Data Discovery Overview

Data Discovery is a reporting and analytics tool overlaying the fibre channel switch, storage, and VMware environments. Discovery deploys into the IT environment as stand-alone collectors that run from the storage management servers that already reside in the Environment. Data is collected for the Discovery without agents (described in "Data Collection" section). Discovery can support a number of storage and switch technologies (described in "Product Support" section).

Collector Description

Discovery utilizes read only access to storage, switch, and VMware components in the environment. Discovery does **not** use agents to collect the data. It leverages read only access to run "show" commands for configuration data and "read" commands for performance data.

Discovery is broken out into 5 types of Collectors (Storage and switch Configuration, Fibre Switch Performance, Array Performance, VMware Collection and File Discovery).

Discovery will need to utilize a Windows 2008, 2012 or 2016, 2019 server for the duration of the collection. The server can be physical or virtual and be either new or existing (management server). The collection server(s) will need to have network access to the servers and components it is to collect.

Minimum requirements of the server hosting the collector:

CPU: **2 Cores**
RAM: **8GB**
Network: **1Gb**
Storage: **10GB (depends on collection sample, see below for sizing)**

Configuration Data Collections

This Collector will leverage native collection CLI's, and APIs based on the type of Storage Array or switch. It will require a user that has **read-only access** or greater to the Storage Array or Switch that Discovery are going to access. Details on supported array and switch types as well as the CLI used to collect is listed below in "Product Support". This collection will be a one-time collection. This is typically run on the native CLI management server, or an existing windows server in the environment.

Storage and Fibre Switch **Requirements:**

1. **Hitachi Block:** Discovery will leverage Hi-Command CLI or Rest API via an automated script from the command suite server srvstorage. Username and Password will need to be changed before running the script.
2. **Hitachi HNAS:** Discovery will leverage an auto-generated script that will SSH into the admin EVS. If the admin EVS is not available, Discovery will leverage the commands in the command spreadsheet and run them using putty to capture the logs. Username and Password will need to be changed before running the script.
3. **Hitachi Content Platform:** Discovery will leverage a Rest API collector for HCP. The IP of the HCP management Interface, username and password have been added to a file called arraylist.txt in the HCP_Collect directory based on the survey sheet.
4. **NetApp:** Discovery will leverage ssh via an automated script to collect data from each NetApp array. IP addresses to each node of 7- mode NetApp and IP address of the cluster management interface for cluster mode Netapp arrays will need to be provided. Username and Password will need to be changed before running the script at time of collection.
5. **NetApp E-series:** Discovery will leverage Storage Manager cli for data collection. We will pull a profile from the array using smcli, or from exporting a full profile from the Storage Manager.
6. **EMC Clariion/VNX:** Discovery will leverage naviseccli via an automated script to collect data from each clarion/VNX array. Username and Password will need to be changed before running the script.
7. **EMC VMAX/Powermax:** Discovery will leverage solutions enabler via an automated script to collect data from each VMAX arrays. The serial numbers of the arrays will be in the script.
8. **EMC Data Domain:** Discovery will leverage the commands in the command spreadsheet and run them using putty to capture the logs.
9. **EMC XtremIO:** Discovery will leverage a Rest API collector. IP of the XMS, username and password have been added to a file called xmslist.txt in the Collector directory based on the survey sheet.
10. **EMC Isilon:** Discovery will leverage ssh via an automated script to collect data from each Isilon array. Username and Password will need to be changed before running the script.
11. **EMC Unity:** Discovery will leverage a Rest API collector. Read-Only user account and password will be needed to access the Unity Array.
12. **EMC Centera:** Discovery will leverage Centera Viewer CLI via Java to collect data from each Centera array. The Username, Password, and IP of the one of the Centera Management IP address will be needed.

13. **HP XP Block:** Discovery will leverage CommandView hdvmCLI via and automated script from the command view server srvstorage. Username and Password will need to be changed before running the script.
14. **HP 3par:** Discovery will leverage ssh via an automated script to collect data from each 3par array. Username and Password will need to be changed before running the script.
15. **HP EVA:** Discovery will leverage SSSU via an automated script to collect data from each EVA array. Username, Password, and EVA name will need to be changed before running the script.
16. **Pure:** Discovery will leverage ssh via an automated script to collect data from each Pure Array channel switch. Username and Password will need to be input before running the script.
17. **IBM SVC/V9000/V7000/V5000/V840/FS900/FS840:** For the SVC Discovery will need to collect the svc.config.backup.xml from the SVC GUI or Discovery will need to SCP this file off. This also applies to the Block side of the V7000, V9000 and V840
18. **IBM Flash Systems:** Discovery will need to collect the svc.config.backup.xml from the SVC GUI or Discovery will need to SCP this file off.
19. **IBM XIV/A9000 (XCLI):** Discovery will leverage xcli in conjunction with the command spreadsheet. Username and password can be changed in the command spreadsheet, which can then be cut and pasted to the command line. The files will need to be sent back to so the host commands can be generated. This is a 2-step process to ensure all hosts are collected.
20. **IBM DS6000/8000:** Discovery will leverage dscli in conjunction with the command spreadsheet. Username and password will need to be changed in the spreadsheet, which can then be cut and pasted on the command line.
21. **Sun 5000 Series:** Discovery will leverage the Command Array Manager (CAM) cli, which is SSCS. We will need the array name list from CAM to be used with the sscs cli.
22. **Xiotech:** Discovery will leverage ssh via an automated script to collect data from each array. Username and Password will need to be changed before running the script.
23. **Brocade Fiber Channel Switches:** Discovery will leverage ssh via an automated script to collect data from each brocade switch. Username and Password will need to be changed before running the script.
24. **Cisco Fiber Channel Switches:** Discovery will leverage ssh via an automated script to collect data from each brocade switch. Username and Password will need to be changed before running the script.
25. **VMware Virtual Centers:** Discovery will leverage the perlAPI via an automated script to collect data from each Virtual Center. Username and Password will need to be input before running the script.

Fibre Channel Performance Module

This Collector will leverage SNMP v2c or SNMP v3 and communicate directly to the fibre channel switch. Running version 2c, Discovery will need the community string for v2c. Using v3, Discovery will need the username and password used to authenticate the v3 switches. Discovery will run the collection from a Windows server as a Windows scheduled task. It is common to use the Windows Management server for the collection of the switches. Discovery pull the switches on a 5-minute interval.

Using Brocade Virtual Fabrics will require a user with access to all virtual fabrics to be collected.

Requires about 25MB a day per Switch from a storage space perspective.

Hitachi Vantara Array Performance Module

Performance Monitor: Discovery Collector will run the export tool from a windows server that has network connectivity to the local Hitachi Service Processor. It will run as windows scheduled task and pull Performance Monitor Data every 24 hours.

Requires about 150MB a day per Array from a storage space perspective.

EMC Array Performance Module

EMC XtremeIO: This collection will leverage the array RestAPI utilizing a read only user account. The collector will be set up as a scheduled task, which will run every hour to collect 5 min intervals from the last hour. The performance files will be stored locally until upload of data to the Discovery portal.

VMware Discovery

Discovery will need a read-only account(s) that has access to all Virtual Center Servers. The performance collection will happen once a day and collect the last 24 hours' worth of performance data. This will run as a scheduled task usually from a management server where Discovery collect array and switch performance. The performance files will be zipped up and stored locally.

Requires about 50 MB a day per Virtual Center from a storage space perspective.

File Scan and Compliance Discovery

File Discovery requires the following Infrastructure to be deployed at the customer location. These scans can be broken into a meta data file scan which is for risk, last access information, and a TCO based on use. PII compliance scans requires additional server infrastructure for scanning of files for sensitive information.

Use Case	Hardware Minimums	Quantity	OS & Base Software	Comments
File Scan	RAM: 32GB Cores: 16 Disk: 300GB	1	Windows Server (2012 R2-2016, 2019) ESS/Stand/DC Edition .NET framework 4.6.2 or above JAVA 8.2 and above	Physical or VM, Install should be on local NTFS disks. Only 1 VM is needed. .NET is required for the Console and StorageX server. *Please have these installed first.* ***Exact Disk size depends on POC size in TB.***
Data Discovery and Compliance	RAM: 32GB Cores: 8 Disk 500GB	4	CentOS* *(vrs. 8.2)* RHEL *(vrs. 7.9, 8.1, 8.2, 8.4)* JAVA 8.2 and above	1 Core Server, 1 Elasticsearch Node, 1 ControlX Server, 1 DSS Server

Product Support

Discovery supports a number of storage and switch technologies as well as VMWare.
Below is a listing of the current **Support Matrix v3.1**

Storage Technologies

Storage Type	CLI Management Tool	Ports	Version
Hitachi VSP, HUS, HUSVM	Windows Management (HCS) Hicommand CLI	2001	5.x or greater/7.x VSP HCS/8.x G1K
Hitachi Block via REST API	REST API	80, 443	Panama2 release or later.
Hitachi AMS	Windows Management/hicommand CLI	2000, 28355	5.x or greater
Hitachi HNAS	SSC.exe/SSH	22	All
Hitachi HCP	RESTAPI/SNMP	UDP 161, 80, 443	All
EMC VMAX, DMX and PowerMax series	Solutions Enabler	Gatekeeper access	5.x or greater
EMC Clariion, CX, VNX series	Navisphere CLI	443, 2163	6.x or greater
EMC Centera	Centera Viewer CLI	22, 7069, 3218, 3682	4.X or greater
EMC Isilon	Direct Connect (SSH)	22	Onefs 7.x or greater
EMC Extreme IO	Rest API	80, 443	4.1x or greater
EMC Unity	Rest API	80, 443	All
HP EVA	SSSU		4.x or greater
HP XP12000, 24000, P9500	HP StorageWorks XP Command View AE	2001	5.x or greater
HP 3Par	HP InformOS (SSH)	22	3.x or greater
IBM SVC - V7000, V9000, V840 “Block" Flash System 900 and 840	Configuration Backup	22	All
IBM XIV / A9000	CLI Management Tool	7778	3.1.x or greater
IBM DS8000, DS6000 Series	DSCLI	1750, 1751	5.x or greater
NetApp E-series / IBM DS 3/4/5000 Series	SMCLI	2463	9.x or greater
NetApp	Direct Connect (SSH)	22	Ontap 6.x or greater
Pure	Purity API (SSH)	22	3.x or greater
Xiotech	SSH	22	All

Switch Technologies

Fibre Switch Type	Management OS	Ports	Version
Cisco	SAN-OS	UDP 161, 22	3.x and greater
Cisco	NX-OS	UDP 161, 22	4.x or greater
Brocade	FabricOS	UDP 161, 22	5.x or greater

VMware

Hypervisor	Management	Ports	Version
VMWare	v-Center	443	4.x or greater