFTP file download because the TFTP server has not been configured (Figure 8).
Figure 8. TFTP File Download Failure

```
Intel(R) Boot Agent XG v2.1.11
Copyright (C) 1997-2000, Intel Corporation

CLIENT MAC ADDR: 00 25 B5 12 34 39  GUID: 6218BF08 B990 11DE 129C 00002EF7867D
CLIENT IP: 172.25.100.122  MASK: 255.255.255.0  DHCP IP: 172.25.100.116
TFTP:
PXE-T01: File not found
PXE-ESW: TFTP Error - File Not found
PXE-MOF: Exiting Intel Boot Agent.
```

**Note:** If the blade does not receive an IP address, check the native VLAN configuration in the service provider and the connectivity and configuration of the DHCP server.

### Step 3: Set Up VMware Auto Deploy

Run VMware-Auto Deploy, which is an application in the vctools directory of VMware vSphere 5. Keep all settings at the defaults; only the IP and credentials for the VMware vCenter need to be provided.

For a successful installation, the Plug-in Manager should show VMware Auto Deploy as registered, and an icon for it should appear under Administration. VMware Auto Deploy by default runs on TCP port 6501 (Figure 9).

**Figure 9.** Plug-in Manager and Administration Displays

Click the Auto Deploy Icon, download the TFTP Boot Zip file, and extract the contents to the tftproot file of the TFTP server (Figure 10).
Figure 10. Download TFTP Boot Zip and Extract Contents

On the blade, reboot with the configuration shown in Figure 11.

Figure 11. Blade Reboot Configuration

You should see the DHCP requests shown in Figure 12.

Figure 12. DHCP Requests Displayed

Note: If the second DHCP request does not receive an IP address with a Microsoft Windows DHCP server, you could be encountering CSCts86689 (Figure 13).
Figure 13. Service Profile Identifiers Read by Auto Deploy

Note the service provider template name, service provider name, and Cisco UCS name that are populated because of Cisco UCS integration with VMware vSphere Auto Deploy.

Step 4: Install VMware vSphere PowerCLI and Configure VMware Auto Deploy

Install VMware vSphere PowerCLI, found in the VMware vSphere 5 distribution.

**Note:** Install or upgrade the PowerShell feature in Microsoft Windows if required. Two icons are created at the end of the installation. Do not use VMware vSphere Power CLI (32bit) for configuration.

Click the VMware vSphere PowerCLI icon to launch VMware PowerCLI and run the following commands (Figure 14):

Set-ExecutionPolicy unrestricted
Connect-viserver <IP of VCenter>
Download the VMware ESX image (.zip) from the vSphere distribution to a local folder and run the `Add-ESXSoftwareDepot` command to add the image to the depot. Run the `Get-EsxImageProfile` command to check the status of the image import operation (Figure 15).

You can now create a deploy rule using the `New-DeployRule` command. The syntax for this command is:

```
New-DeployRule –Name "<name of rule>" –Item "<Image>" –Pattern "<pattern>"
```

A match is performed on the service provider template oemstring "$SPT:ESXFarm"; the escape character (') must be entered before the $ sign.

The `Get-DeployRule` command can be entered to query the rule. Enter `Add-DeployRule -DeployRule "<name of rule>"` to add the newly created rule to the list of rules (Figure 16).
When the blade is rebooted, VMware Auto Deploy should have a match for the oemstring, and the PXE boot of VMware ESXi with the image service provider specified should begin (Figure 17).

After loading is completed, it will show up in VMware vCenter as a standalone host (Figure 18).
The host can be configured according to the properties desired (vSwitch configuration, data stores, security settings, etc.).


**Step 5: Create Host Profile**

At this time, if the host is rebooted all the changes made to it will be lost. To save the settings and use them for other servers to be deployed as part of this cluster, the host profile must be created. Choose Management > Host Profiles and click Create Host Profile to create a host profile from an existing host (Figure 19). Select the host and provide a name for the profile.

**Figure 19.** Create Host Profile

![Create Host Profile](image)

At this time, you can delete this host from VMware vCenter (Figure 20).
Step 6: Configure VMware Auto Deploy for Service Providers to Use Host Profile

Delete the existing deploy rule using the `Remove-DeployRule` command (Figure 21).

You must create a new deploy rule that specifies states the cluster for this host to join and the host profile to be used along with the image and the pattern match (Figure 22):

```
New-DeployRule -Name "<name>" -Item "<image>", "<Cluster to join>", "<Host profile to use>" -Pattern "<pattern>"
```

```
Add-DeployRule -DeployRule "<name>"
```

You can use the `Get-DeployRuleset` command to query the rule created (Figure 23).
When the blade is rebooted, the system is booted using PXE and joins the cluster with the settings as specified in the host profile (Figure 24).

**Figure 24.** Query Rule Created

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Status</th>
<th>Details</th>
<th>Initiated by</th>
<th>vCenter Server</th>
</tr>
</thead>
</table>

Step 7: Deploy Additional Servers to the Cluster

Follow Step 2 (Configure DHCP Server) to add configurations for other service providers created in Step 1. When those service providers are associated, they will be deployed using the common host profile because the match is implemented on the service provider template name.

**Conclusion**

VMware Auto Deploy is a new stateless function that ships with VMware vSphere 5.0. It enables PXE boot of bare-metal hosts and assigns specific configurations to them. In a Cisco UCS operating environment, multiple VMware vSphere servers can be deployed with a single click, using service profiles. By providing a link to the VMware vCenter boot image in the service profile template, the process of deploying multiple VMware vSphere servers can be automated, preventing human errors and deploying servers in a policy-complaint manner. The result is a dramatic savings in time and reduced potential for manual errors.

In the case of incumbent server vendors, the process of deploying multiple VMware vSphere servers is manual and needs to be repeated multiple times, each iteration requiring a server selection, locating a VMware Auto Deploy image, and booting the server. The process is time consuming and error prone. Using VMware Auto Deploy in a Cisco UCS environment, you can easily and quickly deploy VMware vSphere servers using service profile templates—another example of the ease, speed, and agility of Cisco UCS.

To find out more about UCS Service Profiles, please visit:

Create a Service Profile for Cisco UCS Blade


For a basic overview of UCS, please visit:
UCS Overview