The growing number of relational databases on the cloud accentuates the need for data protection and auditing. InfoSphere® Guardium® offers real-time database security and monitoring, fine-grain database auditing, automated compliance reporting, data-level access control, database vulnerability management, and auto-discovery of sensitive data in the cloud. With the Amazon Relational Database Service (RDS), you can create and use your own database instances in the cloud and build your own applications around them. This two-part series explores how to use InfoSphere Guardium to protect database information in the cloud. Part 1 describes how to use InfoSphere Guardium’s discovery and vulnerability assessment with Amazon RDS instances. This tutorial covers how InfoSphere Guardium uses Amazon S3 for backup and restore.

View more content in this series

Overview

InfoSphere Guardium provides robust solution for ensuring the privacy and integrity of trusted information in a data center (SAP, PeopleSoft, Cognos, Siebel, etc.). InfoSphere Guardium solutions for data security and privacy are designed to support this holistic approach, helping an organization protect itself against a complex threat landscape while remaining focused on the business goals.

InfoSphere Guardium extends its support to cloud databases, offering control and protection of sensitive information subject to global laws and regulations in the era of big data. In addition, because some organizations are moving their storage to the cloud as well, InfoSphere Guardium
has been enhanced to include this option for storage of its archive and backup data and configurations. Specifically, in InfoSphere Guardium 9.1, Amazon Simple Storage System (S3) was added as an option for archive and backup. Other storage protocol options include SCP, FTP, Tivoli® Storage Manager (TSM), and EMC Centera. For information about configuring these options, see the Archive, Purge, and Restore topic in the InfoSphere Guardium Information Center (see Resources).

This tutorial describes how Amazon S3 can be used for backup, archive, and restore.

**Understanding InfoSphere Guardium archive, backup, and restore**

This section briefly describes how InfoSphere Guardium archive, backup and restore work and when each of these processes is required. The Resources include links to an IBM Redbooks® and to the Information Center.

Use these options to offload data from the InfoSphere Guardium appliance — including audit data, reports and results, definitions, and configuration files — to support both requirements for forensic data analysis and reporting, and to support disaster recovery scenarios.

We will be referencing options that are available from the InfoSphere Guardium GUI or by running commands from the CLI. Most of the options are available from Data management tab of the administration console in an InfoSphere Guardium collector.

**Figure 1. InfoSphere Guardium Administration Console**

![Image of InfoSphere Guardium Administration Console]

**Archive and backup**

There are two archive operations in InfoSphere Guardium:

- Data archive backs up the audit data that has been captured by InfoSphere Guardium appliance for a given period of time.
- Results archive backs up audit tasks results, including reports, assessment tests, entity audit trail, privacy sets and classification processes as well as the view and sign-off trails and the accommodated comments from workflow processes.
The archive function creates signed, encrypted files that cannot be tampered with. Do not change the names of the generated archive files. The restore operation depends on the file names created during the archiving process.

System backups are used to backup and store all the necessary data and configuration values to restore a server in case of hardware corruption. All configuration information and data is written to a single encrypted file and sent to the specified destination, using the transfer method that the InfoSphere Guardium administrator configured for backups on the appliance (such as FTP or Amazon S3).

**Restore**

Restore operations retrieve the content of data archive, result archive, and system backup files and restore them on the system with minimal effort, at any point in time.

Data and result restore operations can be run from the GUI. Data restore is available from an InfoSphere Guardium collector GUI, as shown in Figure 1. Note that result restore is only available from the Data Management tab of an appliance configured as an aggregator, which is InfoSphere Guardium appliance that collects and merges information from multiple InfoSphere Guardium collectors to facilitate an enterprise view of database usage.

System restore is not available from the InfoSphere Guardium GUI. System restore is used to recover from failure and can be done by using a command-line interface (CLI) command or by using diagnostic tools included with the product. The CLI can be used for all different restore operations.

Archive, backup, and restore are also supported on a Central Manager. A Central Manager is an InfoSphere Guardium appliance assigned to monitor and control an entire deployment, including all collectors and aggregators.

**InfoSphere Guardium Catalog**

InfoSphere Guardium uses catalogs to track where archive and backup files are sent. The catalog enables data and result archive and system backup files to be retrieved and restored on different systems. A separate catalog is maintained on each appliance, and a new record is added to the catalog whenever the appliance archives or backs up information. These catalog entries can be transferred to different appliance in several ways:

- Using Catalog Export and Catalog Import.
- Using Data Restore, since each data restore operation contains the daily archive data including the catalog of that day.
- Using aggregation, which is the process of exporting data on a daily basis from the collector appliances to the aggregator. Aggregation merges the catalogs of the collectors it is monitoring.

When catalog entries are imported from another system, those entries point to files that have been encrypted by the source system. Before restoring or importing any such file, the system shared secret of the system that encrypted the file must be available on the importing system. See
Resources for the Information Center topic that describes the aggregator CLI commands used to back up and restore shared secrets key files from one appliance to another.

**Overview of Amazon S3**

Amazon S3 is a simple storage service that offers software developers a highly scalable, reliable, and low-latency data storage infrastructure at very low costs and eliminates the burden of provisioning and managing hardware.

Using S3 adds another layer of security to the InfoSphere Guardium data, as only users who have Amazon cloud access privilege can work with stored data.

Before you can configure InfoSphere Guardium to back up to the Amazon S3 cloud, you must obtain an Access Key ID and a Secret Access key.

**Figure 2. Amazon S3 architecture**

Amazon S3 allows writing, reading, and deleting objects containing from 1 byte to 5TB of data. Amazon S3 uses containers called **buckets** to store objects. Objects can be retrieved using a unique developer-assigned key. Each object is stored in a bucket and retrieved via a unique developer-assigned key. The number of objects that can be stored is unlimited and can be copied and restored on different buckets on separate geographic area called a **region**, where Amazon services are diversified. It is also possible to change the location of a bucket and transfer it to a different region.

To access Amazon S3, a user must have an **access key ID**, which identifies the user as the party responsible for service requests. The access key ID must be included in each request. This key is not confidential and does not need to be encrypted.

A **secret access key** associated with the access key ID is also required and calculates a digital signature included in the request. Only the registered Amazon Web Services (AWS) user...
should have the secret access key. This key is required for InfoSphere Guardium to access the AWS account. The access key ID and secret access key can be obtained from AWS Security Credentials page on the Amazon website.

Amazon key ID, the secret access key, and the bucket name are required to archive InfoSphere Guardium data and result and backup InfoSphere Guardium system data and system configuration on Amazon S3.

**Figure 3. InfoSphere Guardium architecture**

![InfoSphere Guardium architecture diagram]

**Configuring InfoSphere Guardium to use Amazon S3 storage**

The ability to archive and backup to Amazon S3 is enabled by default as you can see in Figure 4. You can disable and re-enable it using the following CLI commands:

```
store_storage-system amazon_s3 archive on|off
```
Errors related to not having right time zone

Not having system time set correctly causes errors described below. These errors can be viewed in the InfoSphere Guardium Aggregation/Archive log report:

- AmazonServiceException request rejected due to following reason. Error Message: Service unavailableHTTP Status Code: 503AWS Error Code: Service unavailableError Type: ServiceRequest ID: null
- HTTP Status Code: 403 Forbidden
- Error Code: RequestTimeToo-Skewed — Description: The difference between the request time and the server’s time is too large

**Note:** Amazon S3 requires time of all servers making requests to Amazon be within 15 minutes of an Amazon web server’s clock. Setting up all servers to sync their times with an NTP server will help avoid problems later.

To synchronize the time zone, date, and time of the InfoSphere Guardium appliance with a server network protocol, use the following CLI command:

```
store system ntp server
```

When prompted, the details of an accessible NTP server should be provided as shown in Figure 5.

The use of the NTP server should be enabled for the setting to be effective. Use the following CLI command to enable this feature:

```
store system ntp state on
```
Figure 5. CLI commands to set network time protocol

```bash
vx62.guard.swg.usma.ibm.com> store system ntp server
USAGE: store system ntp server
For each server enter either ip or hostname
   Enter up to 3 NTP servers to store:
Enter ntp server: ntp0.swg.usma.ibm.com
Enter ntp server: ntp1.swg.usma.ibm.com
Make sure to use "store system ntp state on" to turn ON the NTP service.
All inspection engines refreshed.
OK
vx62.guard.swg.usma.ibm.com> store system ntp state on
OK
vx62.guard.swg.usma.ibm.com>
```

If a Central Manager is used, the system time needs to be set to current time on central manager and all managed nodes.

This can easily be done by distributing the configuration from central manager to all managed nodes.

Figure 6. Distributing configuration from Central Manager

As shown in Figure 7, the configuration for data archive and result archive and system backup are selected and will be distributed by clicking the **Distribute** button.
Archiving data and results

To reiterate: Use archiving to offload InfoSphere Guardium data (audit records, definitions, metadata with export definitions, etc.) and results (audit tasks results as well as the view and sign-off trails and the comments from workflow processes) from the appliance. The data must be restorable in case forensic analysis is needed or for meeting auditing requirements.

Archiving data

Log on to the InfoSphere Guardium GUI using your access credentials and navigate to Administration Console > Data Management. On the Data Archive configuration page under the Data Management tab of the Administration console, make sure the Archive checkbox is marked, which will enable options for archiving data.

Select the radio button AMAZON S3 to specify Amazon archive specifics such as the bucket name.

Enter your Amazon bucket name, access key ID, secret access key, and a value for Archive data older than. Then click Apply to save the configuration. InfoSphere Guardium will check the bucket name and if it doesn't exist, it will create one automatically using the provided Amazon credential.
A message is displayed indicating the configuration has been saved successfully.

After entering and saving the S3 bucket name and Amazon credentials, the **Run Once Now** and **Modify Schedule** buttons are enabled, as shown in Figure 9. Let's look at each of these options.

**Run Once Now** starts the process instantly.
Figure 9. Running Data Archive

A message is displayed to let you know that the operation started successfully.

To see the backup files on Amazon S3 cloud storage, log in to AWS Management console with your Amazon account credentials and navigate to S3 Static Storage in the Cloud.
Figure 10. Amazon Web Services Console

A list of all existing bucket names is displayed. Click on the bucket name you used for archiving your data.

Figure 11. Amazon S3 Console

The archive files stored on the S3 bucket are shown below.
Figure 12. Archived data files in Amazon S3 bucket

The archive data file name format is as follows:

<day_number>-<hostname.domain>-w<run_date><audit_process_id>-d<data_date>.dbdump.enc

Example:

732423-g1.guardium.com-w20050425.040042-d2005-04-22.dbdump.enc

Note that the archive data filename form an aggregator has an extension of .tgz:

<day_number>-<hostname.domain>-w<run_date><audit_process_id>-d<data_date>.tgz

You can set a new schedule or change the existing schedule to archive just once or on a repeating schedule. The example below shows that we run our data archives every day at 5 a.m.

Figure 13. Schedule definition panel

The data archive process activities and status can be viewed in Aggregation/Archive log report on the InfoSphere Guardium Monitor tab.
**Archiving results**

Archiving results is the same process as archiving data. Make sure to enter a value greater than zero for the field *Archive Result Older than*.

InfoSphere Guardium system has to have result data such as audit tasks result mentioned in the InfoSphere Guardium backup and recovery section to be able to archive the result data.

**Figure 15. Amazon required values for using S3 for result archive**
After saving the configuration, you can run it immediately or schedule the archive, just as you did with the data.

When archiving results is completed the files can be viewed in Amazon bucket.

**Figure 16. Archived results files in Amazon S3 bucket**

![Amazon S3 bucket](image)

The archive results file name format is as follows:

<hostname.domain>-w<run_date><audit_process_id>-d<execution_date>.res.<Guardium_major_release_number>.tar.gz.enc

Example:
g1.guardium.com-w20050425.040042-d2005-04-22.res.90.tar.gz.enc

The result archive process activities can be viewed in the Aggregation/Archive log report on the InfoSphere Guardium Monitor tab.

**Figure 17. Result archive process**

![Aggregation/Archive log report](image)

**Restoring data and results**

Before a restore can be invoked for data and results, the inspection engine has to be stopped, which means that data activity cannot be captured during the restore process. Some organizations do restores on an empty appliance designed just for reviewing historical data, ensuring that ongoing monitoring operations are not affected.

Stopping inspection engine and restarting it can be done by using the following CLI commands:

```bash
stop inspection-core
start inspection-core
```
Figure 18. Inspection engine stop/start CLI commands

```
[root@vxe62 ~]# su - cli
Welcome cli - your last login was Sun Feb 9 12:05:35 2014
vxe62.guard.swg.usma.ibm.com> stop inspection-core
Stopping inspection core
Please do not forget to manually start the Inspection Core after maintenance is done.
ok
vxe62.guard.swg.usma.ibm.com> start inspection-core
Starting inspection core
Started.
ok
vxe62.guard.swg.usma.ibm.com>  
```

Restoring data

Data can also be restored to the same system it was archived from. This is useful in cases that the data of specific date was purged accidentally from the system. This can be done using Catalog Export and Catalog Import Features. These two features transfer catalog entries created during data archive to another system.

To create the data restore search criteria, navigate to the Data Restore option on the Data Management menu. Enter the desired date range and click Search. The system will look for all available archive data files on the selected storage system — in this case on Amazon S3 storage since that is what we used for archiving.

Figure 19. Data restore search

The files displayed below are the data files archived on Amazon S3. Select one or more items from the list and click Restore to start the restore process.
The progress of the restore process can be viewed and monitored by looking at Aggregation/Archive Log report on the InfoSphere Guardium Monitor tab.

Restoring results

Restoring results can only be done from the Investigation Center, an extension of aggregation servers, by an investigation user with the assigned role “inv.” A single InfoSphere Guardium aggregator appliance collects and merges information from multiple InfoSphere Guardium collector appliances. The aggregation process is accomplished by exporting data from collectors on a daily basis. Aggregator extracts each uploaded file and merges it into its internal repository.
As mentioned, the restoring results relies on information stored in the catalog. Therefore, before restoring results from a collector to an aggregator, you must first export the appropriate catalog entries from the collector and import them onto the aggregator.

Also as mentioned, catalog entries point to the files that have been encrypted by the system and before restoring or importing them from another system, the system-shared secret of the system that encrypted the file must be available on importing system. System-shared secret is used to encrypt files exported from the appliance by archive and export activities, and is used to establish secure communication between central managers and aggregators and their managed units.

Result restore is accessible from aggregator GUI. Log on to the GUI using user with investigator (inv) role privileges and navigate to Volume Management > Audit Result Restore, then click Audit Result Restore.

**Figure 22. Audit Result Restore**

In the subsequent search criteria, enter search criteria and click **Search** to get a list of existing backup files that match the criteria.
Figure 23. Results restore search

Click **Restore** to start the restore process for selected items from the search results.

Figure 24. Results restore search results

The restored result page will be updated by the outcome of restore process. It displays the details about the restored file.
**Figure 25. Restored results files**

The log report for the result restore can be viewed by navigating to **Volume Management > Restore Log**.

**Figure 26. Results restore process**

**System backup and restore**

To prepare for disaster recovery, a backup strategy must be in place. As stated in the Deployment Guide for InfoSphere Guardium IBM Redbooks® publication, system backup has two variations:

- Data backup, which contains a snapshot of the entire database, audit data, and all system definitions.
• Configuration backup, which contains only system definitions. There is no audit data and, therefore, it is significantly smaller in size.

Backup should be configured and scheduled to run on a regular basis.

System backup

System backup can be done for system configuration, system data, or both by selecting the backup type checkboxes.

Figure 27. Amazon required values for using S3 for system backup

After entering and saving the S3 bucket name and Amazon credentials, and selecting system backup content, Modify Schedule and Run Once Now become enabled. Click Run Once Now to start the backup process right away or Modify Schedule to schedule this to run regularly.

Figure 28. Running system backup
The system data and configuration files are backed up on Amazon S3 in the bucket named aa-guardium.

**Figure 29. Backup system files in Amazon S3 bucket**

The backup system data and configuration file name format is as follows:

```
<data_date>-<time>-<hostname.domain>-SQLGUARD_CONFIG-<Guardium_major_release_number>.tgz
<data_date>-<time>-<hostname.domain>-SQLGUARD_DATA--<Guardium_major_release_number>.tgz
```

Example:

```
2014-01-27-0206-g1.guardium.com- SQLGUARD_CONFIG-9.0.tgz
2014-01-27-0206-g1.guardium.com- SQLGUARD_DATA-9.0.tgz
```

The system backup process activities can be viewed in Aggregation/Archive log report on the InfoSphere Guardium Monitor tab.

**Figure 30. System backup process**

The system backup can also be done through the CLI using the system backup command. Selecting **Configured Destination** from the menu will backup the system data/configuration files in the configured destination in the GUI.
Restoring system data and configurations

The system restore can be done through the CLI using the system restore command or by using the SQLGuard Diagnostics tool diag. Same as system backup, the Amazon credentials and S3 bucket name are required in order to restore the system.

Figure 32 shows running the restore system from the CLI.

As you can see, Amazon S3 is included in the list of restore options. By selecting that option, you will be guided to enter the Amazon information. Accessing the Amazon S3 bucket allows the system to retrieve all the backup files in that specific bucket and to generate a list of available backup files.
When generating the list of backup files available in Amazon bucket is completed, you are prompted to enter the name of a backup file to restore and select a recovery type by entering a number associated with recovery types.

**Figure 33. System restore process**

When generating the list of backup files available in Amazon bucket is completed, you are prompted to enter the name of a backup file to restore and select a recovery type by entering a number associated with recovery types.

```

Select your recovery type, for most cases use the normal option:

1. normal
2. upgrade

Select the restore type: (q to quit) 1
```

The message below shows a successful restore for the system configuration files.

**Figure 34. System restore result**

```
Select the restore type: (q to quit) 1
pgp: Signature made Mon 27 Jan 2014 02:05:59 AM BST using DSA key ID 2348FF9B
pgp: checking the trustdb
pgp: 3 marginal(s) needed, 1 complete(s) needed, classic trust model
pgp: depth: 0  valid: 1  signed: 0  trust: 0-, 0q, 0n, 0m, 0f, 1u
pgp: Good signature from "Backup Signer <support@guardium.com>"
/bin/tar: Removing leading '/ from member names
To verify the patch level, click on the file installed_patches.html on the fileserver.

Proceeding to shutdown services
Starting data recovery. This process can take a while to complete.
Leave this terminal open until the recovery is complete....
Start to restore turbinedump Mon Jan 27 03:27:00 BST 2014
restore turbinedump completed Mon Jan 27 03:27:11 BST 2014
Update group table Mon Jan 27 03:27:11 BST 2014
Data recovery complete
Proceeding to startup services
Safekeeping xregs
Recovery procedure was successful.
ok
Vx62.guard.swg.usma.ibm.com> 
```

Another option to run system restore is to use the SQLGuard diagnostics tool. To use this method, run the diag command from CLI command line and choose **Perform Maintenance Actions > Transfer backups & System Receiver.**
Figure 35. Running system restore using diag tools

Hint: Use a Linux® terminal to run SQLGuard so you can use the edit menu to cut and paste long strings, such as the secret key. If you use Putty, the Edit menu is not available in the SQLGuard tool.

Summary

Organizations that use InfoSphere Guardium and are moving their data storage to the Amazon Cloud now have the option of using that cloud storage for InfoSphere Guardium, as well. Guardium usage of S3 for backup and recovery provides developer with a simple and inexpensive infrastructure that can be used to store back up and restore any amount large quantity of data, anytime, from anywhere, without limitation.
Resources

Learn

- Learn about the different options for archive, purge, and restore in the InfoSphere Guardium Knowledge Center.
- Learn how to manage backup and archiving in the InfoSphere Guardium Knowledge Center.
- Read about aggregator commands to manage shared secret keys.
- The Amazon Web Services documentation includes information about how to obtain an access key and a secret access key.
- Visit InfoSphere Guardium for links to whitepapers, demos, and more.
- Visit the InfoSphere Guardium Knowledge Center to learn more about using InfoSphere Guardium.
- Watch videos on the InfoSphere Guardium YouTube channel, including videos of what’s new in V9.1.
- Check out the Amazon Web Services documentation for information about AWS services.
- Follow InfoSphere Guardium on Twitter.
- Visit the developerWorks Information Management zone to find more resources for DB2 developers and administrators.
- Stay current with developerWorks technical events and webcasts focused on a variety of IBM products and IT industry topics.
- Follow developerWorks on Twitter.
- Watch developerWorks on-demand demos ranging from product installation and setup demos for beginners, to advanced functionality for experienced developers.

Get products and technologies

- Evaluate IBM products in the way that suits you best: Download a product trial, try a product online, or use a product in a cloud environment.

Discuss

- Participate in the discussion forum for this content.
- Get involved in the InfoSphere Guardium users group on LinkedIn to ask questions and get advice from other users.
- Get involved in the developerWorks community. Connect with other developerWorks users while exploring the developer-driven blogs, forums, groups, and wikis.
About the authors

Azadeh Ahadian

Azadeh Ahadian is a software developer in the Information Management tool organization at IBM's Silicon Valley Lab in San Jose, Calif. She has a Master's degree with more than eight years of industry experience in software application and tools development. Azadeh has been a key contributor and a member of the pureQuery design time team.

Dario Kramer

Dario Kramer has been part of the InfoSphere Guardium team since 2003. Currently, Kramer is the InfoSphere Guardium System Architect and part of the R&D team of InfoSphere Guardium. Kramer led the technical aspects of its cloud integration and participated in the development and design of a number of InfoSphere Guardium modules and applications, and led several development projects. Previously, Kramer held a number of positions in Viryanet, including project manager, sales engineer, and support manager. Kramer has a bachelor's degree in mathematics and statistics.

Indrani Ghatare

Indrani Ghatare has been in software development at IBM for more than 13 years. Currently a member of the R&D team for InfoSphere Guardium, she developed Guardium's archive, backup, and restore capabilities using Amazon S3, monitoring capabilities for MongoDB, universal feed type 3, and many other features. Previously, Ghatare provided several integration solutions using InfoSphere software products. Prior to IBM, she worked at Informix Software Inc. developing Informix Dynamic Server. She holds master's degree in computer science.