Data, Email, and Web Security

v8.5.x

Administering Forcepoint Databases

Revision A

Contents

Introduction

Note

Applies to:

- Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x
- Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x
- Forcepoint Email Security, v8.5.x
- Forcepoint appliances, v8.5.x

Forcepoint security solutions include several databases used to store configuration information, reporting data, and other information.

This paper is designed to help database administrators understand the requirements of the databases used by Forcepoint Web Security, Forcepoint URL Filtering, Forcepoint DLP, and Forcepoint Email Security.

Overview of Forcepoint databases

Note

Applies to:

- Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x
- Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x
- Forcepoint Email Security, v8.5.x
- Forcepoint appliances, v8.5.x

The on-premises Forcepoint security solutions use a variety of databases for different purposes: configuration information, reporting data, URL categorization, fingerprinting, and forensics. Several data formats are used, including SQL, PostgreSQL, and Forcepoint proprietary formats.

These databases include:

Database	Description
Reporting Databases	Web, data, and email security products SQL Server databases that store reporting and logging data for individual Forcepoint security products. The Data reporting database also stores configuration data. See *Understanding the reporting databases*.
Settings Database	Web, data, and email security products PostgreSQL database that stores global configuration and infrastructure settings that affect all Forcepoint Security Manager modules. It is installed automatically on the Forcepoint management server and requires no administrator configuration.
Forcepoint URL Database	Web products only Proprietary database that contains URL categories and protocol definitions, as well as supporting information, such as risk class groupings. A copy of the URL Database resides on each Filtering Service machine. By default, a full update is performed daily. Incremental updates can occur much more frequently if they are enabled on the **Web** > **Settings** > **General** > **Database Download** page in the Forcepoint Security Manager. See Administrator Help for your web protection solution for further details.
RTM Database	Web products only Holds and organizes filtering data for display in Real-Time Monitor. This is an independent database (not hosted on SQL Server) installed with each RTM Client and RTM Server instance. Administrators can specify when Real-Time Monitor captures data on the **Web** > **Settings** > **Reporting** > **Preferences** page in the Forcepoint Security Manager. No other aspect of database behavior is configurable.

Database	Description
Web Forensics Database	Forcepoint Web Security only Stores details about files that may be associated with advanced malware threat activity in your network. Enable or disable the forensics repository and configure its location and size on the **Web** > **Settings** > **Reporting** > **Dashboard** page in the Forcepoint Security Manager. See “Configuring Dashboard reporting data” in the Administrator Help for your web protection solution for details.
Data Fingerprint Database	Forcepoint DLP only Stores data security fingerprints. See *Data Fingerprint Database*.
Data Forensics Database	Forcepoint DLP only Contains information about DLP and discovery transactions that resulted in incidents, such as the contents of an email body, including the From:, To:, and Cc: fields, as well as actual attachments. Transactions can also include web posts, endpoint operations, and discovered as well as other events. For transactions that occurred on a web channel, the forensics might include the URL category property. Configure the size and location of the forensics repository in the Data Security module of the Forcepoint Security Manager. Navigate to the **Settings** > **Deployment** > **System Modules** page and click **Forensics Repository** under the management server.

Related concepts

Understanding the reporting databases on page 6
Data Fingerprint Database on page 4

Data Fingerprint Database

One of the ways that you can classify data in Forcepoint DLP is by “fingerprinting” it. This allows the detection of sensitive information despite manipulation, reformatting, or other modification. Fingerprints enable the protection of whole or partial documents, antecedents, and derivative versions of the protected information, as well as snippets of the protected information whether cut and pasted or retyped.

When you fingerprint data, the fingerprints are stored in the Data Fingerprint Database on the Forcepoint management server and pushed to other Forcepoint DLP components for fast analysis on those machines.

How to tune the performance

To tune performance, you can configure the disk space and cache size of the database on the management server in the Data Security module of the Forcepoint Security Manager.

To do so:

Steps

1) Log onto the Security Manager.
2) Navigate to the **Data** > **Settings** > **Deployment** > **System Modules** page.
3) Select **Primary Fingerprint Repository** under the management server.
4) Adjust the maximum disk space and cache size as needed.

How to configure the settings

Before you begin

Secondary repositories are maintained by Content Gateway and Forcepoint Email Security. You can configure Forcepoint DLP to detect fingerprints from repositories local to those machines (best practice) or from a remote machine, such as the management server.

Periodically, you must synchronize the fingerprints in the secondary repositories with the ones on the management server. How often you synchronize depends on your business needs. For best performance, select a time with low traffic volume.

To configure these settings:

Steps

1) Navigate to the **Settings** > **Deployment** > **System Modules** page.
2) Select **Secondary Fingerprint Repository** under the component or solution of interest.
3) Choose the repository from which to detect fingerprints.
4) Set the maximum cache size and synchronization frequency as needed.

Next steps

See “Configuring the fingerprint repository” in the Forcepoint DLP Administrator Help for instructions on configuring these settings.

Understanding the reporting databases

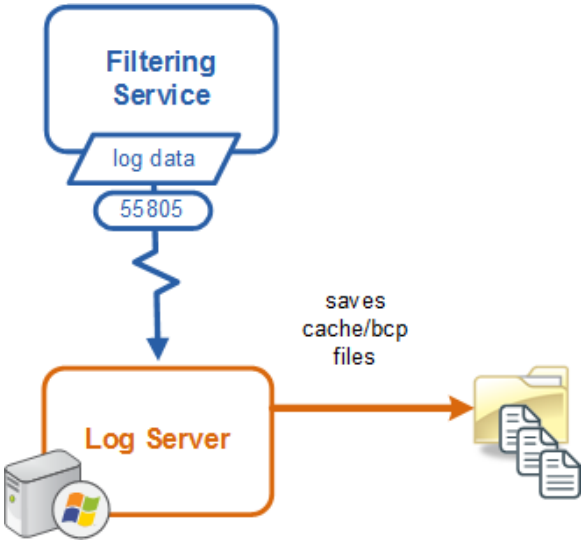
Note

Applies to:

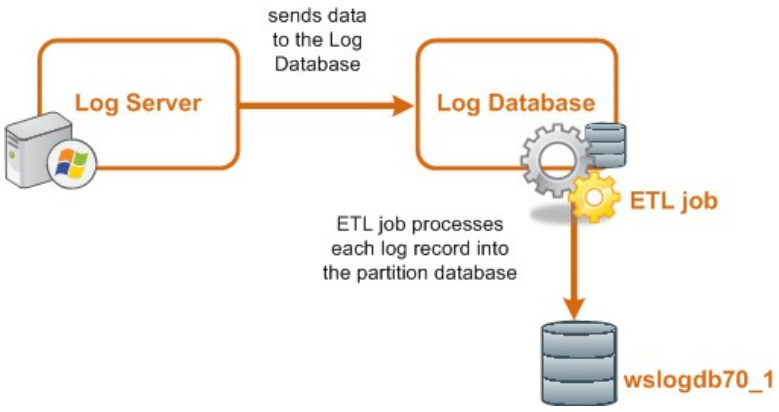
- Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x
- Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x
- Forcepoint Email Security, v8.5.x
- Forcepoint appliances, v8.5.x

Each reporting database stores the logging data collected by a specific Forcepoint security solution: Forcepoint Web Security, Forcepoint URL Filtering, Forcepoint DLP, or Forcepoint Email Security.

- The web protection **Log Database** stores Internet and cloud application request data collected by Log Server, such as the source, destination, time, category, risk class, action (also called disposition), bytes sent and received, and so on for use by web protection reporting tools.
  Log Server receives information about Internet activity from Filtering Service and initially stores it locally:



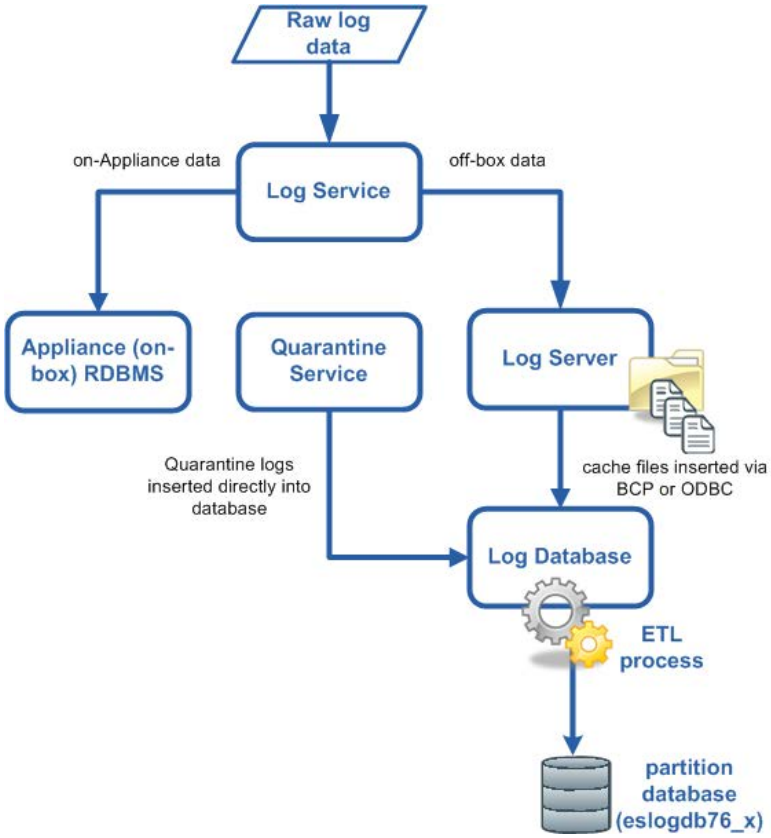
Whenever it is able, Log Server forwards the cache files to the Log Database, where the ETL job processes them into log records in a database partition:



An end user who uses the Filtering Service has no direct or indirect influence over the database. Thus, although the log entry is stored in the Microsoft SQL Server database, the user did not direct its storage and cannot retrieve it.

Prior to v8.3, the only interface to the database itself is the Log Server, reporting services, and Forcepoint Security Manager. But from v8.3 onwards, the cloud app service also can insert the cloud app data to log database. Filtering Service and Content Gateway do not access the database, but instead send information via the Log Server.

- The **Email Log Database** stores records of email traffic and the associated analysis and disposition of that traffic. Forcepoint Email Security reporting uses this information to generate dashboard status charts and email activity reports showing, for example, the size and volume of messages processed, message analysis results, and email source and destination.
  Log and quarantine data are recorded as follows:



- **Data Incident and Configuration Database** stores information about email, web, and other traffic that resulted in data loss prevention (DLP) policy breaches, such as the source, destination, time, status, and severity of each breach. It also stores Forcepoint DLP policy configuration and system settings.

The reporting databases are all hosted by Microsoft SQL Server. They may be hosted by the same installation and instance, or by different installations or instances.

Although Microsoft SQL Server Express may be packaged with your software, most organizations should purchase Microsoft SQL Server Standard or Enterprise. (See *Can I use SQL Server Express?*, for guidance.)

As a best practice, during installation, connect Forcepoint software to a SQL Server instance on another machine. For testing purposes, or in very small networks, it is possible to install SQL Server Express on the Forcepoint management server. For more information, see *Microsoft SQL Server deployment options*.

Related concepts

Microsoft SQL Server deployment options on page 8

Related reference
Can I use SQL Server Express? on page 16

Microsoft SQL Server deployment options

Note

Applies to:

- Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x
- Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x
- Forcepoint Email Security, v8.5.x
- Forcepoint appliances, v8.5.x

In deployments of over 500 users, the best practice is to host the reporting databases on a Standard or Enterprise installation of Microsoft SQL Server. SQL Server typically resides on a dedicated machine. It may host databases for third-party applications, in addition to the Forcepoint databases.

If SQL Server resides on a dedicated machine and there are multiple SQL instances installed, SQL Server Browser must be running.

During Forcepoint software installation, the setup program prompts for the IP address or hostname of the SQL Server machine. It is also possible to specify:

- An instance name
- The port to use for SQL Server communication
- Whether or not to encrypt communication with SQL Server

Find the supported versions of Microsoft SQL Server for your product by selecting the appropriate link in the Certified Product Matrix.

Clustering is supported for all supported versions of SQL Server noted in the Certified Product Matrix linked above (Forcepoint Email Security and Forcepoint Web Security only).

See *Can reporting databases be hosted in a SQL Server cluster?*.

Related concepts
Can reporting databases be hosted in a SQL Server cluster? on page 29

SQL Server Express on the management server

Important

Forcepoint has removed the ability to install SQL Express as an option for new deployments of Forcepoint Security Manager. This change was introduced via a revised version of the Forcepoint Security Installer introduced in July 2019.

The change was required to reduce the risk of deploying SQL Express without the latest security updates. Forcepoint Security Manager still supports and will work with the latest version of SQL Server Express.

You may use SQL Server Express for small deployments, but it must be installed independently.

For smaller deployments (up to 500 users), the reporting databases can be installed on the Forcepoint management server. To do this, select the **Install SQL Server Express on this machine** option, if available, during a Custom installation.

When included, use only the version of SQL Server Express included in the Forcepoint Security Installer. If not included, download and install the supported version of SQL Server Express from Microsoft.

SQL Server Express can run with Forcepoint software on the Windows platforms noted in the Certified Product Matrix.

Reporting database specifications and recommendations

Note

Applies to:	In this topic
■ Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x ■ Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x ■ Forcepoint Email Security, v8.5.x ■ Forcepoint appliances, v8.5.x	■ *Hardware specifications* ■ *RAM* ■ *Disk considerations* ■ *Virtualization*

The performance of Forcepoint reporting solutions is heavily dependent on the SQL Server and the configuration of its underlying resources. The more you invest in the system, the better it will perform.

For optimal results, host the reporting databases on Windows Server 2016 Enterprise Edition and SQL Server 2016 with the latest service pack.

Also consider the factors that follow when designing your database system.

Related concepts

Hardware specifications on page 10
RAM on page 10
Disk considerations on page 11
Virtualization on page 11

Hardware specifications

Use hardware that meets or exceeds Microsoft's recommended (not minimum) hardware requirements for SQL Server.

Microsoft SQL Server 2019 (Forcepoint DLP 8.7.1 and later):

https://docs.microsoft.com/en-us/sql/sql-server/install/hardware-and-software- requirements-for-installing-sql-server-ver15?view=sql-server-ver15

Microsoft SQL Server 2016 and 2017:

https://msdn.microsoft.com/en-us/library/ms143506.aspx

Microsoft SQL Server 2014:

https://www.microsoft.com/en-us/download/details.aspx?id=42299

Microsoft SQL Server 2012:

http://technet.microsoft.com/en-us/library/ms143506(v=sql.110).aspx

RAM

The reports that administrators generate often require that SQL Server be capable of loading, summarizing, and processing large amounts of data across multiple physical databases. Even if your organization doesn't run complex reports, if there are multiple reporting administrators, the system may frequently be asked to generate several reports concurrently. This places high demands on system memory. As a result, increasing RAM may provide noticeable performance improvements.

Keep in mind that the operating system version and/or the version of SQL Server may limit the amount of physical RAM that can be used by the system.

Consult your operating system and SQL Server documentation to ensure that you utilize the maximum physical RAM possible in your chosen system.

See *Other factors that affect the performance of Forcepoint reporting*, for guidance on how the memory demands of Forcepoint reporting may increase or decrease based on how you use it.

Related reference

Other factors that affect the performance of Forcepoint reporting on page 21

Disk considerations

Because database operations are often I/O-intensive, a faster, dedicated disk can improve how the database performs. Use high-performance disks whenever possible to ensure optimum database operation.

SQL Server performs better when there is minimal disk contention. For best performance, place tempdb files, reporting database files, and log files on separate physical drives with dedicated controllers. Follow SQL Server recommendations for installing and configuring SQL Server.

RAID controllers provide disk redundancy and can increase disk throughput by spreading I/O activity across multiple physical disks. If you have a high-demand system, for peak performance, use a RAID 10 configuration managed by a hardware- based RAID controller. (For systems with a lower I/O profile, RAID 5 may provide sufficient performance at a lower cost.)

For best performance:

- Use a hardware RAID controller, not software.
- Max out the RAID controller RAM cache.
- Use multi-channel links between the RAID controller and disk shelf or SAN.
- Separate the transaction logs and database files onto separate disks, arrays, or LUNs.
  This means separate spindles or RAID arrays, not just separate logical drives.

  If you are using a SAN, map the physical and logical LUNs to ensure you are reading and writing key components to separate physical drives. (This is to enable multiple parallel I/O operations.)
- Make sure that different database files are spread across different disks wherever possible. For very large deployments, for example, plan to map wslogdb70_1 to spindle 1, wslogdb70_2 to spindle 2, and so on.
- The disk stripe size (how files are stored across striped arrays) is also critical and is a function of both the typical read/write size and the total database size. Contact Microsoft or a sizing specialist for guidance.
- In large deployments, map different disks, arrays, or LUNs across different hardware controllers to maximize parallel operations and controller cache usage.
- If you are using spinning disks, lower seek times are better.

Virtualization

Forcepoint products can be deployed on virtualized systems. Be aware that the use of virtualization can affect system performance. Expect up to 25% performance degradation.

For specifics on virtualization support, see the article Virtual Machine Support in the Forcepoint knowledge base.

Note that Microsoft has its own support guidelines for SQL Server, which can be found here:

http://support.microsoft.com/?id=956893

How big will my reporting database be?

Note

Applies to:	In this topic
■ Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x ■ Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x ■ Forcepoint Email Security, v8.5.x ■ Forcepoint appliances, v8.5.x	■ *Web protection size estimates* ■ *Email size estimates* ■ *Data size estimates*

When estimating the size of your Forcepoint reporting databases, keep the following factors in mind:

- Forcepoint DLP logs only data violations, which typically represent a small fraction of your total web and email volume.
- Forcepoint Email Security logs detected spam and quarantined email messages, which typically represent up to 90% of your email volume.
- Forcepoint Web Security and Forcepoint URL Filtering log all web transactions, which typically represent about 40% of your total web traffic volume when visits (the default) are recorded rather than hits. See *Web visits, consolidation, and full URL logging*, for details.
- If IPv6 is used in the deployment, the web protection Log Database may increase in size by approximately 4%.

Warning

These are averages calculated across a large number of Forcepoint customers with widely varying deployments. While these numbers are useful for estimation, please analyze your system's actual performance after a few days or weeks to reassess your database sizing needs.

Related concepts

Web visits, consolidation, and full URL logging on page 18

Related reference

Web protection size estimates on page 12
Email size estimates on page 14
Data size estimates on page 15

Web protection size estimates

The table below provides some general rules to help estimate the size of your web protection Log Database, based on whether you are recording visits or hits. and URL hostnames or full URLs. By default, visits and URL hostnames are recorded.

	URL hostnames	Full URL logging

Visits	3 MB per user per month	4.5 MB per user per month
Hits	8 MB per user per month	12 MB per use per month

The size increase that accompanies full URL logging can vary widely based on the type of URL being logged. For example, the following types of traffic tend to generate very long URLs:

- Research database search and results URLs
- Search strings
- Social networking sites
- Online apps and games

Consider the example of a 7500-user organization that wants to store data for three months, using monthly database partition rollover:

- Visits are enabled
- Full URL logging is enabled
- The Internet access policy is permissive
- A high percentage of web traffic includes longer than average URLs (social networking sites and online games)

Based on this information, estimate 6 MB of data stored per user per month.

Because rollover is monthly, it is necessary to allocate enough space to hold up to 4 months of data. (This ensures that it is always possible to report on 3 months of data, even during the first days or weeks of the most recent month.)

To start, therefore, 180 GB would be needed for the database logging partitions (not including the catalog database and AMT partition).

```
6 MB * 7500 users * 4 months = 180 GB
```

In addition, calculate space for the following:

- **Database transaction logs**, whose size depends on the recovery model you choose.
  - For full recovery (not recommended), allow for the same size as the data itself (180 GB in our example).
  - For simple recovery, the log volume depends on the amount of database activity. Allow roughly 30 percent of the size of an active database partition.

  In our example, the total size of 4 full partitions is 180 GB, so each partition is 45 GB in size. So for simple recovery:

  ```
  0.3 * 45 GB = 13.5 GB
  ```

  Note that this space is used only while data is being logged to the database. Inactive partitions, for example, have minimal transaction log activity. Likewise, during time periods in which little or no Internet activity is being recorded, the transaction log is very small.
- **Temporary (tempdb) storage** is used heavily during report generation and when the maintenance job is reindexing the database. The amount of tempdb space required depends heavily on the types of reports being generated.
  In most environments, it is sufficient to allow twice the size of a partition:

  ```
  45 * 2 = 90 GB
  ```
- **Temporary database (tempdb) transaction logs** take up about 20 percent of the tempdb storage size. In our example:
  ```
  0.2 * 90 GB = 18 GB
  ```

The total space required for the web protection Log Database in our example, then, is:

```
180 GB + 13.5 GB + 90 GB + 18 GB = 301.5 GB
```

Tip

As a best practice, record visits rather than hits to reduce significantly the amount of storage space required.

See *Factors that affect reporting database size*, for an explanation of the Forcepoint Web Security visits algorithm, other data reduction options, and the full URL logging setting.

Related reference

Factors that affect reporting database size on page 17

Email size estimates

For best practice, allow 1-3 MB per user per month when estimating the size of your Email Log Database.

For illustration, assume there will be 2 MB of data per user per month for 7500 users. Data will be kept for 3 months, with a scheduled monthly partition rollover.

Because rollover is monthly, it is necessary to allocate enough space to hold up to 4 months of data. (This ensures that it is always possible to report on 3 months of data, even during the first days or weeks of the most recent month.)

This means that you would need 60 GB to start:

```
2 MB * 7500 users * 4 months = 60 GB
```

In addition, you need to calculate space for the following:

- **The database transaction logs**, whose size depends on the recovery model you choose.
  - For full recovery (not recommended), allow for the same size as the data itself (60 GB in our example).
  - For simple recovery, the log volume depends on the amount of database activity. Allow roughly 30 percent of the size of the available data.
    ```
    0.3 * 60 GB = 18 GB
    ```
- **Temporary database (tempdb) storage** is used heavily during report generation. The amount of tempdb space required depends on the reports being run. As a best practice, allow the size of the logged data for temporary storage. In our example:
  ```
  60 GB
  ```
- **Temporary database (tempdb) transaction logs** take up about 20 percent of the tempdb storage size. In our example:
  ```
  0.2 * 60 GB = 12 GB
  ```

The total space required in our example, then, is:

```
60 GB + 18 GB + 60 GB + 12 GB = 150 GB
```

Use the following tables to estimate storage space required for 30 days' worth of data based on email volume.

- The same requirements for partitions, logs, and temporary databases apply.
- The estimates assume an average of 5 recipients per email message.

With the Forcepoint Email Security Hybrid Module enabled:

	10 msg/ user/ day	50 msg/ user/ day	100 msg/ user/ day	200 msg/ user/ day	400 msg/ user/ day
500 users	600 MB	3300 MB	6600 MB	13300 MB	26700 MB

1000 users	1300 MB	6600 MB	13300 MB	26700 MB	53400 MB
2000 users	2600 MB	13300 MB	26700 MB	53400 MB	106800 MB
5000 users	6600 MB	33400 MB	66800 MB	133600 MB	267200 MB

Without the Hybrid Module:

	10 msg/ user/ day	**50 msg/ user/ day**	**100 msg/ user/ day**	**200 msg/ user/ day**	**400 msg/ user/ day**
500 users	500 MB	2600 MB	5200 MB	10400 MB	20900 MB
1000 users	1000 MB	5200 MB	10400 MB	20900 MB	41800 MB
2000 users	2000 MB	10400 MB	20900 MB	41800 MB	83600 MB
5000 users	5200 MB	26100 MB	52200 MB	104500 MB	209000 MB

The calculations used are:

- With the Hybrid Module:
  `(0.01622+0.02348* RECIPIENT)* SPEED * USER`
- Without the Hybrid Module:
  `(0.0384+0.01322* RECIPIENT)* SPEED * USER`

Note that:

- USER represents the total number of users.
- RECIPIENT is the average number of recipients for each message.
- SPEED is the average message volume for each user per day.

Note that the email hybrid service drops or blocks the following types of email before they reach on-premises components:

- Email that comes from known bad (blacklisted) sources
- Email that has a very high spam score in the cloud

This significantly reduces the amount of data stored in the Email Log Database, as shown in the examples above.

Data size estimates

The Data Incident and Configuration Database is different from the Web and Email Log Databases in that it stores both configuration and log data. In addition, it logs only policy violations, not all events. In order to estimate the potential size of the Incident and Configuration Database, consider the following:

Feature	Impact on Database Size
Discovery incidents	14 KB per incident
(not standard when all	Typically, no more than 3.5 GB total
Forcepoint products are	
installed)	
Network and endpoint incidents	10 KB per incident Typically, 30 KB per user per month

Feature	Impact on Database Size
User directory import data	8 KB per record Typically, no more than 1 GB total
Configuration data	200 MB

As a general guideline, it is unlikely that your Data Incident and Configuration Database will require more than 9 GB of storage total, no matter how many users are in your environment.

The volume of configuration and incident data varies based on your use of Forcepoint DLP features as explained in *Factors that affect reporting database size*.

Related reference
Factors that affect reporting database size on page 17

Can I use SQL Server Express?

Note

Applies to:

- Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x
- Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x
- Forcepoint Email Security, v8.5.x
- Forcepoint appliances, v8.5.x

Microsoft SQL Server Express may be bundled into Forcepoint Security Installer for Windows. If it is not, it can be downloaded from Microsoft and installed manually.

Regardless of the resources available on the machine hosting the database engine, Microsoft limits the resources that SQL Server Express can use.

This results in practical limits on the amount of data that can be stored in SQL Server Express, while still maintaining acceptable performance when generating reports and processing log data. For best performance, no more than 30 GB of data should be stored in SQL Server Express across all Forcepoint security products. The sizing recommendations below are based on this assumption.

For an explanation of particular demands that Forcepoint reporting places on available memory, see *Other factors that affect the performance of Forcepoint reporting*.

Use the charts below to see how much data you can store using SQL Server Express based on the Forcepoint product or products you are using. You can then determine if that limit still meets your data retention requirements. If so, you can use SQL Server Express.

Note

Standalone Forcepoint DLP installations can support up to 5000 users with SQL Server Express.

Users	Web*	Web and Data	Data and Email**	Web, Data, and Email
2000+	Do not use	Do not use	Do not use	Do not use

Users	Web*	Web and Data	Data and Email**	Web, Data, and Email
2000	45 days	30 days	Do not use	Do not use
1500	60 days	45 days	Do not use	Do not use
1000	90 days	60 days	Do not use	Do not use
750	4 months	3 months	Do not use	Do not use
500	6 months	4 months	30 days	Do not use
250	1 year	8 months	60 days	30 days
100	1 year+	1 year	5 months	3 months

- * Web: Actual size is based on web requests processed per day, which can be extrapolated from the peak requests per second in your network. If you are not able to determine these numbers, Forcepoint has found that the ratio of users to their peak number of web requests per second (as opposed to average number per second) is 10 to 1. This is an acceptable ratio to use for the purposes of estimating the number of users that this traffic represents in the average network. Note that this calculation is based on a standard curve distribution of traffic throughout the day, where the peak is mid-day and there is almost no traffic throughout the night.
- ** Email: Actual sizing is based on total email messages sent and received per day, including spam. Forcepoint has found that an estimate of 1 message per user per minute (including spam) is an acceptable estimate for the purposes of estimating the email volume generated for any given number of users.

Related reference

Other factors that affect the performance of Forcepoint reporting on page 21

Factors that affect reporting database size

Note

Applies to:	In this topic
■ Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x ■ Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x ■ Forcepoint Email Security, v8.5.x ■ Forcepoint appliances, v8.5.x	■ *Web visits, consolidation, and full URL logging* ■ *Email sizing factors* ■ *Data sizing factors*

Related concepts

Web visits, consolidation, and full URL logging on page 18

Related information
Email sizing factors on page 18
Data sizing factors on page 19

Web visits, consolidation, and full URL logging

Forcepoint Web Security and Forcepoint URL Filtering use proprietary algorithms to reduce the volume of log data in order to achieve a balance between visibility into users' web browsing activity and the size and performance of the Log Database.

- When you enable **visits**, Log Server combines the individual elements that create a web page (such as graphics and advertisements) into a single log record that includes bandwidth information for all elements of the visit.
  When this option is disabled, you instead log **hits**. In this case, a separate log record is created for each HTTP request generated to display different page elements, including graphics, advertisements, embedded videos, and so on. This creates a much larger Log Database that grows rapidly.

  Disabling visits can increase the total amount of data stored in the Log Database by a factor of 2.5.
- To further reduce the size of the database, enable **log record consolidation**. This combines multiple, similar Internet requests into a single log record, reducing the granularity of reporting data.
- By default, web protection products log only the URL hostname for each request, instead of the full URL. Storing the full URLs provides more visibility into where users are going within a particular website, but increases the Log Database storage demands.
  Enabling full URL logging can increase the size of each record by 50%.

For information about more ways to either reduce the size of the Log Database or increase the amount of data recorded, refer to: Log Database sizing guidance.

Email sizing factors

Email hybrid service

The Forcepoint Email Security hybrid service (included with the Hybrid Module) drops email that comes from known bad (blacklisted) sources and blocks email with a very high spam score in the cloud before it ever reaches the email appliance. This reduces the amount of data stored in the Email Log Database for reporting by 30 MB per user per month.

Above average email traffic: recipients, quarantined messages, or spam

The sizing guidelines above are based on the following assumptions about the email traffic handled by Forcepoint Email Security. These assumptions are derived from the average email traffic pattern of Forcepoint customers over time.

- There are an average of five recipients for each email message.

- When the email hybrid service is not enabled, the ratio among spam messages, infected messages, and clean messages is 85-to-1-to-14.
- The email hybrid service scans only inbound email traffic, and it can block 25% of spam.
- All outbound and internal email messages are clean.

Note that Forcepoint Email Security counts the number of recipients for each message rather than the number of messages sent. Each recipient is counted as a transaction.

If the pattern of email traffic in your organization exceeds these averages, your storage capacity will vary.

Data sizing factors

Number of discovery incidents

Before you begin

Forcepoint DLP limits the number of discovery incidents that can be stored in the Data Incident and Configuration Database in order to prevent improperly configured discovery policies from flooding the database. By default, this limit is set to 1 million incidents. **If you are using SQL Server Express, you should reduce this number to 250,000**.

To do this:

Steps

1) Log onto the Forcepoint Security Manager.
2) Go to the **Data** > **Settings** > **General** > **Reporting** page.
3) Select the **Discovery** tab.
4) Adjust the Maximum Discovery Incidents field.

Next steps

Refer to “Setting preferences for discovery incidents” in the Forcepoint DLP Administrator Help for more information.

Note

Forcepoint Data Discovery is not included in combined web, data, and email security solutions. It is an add-on feature that requires a separate subscription.

Forcepoint DLP supports up to 500,000 incidents per partition. See the Forcepoint DLP Administrator Help for more information.

Rate of network and endpoint incidents

The rate of network and endpoint incidents detected varies widely across Forcepoint customers. The sizing guidelines above are based on an average incident rate of 1 per user every 10 days (an incident is a policy violation). For best practice, periodically review the actual incident rate in the database to gauge how closely your environment matches this average, and then adjust your database storage requirements based on the actual data in your environment.

Do this by examining the Incident Trends report found in the Data Security module of Forcepoint Security Manager under **Main** > **Reporting**.

Note

Forcepoint DLP Endpoint is not included in combined web, data, and email security solutions. It is an add-on feature that requires a separate subscription.

The Forcepoint DLP database stores data in partitions per each calendar quarter. You can have 1 active partition for the current quarter.

If you are using Microsoft SQL Server Standard or Enterprise for your reporting database, you can have up to 8 online partitions (approximately 2 years), but if you are using SQL Server Express, you can have only 4 (approximately 1 year). (Online partitions are partitions that can be used to show reports and log data.)

For both databases, you can have up to 12 archived partitions representing 3 years of records, and 4 restored partitions (1 year).

Partition type	Microsoft SQL Server Standard or Enterprise	Microsoft SQL Server Express
Active	1 partition (current quarter)	1 partition (current quarter)
Online	up to 8 partitions (2 years)	up to 4 partitions (1 year)
Restored	up to 4 partitions (1 year)	up to 4 partitions (1 year)
Archived	up to 12 partitions (3 years)	up to 12 partitions (3 years)
Total available managed partitions	**25**	**21**

Refer to “Archiving incident partitions” in the Forcepoint DLP Administrator Help for more information on archiving. For instructions on setting the maximum disk space allowed for the incident archive, refer to “Configuring the incident archive.”

Size of user directory import

To support user-based policy and reporting, Forcepoint DLP imports entries from your user directory—such as Active Directory or Domino—into the Configuration Database. Depending on the size and design of your user directory, this can result in database space being consumed by entries that are not needed by Forcepoint DLP. To reduce the number of imported user directory entries:

- Configure Forcepoint DLP to import entries from a more specific root context that is deeper in the tree than the directory’s root context.
- Restrict the user attributes that are imported by specifying specific attributes to import; or eliminate them altogether by disabling the import of user attributes.

How to configure user directory settings

Steps

1) Log onto the Forcepoint Security Manager.
2) Go to the **Data** > **Settings** > **General** > **User Directories** page.
3) Select the user directory to edit.
4) Modify or add a root naming context.
5) Modify the user attributes settings.

Next steps

Refer to the “Adding or editing user directory server information” section in the Forcepoint DLP Administrator Help for information on configuring these settings.

Other factors that affect the performance of Forcepoint reporting

Applies to:	In this topic
■ Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x ■ Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x ■ Forcepoint Email Security, v8.5.x ■ Forcepoint appliances, v8.5.x	■ *Factors affecting Web reporting performance* ■ *Factors affecting Email reporting performance* ■ *Factors affecting Data reporting performance*

Related concepts

Factors affecting Data reporting performance on page 24

Related information

Factors affecting Web reporting performance on page 22
Factors affecting Email reporting performance on page 23

Factors affecting Web reporting performance

Users' web browsing behavior

Web browsing behavior varies widely from organization to organization. Periodically review your database performance, your reporting needs, and the actual data in the database so you can identify ways to reduce the demands on your reporting system.

Selective logging

Forcepoint Web Security and Forcepoint URL Filtering allow you to reduce the demands on your reporting system by not logging traffic to websites in selected categories. For example, online retailers might disable logging for Shopping categories. This can result in a large reduction in the amount of data that has to be stored and managed.

For information on configuring selective logging, refer to "Configuring how requests are logged" in the Administrator Help for your web protection solution.

Use of Detail Reports over long time periods

Reports of web browsing activity over long time periods (weeks and/or months) require much more memory, processor time, and disk I/O to generate. For better performance, run summary reports across long time periods on a regular schedule, then use detail reports only for investigating specific users in shorter time periods. If you have business requirements that demand generating detail reports across a large time window, you can:

- Schedule the reports to run during low-activity periods in your network.
- Invest in more hardware resources for your reporting system.

Number of scheduled reports or number of delegated reporting administrators

If you have several delegated administrators that use reporting each day or create several scheduled reports to run each night, this can degrade the performance of your reporting tools. If you meet these usage profiles, consider investing in more hardware resources for your reporting system: more RAM, faster disks, faster CPUs, and higher-end versions of SQL Server and Windows that support more hardware.

Geographical location of users

If you have users distributed among multiple physical locations and your business does not require unified reporting across all users, consider deploying separate Log Server and Log Database instances in each location.

Calculation of Internet browse time

By default, a database job calculates Internet browse time at 2 a.m. for the previous day's activity. This is a memory-, processor-, and disk I/O-intensive activity. If you don't use Internet browse time to manage your users'

web browsing activity, consider disabling Internet browse time to improve the performance of your reporting system.

For information on configuring Internet browse time, refer to “Configuring Internet browse time options” in the Administrator Help for your web protection solution.

Partitioning

The Log Database is segmented into partitions for easier data management. Depending on the time period covered by a report, Forcepoint software may need to query multiple partitions. This may make report generation less efficient.

By default, a new Log Database partition is created when the current partition size reaches 5 GB (3 GB if you use SQL Server Express). (With Standard and Enterprise versions of SQL Server, you can also configure the Log Database to roll over at a specific time interval.)

Review the size and content of the partitions in the database after your system has been installed and receiving data for a few days, then tune the partitioning configuration (rollover size or time period) accordingly.

For information on managing partitions, refer to “Configuring database partition options” in the Administrator Help for your web protection solution.

Note

Due to Microsoft SQL Server restrictions, you are limited to 70 active log database partitions. You may retain more than 70 partitions, but only 70 may be active (enabled) at any one time.

Hybrid service users

The sizing guidelines in this document include logs generated by users managed by the Forcepoint Web Security hybrid service (Hybrid Module). When sizing your reporting system, do not forget to include those users.

Configuration options that affect Log Database sizing, including selective logging, logging visits, and full URL logging, also apply to hybrid log records, so no special consideration needs to be made for those users.

Factors affecting Email reporting performance

Number of scheduled or custom presentation reports

If you create a large number of scheduled reports to run each night (more than 10) or use a large number of custom presentation reports (more than 10) each day, this can affect reporting performance. In particular, the following 3 reports place high demands on system resources:

- Top *n* External Recipients by Message Volume
- Top *n* External Recipients by Message Size
- Top *n* Data Loss Prevention Violations by Volume

If you meet these usage profiles, consider investing in more hardware resources for your reporting system: more RAM, faster disks, faster CPUs, and higher-end versions of SQL Server and Windows that support more hardware.

Partitioning

The Email Log Database is segmented into partitions for easier data management. Depending on the time period covered by a report, your reporting tools may need to query multiple partitions. Running such reports may be inefficient.

By default, a new Email Log Database partition is created after 5 GB of data has been stored in a single partition. You can also configure Forcepoint Email Security to create a new partition based on a weekly or monthly time interval.

Review the size and content of the partitions in the database after your system has been installed and receiving data for a few days, then tune the partitioning configuration accordingly.

For information on managing partitions in Forcepoint Email Security, refer to “Configuring Log Database options” in Forcepoint Email Security Administrator Help.

Note

Due to Microsoft SQL Server restrictions, you are limited to 70 active log database partitions. You may retain more than 70 partitions, but only 70 may be active (enabled) at any one time.

Factors affecting Data reporting performance

Forcepoint DLP automatically archives database partitions containing older data to reduce storage requirements and maintain a high performance reporting experience. To further reduce the data storage requirements, you can choose to create archive partitions sooner and keep fewer concurrent restored archives.

Refer to “Archiving incident partitions” in the Forcepoint DLP Administrator Help for more information on archiving.

Forcepoint reporting database FAQs

Applies to:	In this topic
■ Forcepoint Web Security and Forcepoint URL Filtering, v8.5.x ■ Forcepoint DLP, v8.5.1, v8.6.x, v8.7.x, v8.8.x, v8.9.x, v9.0, v10.x ■ Forcepoint Email Security, v8.5.x ■ Forcepoint appliances, v8.5.x	■ *Which database tools are required or used?* ■ *Which permissions are required?* ■ *Which database jobs are run?* ■ *How does the installer set up each database?* ■ *How big should the database partitions be?* ■ *How many partitions can be accessed at the same time?* ■ *How do I configure partition rollover?* ■ *What if I need more partitions to run reports?* ■ *Do the reporting databases use named instances?* ■ *Can reporting databases be hosted in a SQL Server cluster?*

Related concepts

Which database tools are required or used? on page 25
Which permissions are required? on page 25
Which database jobs are run? on page 26
How does the installer set up each database? on page 26
How big should the database partitions be? on page 28
How do I configure partition rollover? on page 28
What if I need more partitions to run reports? on page 29
Do the reporting databases use named instances? on page 29
Can reporting databases be hosted in a SQL Server cluster? on page 29
How many partitions can be accessed at the same time? on page 28

Which database tools are required or used?

Forcepoint reporting components connect to the SQL Server database engine as clients and perform standard Transact-SQL commands and stored procedures.

Forcepoint Web Security and Forcepoint Email Security may use 2 database utilities:

- **bcp** to use bulk insertion for adding logs to the database.
- **osql** to run SQL scripts during Log Database installation.

Which permissions are required?

During Forcepoint DLP installation, modification, or repair, the account used for database creation and access needs **sysadmin** server role membership. Also, **Backup database** permission on the **URL** database is required for installation only. After installation, the account privileges can be reduced to the **db_owner** of the newly created databases, and no access to any other user database except system databases such as UEL, tempdb, and model is required. Additionally, the **dbcreator** server role should be granted to enable backup and restore functionality.

If you're using SQL Server to install the Web Log Server and Email Log Server, the user account that owns the reporting database must:

- Be a member of the **dbcreator** server role
- In the **msdb** database:
  - Have membership in the **db_datareader** role
  - Have membership in one of the following roles:
    - SQLAgentUser Role
    - SQLAgentReader Role
    - SQLAgentOperator Role

For SQL Server Express, the user account requires the **sysadmin** server role.

See the Certified Product Matrix for supported versions of SQL Server.

Which database jobs are run?

The following database jobs are installed with the Web Log Database and Email Log Database:

- The Extract, Transform, and Load (ETL) job runs continually, receiving data from Log Server, processing it, and then inserting it into the partition database. When trend data (Web) retention is enabled, the ETL job is also responsible for inserting trend data into the catalog database.
  The ETL job must be running to process log records into the Log Database.
- The database maintenance job performs database maintenance tasks. This job runs nightly, by default. (Web) Once data is processed and moved to the database tables used by the Cloud App report, the maintenance job is also responsible for deleting cloud apps data that is more than 2 days old from temporary log database tables.

ETL jobs are run, then re-run 10 seconds after they finish for SQL Server Standard and Enterprise. For SQL Server Express, 60 seconds elapse between completion of one job and start of the next.

Maintenance jobs are run once every night by default. The jobs are run automatically. The Web Log Database also installs the following jobs:

- The Internet browse time (IBT) job analyzes data and calculates browse time for each user. The IBT database job is resource intensive, requiring significant server resources. This job runs nightly.
- When trend data retention is enabled, the trend job uses daily trend data created by the ETL job to update weekly, monthly, and yearly trend records for use in presentation reports. This job runs nightly.
  Even when trend data retention is disabled, the trend job processes data from the threats (AMT) partition to provide trend data on the Threats dashboard.
- The Advanced Malware Threat (AMT) ETL job receives, processes, and inserts data into the threats partition database. Only log records that include a severity ranking are recorded in the threats partition. Data from this partition is used to populate the Threats dashboard.
  The AMT ETL job also populates the database tables used to provide the data for all application reports and the Advanced File Analysis report.

When configuring the start time for the (Web and Email) maintenance job and the (Web) Internet browse time job, consider system resources and network traffic. These jobs can be resource intensive and time consuming, so they can have a negative impact on logging and reporting performance. When trend data (Web) retention is enabled, the trend job is run, by default, at 4:30 AM. Try to avoid starting other jobs at time that might overlap with the trend job.

Both Log Databases require either the SQL Server Agent service (SQL Server Standard or Enterprise) or Service Broker (SQL Server Express) to run database jobs.

Note

Forcepoint DLP does not provide database maintenance; therefore it is recommended that users running large databases perform independent maintenance tasks.

How does the installer set up each database?

The reporting databases should allow TCP and trusted-mode connections from the Forcepoint management server, Email Log Server, and Web Log Server, as well as from any email-capable appliance.

Web Log Database

By default, the web protection Log Database includes one catalog database, one standard logging partition database, and one threats (AMT) partition database. Typically, multiple standard logging partition databases are created as Internet activity is recorded.

- The catalog database provides a single connection point for the various components that need to access the Log Database: Log Server, presentation reports, and investigative reports configuration. It also contains definitions for the following:
  - Category names
  - Risk classes
  - Users
  - User-to-group mapping
  - Database job information

  The catalog database also maintains a list of all the database partitions.
- Standard logging partitions store the individual log records of Internet activity. New partitions are created based on size (3 or 5 GB, by default) or date interval.
- The threats (AMT) partition stores information about requests that have been assigned a severity level, and is used to populate the Threats dashboard.

Email Log Database

The Email Log Database includes one catalog database and (initially) a standard logging partition.

- The catalog database provides a single connection point for the various components that need to access the Log Database: Log Server, the Forcepoint Email Security quarantine service, and the Email Security module of the Forcepoint Security Manager (presentation reports, dashboard, log database configuration page). It also includes definitions for the following:
  - Forcepoint Email Security actions
  - Mail direction
  - Message type
  - DLP severity level
  - Email appliance mapping
  - Email policies
  - Rules
  - Viruses
  - DLP policy names
  - IP addresses
  - Email addresses
  - Domains
  - Database jobs

  The catalog database also maintains a list of all the database partitions.
- Database partitions store the individual log records, including connection log, message log, policy log, delivery log, status log, and hybrid service status log. New partitions are created based on size (5 GB, by default) or date interval.

How big should the database partitions be?

For Web, see *Partitioning*.

For Email, see *Partitioning*.

For Data, see *Rate of network and endpoint incidents*.

Related concepts
Partitioning on page 23
Partitioning on page 24

Related reference
Rate of network and endpoint incidents on page 20

How many partitions can be accessed at the same time?

Forcepoint DLP maintains incident partitions independently of the database engine, based on quarters (3-month periods). By default, SQL Server Express maintains 8 partitions that are online simultaneously, and other SQL Server editions maintain 12 partitions online. You can choose to move any number of partitions online simultaneously as long as your disk space and SQL Server database permit it.

With web and email security solutions, you can access all enabled partitions.

How do I configure partition rollover?

With web and email security solutions, partition rollover can occur automatically when partitions reach a specified size or (SQL Server Standard or Enterprise) date.

- When partition rollover is based on size, all log records are inserted into the most recent active partition that satisfies the size rule. When the partition reaches the designated maximum size, a new partition is created.
- When partition rollover is based on date, new partitions are created according to the established cycle. For example, if the rollover option is monthly, a new partition is created as soon as any records are received for the new month. Log records are inserted into the appropriate partition based on date.

Partition rollover can also be initiated manually.

For information about configuring automatic or manual rollover, see:

- “Configuring database partition options” in the Administrator Help for your web protection solution.
- “Configuring Log Database options” in the Forcepoint Email Security Administrator Help .

For Data solutions, partition rollover is configured on the **Data** > **Settings** > **General** > **Archive Partitions** page in the Forcepoint Security Manager. Here, you configure when to create an archive partition and when to restore it. For instructions, refer to “Archiving incident partitions” in the Forcepoint DLP Help.

What if I need more partitions to run reports?

For web and email security solutions, the available Log Database partitions, both enabled and disabled, are listed on the **Settings** > **Reporting** > **Log Database** page in the respective Web Security and Email Security modules of the Forcepoint Security Manager. To include data from a disabled partition, first enable it, then run the report. You can use this page to disable the partition again once you have retrieved the desired data.

For Forcepoint DLP, when you want to run a report and some or all of the data you want is stored in an offline partition, you must bring that partition online, or the generated report will not contain all the data you need.

Do the reporting databases use named instances?

If you are using SQL Server Standard or Enterprise to host your Forcepoint reporting databases:

- The web protection Log Database can be hosted by one instance, and the Email Log Database and Data Incident and Configuration Database can be hosted elsewhere (on another server or instance).
- All 3 reporting databases can be hosted by the same SQL Server instance.

Can reporting databases be hosted in a SQL Server cluster?

SQL Server clustering may be used with all supported standard and enterprise versions of Microsoft SQL Server for failover or high availability (Forcepoint Email Security and Forcepoint Web Security only).

If your organization uses a SQL Server cluster to provide failover for your database servers, the Forcepoint reporting databases can be hosted by the cluster if:

- A virtual IP address is assigned to the cluster.
- The data managed by SQL Server is housed on a reliable shared disk array.

When you install reporting components in a network that uses a SQL Server cluster, it is imperative that the cluster's **virtual IP address** is used to configure the reporting database connection. When this is done, reporting data is sent to SQL Server via the virtual IP address.

If you configure reporting components (like Web and Email Log Server) to use the IP address of an individual node in the cluster, they cannot take advantage of the failover protection of the cluster.

- If the configured node becomes unavailable, reporting components cannot process data into the reporting database.
- For web and email security solutions, log cache files continue to be saved on the Log Server machine. These files can build up quickly and fill the disk, causing additional problems.

When failover occurs, reporting components must wait briefly while the secondary SQL Server is made primary. When SQL Server begins accepting data over the virtual IP address again, reporting data is once again sent successfully.

This pause in recording data occurs both when failover occurs in a SQL Server cluster and when a standalone SQL Server installation fails and is later brought back online. Any records that were **actively being processed** into the reporting database when the primary SQL Server fails are lost.

- For Web and Email solutions, if you are using BCP log insertion, the loss is generally a full batch (log cache file) of filtering records.
- With ODBC log insertion, fewer records may be lost.

Published 10 April 2025