
Micro Focus Security ArcSight Logger

Software Version: 7.0.1

Configuration and Tuning Best Practices

Document Release Date: February, 2020

Software Release Date: February, 2020



Legal Notices

Micro Focus
The Lawn
22-30 Old Bath Road
Newbury, Berkshire RG14 1QN
UK

<https://www.microfocus.com>

Copyright Notice

© Copyright 2020 Micro Focus or one of its affiliates

Confidential computer software. Valid license from Micro Focus required for possession, use or copying. The information contained herein is subject to change without notice.

The only warranties for Micro Focus 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. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein.

No portion of this product's documentation may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or information storage and retrieval systems, for any purpose other than the purchaser's internal use, without the express written permission of Micro Focus.

Notwithstanding anything to the contrary in your license agreement for Micro Focus ArcSight software, you may reverse engineer and modify certain open source components of the software in accordance with the license terms for those particular components. See below for the applicable terms.

US Governmental Rights. For purposes of your license to Micro Focus ArcSight software, "commercial computer software" is defined at FAR 2.101. If acquired by or on behalf of a civilian agency, the US Government acquires this commercial computer software and/or commercial computer software documentation and other technical data subject to the terms of the Agreement as specified in 48 CFR, 12.212 (Computer Software) and 12.211 (Technical Data) of the Federal Acquisition Regulation ("FAR") and its successors. If acquired by or on behalf of any agency within the Department of Defense ("DOD"), the US Government acquires this commercial computer software and/or commercial computer software documentation subject to the terms of the Agreement as specified in 48 CFR, 227.7202-3 of the DOD FAR Supplement ("DFARS") and its successors. This US Government Rights Section 18.11 is in lieu of, and supersedes, any other FAR, DFARS, or other clause or provision that addresses government rights in computer software or technical data.

Trademark Notices

Adobe™ is a trademark of Adobe Systems Incorporated.

Microsoft® and Windows® are US registered trademarks of Microsoft Corporation.

UNIX® is a registered trademark of The Open Group.

Documentation Updates

The title page of this document contains the following identifying information:

- Software Version number
- Document Release Date, which changes each time the document is updated
- Software Release Date, which indicates the release date of this version of the software

To check for recent updates or to verify that you are using the most recent edition of a document, go to:

[ArcSight Product Documentation on the Micro Focus Security Community](#)

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://communitysoftwaregrp.com/t5/ArcSight-Product-Documentation/ct-p/productdocs

Contents

- Introduction 7
 - About Logger 7

- Chapter 1: Input and Output Components 8
 - Web Connections 8
 - Connectors 9
 - Receivers 9
 - Devices 10
 - Device Groups 10
 - Forwarders 10
 - Improving Forwarder Performance 11

- Chapter 2: Storage Components 13
 - Storage Volume 13
 - Storage Groups 13
 - Storage Group Space 14
 - Storage Rules 14

- Chapter 3: Notifications 15
 - Real-time Alerts 15
 - Saved Searches and Alerts 16
 - Time Settings and Scheduled Tasks 17

- Chapter 4: Event Archives 18
 - Restoring Archives 19

- Chapter 5: Disk Space and Database Fragmentation 20
 - Removing Old Files to Restore Disk Space 21
 - hprof Files 21
 - Saved Searches 21
 - Reports 21

Chapter 6: Search	22
Exporting Search Results	22
Indexing	22
Using the Lookup Search Operator	22
Deleting Lookup Files	23
Replacing Existing Lookup Files	23
Maximum Number of Lookup Entries	23
Lookup Search Performance	23
Improving Search Performance	24
System Configuration and Data Organization	24
High Event Input	25
Size and Distribution of Search Data	25
Search Time Frame	26
Number of Events that Match the Search	26
Regular Expressions Within Queries	26
Complexity of the Search Query	27
Boolean Operators Within the Search	27
Indexed Fields within the Search	27
Super-Indexed Fields within the Query	28
Concurrent Searches, Reports, and Forwarders	29
The Size and Type of Events	29
Logger Options that Affect Search	29
Other Factors that Can Affect Search Speed	29
Chapter 7: Peer Loggers	30
Authentication	30
Using the CEF Search Operator	30
Improving Peer Search Performance	31
Chapter 8: Reports	33
Improving Report Performance	33
Report Timeout Settings	36
Improving Performance of Distributed Reports	36
iPackager Report Backup	37
Chapter 9: System Administration	38
Authentication	38

Network Interface Cards (NICs)	38
User Groups and Search Group Filters	39
System Health	39
License	39
Chapter 10: Web Services	40
Using Special Characters in Regex Queries	40
Chapter 11: Logger on Logger	41
Calculating Logger Raw Events Size and Compression	41
Average Raw Event Size for Licensing	41
Send Documentation Feedback	44

Introduction

This guide provides some best practices obtained from existing customers, field engineers, and ArcSight development and QA groups. It identifies and describes Logger configuration components that can influence its performance and provides recommendations for obtaining optimal performance on a Logger system.

The information in this guide applies to all Micro Focus ArcSight (Logger) 7.0.1 appliance and software models, except where specifically noted.

Additional guidelines and instructions are included in the applicable sections of the Logger 7.0.1 Administrator's Guide, available from the [Micro Focus Security Community](#).

About Logger

Logger is a log management solution that is optimized for extremely high event throughput, efficient long-term storage, and rapid data analysis. Logger receives and stores events; supports search, retrieval, and reporting; and can optionally forward selected events.

Logger is built for fast event insertion and forwarding, and high performance search and analysis. However, when these activities occur simultaneously, Logger components compete for resources and can affect Logger's performance. Other factors that affect Logger performance include the network environment, the complexity of the functions you are performing, the Logger type, and how you have Logger configured.

Many factors can affect Logger's search speed and scan rate, as well. Factors include, among other things, the size of the data set to be searched, the complexity of the query, and whether the search is distributed across peers.

When deploying and configuring Logger or troubleshooting it to achieve optimum performance, follow the guidelines discussed in this guide. If you need additional guidance, contact Micro Focus ArcSight Customer Support.

Chapter 1: Input and Output Components

The following sections discuss factors to consider and provide guidelines for configuring Logger input and output components.

- [Web Connections](#) 8
- [Connectors](#) 9
- [Receivers](#) 9
- [Devices](#)10
- [Device Groups](#)10
- [Forwarders](#)10

Web Connections

Prior to 7.0 version, Loggers were limited to only 250 connections. Now, Logger supports up to 1000 simultaneous HTTPS connections. These connections can be from Web browsers connecting to the Logger Web UI, from connectors to SmartMessage receivers configured on the Logger, from API clients, and from peer Loggers.

To review the connections (including peer searches and SmartConnectors), run the following Linux command:

```
netstat -atlnp | grep <port>
```

Tip: <port>: Is the port that connects to the UI.

The established connections are the ones we are interested in. To get a count of the established connections to your Logger, run the following Linux command:

```
netstat -ntlap | grep 443 | grep httpd | grep ESTABLISHED | wc -l
```

Apache supports up to 1000 simultaneous connections by default (as defined by the MaxClients value in the httpd.conf file). Logger's HTTPS connections are a subset of these. To determine the number of Apache processes currently running on your system, run the following Linux command:

```
ps aux |grep httpd
```

Connectors

Connectors can send events to Logger varying greatly in their peak throughput. Simple connectors sending smaller events will have a higher throughput than more complex connectors sending larger events.

Logger supports more than 4000 simultaneous HTTPS connections. If you have a large number of connectors connecting individually to SmartMessage receivers on Logger, consider aggregating connectors. Refer to the *SmartConnector User's Guide* for more information.

To show all connected SmartConnectors:

Run a search like the following to list the SmartConnectors connecting to Logger.

```
...| top agentHostName
```

Receivers

There is no limitation on the number or type of receivers, or its maximum throughput. However, adding more than 40 to 50 receivers may affect performance. A high incoming event rate and large event size can affect the performance of a receiver. The recommended maximum total events per second (EPS) incoming rate is 15K. The connectors that send events to the Logger may have limits on their throughput.

To monitor the event flow on each receiver through alerts and for better granularity in searches, use an individual receiver for each connector sending events to Logger. Do not reuse deleted receiver names. Logger retrieves old device information when conducting a device search.

As with other considerations related to scope, questions about how to configure receivers and how many connectors should send data to any given receiver, are best answered when taking the entire environment into consideration (data type, usage requirements, and so on). The Micro Focus ArcSight Professional Services team can do full scoping of such scenarios. Contact your local Micro Focus ArcSight Sales representative for more details.

To show all connected Receivers:

Run the following Linux command to list all the Receivers connecting to Logger.

```
sort logger_receiver.properties | grep enabled
```

Enabled Receivers are listed as "True."

Devices

A device is a named event source, comprising of an IP address or hostname of the event sender and the name of the receiver that receives the event. Therefore, a host or connector that sends events to two different receivers on the same Logger is recognized as two different devices.

Device Groups

Device groups classify events received from various devices. For example, device A and device B events could be stored in Device Group AB and device C events could be stored in Device Group C. There is no limit on the number of device groups on a Logger.

You can write storage rules that direct events from specific device groups to storage groups. Also, you can include device groups in queries to limit the data set that Logger must scan, thus resulting in faster searches.

Forwarders

Forwarders send events received on Logger to specific destinations such as ESM, other connectors, or other Loggers. Logger uses its onboard connector when forwarding events to ESM. You can forward all events, use out-of-box filters, or write queries to forward only specific events. You can forward events continuously in real-time or only forward events for a specified time range.

The rate at which a forwarder send events depends on a number of factors including the number of forwarders, the size of the events, and the complexity of the query used to filter the events. Larger events and more complex queries can lower the events per second (EPS) out rate.

A Logger without filters can forward from 10K to 16K EPS, depending upon the forwarder type. TCP Forwarders have higher EPS rates than Forwarders using UDP or the onboard connector. Forwarding CEF events from Logger to a Syslog destination provides better throughput than forwarding to other destinations.

When filtering events for forwarding, Logger must evaluate each event against the query to determine whether to forward it to the destination. This slows down the forwarding rate. The more complex the query is, the slower the forwarding rate. (A complex query typically includes a number of Boolean expressions, such as a regular expression with multiple OR operators.)

Improving Forwarder Performance

The following guidelines can help optimize the forwarding rate.

When configuring forwarders:

- Reduce EPS in rate (into the receivers) through filtering and aggregation.
- Increase the cache for the forwarder to 10G. This will help prevent dropped events if the destination is down.
- Increase EPS throughput when adding additional forwarders. Adding a second forwarder could increase the forwarding rate by 20-30%.

Caution: Do not add more than five forwarders as it may reduce the Logger's performance. Instead, add another Logger to distribute the forwarding load. Forwarders have to compete for the same Logger resources and onboard connector in high EPS situations or where other resource-intensive features are running in parallel (alerts, reports, and several search operations) with a complex forwarding filter.

- Ensure that the forwarder's destination can keep up with the forwarded events. Otherwise, add another forwarding destination.
- Adding a second logical ESM destination to increase the outbound EPS limit may affect Logger's performance.
- To increase the forwarding rate to an ESM destination, create a secondary ESM destination with a secondary forwarder.
- To increase the outbound EPS limit for forwarders with filters, move the filtering operation from the Logger forwarder to the source connectors and devices. Doing so removes the need to filter events, and you can then forward all events.
- Avoid forwarding events across a wide-area network (WAN).

When forwarding to ESM:

- Disable event aggregation (from the ArcSight Manager).
- Make sure the "preserve raw events" is turned off for the connector (Logger's Forwarder connector in ESM). This is also set at the ESM destination.
- Disable real-time alerts on Logger and use rules/alerts within ESM instead.
- Disable basic aggregation for Logger's forwarding connector because it is resource intensive. Basic aggregation is set in the ArcSight Console.
- Disable DNS lookup on the Forwarder connector in ESM.
- Use one forwarder and apply a filter-out filter on the connector resource in ESM to exclude data that you do not want to forward.

- While adding additional forwarders can increase EPS throughout to ESM, configure only one ESM destination for each ESM server. Each additional ESM destination shares memory with all configured ESM destinations, which can cause contention and potential connector failure if oversubscribed. As a workaround, you can increase the Logger on-board connector from 256MB to 512MB from the ESM console, or logically separate the events into several active channels once they arrive to the ESM.
- When separating incoming filtered events on ESM, use Active Channel filters instead of creating multiple Active Channels on multiple incoming Logger connectors.
- Separate the events from the source connectors into two streams, each one of them going to a dedicated receiver. Use one stream for the events that need to be forwarded to ESM and the other stream for the events that do not need to be forwarded. Then, define a filter condition on the device or device group receiving the events from the first stream. Doing so enables you to configure an efficient filter condition.
- To forward events from Logger to ESM, use a Syslog connector to send events to Logger. If a different method such as Netcat is used, the events are forwarded to Logger but not to ESM.

When writing forwarder queries:

- Follow the guidelines for query performance and scan rate provided in ["Improving Search Performance" on page 24](#) when writing queries for forwarders.
- When filtering events in Regular Expression queries, use the metadata query terms **_storageGroup** and **_deviceGroup**. Including storage groups in searches is more efficient than including device groups.
- Make queries as simple as possible. Simplify regex filters at the forwarders.
- Use Unified Query forwarders as much as possible. In most cases, Unified Query based forwarding is faster than Regular Expression based forwarding. Convert Regex filters to Unified filters where possible.
- If you want to forward all events that have the same data at the beginning or end of an event, anchor the regular expression in the forwarding filter for efficient filtering. For example, if you want to forward all events that start with "CEF," use "^CEF" in the regular expression instead of "CEF" because "^CEF" will match the first three characters of the event, and if a match is found, the event will be forwarded. If you use "CEF" in the query, Logger will scan the entire event for the string "CEF."

Chapter 2: Storage Components

The following sections provide guidelines for configuring Logger storage components.

- [Storage Volume](#)13
- [Storage Groups](#)13
- [Storage Rules](#)14

Storage Volume

Storage volume defines Logger's primary storage space. Although you can increase the size of an initially defined storage volume, follow these guidelines for optimal use of available storage space and expected performance.

Micro Focus ArcSight recommends to use NFS for archive storage. Using NFS as primary storage may cause sub-optimal performance and reliability.

- You can increase the size of a Storage Volume, but you cannot decrease it. Each Logger model has a maximum allowed Storage Volume size.
- On a SAN Logger appliance, make sure that you allocate the maximum size logical unit number (LUN) during initial Logger setup. Logger cannot detect a resized LUN. Therefore, if you change the LUN size after it has been mounted on a Logger, Logger may not recognize the new size.

Storage Groups

Storage groups enable you to implement different retention policies. Therefore, data stored in one storage group can be held for longer or shorter time than another group.

Note: The names of the Internal Storage Group and Default Storage Group cannot be modified. User-created storage groups can be renamed if necessary.

In many cases, storage group retention policies are dictated by compliance requirements, such as PCI. However, such requirements might not be met if the storage groups fill up, because the oldest events could be purged automatically to make room for incoming events, even if they are still within the retention period. Even though you set the Default Storage group to 365 days retention, you should not simply assume that all the data will still be there on day 365 in a growing environment. As your environment grows, it is important re-scope your requirements for receivers, forwarders, retention policies, number of Loggers, and so on, accordingly. The Micro Focus ArcSight Professional Services team can do full scoping of such scenarios. Contact your sales representative for more details.

On the other hand, set the retention policy period for a date beyond the maximum live data age.

Storage Group Space

To ensure better control of storage group retention and disk space utilization, do not allow your storage group utilization to increase above 90%. As storage groups near 99% utilization, they start running out of disk space, which reduces the performance of searches due to increasing fragmentation.

Tip: Configure alerts to notify the appropriate users when the Storage Group usage gets too high and defragment the database at those times. Additionally review your archive setup and retention policy, and confirm that it is set up correctly. For more information, see ["Notifications" on page 15](#), ["Disk Space and Database Fragmentation" on page 20](#), and the Logger Administrator's Guide.

Storage Rules

Storage rules direct events from specified device groups to specific storage groups. Use storage rules to direct events to the correct storage group. You could set up up to 40 storage rules to store events in storage groups with different retention periods.

A storage group will no longer save events after it has reached the maximum capacity. Despite the over limit, the storage rules will continue sending events to the storage group. To check the allocated and used space, go to **Configuration > Storage Groups**.

Chapter 3: Notifications

You can set up alerts to be triggered by specified events or event patterns and, optionally, to send notifications to previously configured destinations such as email addresses or SNMP servers. Logger provides two types of alerts, real-time alerts and saved search alerts.

Tip: . Your system uses the Simple Mail Transfer Protocol (SMTP) and you can choose to set authentication and TLS to send email notifications such as alerts and password reset emails.

The following sections discuss factors to consider and provide guidelines for configuring alerts on Logger.

- [Real-time Alerts](#)15
- [Saved Searches and Alerts](#)16

Real-time Alerts

Alerts are triggered in real time. That is, when a specified number of matches occur within the specified threshold, an alert is immediately generated. Although any number of real-time alerts can be defined, a maximum of 25 real-time alerts can be enabled at one time. To enable an additional alert, you will need to disable a currently enabled alert.

Note: If you have the maximum number of alerts enabled, and the receiver EPS is higher than 30k, you may see some slow-down in the receiver EPS to prevent slower search times.

You can use preconfigured filters to specify event patterns when creating alerts.

Tip: Save a copy of a preconfigured filter and edit the copy to meet your business needs (or just write your own.) Refer to the Logger Administrator’s Guide for more information.

The particular filters available depend on your Logger version and model, but may include:

- System Alert - Disk Space Below 10% (CEF format)
- System Alert - Root Partition Free Space Below 10% (CEF format)
- System Alert - Storage Group Usage Above 90% (CEF format)

Use the system filters for real-time alerts to quickly find and handle system or hardware issues. Create saved-search alerts for other things, such as log source alerts.

Real-time alerts can affect system performance, especially if many other resource-intensive features are running on Logger in parallel.

Saved Searches and Alerts

Scheduled Search/Alerts (saved search alerts) are triggered at scheduled intervals. That is, when a specified number of matches occur within the specified threshold, an alert is triggered at the next scheduled time interval. For example, if a saved search alert is set to trigger every hour when five matches occur in sixty seconds and if five matches occur between 12:05 PM to 12:06 PM, the alert will be triggered at 1:00 PM. Refer to the *Logger Administrator's Guide* for more information on alert triggers and notifications.

Although you can define any number of saved search alerts, a maximum of 50 can run concurrently. Contact support if you need to change this number.

To ensure system performance, a maximum of 200 alerts are allowed per saved search alert job. Therefore, if a saved search alert job triggers more than 200 alerts, only the first 200 alerts are sent out for that job iteration; the rest are not sent. Additionally, the job is aborted so it does not trigger more alerts for that iteration and the status for that job is marked "Failed" in the **Finished Tasks** tab (**Configuration | Scheduled Tasks > Finished Tasks**). The job runs as scheduled at the next scheduled interval and alerts are sent out until the maximum limit is reached.

When creating or editing a Scheduled Search or Alert, be sure to set the number of days after which to delete the file, so that old Saved Search files do not accumulate. This will help to conserve disk space. For more information, see ["Disk Space and Database Fragmentation" on page 20](#).

Time Settings and Scheduled Tasks

Precise time stamping of events is critical for accurate and reliable log management. The times displayed for Logger operations such as searches, reports, and scheduled jobs are in the Logger's local time zone.

Follow these guidelines to ensure accuracy of time and optimal scheduled task handling:

- Use an NTP server instead of manually configuring time and date on your system.
- If your Logger is in a time zone that observes daylight saving time (DST), avoid scheduling tasks to run during the hour that is lost or gained at the start and end of DST. Scheduled operations such as reports, event archives, and file transfers are affected when system time is adjusted on the Logger at the start and end of the DST:
 - Operations scheduled for the hour lost at the start of DST (in early Spring) will not run on the day of time adjustment.
 - Operations scheduled for the hour gained at the end of the DST (in late Fall) will run at Standard time instead of the DST time.

To avoid confusing results when the system time zone is set to **/US/Pacific-New**, set the system time zone to a specific region, such as **/America/Los_Angeles**.

Chapter 4: Event Archives

Event archives enable you to save the events for any day in the past, not including the current day.

The primary function of event archives is to allow for long-term storage of events that are not stored locally on Logger (outside any storage group retention policy). An added advantage is that in the event of total data loss, such as in the case of appliance failure, any data that is archived will still be accessible over NFS/CIFS. This does not, however, fulfill the requirements for full disaster recovery because event archives do not contain indexes and therefore, searches and reports, run on archive data, will run much more slowly, possibly timing out. Refer to the Logger Administrator's Guide for information on how to increase the client time out and the database connection timeout.

For full disaster recovery planning, consider having multiple Loggers in a High Availability setup with events being dual-fed from the connectors. The Micro Focus ArcSight Professional Services team can do a full scoping of such scenarios. Contact your local sales representative for more details.

If moved from their original location, archives from earlier versions cannot be loaded on to the Logger. If you need to delete the archives, use the Logger user interface to do so (**System Admin > Storage > Remote File Systems**). Any attempt to load or delete an old archive will look for the original remote archive location. If this was deleted it will need to be added back again with the same name, even if the archive itself was moved to another server.

Follow these guidelines for optimal performance when archiving events:

- Archive during off-peak hours.
- Only one manual archive job can run at a time. However, a scheduled archiving operation can run in parallel with a manual job.
- For a manual archive operation, do not archive too many days or storage groups worth of data at a time. If you have a large data set to archive, archive in smaller chunks to prevent a negative impact on Logger's performance.
- Combine daily scheduled event archiving with daily scheduled configuration backups. Without a daily configuration backup, event archives from the previous day will not be usable in the event of total data loss. In such scenarios, a restore of a previous configuration backup from an earlier week/month will only allow access to event archives up until that point.
- A location of the Remote Path must be used for one system only. If multiple systems mount the same remote path and write to it, the location data will be corrupted or deleted.

Restoring Archives

Events are not copied back to local storage when event archives are loaded. Instead, a pointer to the archive is activated and it is included in queries.

When loading event archives that have been archived offline, but still have not been affected by the system's retention policy, Logger searches against the loaded archive instead of the same data that is local to Logger resulting in a much slower search.

Note: Even though an archive has been created, you cannot load an archive for data that is still in current storage. Loading the archive will fail if that data has not already passed its retention date and been aged out of current storage.

While there is no limit to how many archives can be loaded, the size of the metadata table increases causing slower queries. If you load a large number of archives, searches on the regular data may be slower. How much slower, depends on how much data is in the archives and also on how much regular indexed data is in the system.

Tip: If you have a lot of archive material to restore, a freshly-installed Logger that has had a Configuration Backup applied may provide the fastest restoration. Remember to attach the same archive mount names to the new Logger.

Chapter 5: Disk Space and Database Fragmentation

Sufficient disk space on Logger is important for all functionality to work correctly. It is important to ensure that at least 50% of the root disk (/) is free for usage by the system as and when needed.

Note: Do not confuse disk space usage under the root disk (/) with usage under **/opt/data** where events are stored. The area under **/opt/data** is always 100% full when pre-allocation is configured during initialization.

As the Logger database expands, more indexing is required and there are more events to scan. This can result in decreased search speed. To help maintain and improve search speed as your database grows, defragment the database annually. You should also run a defragmentation if you observe a slow-down or if you see a message in the UI or in the postgres log that recommends doing so.

Tip: You can configure alerts to notify the appropriate users when the free space gets too low, and defragment the database at those times. See ["Notifications" on page 15](#) and the [Logger Administrator's Guide](#) for more information.

Removing Old Files to Restore Disk Space

hprof Files

On Software Logger and Logger Appliances that have SSH enabled, you can get back some disk space by removing old **hprof** files. You can find and remove the **xxx_yyy.hprof** files from the **/current/arcsight/logger/** directory.

Saved Searches

You can delete custom saved searches as well as old instances of the search output that have accumulated over time. You can delete published instances of a saved search or alert from the **Configuration | Search > Saved Search Files** page. You can delete an unnecessary saved search or alert itself from the **Configuration | Search > Scheduled Searches/Alerts** page.

Reports

You can delete custom reports as well as old instances of the report output that have accumulated over time. Please be certain that you want to remove these old reports, and do so carefully. You can delete published instances from the **List Published Outputs** page, accessed by right-clicking the report in the Report Explorer. You can delete an unnecessary report itself by using the right-click menu in the Report Explorer.

Chapter 6: Search

The following sections discuss factors to consider and provide guidelines for searching and search performance.

- [Exporting Search Results](#) 22
- [Indexing](#) 22
- [Using the Lookup Search Operator](#) 22
- [Improving Search Performance](#) 24

Exporting Search Results

By default, Logger retrieves and export up to 1 million matching results when conducting an active search. To modify the amount of search results up to 10 million, go to **Configuration > Search Options**. Make sure to modify the **Max hits of Search UI** field for the **Classic Search** page and **Max hits of Search API** for **Search** page and **API**. Logger must be restarted/ rebooted after this action.

The performance of an export operation depends on the size of the search data set results. When a very large set of search results is exported, you may observe sub-optimal export performance.

Indexing

To avoid performance degradation in certain scenarios, it is recommended to index only the fields necessary for your environment (queries). Once a field has been indexed, it cannot be removed it.

To check the indexed fields in your system, open the **Configuration > Search > Default Fields** page and look for the fields with a check mark  in the **Indexed** column.

Tip: Allow time between adding a field to the index and using it in the search query. If Logger is in the process of indexing a field and you use that field to run a search query, the search performance for that operation will be slower than expected.

Using the Lookup Search Operator

The lookup search operator enables you to increase the data from an external file uploaded into Logger. After the valid lookup file has been updated in Logger, you can search for it using the lookup search operator.

Deleting Lookup Files

You can delete a lookup file after the search session using the file is terminated or times out (after 15 minutes of inactivity), clearing the file from cache. Otherwise, Logger will display an error message indicating the lookup file is being used and cannot be deleted. To terminate the search session, run a search without a lookup file in the same search window.

Replacing Existing Lookup Files

Lookup files should be uploaded and deleted only through the Logger UI. If you manually replace a lookup file from the appliance or a computer with the Software Logger installed, some summary information displayed in the lookup file UI page will be out of sync resulting in an improper function of the lookup search.

Maximum Number of Lookup Entries

The maximum number of supported lookup entries (an individual, comma-separated value in the lookup file) is 5,000,000. If a lookup file used in the search has 4 columns and 10 rows, the total number of lookup entries to be loaded into memory is 40 (4x10). The maximum number of rows loaded for lookup depend on the number of columns in the lookup file.

For example, if a lookup file contains 500 columns, the maximum number of rows allowed for lookup will be $5,000,000/500 = 10,000$ rows, and any subsequent rows will not be used. On the other hand, if the table has only four columns, the maximum number rows allowed for lookup will be $5,000,000/4 = 1,250,000$ rows.

Lookup Search Performance

In addition to the number of Lookup entries, Lookup Search performance depends on the same factors as other searches. Follow the advice provided in ["Improving Search Performance" on the next page](#) for fastest results.

Improving Search Performance

Many factors can affect search speed and scan rate. The amount of time a search requires depends on, among other things, the size of the data set to be searched, the complexity of the query, and whether the search is distributed across peers.

Loggers are different from each other. Loggers with the same version, hardware model, platform, and configuration can differ on their data. The load upon each system varies greatly from moment to moment, so there is no single "right" value for query or forwarding speed.

The following guidelines can help optimize search performance.

- [System Configuration and Data Organization](#) 24
- [High Event Input](#)25
- [Size and Distribution of Search Data](#)25
- [Search Time Frame](#)26
- [Number of Events that Match the Search](#) 26
- [Regular Expressions Within Queries](#) 26
- [Boolean Operators Within the Search](#)27
- [Indexed Fields within the Search](#)27
- [Super-Indexed Fields within the Query](#)28
- [Concurrent Searches, Reports, and Forwarders](#)29
- [The Size and Type of Events](#)29
- [Logger Options that Affect Search](#)29
- [Other Factors that Can Affect Search Speed](#) 29

System Configuration and Data Organization

Search performance can be affected by your environment set up and the way your data is organized.

- Ensure network speed is optimal.
- Configure peer Loggers on the same subnet.
- Partition the data to search in chunks rather than all at once.
 - Use peers to distribute the data.
 - Use storage groups to divide the data and then use these storage groups in your search.

For example, if you are searching for events from a source, you can exclude the Internal Event Storage Group where all internally-generated ArcSight events are sent. However, if you have a storage group of only Windows events, you can specify the search for that particular storage group. This will speed the search by eliminating places where Logger looks for events. For more information, see ["The Size and Type of Events" on page 29](#).

High Event Input

When the event input is high, indexing can lag behind. As a result, the search defaults to a slower non-indexed search.

Tip: The Global Summary includes the date and time of the most recently indexed data. To check if your index is up to date, compare the Global Summary date and time with the Logger system time, which is on the same page. The Global summary will display the following format: **There are 22,743 events indexed from 2015/03/12 20:15:01:546 EDT to 2015/03/13 17:05:16:375 EDT.** In system time, the timestamp can be found by hovering the upper right-hand corner of the Logger Summary page. If the two timestamps are nearly equal, then Logger indexing is working.

To avoid this issue, run a fixed-time search that does not include the last two minutes.

- If this is a recurring problem, make sure that your environment is sized correctly. The Micro Focus ArcSight Professional Services team can do full scoping of such scenarios. Contact your local Micro Focus ArcSight Sales representative for more details.
- On the Connector, turn aggregation on to lower the number of duplicate events. This will also lower the EPS rate.
- Use the Search Analyzer tool to determine if the fields used in your query are indexed. See the Logger Administrator's Guide for details.

Size and Distribution of Search Data

Restricting searches to specific storage groups or peers decreases the number of events to search because storage group or peer filter is applied before the query is executed. If there are fewer events to scan, as is usually the case when looking at a single storage group rather than all of them, the result returns more quickly.

Use metadata query terms **_storageGroup** and **_peerLogger** to limit the number of events that must be scanned.

Tip: Including storage groups and peers in search queries is more efficient than including device groups. Use storage groups and peers in the query as much as possible, to reduce the amount of data searched.

- To limit the search to the Logger at **192.0.2.9**, use the following:
_peerLogger IN ["192.0.2.9"]
- To limit the search to the default storage group, use the following:
_storageGroup IN ["Default Storage Group"]

Search Time Frame

Searching against a longer time frame takes longer than searching against a shorter time frame, since there are more events to search. For faster results, limit the search to a shorter time frame.

Search based on event time

Logger recommends to not use the end field as part of the event query. Instead, change the search type to “event occurred time”.

Logger recommends to send old events during the lowest EPS ingestion periods. While sending the old events, use a receiver that is not getting a current time event. Additionally, send the oldest events to a separate storage group.

Number of Events that Match the Search

Searches that result in a high number of matching events will be slower than searches with a low event match since there will be fewer events to load in the memory.

- If the search results returns with large number of matches, modify the search query to make it more specific.

For example:

Instead of: **authentication | where name CONTAINS "failure"**

Use the following:

authentication AND name CONTAINS "failure"

- Use metadata query term **_deviceGroup** to reduce the result set to certain device groups for each search condition.

For example:

To limit the search to the smart device group on the Logger at **192.0.2.9**,

Use the following: **_deviceGroup IN ["192.0.2.9 [smart]"]**

- Write queries that take advantage of superindexes. For information on how to write super-indexed field queries optimally, including examples, refer to the topic “Searching for Rare Field Values” in the Logger Administrator’s Guide.

Regular Expressions Within Queries

Regular expressions do not utilize indexing, so queries containing regular expressions (and search operators that result in regular expression-type parsing, such as REX) can slow down operations. To optimize search speed when using regular expressions in queries, make sure the data set that the regular expression needs to scan is small.

To control the data set:

- Precede the regular expression with a search term that reduces the data set size. For example, to extract the IP address from all events that contain the words "telnet" and "failed", use these words as the full-text search terms to reduce the data set that the following regular expression will need to scan:

```
telnet failed|regex ="(\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3})"
```

- Use metadata such as device group, storage group, and peers instead of Boolean operators to filter events, where possible.

Tip: Including storage groups and peers in search queries is more efficient than including device groups. Use storage groups and peers in the query as much as possible, to reduce the amount of data searched.

- Specify an indexed structured search that reduces the data set size before using a regular expression query term.

Complexity of the Search Query

The search speed can vary depending on the query's complexity. Reduce the complexity of your queries where possible. Use simple operators like **replace** and **rename** and reduce the use of complex operators such as **rex**, **sort**, and **chart**.

Boolean Operators Within the Search

Search speed can also vary depending on the search conditions used. A query that includes **OR** or **AND** operators takes longer to process. The **OR** operator is particularly resource intensive because it requires the regular expression to scan the text of each event multiple times. To determine if this is happening, reduce the number of **OR** and **AND** operators and run the searches again.

Using **AND** and **OR** with the **=** operator can be very powerful when searching super-indexed fields. However, to obtain the greatest search speed improvement, you must use them carefully. For information on how to write super-indexed field queries, including examples, consult the *Logger Administrator's Guide*.

Indexed Fields within the Search

Searches where all fields are indexed are faster than searches with non-indexed fields. Use queries where all fields are indexed as much as possible. A list of the default fields, along with their index status is available on the **Default Fields** tab (**Configuration > Search > Default Fields**).

- To speed up a non-indexed search, combine index field based search or full text search with non-indexed field search.

For instance, some fields cannot be indexed. The query "**<non indexed> CONTAINS "username"**" would slow the search process. Instead, enter the following command, so the query is not slowed by the non-indexed field:

```
name = "TCP_MISS" | where <non indexed item> CONTAINS "username"
```

Even though a search query includes only indexed fields, you might not realize the performance gain you expect in these situations:

- When you perform search on data in a time range in which a currently indexed field (included in the query) was non-indexed, the query will run at the speed you would expect if the field was not indexed. This is because new indexing information is not applied to previously stored events.

For example, you index the "port" field on August 13th at 2:00 PM. You run a search on August 14th at 1:00 PM. to find events that include port 80 and occurred between August 11th and August 12th. The "port" field was not indexed between August 11th and the 12th. As a result, the search defaults to a slower, non-indexed search.

- When a query that includes indexed field is performed on archived events, the query runs slower than when the data was not archived. This occurs because the index data is not archived with events. As a result, the search defaults to a slower non-indexed search.
- When you include a field in your search query that Logger is in the process of indexing, the query will run slowly. This issue is discussed in ["High Event Input" on page 25](#).

Super-Indexed Fields within the Query

To search for uncommon values in IP addresses, host name, and user name fields, use superindexing fields for faster search speeds. Superindexes rule out chunks of data from your search and return search results very quickly when there are few or no hits.

Fields With SuperIndexes

destinationAddress	destinationHostName	destinationPort
destinationUserId	destinationUserName	deviceAddress
deviceEventClassId	deviceHostName	deviceProduct
deviceVendor	sourceAddress	sourceHostName
sourcePort	sourceUserId	sourceUserName

For faster searching on super-indexed fields, follow these guidelines:

- Use only the equal sign operator (=)
- Use the AND operator (**AND**) only with non-super-indexed fields

For more information on how to write super-indexed field queries optimally, including examples, consult the [Logger Administrator's Guide](#).

Concurrent Searches, Reports, and Forwarders

Searches, forwarders, and reports all use the same search engine. Heavy loads on the system (such as a high incoming EPS, forwarding with filtering, and multiple search and report operations going on in parallel) will impact Logger's performance and response time. Each search you run, consumes the Logger storage space and CPU bandwidth.

Spread resource-intensive tasks to off-peak hours as much as possible. Schedule searches and reports to run at a low load time on the system or reduce the load when your searches or reports need to run.

The Size and Type of Events

Searches against small size events, such as syslog (where the event size varies from 1K-1.5K) will be faster than events with larger size such as Blue Coat events (where the event size varies from 2.5K-4K). This behavior will be more noticeable when the search is a non-indexed search.

Logger Options that Affect Search

If the **Discovered Field** and **Summary Field** options are enabled, the system will try to populate these fields during the search, which can slow it. This becomes more noticeable when there is high event match. For faster results, disable these options.

Other options that might affect search speed include:

- **Secondary Delimiter Support**—Turn it off to improve performance (Configuration | Search > **Search Options**)
- **Source type support**—Use specific source types to improve performance.

Other Factors that Can Affect Search Speed

- Logger version
- Appliance and model
- The number of events already in the system
- Ingestion rate (insertion rate)

Chapter 7: Peer Loggers

The following sections discuss factors to consider and provide guidelines for configuring and searching across peers. Refer to the Logger Administrator's Guide for the number of peers supported on your Logger release.

- [Authentication](#)30
- [Using the CEF Search Operator](#) 30
- [Improving Peer Search Performance](#)31

Authentication

For security reasons, Micro Focus ArcSight recommends that you use authorization IDs to establish peer relationships.

- If the remote Logger is configured for SSL Client authentication (CAC), you must configure an authorization ID and code on the initiator Logger.
- If user name and password are used for authenticating to a remote peer Logger, the credentials are only used one time, during the peering relationship set up. After a relationship has been established, the credentials are not saved (on the Peer Loggers page) and the peers do not authenticate periodically. Therefore, if the user name or password used to establish a relationship is changed at a later date or the user name is deleted, peering relationship is not broken. However, if you delete the peering relationship or it breaks for other reasons, you will need to enter the updated credentials to re-establish the relationship.

Using the CEF Search Operator

With Logger versions 5.2 and later, you do not need to explicitly extract the CEF fields and then apply other search operators to those fields. The CEF operator is implicit. You can specify the event fields directly in queries.

For example, to find the top values in the message field, instead of `... | cef message | top message`,

use the following:

`... | top message`

When you run a peer search, initiate queries that do not explicitly use the CEF operator from a Logger running version 5.2 or later. A query that does not use CEF defined fields will run if the query is initiated on a Logger running version 5.2 or later. However, if the query is initiated on a 5.1 or earlier Logger version

(before CEF was deprecated), it will fail. For more information, see ["Improving Search Performance" on page 24](#).

Improving Peer Search Performance

Searches done across peer Loggers are done locally on the peer rather than the Logger initiating the search. The following guidelines can help optimize the performance of peer searches.

When configuring peers:

- Search head is a peered Logger that is strictly used for searching; any other activity must be done by a node. Therefore, set up your architecture so that no data is sent to the search head Loggers. Search heads require a minimum of 16 GB of RAM; however, 32 GB is highly recommended.
- You can configure up to 100 peers for a Logger. A maximum of 10 users can log in to the search head and run searches across 10 nodes at the same time. You can scale this to enable 100 users to run concurrent searches by setting up 10 search heads.
- If you set up a number of peers on a local network to horizontally scale out the system, be sure to configure them identically. For example, they must all have the same storage groups and search options.
- If you added custom schema fields to your Logger schema, those same fields must exist on all peers. Otherwise, a search query containing those fields will return an error when run across peers.
- The time and date of the system on which the software Logger is installed must be set correctly with respect to its time zone to peer with other Loggers. Micro Focus ArcSight recommends that you configure the Logger system to synchronize its time with an NTP server regularly.

When running searches across peers:

- Follow the guidelines for query performance and scan rate provided in ["Improving Search Performance" on page 24](#) when writing queries for searches across peers.
- Peer search speed improvements gained by using search heads apply only to searches run through the user interface. Using search heads does not improve the speed of scheduled searches or searches run through Logger Web Services.
- Ensure that the device and storage groups specified in the query exist on all peers. Peers on which a device or storage group does not exist are skipped.
- Make sure that event fields on ALL peers are indexed for the time range specified in a query. If an event field is indexed on a local Logger but not on its peers for a specific time range, a distributed search will run at optimal speed on the local Logger, however will run slower on the peer Loggers. Therefore, the search performance in such a setup will be slow.
- For peers with different schema, make sure that your searches and reports only involve fields that have the same name and data type on all peers. Otherwise, the search or report will fail.

- When peers of mixed Logger versions are involved in the same search, the search features you can use are determined by capabilities of the peer with the earliest, and therefore most limited, version.
- Using search heads enables faster peer searches that particularly use aggregation operators, such as chart, sort, and top. To improve search results, specify in your query all peers to be searched and exclude Local Logger.

For details of available capabilities, such as available search operators, refer to the release notes of the earliest peer Logger.

Chapter 8: Reports

Reports that must process very large data sets can be resource intensive. Micro Focus ArcSight recommends schedule your reports whenever possible, so that most reports are run during periods of light load. To preserve disk space and accumulate the minimum quantity of old reports, make sure your published report **Expires On** a specific date. For more information, see "[Disk Space and Database Fragmentation](#)" on page 20.

When isolating a user to a specific report or report folder, the report rights must include all associated rights to that report folder within its folder path. For example, if report folder: **Anti-Virus** is the target report folder, you must also include the rights to report folder: **Device Monitoring** Refer to the Logger Administrator's Guide for instructions.

The following sections discuss factors to consider and provide guidelines for reporting.

- [Improving Report Performance](#)33
- [Report Timeout Settings](#) 36
- [Improving Performance of Distributed Reports](#)36
- [iPackager Report Backup](#) 37

Improving Report Performance

The following guidelines can help optimize report performance.

When running reports:

- Run a maximum of 10 scheduled concurrent reports.
- The number of concurrently running reports depends on system configuration, browser limit, and number of rows and column per report. For instance, 5 concurrent Logger Search Smart Reports with 10k rows and 42 columns in multiple tabs are completed successfully on a single browser.
- Logger can render up to 20k rows per SMART report before performance is compromised. A warning message after 10k row is actually displayed to warn about this limitation.
- Before running ad-hoc reports, ensure that published reports and saved searches are kept ONLY for as long as required. For example, running 10 reports that each generate 1 GB files will utilize 10 GB of space that could otherwise be used by the system.

Low disk space in the following directories can be a result of large and potentially unnecessary (or unused) CSV exports and published reports.

- On Appliances: **/root**
- On Software Loggers: **\$ARCSIGHT_HOME** (Software Logger installation path)

Tip: When scheduling published reports, Micro Focus ArcSight recommends that you change the retention period to 1 week after generation. To do this, use the following option on the **Add Report Job** page:

Valid Upto *<N>* *<Unit of time>* After Generation

- Running large reports can take up a lot of space temporarily. Reports can fail if space is limited.

Tip: Configure alerts to notify the appropriate users when free space is scarce. For more information, see ["Notifications" on page 15](#), ["Disk Space and Database Fragmentation" on page 20](#), and the *Logger Administrator's Guide*.

- To run reports with millions of events more efficiently, increase the memory heap size for the report engine. For more information, see *Java Memory Allocation* section in [Logger Administrator's Guide](#) or [Contact Support](#).

Tip: You can increase the memory size based on the memory available in your environment. Take in consideration the memory allocated for other Logger process before proceeding.

- Specify a scan limit for reports run manually. The default value is 100 000. When you specify a scan limit, the latest *N* events are scanned. This results in faster report generation and is beneficial when you want to process only the latest events in the specified time range instead of all the events stored in Logger.
- In addition to the search fields, all fields displayed in the report should be indexed. In addition to the fields in the WHERE clause of the query, the fields in the SELECT clause also need to be indexed. A list of the default fields, along with their index status is available on the **Default Fields** tab (**Configuration | Search > Default Fields**).

When writing report queries:

- Follow the guidelines for query performance and scan rate provided in ["Improving Search Performance" on page 24](#) when writing queries for reports.
- Use the where clause to specify conditions to narrow down your results, for example you could use queries like the following:
where <fieldName>="<value>"
- Select specific fields, avoid patterns that will return too many hits.
 - Use queries like the following:
Select <fieldName1>, <fieldName2> ... from events
 - Avoid queries like the following:
Select * from events.
 - For large reports, add a filter that specifies the fields of interest, such as the following, to the SQL before the sort and order condition:


```
select events.arc_sourceAddress, events.arc_destinationAddress,  
events.arc_destinationPort,  
events.endTime  
from events where events.arc_destinationPort >22 and  
events.arc_categoryOutcome="/Failure" ;
```
- Avoid aggregation operations over large data sets.
- Avoid self-joins.
- Avoid sub-queries such as:
Select * from <tableName> where <fieldName> in (select ...)
- Avoid using **order by** with a large amount of data, as that will take a long time. Limit the number of rows you want to order by.
 - Use queries like the following:
Select ... from <tableName> group by <fieldName> order by <fieldName>
 - Avoid queries like the following:
Select ... from <tableName> order by <fieldName>
- Avoid sorting on entire fields that are very large, because that will use a lot of disk space. Use a substring of a field instead of the full length of the field if the substring is good enough.
In the examples below, we use the **name** field, which is 512 characters long.
 - Use queries like the following:
SELECT * FROM <tableName> order by substr(name,1, 64) LIMIT 50;
 - Avoid queries like the following:
SELECT * FROM <tableName> order by name LIMIT 50;
- Avoid using **group by** unless it is necessary. Change the query to order by a short field instead of a long one. (**group by** also uses the same fields as **order by**.)

- Limit the number of rows to sort. Use queries like the following:

```
Select ... from <tableName> group by <fieldName> order by <fieldName>or  
Select ... from <tableName> where <fieldName>= ... group by <fieldName>  
order by <fieldName>
```
- Avoid queries like the following:

```
Select ... from <tableName> order by <fieldName>
```
- Avoid using **order by** directly on the event table. Use queries like the following:

```
Select ... from <tableName> where <fieldName> =... order by <fieldName>
```

Report Timeout Settings

If your report is timing out, you can increase the **DATABASE_TIMEOUT** and the **HTML_VIEWER_TIMEOUT**. However, increasing the Report Timeout above the default setting of four hours puts additional load on the system, because spacing out of reports over any given day becomes more difficult, particularly since manual reports and searches compete for resources. Refer to the Logger Administrator's Guide for information about time-outs that can affect long-running reports.

Although you can increase the default timeout settings for scheduled reports, Micro Focus ArcSight recommends that you optimize the report query instead.

For example, if you do not need all the events or too many will be returned, you can get a sample by using a scan limit. When you specify a scan limit, the latest **N** events are scanned. The default scan limit is zero, which means all events.

You can also add the following clause to the bottom of your query: **LIMIT N**, where **N** is the number of events.

Improving Performance of Distributed Reports

Distributed reports include matching events from the specified peers of the originating Logger.

Use the following guidelines to help optimize the performance of distributed reports:

- Avoid running a distributed report on a Wide Area Network (WAN) link.
- Avoid running more than three concurrent distributed reports.
- When writing queries for distributed reports, follow the guidelines for query performance and scan rate provided in ["Improving Search Performance" on page 24](#).
- If you are running the report on a very large data set and the performance of the report is not optimal, reduce the size of the data set.
- Use pushed functions in **WHERE** clauses. SQL functions that are pushed to peers include:
 - String functions:

`char_length, char, concat, insert, lcase, left, length, locate, lpad, ltrim, replace, right, rpad, rtrim, strcmp, substr, trim, ucase`

- Numeric functions:

`abs, ceiling, floor, round, sign, truncate`

- Date/time functions:

`cast, dayofmonth, hour, minute, month, second, str_to_date, time_to_sec, unix_timestamp, year`

iPackager Report Backup

Although the iPackager utility is primarily to allow quick distribution of reports to multiple Loggers, it can be used as a report backup tool. You can package all or selected reports and report objects residing on a Logger. This package can be later imported on a different Logger installation at any time.

When writing reports for export to other systems, use the default group, and device-independent syntax, so the system content will not be overwritten and the report will run on other systems after distribution.

Chapter 9: System Administration

The following sections discuss factors to consider and provide guidelines for Logger System Administration.

Note: A Logger Appliance with a failed hard drive will display a warning message. Micro Focus ArcSight strongly recommends that you contact support immediately to get the drive replaced.

- [Authentication](#)38
- [Network Interface Cards \(NICs\)](#)38
- [User Groups and Search Group Filters](#)39
- [System Health](#)39
- [License](#)39

Authentication

If you are using LDAP or RADIUS authentication, Micro Focus ArcSight strongly recommends configuring a backup LDAP/RADIUS server to help ensure uninterrupted access to Logger.

Network Interface Cards (NICs)

Hostname

When setting the IP addresses for the Network Interface Cards (NICs), hostname, and default gateway for your system, make sure that your Domain Name Service (DNS) can resolve the host name you specify to your system's IP address. Performance is significantly affected if the DNS cannot resolve the host name. The **Hostname** in the Certificate Signing Request (CSR) must be the same as the system host name.

Interface

To improve the NIC's performance using more than 1 network interface, enable the automatically route outbound packets. You can also create an alias for each IP address. However, you cannot modify its speed.

Bonding and Trunking

To avoid data loss, Logger must stop receiving events before starting bonding and trunking process. Perform the bonding and trunking only on eno interfaces. If you use a configuration mode different than 0,

1, 4, or 5, you will experience appliance instability or network configuration damages. Redeploy the appliance or system restore will be necessary.

User Groups and Search Group Filters

Implementing Logger users and groups to view only specific events can have performance implications, depending on the filters used to determine the events that the users can see.

- Use indexed search queries (Unified Queries) as much as possible. In most cases, unified query-based searches are faster than regex-based searches.
- Duplicate parameters (storage group and search group) are not supported for unified queries. In a Logger search report, the report category filter will prevail over the search group filter.
- To filter events in regular expression queries, use metadata instead of Boolean operators as much as possible. Including storage groups and peers in searches is more efficient than including device groups. Use storage groups in the query as much as possible, to reduce the amount of data searched.

System Health

To monitor Logger's health and performance, review the system health events by using Simple Network Management Protocol (SNMP) or Logger search. For more information, see ["Notifications" on page 15](#) and the Logger Administrator's Guide.

License

Logger recommends the user to transition to EPS license before GB Standalone is not longer supported. For more information, contact your sales representative.

If a license is installed, make sure to restart your Logger after the license update. Logger cannot longer display GB data in the license usage page if the user selects EPS license after upgrade. Despite the ArcMC management option is enabled in the page, Logger will behave as Standalone license if no ArcMC information is added to the Logger.

Chapter 10: Web Services

The Logger Service Layer exposes Logger functionalities as Web services. By consuming the exposed Web services, you can integrate Logger functionality in your own applications. Using the Web service APIs, you can create programs that execute searches on stored Logger events or run Logger reports, and feed them back to your third-party system.

Using Special Characters in Regex Queries

To run queries such as `|regex "|` (or other special characters) when doing a Logger search, turn on base64 encoding on the Logger side and use base64 decoding on the client side.

To turn on base64 encoding on the Logger side:

Add the following line to the `/userdata/logger/user/logger/logger.properties` file and the `/userdata/logger/user/logger/logger_webservices.properties` file. (Create this file if it does not exist.)

```
api.search.base64encode=true
```

To use base64 decoding on the client side:

Add the base64 decoding as shown in the highlighted location in the `runSearch()` method of the Web service client:

```
Tuple[] tuples = searchService.getNextTuples(...);
    for (Tuple tuple : tuples) {
        String[] arr = tuple.getData();
        for(int j=0; j< header.length; j++){
            arr[j] = new String(Base64.decode(arr[j]));
// <= Add this line to decode the received string using base 64.
        }
    }
```

Chapter 11: Logger on Logger

Logger has several features that you can use to get more information about how Logger is doing and how it is being used.

- [Calculating Logger Raw Events Size and Compression](#)41
- [Average Raw Event Size for Licensing](#)41

Calculating Logger Raw Events Size and Compression

To see usage data on your Logger, log into the Logger UI and open **Configuration > Advanced > License Usage**.)

You can use agent events statistics to determine events size and event count for the data collected from Smart Connectors. You need to collect data for at least 24 hours (or more) to determine the daily data usage and the average raw events.

The compression rate can vary depending on several factors, such as the following:

- The size of events
- The type of events
- The uniqueness of fields in the events
- The EPS rate

Because of these and other variables, it is hard to predict the compression rate. The value differs from Logger to Logger. Based on what we have seen, the average compression rate ranges from 8-10x.

Average Raw Event Size for Licensing

To calculate the daily data for the raw events, you can use the event information generated from the Smart Connector (**deviceEventClassId =agent:050**).

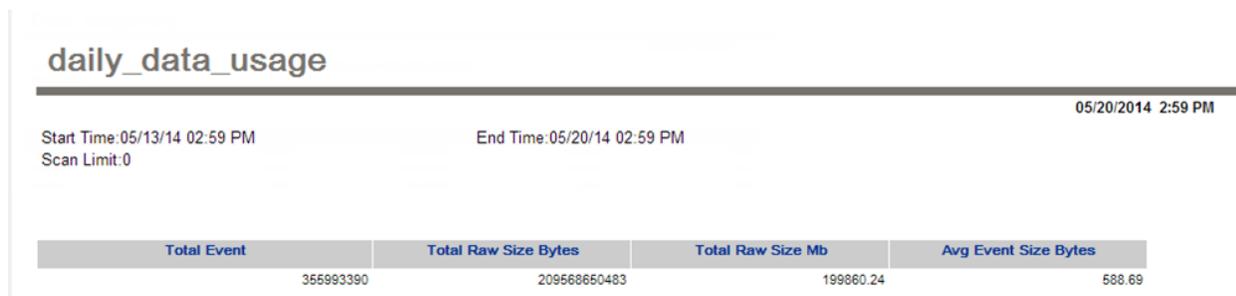
Use the following fields:

- **deviceCustomNumber3**: The number of non-internal events seen by this component since the last internal event.
- **deviceCustomString4**: The number of characters in the raw events in the non-internal events seen by this component since the last internal event.

To calculate the daily data for raw events:

Log into the Logger UI and open **Reports > Design> Queries> Design**. Save the following query. Then create and run a new report with the new query.

```
select sum(events.arc_deviceCustomNumber3) as "Total_event", sum(events.arc_
deviceCustomString4) as "Total_raw_size_bytes",
sum((events.arc_deviceCustomString4)/1048576) as "Total_raw_size_MB",
(sum(events.arc_deviceCustomString4)/sum(events.arc_deviceCustomNumber3)) as
"avg_event_size_bytes"
from events
where events.arc_deviceEventClassId = 'agent:050'
```



This daily_data_usage report covered a 24-hour period and got the average raw event size and total event count and data usage.

To get a list for events per day over time:

Login to the Logger UI and open **Reports > Design> Queries> Design**. Save the following query. Then create and run a new report with the new query:

```
select DATE_FORMAT(events.arc_deviceReceiptTime,"%Y-%m-%e") as "date",
sum(events.arc_deviceCustomNumber3) as "Total_event", sum(events.arc_
deviceCustomString4) as "Total_raw_size_bytes",
sum((events.arc_deviceCustomString4)/1048576) as "Total_raw_size_MB",
(sum(events.arc_deviceCustomString4)/sum(events.arc_deviceCustomNumber3)) as
"avg_event_size_bytes"
from events
where events.arc_deviceEventClassId = 'agent:050'
group by date
```

Data_usage/day

05/21/2014 11:48 AM

Start Time: Tue May 13 00:00:00 PDT 2014
Scan Limit: 0

End Time: Wed May 14 23:59:59 PDT 2014

Date	Total Event	Total Raw Size Bytes	Total Raw Size Mb	Avg Event Size Bytes
2014-05-13	72078051	37055875753	35339.24	514.11
2014-05-14	77129190	38341716456	36565.51	497.11
2014-05-15	471215	291094059	277.61	617.75

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 computer, click the link above and an email window opens with the following information in the subject line:

Feedback on Configuration and Tuning Best Practices (Logger 7.0.1)

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 arcsight_doc@microfocus.com.

We appreciate your feedback!