HOW TO ENHANCE PIVOTAL’S BOSH BACKUP & RESTORE WITH DATA DOMAIN & BOOSTFS

Simplify Pivotal Cloud Foundry backups by integrating with efficient, highly reliable, dedicated backup storage

This whitepaper describes how to design your Pivotal Cloud Foundry backup and restore implementation to maximize performance and minimize cost. Dell EMC’s Data Domain provides significant reduction in storage requirements for Pivotal Cloud Foundry backups made via BOSH Backup and Restore (BBR), while also enabling impressive reduction in backup transfer time.

September, 2017
Table of Contents

Executive Summary  ............................................................................................................................................................... 3
Why Data Domain and BoostFS? .......................................................................................................................................... 4
What is BOSH Backup & Restore? ........................................................................................................................................ 5
How to leverage Data Domain for your BBR-based backups ................................................................................................ 6
Restoring PCF using BBR backup artifacts ........................................................................................................................... 7
Our internal testing results ..................................................................................................................................................... 8
Anticipated future evolution .................................................................................................................................................... 8
Conclusion .............................................................................................................................................................................. 9
References ............................................................................................................................................................................. 9
Executive Summary

Pivotal has developed a solution used to back up the Pivotal Cloud Foundry (PCF) platform which they call BOSH Backup and Restore (BBR). The main purpose of BBR is to enable the recovery of the PCF platform in case of corruption or disaster. Pivotal has released BBR into the open source Cloud Foundry (CF) community, and there is an expectation that this solution will be widely adopted in the PCF customer base.

Dell EMC, as a leader in data protection and innovation, would like to provide the following whitepaper as a guide to be used to enhance the native BBR solution with the industry leading power of Data Domain. This whitepaper outlines a solution based on Data Domain (DD) and DD Boost File System (BoostFS) that solves some of the complexities posed by BBR backups. This solution can be used with physical Data Domain appliances in on-premises PCF deployments as well as on any public or private cloud environment where Data Domain Virtual Edition (DD VE) is available. The solution focuses on moving the BBR backup copies directly to Data Domain, enabling data reduction, more secure backups, and the removal of backup data from the PCF environment itself. Moreover, the solution does not require changing the basic workflows of BBR. Internal testing of this solution shows a 75% reduction in storage due to data reduction, and a 20% reduction in the transfer time incurred per backup. These benefits are easily obtainable for any current PCF customer through simple configuration changes and the utilization of Data Domain or Data Domain VE. The solution outlined in this whitepaper is targeted to PCF environments, but can easily be applied to other Cloud Foundry environments as well.

This whitepaper is meant to illustrate how Data Domain could be leveraged for usage with PCF and BBR. DellEMC and Pivotal are actively testing this solution with interested customers. If you would like to try out the product, please contact Feedback_DD_BBR@emc.com. In the meantime, please note that this is not an officially supported solution and should not be used in a production environment at this time.
Why Data Domain and BoostFS?

Data Domain is the ideal backup store for BBR backups as the industry leader in the space of purpose-built backup appliances. Utilizing efficient de-duplication, Data Domain reduces the storage footprint of backups by intelligently identifying unique data in an incoming data stream. When used with BBR, Data Domain provides a similar advantage to incremental backups by saving only new or changed data after the initial full backup.

Data Domain offers a number of compelling advantages as the protection storage leader:

- Flexible deployment through Data Domain Virtual Edition (DD VE) on vSphere and Hyper-V, plus Azure and AWS clouds.
- Capability to natively-tier data to cloud object store provides modern and cost efficient long-term retention
- Industry leading storage reduction through variable-length deduplication and Stream Informed Segment Length processing
- Data Invulnerability Architecture with multiple means to verify and protect data
- Secure multi-tenancy allows for high system utilization

BoostFS allows a file system to be mounted and used by applications, where the storage is provided by Data Domain. The storage is accessed via a network using the DD Boost protocol which offers increased performance for faster, more efficient backups to Data Domain and DD VE instances. BoostFS provides a distributed deduplication process, enabling less data to be transferred between the host and Data Domain. Additional benefits of not retransmitting duplicate data include:

- Reduction in backup transfer time
- Reduction in bandwidth requirement

BoostFS is free to use for customers with existing DD or DD VE appliances. The addition of BoostFS to an existing BBR setup is extremely easy to accomplish and does not impact how BBR functions operationally. Using BoostFS with Data Domain or DD VE in combination for BBR backups results in:

- Greatly reduced backup storage usage due to efficient de-duplication
- Reduced storage demands on the Jumpbox VM running BBR as backup artifacts are no longer stored locally
- Faster backups due to the reduced amount of data that needs to be copied from the Jumpbox to Data Domain
- Reduced network bandwidth used between Jumpbox and Data Domain/DD VE storage

“Data Domain is the ideal backup store for BBR backups, as the industry leader in the space of purpose-built backup appliances.”
What is BOSH Backup & Restore?

Pivotal’s BBR utility provides a framework for the backup and restore of BOSH deployments and BOSH directors in PCF platforms. A command line tool is provided to invoke the backup and restore jobs, which triggers backup and restore functionality within BOSH deployed elements. During a backup, the BBR tool collects backup artifacts and transfers them to a default location on a virtual machine referred to in the diagram below as the jumpbox where BBR is installed and running.

The Pivotal BBR team has provided the BBR backup-restore scripts needed to backup persistent data associated with PCF Elastic Run Time (ERT), as well as for BOSH director deployment. The data associated with the deployed applications on the PCF platform is not currently backed up unless that data is located in a BOSH deployed service that has implemented BBR supported backup and restore functionality.

Individual BOSH releases are recommended to develop their own backup and restore implementations by adhering to the contract defined by BBR for the backup and restore interfaces. Once implemented, these interfaces are executed in a well-defined sequence by the BBR orchestrator. The semantics of the actual backup and restore interfaces are left to the BOSH deployment release authors who know best about the required steps for the backup and restore. These include pre-backup locks, post-backup unlocks as well as pre-restore and post-restore activities. BBR ensures that with the appropriate implementation of backup/restore interfaces, deployments with distributed job-instances can maintain integrity of the backup and restored data.

The following workflow, provided in the Pivotal Documentation, illustrates the BBR workflow for the PCF platform operator:

1. Ensure that the PCF platform is of version 1.11 or above
2. Install the BBR binary on PCF Jumpbox
3. Execute the BBR binary on the Jumpbox specifying the PCF ERT or BOSH-director as the component to be backed up
a. The BBR binary identifies the backup scripts associated with all jobs specified in the deployment, executes them in correct order, and copies the backup artifacts from the deployments to the Jumpbox.

4. Transfer the backup artifacts from the Jumpbox to a secondary backup store

The expectation is that all data-service components deployed by BOSH will eventually support backup and restore scripts.

In the workflow described above, each backup is a full backup, initially stored on the Jumpbox itself. Depending on the size of the PCF deployment, the Jumpbox could require large amounts of storage. In addition, one may need to store multiple backups on secondary storage to satisfy availability requirements further increasing the storage footprint. In the following sections, we describe how to leverage the power of Data Domain and simplicity of BoostFS to reduce the storage requirements on the Jumpbox, and reduce overall backup storage usage as a consequence. In addition, this eliminates the need for a user to move the backup from the jumpbox to the secondary storage which is a tidier and more secure solution.

How to leverage Data Domain for your BBR-based backups

The following illustration shows the BBR workflow using BoostFS on the Jumpbox and DD VE as the remote backup copy storage for a PCF deployment on a public cloud.

The red dashed ovals indicate the changes required to use BoostFS and DD VE for the BBR workflow. Begin with an established PCF deployment in an environment that matches the pre-requisites outlined below.

This whitepaper is meant to illustrate how DD could be leveraged for usage with PCF and BBR. It is not an officially supported solution and should not be used in a production environment at this time. This solution is still under test with DellEMC and Pivotal; consider this for use in your sandbox or development environments initially.

The diagram below represents our test environment. PCF and DD VE are both available on a variety of clouds as well as on-premises, and the solution should work on any of them as long as the Data Domain environment is network accessible from PCF.
Pre-requisites:
1. Pivotal Cloud Foundry version 1.11 or later installed on-premises or in a supported cloud environment. Go to https://docs.pivotal.io/pivotalcf/installing/index.html for more information on installing PCF.

Installation Steps:
   1. Note: The OS for the jumpbox will need to support File system in User SpacE (FUSE) for BoostFS.
3. Establish network connectivity to your Dell EMC Data Domain system and your Jumpbox. Depending on security rules, this may require opening network ports or creating routes to allow the Jumpbox to communicate with the Data Domain system.
4. On the Data Domain, create the storage unit that will be used as the mount point to store the backups using either the CLI or GUI interface for BoostFS.
5. On the Jumpbox, mount the storage-unit created in step-4.
   1. ssh to the Jumpbox
   2. Execute `boostfs lockbox set -d <dd IP address> -s <Storage-unit name> -u <authorized-user-name>.
   3. Create a folder on the Jumbox to be used for the mount point. (e.g. mkdir ddbase_backup).
   4. Execute `boostfs mount <folder-name> -d <dd IP address> -s <storage-unit-name>.
6. The BoostFS mount point can now be used as a destination for BBR by executing BBR commands from within the BoostFS mounted folder.

This process provides a simpler and more secure workflow by eliminating the need to manually copy the backup to secondary storage in the original workflow described above.

Restoring PCF using BBR backup artifacts
Restoring your PCF instance is a complex process. The documented procedure describes a complex process which includes preparation, deployment, and restoration of the environment. The restoration process happens in 3 general phases:

- Rebuilding the infrastructure for your environment
- Transfer backup artifacts to jumpbox
- Restoring the environment from the backup artifacts

With our new solution using a Data Domain enabled mount point; it is no longer necessary to transfer the backup artifacts to the jumpbox. The restoration steps can directly use the BoostFS mount point to access previous backup artifacts once it is mounted on the Jumpbox.

All BBR backups are full backups of the environment, but with BoostFS and Data Domain the deduplication process makes a full backup perform like an incremental one. With the large size of the files being restored, alleviating the need to copy the artifacts back to the jump box from secondary storage saves time and network costs. This may result in a faster restoration of the director and elastic runtime. However, given the length of time required for other steps in the restore procedure, this time gain is not significant on its own. The main focus of this solution is to enhance the backup process, not the restoration.

The full documented restore procedure can be found at https://docs.pivotal.io/pivotalcf/customizing/backup-restore/restore-pcf-bbr.html
Our internal testing results

This solution was tested and validated with multiple backups of PCF deployments in a public cloud service provider. Both BBR direct backups and DD enabled backups were tested, using DD VE. The same cloud storage was used as the ultimate target for all backup artifacts. Our test setup had a very modest 15GB of data to be backed up. As subsequent backups increase efficiency due to de-duplication, we were able to see up to a 75% reduction in backup space used per backup following the initial copy.

<table>
<thead>
<tr>
<th>Backup</th>
<th>Cumulative Backup Size in GB (No DD)</th>
<th>Cumulative Backup Size in GB (with DD)</th>
<th>Data Change % ***</th>
<th>Compression Ratio with DD</th>
<th>Compression Percent with DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backup-1</td>
<td>34</td>
<td>10.6</td>
<td>100.0%</td>
<td>1.3</td>
<td>24.3%</td>
</tr>
<tr>
<td>Backup-2</td>
<td>28.2</td>
<td>11.1</td>
<td>1.4%</td>
<td>2.5</td>
<td>60.6%</td>
</tr>
<tr>
<td>Backup-3</td>
<td>45.5</td>
<td>13.7</td>
<td>7.6%</td>
<td>3.3</td>
<td>69.9%</td>
</tr>
<tr>
<td>Backup-4</td>
<td>64.1</td>
<td>15.3</td>
<td>5.7%</td>
<td>4.2</td>
<td>76.1%</td>
</tr>
</tbody>
</table>

***Note: Data Change % calculation is an estimate done based on the backup size change. A fair assumption as PCF platform changes were done by adding applications/services.

Additionally, we saw a 20% reduction in transfer time required to complete those subsequent backups. Due to the method by which BBR creates and collects backup artifact, BoostFS is only able impact the tail end of the backup cycle, resulting in the 20% gain. Extrapolating the data reduction and time benefits for a series of full backups for large scale customers, the gains provided by Data Domain quickly become significant to operations.

Anticipated future evolution

In theory, this solution could be enhanced, extended, or applied in the following ways:

- We expect vendors currently providing BOSH deployed services to extend their solutions to include BBR capable backup and restore functionality, increasing both the level of protection, and the size of data being protected, and making integrations with Data Domain even more valuable
- BBR itself could be enhanced to allow for backup artifacts to move directly to backup storage, bypassing the Jumpbox completely, possibly adding incremental backups
- A similar backup strategy could be provided for other services such as the Pivotal Container Service (PKS) or natively to Kubernetes deployments
- Cross cloud artifact replication to enable cloud independence and workload portability based on geographical needs and infrastructure availability
Conclusion

Our goal is to ensure the best possible user experience for customers with both PCF and DD installations. As a start, we have provided this guide on how to integrate Data Domain into the Pivotal BOSH Backup and Restore process. This solution brings the benefit of data reduction, increased performance, greater security, and efficiency to your PCF backups without changing the basic workflows of BBR. Further improvement can be made in the future through collaboration with both the Pivotal and Dell EMC Data Protection teams.

References

- BOSH Backup and Restore: [https://docs.pivotal.io/bbr/](https://docs.pivotal.io/bbr/)
- DD Boost Everywhere – File System Plug-in Integration Guide.pdf: [https://community.emc.com/docs/DOC-55465](https://community.emc.com/docs/DOC-55465)

This whitepaper is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind,