Hewlett Packard
Enterprise

HPE Security ArcSight Investigate

Software Version: 2.01

ArcSight Investigate Backup and Restore Tech Note

February 1, 2018

Legal Notices

Warranty

The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

The network information used in the examples in this document (including IP addresses and hostnames) is for illustration purposes only.

HPE Security ArcSight products are highly flexible and function as you configure them. The accessibility, integrity, and confidentiality of your data is your responsibility. Implement a comprehensive security strategy and follow good security practices.

This document is confidential.

Restricted Rights Legend

Confidential computer software. Valid license from Hewlett Packard Enterprise required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Copyright Notice

Follow this link to see a complete statement of copyrights and acknowledgements:
https://community.softwaregrp.com/t5/Discussions/Third-Party-Copyright-Notices-and-License-Terms/td-p/1589228

Support

Contact Information

Phone	A list of phone numbers is available on the Technical Support Page: https://softwaresupport.softwaregrp.com/support-contact-information
Support Web Site	https://softwaresupport.softwaregrp.com/
ArcSight Product Documentation	https://community.softwaregrp.com/t5/ArcSight-Product-Documentation/ct-p/productdocs

Contents

Introduction

This tech note describes the following:

- Backup and restore of the Vertica database
- Backup and restore of Investigate user management and search datastores.

You can restore any of this data as necessary.

Backup the Vertica database

When to perform a backup

- Prior to a Vertica upgrade
- Prior to adding or removing a Vertica node
- After adding or removing a Vertica node
- After recovering from a crash
- Routinely

Vertica backup requirements

General requirements

- Extra storage

  While a backup is taking place, the backup process can consume additional storage. The amount of space consumed depends on the size of your catalog and any objects that you drop during the backup. The backup process releases this storage once the backup is complete.
- Vertica-to-Vertica version

  Backups can only be restored to the same version of Vertica from which the original backup came. For instance, you cannot backup Vertica 8.01 and restore it to a Vertica 8.10.
- Stop ingesting events

  Ingesting events into the database (through the Investigate Scheduler) during backup may result in the most recently ingested events not being backed up. In order to ensure that all events are backed up, you should stop ingestion prior to starting the backup.
- Backup host

  For best network performance, each Vertica node should have its own backup host.
- Directory usage

  Use one directory on each Vertica node to store successive backups.

Backup locations

Vertica supports the following locations where backups can be saved:

- Local folder on the Vertica node
- Remote server

Backup host file system

Backups can only be performed on the following file system types:

- ext3
- ext4
- NFS

Required storage

It is recommended that each backup host have space for at least twice the database node footprint size. Consider your long-term backup storage needs.

If you are using a single backup location, the estimated storage space needed for the Vertica cluster can be calculated through the following Vertica operation:

```
dbadmin=> select sum(used_bytes) as total_used_bytes from v_monitor.storage_containers;
 total_used_bytes
------------------
       5717700329
(1 row)
```

If you are using multiple backup locations, one per node, then the storage needed by each node can be calculated through the following Vertica operation:

```
dbadmin=> select node_name, sum(used_bytes) as total_used_bytes from v_
monitor.storage_containers group by node_name;
        node_name       |   total_used_bytes
------------------------+--------------------
 v_investigate_node0002 | 1906279083
 v_investigate_node0003 | 1905384292
 v_investigate_node0001 | 1906036954
(3 rows)
```

Backup host preparation

- Multiple hosts can be used to backup a Vertica cluster. Each host must be prepared prior to backup.
- Remote backup hosts must have SSH access and password-less SSH setup from Vertica node 1 in order for the database administrator to access the hosts (see "Setting up password-less SSH" below).
- If one host is the backup destination for multiple Vertica nodes, then increase the maximum SSH connections on the backup host by increasing the `MaxStartups` parameter in `/etc/ssh/sshd_config`. The `MaxStartups` number should be greater than the number of nodes in the Vertica cluster.

Setting up password-less SSH

1. Login to the backup server
2. Create user `$db_admin`

   `$db_admin` is the administrator for the Vertica cluster.
3. Ensure `$db_admin` has write permission to the dedicated directory where the backup will be stored.
4. Login to Vertica node 1 as root.
5. Become the Vertica database admin.

   ```
   # su -l $db_admin
   ```
6. Setup password-less SSH for all backup servers.

   ```
   # ssh-keygen -t rsa
   ```

   ```
   # ssh-copy-id -i ~/.ssh/id_rsa.pub $db_admin@$back_up_server_ip
   ```

Backing up Vertica

About

Backup is performed by the database admin (`$db_admin`).

Procedure

1. Login to Vertica cluster node 1 as root.
2. Generate a backup configuration file.

   ```
   # su -l $db_admin
   ```

   ```
   # /opt/vertica/bin/vbr --setupconfig
   ```

General configuration options:

- Restore Points — Default is 52, assuming a weekly backup for one year.

  Multiple restore points gives you the option to recover from one of several backups. For example, if you specify 3, you have 1 current backup, and 3 backup archives. Vertica stores the value you enter as the `restorePointLimit` parameter in the `vbr` configuration file.

- Password Save — The backup configuration can optionally save the database admin password to avoid prompting in the future.

- Advanced Options — These options allow additional security measures. It is recommended that the default options are used.

Per node configuration options:

- Backup Host Name — For each Vertica node, you will be prompted to specify the backup host. This host can be either the local machine or a remote host.

- Backup Directory — For each backup host, the directory where the backup will be stored must be specified

Example:

```
Warning: This setup tool is deprecated, and will be removed in a future
version. Please use config file samples we provide in
/opt/vertica/share/vbr/example_configs instead of using this tool.

Snapshot name (backup_snapshot): vertica_backup

Number of restore points (1): 52

Specify objects (no default):

Object restore mode (coexist, createOrReplace or create)
(createOrReplace): createOrReplace

Vertica user name (dbadmin): $db_admin

Save password to avoid runtime prompt? (n) [y/n]: n

Node v_investigate_node0001

Backup host name (no default): [BACKUP HOST 1 IP]

Backup directory (no default): /opt/vertica/backup1

Node v_investigate_node0002

Backup host name (no default): [BACKUP HOST 2 IP]

Backup directory (no default): /opt/vertica/backup2

Node v_investigate_node0003
```

```
Backup host name (no default): [BACKUP HOST 3 IP]

Backup directory (no default): /opt/vertica/backup3

Change advanced settings? (n) [y/n]: n

Config file name (vertica_backup.ini):

Saved vbr config to vertica_backup.ini.
```

The `vertica_backup.ini` file is created in `/home/$db_admin`.

Note: The config file is needed for all future backup and restore operations. Save it in a safe place.

```
# cat ./vertica_backup.ini

[Misc]

snapshotName = vertica_backup

restorePointLimit = 52

objectRestoreMode = createOrReplace

[Database]

dbName = investigate

dbUser = analyst

dbPromptForPassword = True

[Transmission]

[Mapping]

v_investigate_node0001 = [BACKUP HOST 1 IP]:/opt/vertica/backup1

v_investigate_node0002 = [BACKUP HOST 2 IP]:/opt/vertica/backup2

v_investigate_node0003 = [BACKUP HOST 3 IP]:/opt/vertica/backup3
```

3. Initialize the backup locations.

```
# /opt/vertica/bin/vbr --task init --config-file vertica_backup.ini

Initializing backup locations.

Backup locations initialized.
```

4. Stop the Investigate scheduler.

```
# exit
```

```
# cd /root/install-vertica

./kafka_scheduler stop
```

Stopping the Investigate Scheduler from writing data to the Vertica database ensures that you do not lose events during backup.

5. Backup Vertica data.

```
# su -l $db_admin

# /opt/vertica/bin/vbr --task backup --config-file vertica_backup.ini

Starting backup of database investigate.

Participating nodes: v_investigate_node0001.

Enter vertica password:

Snapshotting database.

Snapshot complete.

Approximate bytes to copy: 270383427 of 270383427 total.

[==================================================] 100%

Copying backup metadata.

Finalizing backup.

Backup complete!
```

6. Verify that the backup files were written to the backup locations.

```
# ssh [BACKUP HOST 1 IP] ls /opt/vertica/backup1

backup_manifest

Objects

Snapshots

# ssh [BACKUP HOST 2 IP] ls /opt/vertica/backup2

backup_manifest

Objects

Snapshots

# ssh [BACKUP HOST 3 IP] ls /opt/vertica/backup3

backup_manifest
```

```
Objects

Snapshots
```

Incremental Vertica database backups

About

Incremental backups use the same setup as a full backup and just back up what changed from the previous full backup. When a full backup is executed using the same configuration file, subsequent backups are incremental. When you initiate an incremental backup, the `vbr` tool displays a backup size that is a portion of the total backup size. This portion represents the delta changes that will be backed up by the incremental backup.

Procedure

Run the following commands:

```
# /opt/vertica/bin/vbr --task backup --config-file vertica_backup.ini

Starting backup of database investigate.

Participating nodes: v_investigate_node0001.

Snapshotting database.

Snapshot complete.

Approximate bytes to copy: 2680 of 270455624 total.

[==================================================] 100%

Copying backup metadata.

Finalizing backup.

Backup complete!
```

Verifying the integrity of the Vertica database backup

About

The `full-check` option is used to verify the integrity of the Vertica database backup snapshots, which reports the following:

- Incomplete restore points
- Damaged restore points
- Missing backup files
- Unreferenced files

Note: Backup files generated by the interrupted process will remain in the backup location and subsequent backups will resume where the interrupted backup left off.

Backup operations are atomic, so interrupted backup operations will not affect previous backup files.

Procedure

- Run the following command:

```
# /opt/vertica/bin/vbr --task full-check --config-file vertica_backup.ini

Checking backup consistency.

List all snapshots in backup location:

Snapshot name and restore point: backup_snapshot_20180116_172347, nodes:
['v_investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172253, nodes:
['v_investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172236, nodes:
['v_investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172310, nodes:
['v_investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172158, nodes:
['v_investigate_node0001'].

Snapshots that have missing objects(hint: use 'vbr --task remove' to delete
these snapshots):

Backup locations have 0 unreferenced objects

Backup locations have 0 missing objects

Backup consistency check complete.
```

Manage existing backups

Viewing available backups

- Run the following command:

```
# /opt/vertica/bin/vbr --task listbackup --config-file vertica_backup.ini
```

```
backup                          backup_type  epoch   objects   nodes
(hosts) file_system_type
```

```
vertica_backup_20180104_142326   full         29               v_
investigate_node0001(10.12.57.27)   [Linux]
```

The backup name is comprised of the snapshot name and backup timestamp.

Example:

snapshot name: `vertica_backup`

timestamp: `20180104_142326`

Deleting a backup

About

Use only the `vbr` tool to delete a backup.

Procedure

Run the following commands:

```
# /opt/vertica/bin/vbr --task remove --config-file /backup/vertica_backup.ini
--archive 20180104_142326
```

```
# 20180104_142326 is the backup timestamp
```

```
Removing restore points: 20180104_142326
```

```
Remove complete!
```

```
# /opt/vertica/bin/vbr --task listbackup --config-file vertica_backup.ini
```

```
backup                backup_type   epoch   objects   nodes(hosts)
              file_system_type
```

Restore Vertica data

Vertica restoration requirements

- Backups can only be restored to the same version of Vertica that the original backup came from. For instance, you can not backup Vertica 8.01 and restore it to a Vertica 8.10
- Restore to a cluster that is identical to the cluster from which the backup originated.

Prior to restoring a Vertica cluster, the cluster must meet the following conditions:

- Target database is already created and can be empty
- Target database name matches the backup database name
- Target database is stopped
- All Vertica nodes in the target cluster are up
- All Vertica nodes in the target cluster have identical names to the original backup

Restoring the Vertica database

About

Data restorations is performed by the database admin.

Procedure

1. Rebuild a Vertica cluster identical to the original cluster.
2. Login to Vertica node 1.

   ```
   # su -l $db_admin
   ```

   `$db_admin` has password-less SSH to the `$db_admin` of backup host (see "Setting up password-less SSH" on page 7).
3. Copy `vertica_backup.ini` to `/home/$db_admin`.
4. Stop the database.

   ```
   # /opt/vertica/bin/admintools -t stop_db -d investigate -p *****

   Connecting to database

   Issuing shutdown command to database

   Database investigate stopped successfully
   ```

5. Restore the backup data.

```
# /opt/vertica/bin/vbr --task restore --config-file vertica_backup.ini

Starting full restore of database investigate.

Participating nodes: v_investigate_node0001, v_investigate_node0002, v_
investigate_node0003.

Restoring from restore point: investigate_backup_20180110_010826

Determining what data to restore from backup.

[==================================================] 100%

Approximate bytes to copy: 2246248425 of 2246250258 total.

Syncing data from backup to cluster nodes.

[==================================================] 100%

Restoring catalog.

Restore complete!
```

6. Start the database.

```
# /opt/vertica/bin/admintools --task start_db -d investigate -p *****

Starting nodes:

v_investigate_node0001 (127.0.0.1)

Starting Vertica on all nodes. Please wait, databases with a large catalog
may take a while to initialize.

Node Status: v_investigate_node0001: (DOWN)

Node Status: v_investigate_node0001: (DOWN)

Node Status: v_investigate_node0001: (DOWN)

Node Status: v_investigate_node0001: (DOWN)

Node Status: v_investigate_node0001: (UP)

Database investigate started successfully
```

7. Start the Investigate Scheduler.

```
# exit

# cd /root/install-vertica

# ./kafka_scheduler start
```

Backing up Investigate management and search datastores

About

We recommend that you identify a back up location that is not under the `/opt/arcsight` directory. Identify a local folder on the system or a remote location.

This procedure uses the `/opt/investigate/backup` directory as an example.

Procedure

1. Logout of any open Investigate sessions.
2. SSH to the Kubernetes cluster master node 1.
3. Run the following commands:

   ```
   # cd /opt/arcsight/volumes/investigate/
   ```

   ```
   # mkdir /opt/investigate/backup
   ```

   ```
   # cp –R * /opt/investigate/backup
   ```

   ```
   # diff -r -s /opt/investigate/backup/mgmt
   /opt/arcsight/volumes/investigate/mgmt
   ```

   ```
   # diff -r -s /opt/investigate/backup/search
   /opt/arcsight/volumes/investigate/search
   ```

   A message should state that all files are identical. If they are not identical, repeat the procedure.

Restore management and search data

Restoring Investigate management and search datastores

About

When restoring the Investigate management and search datastores, retain the original directory structure under `/opt/arcsight/volumes/investigate/`.

The management datastore will be restored to the `/opt/arcsight/volumes/investigate/mgmt/db/` directory. The search datastore will be restored to the `/opt/arcsight/volumes/investigate/search` directory.

Prerequisite

Back up the Investigate management and search datastores to the `/opt/investigate/backup` directory on the Kubernetes master node (see "Backing up Investigate management and search datastores" on the previous page).

Procedure

1. Logout of any open Investigate sessions.
2. SSH to the Kubernetes master node and then run the following commands:

   ```
   # cd /opt/investigate/backup
   ```

   ```
   # cp –R search/* /opt/arcsight/volumes/investigate/search
   ```

   Reply `yes` to overwrite files and folders.

   ```
   # cd /opt/arcsight/volumes/investigate/mgmt/db/
   ```

   ```
   # rm - rf h2.lock.db
   ```

   ```
   # cp /opt/investigate/backup/mgmt/db/h2.mv.db   .
   ```

   Reply `yes` to overwrite files and folders.

   ```
   # diff -r -s /opt/arcsight/volumes/investigate/mgmt/db/h2.mv.db
   /opt/investigate/backup/mgmt/db/h2.mv.db
   ```

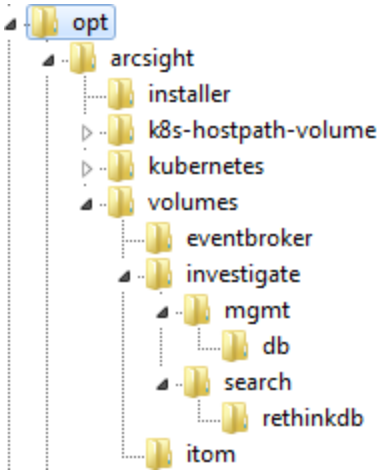
   ```
   # diff -r -s /opt/investigate/backup/search
   /opt/arcsight/volumes/investigate/search
   ```

A message should state that all files are identical. If they are not identical, repeat the procedure.

3. Change the permission of the Investigate directory.

```
# chown 1999:1999 -R /opt/arcsight/volumes/investigate/
```

The directory structures should look similar to the following:



Refreshing the datastore configuration

1. SSH to the Kubernetes master node if not already there.
2. Find the name of the management pod.

```
# kubectl get pods -n arcsightinvestigate1 | grep management
hercules-management-*   2/2       Running   0          1h
```

3. Delete the user management pod.

```
# kubectl delete pod <hercules-management-*> -n arcsightinvestigate1
```

4. Ensure a new management pod is created and wait for the pod in "Running" state.

```
# kubectl get pods -n arcsightinvestigate1 | grep management
hercules-management-*   2/2       Running   0          1h
```

5. Delete the rethinkdb pod.

```
# kubectl delete pod hercules-rethinkdb-0 -n arcsightinvestigate1
```

6. Ensure a new rethinkdb is created and wait for the pod in "Running" state.

```
# kubectl get pods -n arcsightinvestigate1 | grep rethinkdb
hercules-rethinkdb-0                  1/1       Running   0          1h
```

7. Open Investigate (`http://<master_ip>`).
8. At the login prompt, log in as an existing user.

   Note: A "502 bad gateway error" may appear when you launch Investigate. Refresh the page after about 30 seconds. As long as the pod is running, the error should go away and Investigate will load normally.

   After logging in, if you see your previously created searches and users, no further steps are necessary.

Potential issues during backup and restore

Vertica downtime exceeds the retention time for the Kafka cluster

If the Vertica cluster downtime exceeds the retention time for the Kafka cluster, there is a chance that the Vertica-stored Kafka offset is no longer present in the Event Broker cluster. In this case, the scheduler will not be able to consume new data.

Basic steps:

1. Determine the last Kafka offset read by the Scheduler.
2. Confirm the Scheduler is no longer copying data.
3. Reset the Scheduler.

Step 1: Determine the last Kafka offset read by the Scheduler

Each Vertica node in the cluster will copy data from one or more Kafka partitions. In order to see the source Kafka topic, as well as the last offset copied from each partition, run the Vertica SQL commands below. This query also identifies the number of partitions. Use that value as part of the limit clause in the second query below.

```
dbadmin=> select * from investigation_scheduler.stream_sources;
id  |      source      | cluster | partitions | enabled
----+------------------+---------+------------+---------
1   | eb-internal-avro |       1 |          1 | t
(1 row)
```

```
dbadmin => select frame_start, DISTINCT source_partition, end_offset from
investigation_scheduler.stream_microbatch_history order by frame_start desc
limit 1;

frame_start              | source_partition | end_offset

-------------------------+------------------+------------

2018-01-16 18:19:06.381  |                0 |          12
```

If the end_offset + 1 no longer exists in the Kafka cluster for the topic's source_partition, then it is likely that the scheduler will no longer be able to read from Kafka until it has been given a valid offset.

Step 2: Confirm the Scheduler is no longer copying data

About

You can confirm whether the scheduler is copying data or not by checking the status and examining the last copied offset in the microbatch status. If the offset number is not increasing, then the scheduler can no longer find the valid offset and needs to be reset.

Procedure

- Check the Scheduler offsets.

```
# ./kafka_scheduler status

...

'investigation_scheduler' scheduler last 10 microbatch status:

       frame_start       |   source_name    | start_offset | end_offset |
end_reason   | partition_bytes | partition_messages

-------------------------+------------------+----+----+---------------+-----
------------+--------------------

2018-01-17 09:29:01.604 | eb-internal-avro |  9 |  9 | END_OF_STREAM |
         0 |                 0

2018-01-17 09:28:51.595 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |               0 |                 0

2018-01-17 09:28:41.586 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |               0 |                 0

2018-01-17 09:28:31.577 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |               0 |                 0
```

```
2018-01-17 09:28:05.824 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

2018-01-17 09:27:55.524 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

2018-01-17 09:27:45.515 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

2018-01-17 09:27:35.507 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

2018-01-16 18:19:06.381 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

2018-01-16 18:18:56.374 | eb-internal-avro |           9 |          9 |
END_OF_STREAM |              0 |                 0

(10 rows)
```

Step 3: Resetting the Scheduler

About

The Event Broker cluster may continue to receive data. If the Vertica cluster down time exceeds the Event Broker data retention time, there is a chance that the Vertica offset will no longer be valid. The Scheduler must be reset.

To reset the Investigate Scheduler, it needs to be deleted and then recreated.

Procedure

- Execute the following commands:

```
# ./kafka_scheduler delete

Are you sure that you want to DELETE scheduler metadata (y/n)?y

Terminating all running scheduler processes for schema: [investigation_
scheduler]

scheduler instance(s) deleted for 192.214.138.94

bash: /root/install-vertica/kafka_scheduler.log: No such file or directory

scheduler instance(s) deleted for 192.214.138.95

bash: /root/install-vertica/kafka_scheduler.log: No such file or directory

scheduler instance(s) deleted for 192.214.138.96

db cleanup: delete scheduler metadata
```

```
# ./kafka_scheduler create
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

create scheduler under: investigation_scheduler

scheduler: create target topic

scheduler: create cluster for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler: create source topic for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler: create microbatch for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler instance(s) added for 192.214.138.94

scheduler instance(s) added for 192.214.138.95

scheduler instance(s) added for 192.214.138.96
```

Send Documentation Feedback

If you have comments about this document, you can contact the documentation team by email. If an email client is configured on this system, click the link above and an email window opens with the following information in the subject line:

Feedback on ArcSight Investigate Backup and Restore Tech Note (Investigate 2.01)

Just add your feedback to the email and click send.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to arc-doc@hpe.com.

We appreciate your feedback!