IBM Operations Analytics Log Analysis

Performance and tuning guide

Version 1.3

IBM Operations Analytics Log Analysis

Performance and tuning guide

Version 1.3

Note
Before using this information and the product it supports, read the information in "Notices," on page 23.

Edition notice

This edition applies to IBM Operations Analytics Log Analysis and to all subsequent releases and modifications until otherwise indicated in new editions.

References in content to IBM products, software, programs, services or associated technologies do not imply that they will be available in all countries in which IBM operates. Content, including any plans contained in content, may change at any time at IBM's sole discretion, based on market opportunities or other factors, and is not intended to be a commitment to future content, including product or feature availability, in any way. Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only. Please refer to the developerWorks terms of use for more information.

Contents

Chapter 1. About this publication

This guide contains information about how to use IBM® Operations Analytics Log Analysis.

Audience

This publication is for users of the IBM Operations Analytics Log Analysis product.

Publications

This section provides information about the IBM Operations Analytics Log Analysis publications. It describes how to access and order publications.

Accessing terminology online

The IBM Terminology Web site consolidates the terminology from IBM product libraries in one convenient location. You can access the Terminology Web site at the following Web address:

http://www.ibm.com/software/globalization/terminology.

Accessibility

Accessibility features help users with a physical disability, such as restricted mobility or limited vision, to use software products successfully. In this release, the IBM Operations Analytics Log Analysis user interface does not meet all accessibility requirements.

Accessibility features

This information center, and its related publications, are accessibility-enabled. To meet this requirement the user documentation in this information center is provided in HTML and PDF format and descriptive text is provided for all documentation images.

Related accessibility information

You can view the publications for IBM Operations Analytics Log Analysis in Adobe Portable Document Format (PDF) using the Adobe Reader.

IBM and accessibility

For more information about the commitment that IBM has to accessibility, see the IBM Human Ability and Accessibility Center. The IBM Human Ability and Accessibility Center is at the following web address: http://www.ibm.com/able (opens in a new browser window or tab)

Tivoli technical training

For Tivoli® technical training information, refer to the following IBM Tivoli Education Web site at http://www.ibm.com/software/tivoli/education.

Providing feedback

We appreciate your comments and ask you to submit your feedback to the IBM Operations Analytics Log Analysis community.

Conventions used in this publication

This publication uses several conventions for special terms and actions, operating system-dependent commands and paths, and margin graphics.

Typeface conventions

This publication uses the following typeface conventions:

Bold

- Lowercase commands and mixed case commands that are otherwise difficult to distinguish from surrounding text
- Interface controls (check boxes, push buttons, radio buttons, spin buttons, fields, folders, icons, list boxes, items inside list boxes, multicolumn lists, containers, menu choices, menu names, tabs, property sheets), labels (such as **Tip:**, and **Operating system considerations**:)
- Keywords and parameters in text

Italic

- Citations (examples: titles of publications, diskettes, and CDs
- Words defined in text (example: a nonswitched line is called a *point-to-point line*)
- Emphasis of words and letters (words as words example: "Use the word *that* to introduce a restrictive clause."; letters as letters example: "The LUN address must start with the letter *L*.")
- New terms in text (except in a definition list): a *view* is a frame in a workspace that contains data.
- Variables and values you must provide: ... where *myname* represents....

`Monospace`

- Examples and code examples
- File names, programming keywords, and other elements that are difficult to distinguish from surrounding text
- Message text and prompts addressed to the user
- Text that the user must type
- Values for arguments or command options

Chapter 2. Introduction

Use this guide to help you to tune Log Analysis and the related servers to help optimize performance.

You can deploy Log Analysis in various scenarios such as for a trial, as a proof of concept or a full production deployment. For more information, see Deployment options.

This documentation assumes that you deployed Log Analysis in a production like scenario. It assumes that you install Log Analysis, the Indexing Engine, and Logstash on separate servers to mimic a typical commodity or production installation. If you install the Indexing Engine or any of the other Log Analysis components on more than one server, you need to tune each server as described in this documentation.

The guide also assumes that you deploy the scalable data architecture. For more information, see Deploying scalable data collection architecture.

The settings in documented here are based on test systems that approximate the conditions of a production deployment. The benefits that are accrued from this tuning can vary depending on the specifics of your environment and installation.

Chapter 3. Hardware considerations and configurations

Before you tune Log Analysis and the related components, you need to size and tune your hardware and operating system.

Sizing hardware

To ensure optimal performance of medium and large-scale deployments of Log Analysis, ensure that you meet the hardware specifications.

If you do not follow the recommendations here and instead use a virtual environment, ensure that the CPU and memory are dedicated and reserved for optimal performance.

For more information about the specifications for Log Analysis, see Hardware and software requirements.

For more information about the hardware requirements for scalable data collection, see Sizing hardware for scalable data collection.

Disabling hyperthreading

Tests show that using hyperthreading is not advantageous to performance and it can hinder optimal performance levels. If you enabled it, you need to disable it.

Before you begin

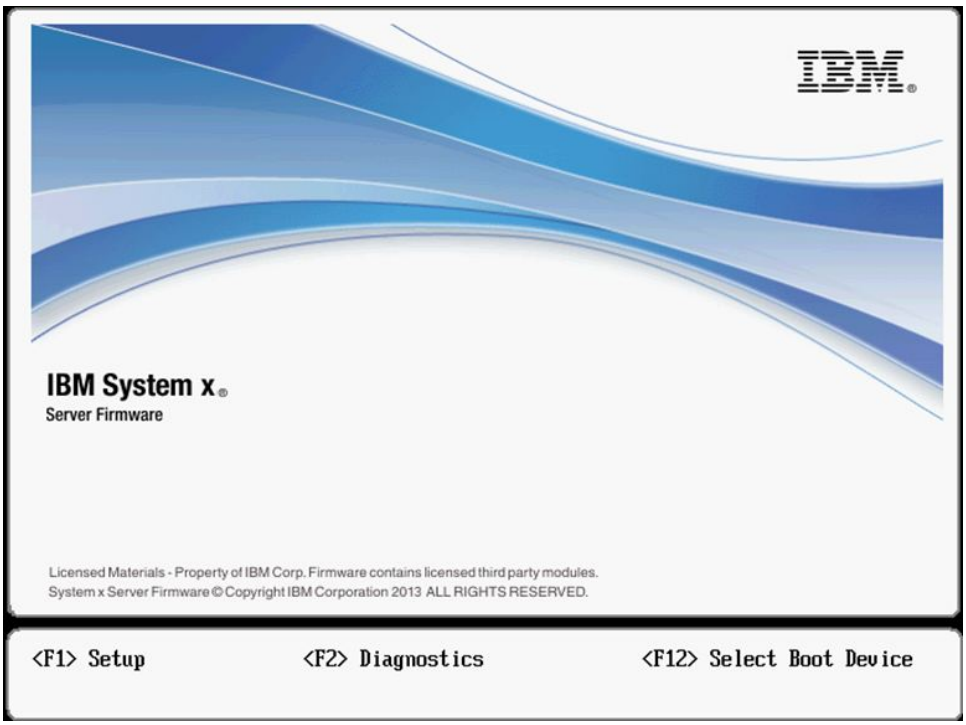
Ensure that you have KVM access for the server that you are configuring.

About this task

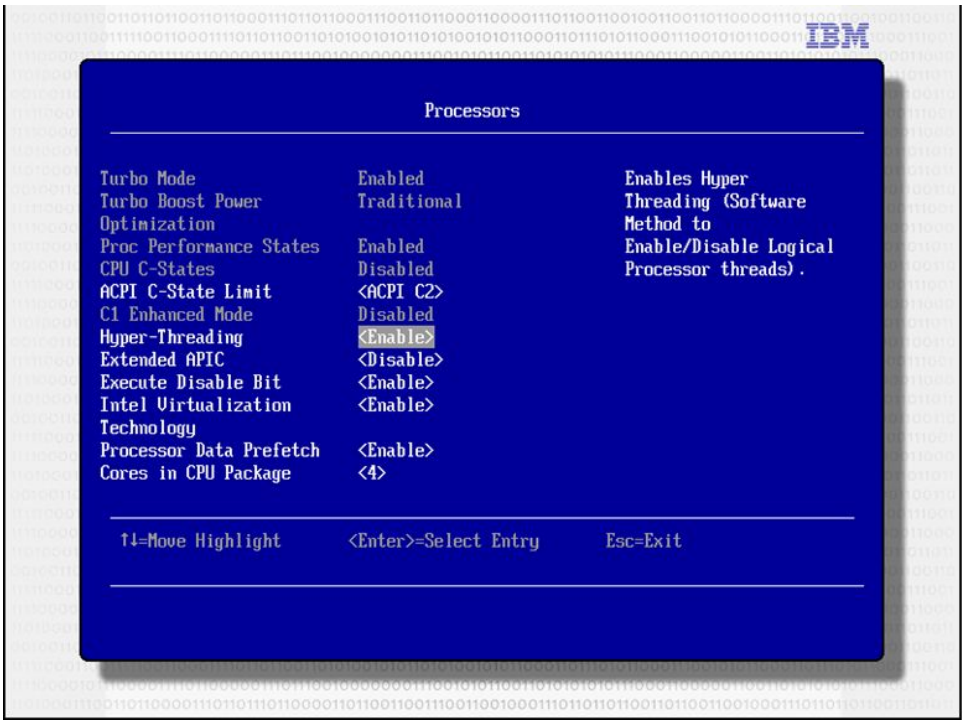
This procedure describes how to disable hyperthreading on the IBM xSeries hardware only. If you do not use this hardware, refer to the manual for your specific hardware for instructions about how to disable hyperthreading on your server.

Procedure

1. Restart the server.
2. When the launch screen displays, press `F1`.



3. The **System Configuration and Boot Management** screen displays as follows. To disable hyperthreading, select **System settings** > **Processors** > **Hyper-threading** > **Disable**.



4. To return to the main menu, press Esc twice. Select **Save Settings**.
5. Select **Exit** and restart the server.

Run network cards and switches at full duplex

To optimize performance, you need to enable full duplex mode for the network cards and switches that are used by the Log Analysis servers.

Run your network cards and switches at full duplex at the highest supported speed. Full duplex mode is much quicker than half duplex mode. Ensure that the network speed of your adapter, cables, switches, and other devices can accommodate the required throughput.

For a minimum speed of 1 Gbps (billions of bits per second), we recommend that you use full duplex Ethernet for a production deployment of Log Analysis.

Reducing the `TIME_WAIT` parameter for socket connections

TCP socket connections in Linux operating systems are set to 60 seconds by default. You can decrease this to optimize socket level performance by decreasing the time that is spent in the `TIME_WAIT` state.

About this task

To check the current `TIME_WAIT` value, run the following command:

```
$ cat /proc/sys/net/ipv4/tcp_fin_timeout
```

The default value is 60 seconds.

You need to complete this task on each server where Log Analysis or its components are installed.

Procedure

1. Use the root user ID to log in to the server where Log Analysis server or its component is installed.
2. To temporarily change the value of the `TIME_WAIT` parameter, run the following command:

   ```
   # echo 15 > /proc/sys/net/ipv4/tcp_fin_timeout
   ```
3. To permanently change the value of the `/proc/sys/net/ipv4/tcp_fin_timeout` parameter to 15 seconds, change the value of the `net.ipv4.tcp_fin_timeout` parameter in the `sysctl.conf` file to 15. For example:

   ```
   net.ipv4.tcp_fin_timeout=15
   ```
4. To ensure that the new value is implemented, you must restart the networking service. To restart the networking service, run the following command:

   ```
   # service network restart
   ```
5. To validate the result, run the following command:

   ```
   $ cat /proc/sys/net/ipv4/tcp_fin_timeout
   ```

   Ensure that the returned value is 15 seconds.

Chapter 4. Tuning IBM Operations Analytics - Log Analysis

Read the following section to find out how to optimize IBM Operations Analytics - Log Analysis.

Tuning the Log Analysis server

To tune your Log Analysis server, you need to tune two settings, `nproc` and `nofile`.

Procedure

1. To ensure that enough file descriptors and socket connections are available for the non-root user, switch users to the user who is running Log Analysis and enter the following command:

   ```
   ulimit -n
   ```

   If the value returned is `1024` for the non-root user, you need to increase the `nofile` limit.

2. To increase the `nofile` limit, log in as a root user and add the following section to the `/etc/system/limits.conf` file:

   ```
   ### Increase the count of file handles/socket connections for
    the non-root Log Analytics user ID
   ### Change "loganalytics" to the User ID which will run Log Analytics
   loganalytics  hard nofile  10000
   loganalytics  soft nofile  10000
   ```

   Save and exit the file. Stop and restart the Log Analysis server. Switch to the user for whom you increased the limit and enter the `ulimit -n` command to verify that setting is updated.

3. To ensure that ample threads and processes are available for the non-root user, switch users to the user who is running Log Analysis and enter the following command:

   ```
   ulimit -u
   ```

   If the value returned is `1024` for the non-root user, you need to increase the `nproc` limit.

4. To increase the `nproc` limit, log in as a root user and add the following section to the `/etc/system/limits.conf` file:

   ```
   ### Increase the count of threads/processes for the
   non-root Log Analytics user ID
   ### Change "loganalytics" to the User ID which will run solr
   loganalytics  hard nproc  5000
   loganalytics  soft nproc  5000
   ```

   Save and exit the file. Stop and restart the Log Analysis server. Switch to the user for whom you increased the limit and enter the `ulimit -u` command to verify that setting is updated.

5. To tune the Log Analysis Java virtual machine (JVM), edit the `<HOME>/IBM/LogAnalysis/wlp/usr/servers/Unity/jvm.options` as follows to increase the limit from 1 GB to 8 GB:

   ```
   # -Xms1g
   -Xms8g
   # -Xmx1g
   -Xmx8g
   ```

   This setting changes the value from the default of 1 GB to 8 GB. Save and exit. Stop and restart the Log Analysis server.

Tuning the Indexing Engine servers

To tune your Indexing Engine server, you need to tune two settings, `nproc` and `nofile` and the Java virtual machine.

About this task

The Indexing Engine servers use Apache Solr to index data. For more information, see Configuring the Indexing Engines.

Repeat these steps for each Indexing Engine server in your landscape.

Procedure

1. To ensure that the number of file descriptors and socket connections that are available for the non-root user, switch users to the user who is running Apache Solr and enter the following command:

   ```
   ulimit -n
   ```

   If the value returned is `1024` for the non-root user, you need to increase the `nofile` limit.

2. To increase the `nofile` limit, log in as a root user and add the following section to the `/etc/system/limits.conf` file:

   ```
   ### Increase the count of threads/processes for the non-root SOLR user ID
   ### Change "solr" to the User ID which will run solr
   solr  hard nofile  20000
   solr  soft nofile  20000
   ```

   Save and exit the file. Stop and restart the Indexing Engine server. Switch to the user for whom you increased the limit and enter the `ulimit -n` command to verify that setting is updated.

3. To ensure that there are ample threads and processes available for the non-root user, switch users to the user who is running Apache Solr and enter the following command:

   ```
   ulimit -u
   ```

   If the value returned is `1024` for the non-root user, you need to increase the `nproc` limit.

4. To increase the `nproc` limit, log in as a root user and add the following section to the `/etc/system/limits.conf` file:

   ```
   ### Increase the count of threads/processes for the non-root SOLR user ID
   ### Change "solr" to the User ID which will run solr
   solr  hard nproc  5000
   solr  soft nproc  5000
   ```

   Save and exit the file. Stop and restart the Indexing Engine server. Switch to the user for whom you increased the limit and enter the `ulimit -u` command to verify that setting is updated.

5. Increase the maximum and minimum values for the JVM from the default of 1 GB. For example, if you stream 500 GB of data daily, increase these values to 16 GB for optimal performance.

   To increase these values, add the following parameters in the `#SOLR Specific properties` section of the `<HOME>/IBM/LogAnalysis/wlp/usr/servers/Unity/apps/Unity.war/WEB-INF/unitysetup.properties` file:

   ```
   #Minimum Java Heap size for running SOLR, in MB
   # MIN_SOLR_HEAP_SIZE=1024
   MIN_SOLR_HEAP_SIZE=16192
   ```

```
#Maximum Java Heap size for running SOLR, in MB
# MAX_SOLR_HEAP_SIZE=1024
MAX_SOLR_HEAP_SIZE=16192
```

Save and exit the file.

6. If you observe high levels of disk read/write activity as part of a high production workload, consider enhancing the disk subsystem to target less than or equal to 50% disk busy. Set an appropriate disk read/write (input/output operations per second (IOPS)) to accommodate the specific requirements of your environment.

Increasing the number of shards

To balance the data loading and search performance, increase the number of shards to equal half the number of physical cores on theLog Analysis server.

About this task

- Ensure that you use physical cores.
- Ensure that hyperthreading is disabled.
- Ensure that you increase the value so that it is half the number of the total physical cores that are available on the servers where you installed the Indexing Engine component. For example, if you use a commodity installation that uses eight physical cores, you need to set `INDEX_NUM_SHARDS` to 4. If you use a production installation that uses 16 physical cores, you need to set `INDEX_NUM_SHARDS` to 8.
- You only need to change the `INDEX_NUM_SHARDS` on the server where Log Analysis is installed. The changes are automatically propagated in any remote Indexing Engine servers.

Procedure

1. Stop the server where Log Analysis is installed.
2. Open the `<HOME>/IBM/LogAnalysis/wlp/usr/servers/Unity/apps/Unity.war/WEB-INF/unitysetup.properties` file.
3. Change the `INDEX_NUM_SHARDS` so that it is half the number of the total physical cores that are available on the servers where you installed the Indexing Engine component. For example, if you use eight physical cores, set it to 4:

   ```
   INDEX_NUM_SHARDS=4
   ```

4. Save the `<HOME>/IBM/LogAnalysis/wlp/usr/servers/Unity/apps/Unity.war/WEB-INF/unitysetup.properties` file.
5. Restart the Indexing Engine server.

Data ingestion

To optimize performance, it is important that you use the correct method of data ingestion.

Each type of data ingestion is optimized for different scenarios. For example, to ingest a batch of log data for test purposes, use the data collector. If you want to stream log file information, use the IBM Tivoli Monitoring Log File Agent.

For more information, see Loading and streaming data

To improve performance, you can install the EIF Receiver and the IBM Tivoli Monitoring Log File Agent on remote servers. If you are running a test environment, you may not need to install remote instances. However, if you are setting up a larger environment, you need to install the EIF Receiver and IBM Tivoli Monitoring Log File Agent components on remote servers. For more information, see Streaming data from multiple remote sources across a network.

Considerations when using the IBM Tivoli Monitoring Log File Agent

Before you configure the IBM Tivoli Monitoring Log File Agent to ingest data, update the IBM Tivoli Monitoring Log File Agent to ensure that the configuration is appropriate to the log file that you are likely to ingest.

Log file size

If your log files are likely to exceed 50 MB, increase the size of the IBM Tivoli Monitoring Log File Agent cache: In the appropriate configuration file, set `BufEvtMaxSize=102400`. For WAS log files, update `<HOME>/IBM/LogAnalysis/IBM-LFA-6.30/config/lo/WASInsightPack-lfawas.conf`. For DB2 log files, update `<HOME>/IBM/LogAnalysis/IBM-LFA-6.30/config/lo/DB2InsightPack-lfadb2.conf`.

You must delete the appropriate existing cache file. For WAS log files, delete `<HOME>/IBM/LogAnalysis/logs/lfa-WASInsightPack.cache` and for DB2 log files, delete `<HOME>/IBM/LogAnalysis/logs/lfa-DB2InsightPack.cache`

For very large log files, update the cache size of the EIF receiver. In the `<HOME>/IBM/LogAnalysis/UnityEIFReceiver/config/eif.conf` file, increase the value of the `BufEvtMaxSize` property.

For WAS, update `<HOME>/IBM-LFA-6.30/config/lo/WASInsightPack-lfawas.conf` file. DB2 update `<HOME>/IBM-LFA-6.30/config/lo/DB2InsightPack-lfadb2.conf` file.

If you make any changes to the configuration, you must restart the service for the changes to take effect. To restart the service, from the `<HOME>/IBM/LogAnalysis/utilities` directory, run the following commands:

- `unity.sh -stop`
- `unity.sh -start`

Maximum log line length

The IBM Tivoli Monitoring Log File Agent monitors each log file line. The default maximum line length that can be processed by the IBM Tivoli Monitoring Log File Agent is 4096 bytes. This is equivalent to 4096 ASCII characters. This limitation is related to the log line and not the log record. If a log record consists of multiple log lines, such as in the case of a stack trace, the limit applies to each line. This is a limitation of the IBM Tivoli Monitoring Log File Agent and does not apply if you use an alternative data collection mechanism.

Performance implications of using the IBM Tivoli Monitoring Log File Agent

Loading logs using the IBM Tivoli Monitoring Log File Agent is a CPU bound process. If your system does not meet the minimum requirements you will need to increase the `MaxEventQueueDepth`. On some systems, altering this value may produce a noticeable impact on performance. This will buffer additional IBM Tivoli

Monitoring Log File Agent events while they are waiting to be processed. The required value for `MaxEventQueueDepth` may vary depending on the size of the rolled log and the number/speed of your CPU's. If you choose not to increase this value, then older events may be replaced on the event queue by newer events and not sent to the IBM Operations Analytics Log Analysis server.

To minimize the chance of data loss due to CPU bottlenecks, and to reduce the latency between when a log record is written to the file and when it is loaded, we recommend that the maximum size of a log be small enough so that you system does not fall behind while processing the logs.

`TIME_WAIT` parameter for socket connections

Tune the `TIME_WAIT` parameter for socket connections as described in this document. For more information, see "Reducing the `TIME_WAIT` parameter for socket connections" on page 7,

Run network cards and switches at full duplex

Run network cards and switches at full duplex. For more information, see "Run network cards and switches at full duplex" on page 7.

Configuring receiver buffer size and timeout

When collecting data using the IBM Tivoli Monitoring Log File Agent (LFA) and Tivoli Event Integration Facility (EIF) Adapter flow, you might need to change the rate at which events are flushed to the generic receiver for indexing. Incoming events are buffered at the EIF receiver side.

About this task

To improve overall IBM Operations Analytics Log Analysis performance, you can configure the buffer size and timeout period to match the rate of incoming events. When the event rate increases, increase the buffer size and decrease the timeout period. When the event rate decreases, decrease the buffer size and keep the timeout interval at the default value or increase it, depending on the event rate.

The following steps are the same for remote and local installations of the EIF unless stated otherwise.

Procedure

To change the buffer size and timeout parameters:

1. Stop IBM Operations Analytics Log Analysis:
   - If you use a local installation of the EIF, use the following command to stop IBM Operations Analytics Log Analysis:

     ```
     <HOME>/utilities/unity.sh -stop
     ```
   - If you use a remote installation of the EIF, use the `eifutil.sh -stop` command to stop the instances.
2. Open the configuration file for editing:
   - If you use a local installation of the EIF, open the `unity.conf` file in the `<HOME>/UnityEIFReceiver/config/` directory.
   - If you use a remote installation of the EIF, the `unity.conf` file is in the `<remote_deployment_location>/LogAnalysis/DataForwarders/EIFReceivers/`

<*eif_inst_#*>/config/unity.conf directory. Where <*remote_deployment_location*> is the directory on the remote machine where you deployed the EIF instance. <*eif_inst_#*> is the folder used for the specific remote EIF instance.

3. Change the Timeout and Buffer Size parameters to suit your operating environment:

```
#Timeout in Seconds
logsource.buffer.wait.timeout=10
#Buffer Size in Bytes
logsource.max.buffer.size=250000
```

4. Save your changes.
5. Start IBM Operations Analytics Log Analysis:
   - If you use a local installation of the EIF, use the following command to start IBM Operations Analytics Log Analysis:

     ```
     <HOME>/utilities/unity.sh -start
     ```

   - If you use a remote installation of the EIF, use the `eifutil.sh -start` command to start the instances.

Results

With higher buffer sizes, notice that it takes a longer time to fill the buffer with events and for batches to be posted to the receiver.

Chapter 5. Tuning scalable data collection components

If you use the recommended scalable data collection architecture, you need to tune the components.

For more information about scalable data collection, see Deploying scalable data collection architecture.

The following table summarizes the tuning settings for each component. These recommendations assume that your deployment is a production deployment and that you stream 500 GB of data daily:

Table 1. Tuning summary

Setting	Recommended value
Log Analysis server	
Number of files as specified in the `nofile` value.	10000
Number of processes and threads as specified in the `nproc` value.	5000
Maximum value for JVM	8 GB
Minimum value for JVM	8 GB
Indexing Engine servers	
Number of files as specified in the `nofile` value.	20000
Number of processes and threads as specified in the `nproc` value.	5000
Minimum heap size	16 GB
Maximum heap size	16 GB
Number of shards	8. This setting is based on a server with 16 physical cores.
HAProxy	
`nbproc`	2 or more depending on your environment. Set this value to half the number of physical cores on your server. For example, if you have 16 cores, set this value to 8.
Number of files as specified in the `nofiles` value.	25000
Receiver Logstash instance	
Number of files as specified in the `nofile` value.	25000
Number of processes and threads as specified in the `nproc` value.	5000
Maximum and minimum heap size for the Java virtual machine	8 GB
Apache Kafka and Apache ZooKeeper	
Number of files as specified in the `nofile` value.	5000

Table 1. Tuning summary (continued)

Setting	Recommended value
Maximum and minimum heap size of the JVM used by Apache ZooKeeper.	1 GB
Maximum and minimum heap size of the JVM used by Apache Kafka.	4 GB
Log retention period in hours.	24 hours
Sender Logstash instance	
Maximum and minimum heap size for the Java virtual machine	8 GB

Tuning HAProxy

To ensure optimal performance, tune the HAProxy component of your scalable data collection architecture.

About this task

For more information about installing HAProxy, see Installing and configuring HAProxy.

Procedure

1. Check if HAProxy is running in single thread mode. It runs in this mode by default. Log in as the root user and enter the following command:

   ```
   top -p "$(pgrep -d ',' haproxy)"
   ```

   If a single Process Identifier (PID) is returned, this result means that HAProxy is running in single-threaded mode. For a production environment, set the number of HAProxy instances to half the number of physical processor cores on the server, targeting no more than 50% overall processor utilization on the server.
2. To increase the number of HAProxy processes, for example, on a server with eight physical processors, edit the `haproxy.cfg` file, adding the following parameter. Set the value to four HAProxy instances.

   ```
   nbproc 4
   ```

   Setting this option to four quadruples the number of HAProxy instances to four (of the eight available physical processors), targeting an upper boundary of 50% total processor utilization on the HAProxy server.
3. To help you to make the correct number of socket connections available, enable the HAProxy statistics page. Add the following section in the `haproxy.cfg` file:

   ```
   ############################################################
   ## Listening for HAProxy Stats page
   ############################################################
   listen stats
   bind :9000
   mode http
   stats enable
   stats hide-version
   stats realm HAProxy\ Statistics
   stats uri /haproxy_stats
   ```

   Stop and restart HAProxy. This setting binds the local host to the `9000` port and enables the statistics page.

4. To open the statistics page, open a browser and enter `http://<localhost>:9000/haproxy_stats`.

   Compare the values in the **Sessions "Limit"** column to the values in the **Sessions "Cur"** and **Sessions "Max"** columns. Ensure that the values for the current and maximum sessions do not exceed the limit that is specified in the **Sessions "Limit"** column to prevent impediments in the data flow.

   If you want to raise the limit, edit the `haproxy.cfg`. For example, to increase it to 5000, edit the file as follows:

   ```
   # maxconn 256
   maxconn 5000
   ```

   Save and exit the file. Restart HAProxy to ensure that the change takes effect. Use the statistics page to verify that the limit is increased. For production environments, ensure that the limit is 20% larger than the values in the **Sessions "Cur"** and **Sessions "Max"** columns.
5. Switch users to the user who is running the HAProxy instance and enter the following command:

   ```
   ulimit -n
   ```

   If this command returns a value of `1024` for a non-root user, a limit is imposed on file descriptors or socket connections for the non-root user.
6. To address the restriction that is mentioned in the previous step, log in as a root user and add the following section to the `/etc/system/limits.conf` file:

   ```
   ### Increase the count of file handles/socket connections
   ### for the non-root haproxy user ID
   ### Change haproxy to the User ID which will run haproxy
   haproxy  hard nofile  25000
   haproxy  soft nofile  25000
   ```

   Set the `nofile` value to `25000` for every 500 GB of data that you load daily. Save the file.
7. Switch user to the user ID for whom you increased the limit and enter the `ulimit -n` command to verify that the setting is updated. Stop and restart HAProxy.

Tuning the Receiver cluster

Before you use scalable data collection architecture, tune the Logstash cluster that you use to receive data from HAProxy.

About this task

For more information about installing the Receiver cluster, see Configuring the Receiver cluster.

For performance reasons it is recommend that Logstash be installed on a different server than IBM Operations Analytics Log Analysis. Logstash is processor, memory, and disk intensive if the annotation and indexing functions are utilized.

Repeat this task for each server in your Receiver cluster.

Procedure

1. Ensure that enough socket connections are available for the non-root user who runs Logstash. To test this limit, switch user to the non-root user who runs Logstash and enter the following command:

```
ulimit -n
```

If the value that is returned is 1024 or less, increase the nofile limit.

2. To increase the nofile limit, add the following to the etc/system/limits.conf file. For Logstash server that streams 200 GB of data daily, increase the limit to 25000

```
### Increase the count of file handles/socket connections for
the non-root logstash user ID
### Change "logstash" to the User ID which will run logstash
logstash  hard nofile  25000
logstash  soft nofile  25000
```

Save and exit the file. Enter the ulimit -n command to verify that the setting is updated for the non-root user.

3. Ensure that enough processes are available for the non-root user who runs Logstash. To test this limit, switch user to the non-root user who runs Logstash and enter the following command:

```
ulimit -u
```

If the value that is returned is 1024 or less, increase the nproc settings.

4. To increase the nproc limit, add the following to the etc/system/limits.conf file. For Logstash server that streams 200 GB of data daily, increase the limit to 5000

```
### Increase the count of threads/processes for the
non-root logstash user ID
### Change "logstash" to the User ID which will run logstash
logstash  hard nproc  5000
logstash  soft nproc  5000
```

Save and exit the file. Stop and restart the Logstash server. To verify that the setting is updated, switch to the user who you updated the setting for and enter the ulimit -u command.

5. Increase the minimum and maximum heap sizes for the Java virtual machine that is used by Logstash from the default value of 1 GB to 8 GB. To increase the heap size, add the following to after the "plugin_path"="${LOGSTASH_HOME}/logstash-scala/" section in the <*install_dir*>/utilities/logstash-util.sh file:

```
### Increase the global logstash heap size to 8 GB
export LS_HEAP_SIZE="8g"
```

Locate the following section in the same file:

```
# There are no JAVA_OPTS set from the client, we set a predefined
# set of options that think are good in genera
```

Add the following lines:

```
JAVA_OPTS="$JAVA_OPTS -Xms8g"
JAVA_OPTS="$JAVA_OPTS -Xmx8g"
```

Tuning Apache Kafka and Apache ZooKeeper

Before you can use the Apache Kafka and Apache ZooKeeper components of your scalable data collection architecture, tune these components to optimize performance.

About this task

You install Apache Kafka and Apache ZooKeeper as part of your scalable data collection architecture. It consists of one or more brokers. The first broker contains a Apache ZooKeeper instance unless you configure it otherwise. For a production environment, install at least two brokers. For more information about configuring brokers, see Configuring Apache Kafka brokers.

Procedure

1. Switch users to the user who runs Apache Kafka and enter the following command:

   ```
   -ulimit n
   ```

   If the value is 1024 or less, you need to increase the `nofile` limit to accommodate more file descriptors and socket connections.

2. To increase the `nofile` limit, log in as a root user and add the following section to the `/etc/system/limits.conf` file. This example assumes that you are running a production deployment where you stream 500 GB of data daily:

   ```
   ### Increase the count of file handles/socket connections for
   the non-root kafka user ID
   ### Change "kafka" to the User ID which will run kafka/zookeeper
   kafka  hard nofile  5000
   kafka  soft nofile  5000
   ```

   Save and exit the file. Stop and restart the Apache Kafka. Switch to the user for whom you increased the limit and enter the `ulimit -n` command to verify that setting is updated.

3. Increase the default minimum and maximum heap sizes for the Java virtual machine that is used by Apache ZooKeeper from 512 MB to 1 GB. To increase these values, add the following lines to the `KAFKA_LOG4J_OPTS` section of the *<kafka_dir>*`/bin/zookeeper-server-start.sh` file:

   ```
   if [ "x$KAFKA_LOG4J_OPTS" = "x" ]; then
    export KAFKA_LOG4J_OPTS="-Dlog4j.configuration=file:$base_dir
   /../config/log4j.properties"
   fi
   ```

   Add or modify the following lines to increase these values to 1 GB:

   ```
   if [ "x$KAFKA_HEAP_OPTS" = "x" ]; then
    export KAFKA_HEAP_OPTS="-Xms1G -Xmx1G"
   fi
   ```

   Save and exit the file. Restart Apache ZooKeeper to ensure that the changes are picked up.

4. Increase the default minimum and maximum heap sizes for the Java virtual machine for Apache Kafka from 1 GB to 4 GB. To increase these values, add the following lines to the `KAFKA_LOG4J_OPTS` section of the *<kafka_dir>*`/bin/kafka-server-start.sh` file:

   ```
   if [ "x$KAFKA_LOG4J_OPTS" = "x" ]; then
    export KAFKA_LOG4J_OPTS="-Dlog4j.configuration=file:$base_dir
   /../config/log4j.properties"
   fi
   ```

   Add or modify the following lines to increase these values to 4 GB:

   ```
   if [ "x$KAFKA_HEAP_OPTS" = "x" ]; then
    export KAFKA_HEAP_OPTS="-Xms4G -Xmx4G"
   fi
   ```

Save and exit the file. Restart Apache Kafka to ensure that the changes are picked up. Ensure that the Apache ZooKeeper instance is running before you start Apache Kafka.

Repeat this step for each Apache Kafka broker.

5. Tune the log retention settings in Apache Kafka. By default, it maintains a cache of data for 168 hours for recovery purposes. This data is stored on disk and can cause space management issues. In a production environment, reduce this setting to 24 or 48 hours depending on your specific needs.

   For example, to reduce this setting to 24 hours, change the `log.retention.hours` parameter's value from `168` to `24` in the `<kafka_dir>/config/server.properties` file. For example:

   ```
   log.retention.hours=24
   ```

   Save and exit the file. Restart Apache ZooKeeper.

   Repeat this step for each Apache ZooKeeper broker.

Tuning the Sender cluster

Before you use scalable data collection architecture, tune the Logstash cluster that you use to send data to Log Analysis.

About this task

For more information about installing and configuring the Sender cluster, see Configuring the Sender cluster.

For performance reasons it is recommend that Logstash be installed on a different server than IBM Operations Analytics Log Analysis. Logstash is processor, memory, and disk intensive if the annotation and indexing functions are utilized.

For most production systems, the default value of `1024` is enough for both the `nproc` and `nofile` values. However you do need to increase the heap size for the Java virtual machine.

Procedure

1. Increase the minimum and maximum heap sizes for the Java virtual machine that is used by Logstash from the default value of 1 GB to 8 GB. To increase the heap size, add the following to after the `"plugin_path"="${LOGSTASH_HOME}/logstash-scala/"` section in the `<install_dir>/utilities/logstash-util.sh` file:

   ```
   ### Increase the global logstash heap size to 8 GB
   export LS_HEAP_SIZE="8g"
   ```

2. Locate the following section in the same file:

   ```
   # There are no JAVA_OPTS set from the client, we set a predefined
   # set of options that think are good in genera
   ```

   Add the following lines:

   ```
   JAVA_OPTS="$JAVA_OPTS -Xms8g"
   JAVA_OPTS="$JAVA_OPTS -Xmx8g"
   ```

3. Save and exit the file.
4. Stop and restart Logstash to ensure that the setting is updated.

Chapter 6. Tuning and configuring Mozilla Firefox

IBM Operations Analytics Log Analysis supports Mozilla Firefox Extended Support Release (ESR).

Creating dedicated Firefox profiles

To further optimize the performance of IBM Operations Analytics Log Analysis, you can create a dedicated Firefox profile that is used only for IBM Operations Analytics Log Analysis.

Before you begin

Ensure that you have the required permissions within your organization to create new Firefox profiles.

About this task

In most cases, the default Firefox profile is adequate for IBM Operations Analytics Log Analysis. However, in some cases, the performance of Firefox can be diminished. This diminished performance can reduce the overall performance of IBM Operations Analytics Log Analysis. Firefox performance can diminish for various reasons, such as third-party extensions or plug-ins and customizing of the default profile.

Procedure

1. Open a command-line window and navigate to the location of the Firefox executable file.
2. Open the **Firefox Profile Manager**. For Windows operating systems, run the following command:

   ```
   firefox.exe -p
   ```

   For UNIX or Linux operating systems, run the following command:

   ```
   firefox -p
   ```
3. To start the **Create Profile** wizard, click **Create Profile** > **Next**.
4. Enter a new profile name, for example SCALA. You can use the default location for settings and preferences.
5. To start Firefox, click **Finish**.
6. When Firefox opens, it prompts the user to select the profile that is used for the new session. Select the profile that you created in the previous steps.

Results

For more information about Firefox profiles, see Use the Profile Manager to create and remove Firefox profiles.

Configuring Firefox for extensive result sets and charts

If you want to use IBM Operations Analytics Log Analysis to process many facets or very detailed charts, you can configure Firefox to reduce the number of Java Script timeout messages that are displayed on the Firefox UI.

About this task

If you use IBM Operations Analytics Log Analysis to display many facets or a very detailed and intricate chart, you may notice that Firefox generates a number of error messages about a Java Script timeout. To continue working, the IBM Operations Analytics Log Analysis user must confirm this message.

You can change the default settings in Firefox so that the number of these messages is reduced and overall performance is optimized.

Procedure

1. Log in to Firefox. If you created a dedicated Firefox profile for IBM Operations Analytics Log Analysis, use it.
2. Open a new tab. Enter `about:config` in the **Search** bar and press enter. To continue, click the **I'll be careful, I promise** button.
3. Enter `dom.max script run time` in the **Search** bar. Wait for the setting to display in the **Preferences** table. The default value is 10 seconds.
4. To change the default setting 10 - 60 seconds, double-click on **dom.max script run time** and change the value 10 - 60 seconds.

Appendix. Notices

This information was developed for products and services that are offered in the USA.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
United States of America

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:
INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those

websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
2Z4A/101
11400 Burnet Road
Austin, TX 78758 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at www.ibm.com/legal/copytrade.shtml.

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

IBM Online Privacy Statement

Privacy Policy Considerations

IBM Software products, including software as a service solutions, (“Software Offerings”) may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user, or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session and persistent cookies that collect each user's user name and password for purposes of session management, authentication, enhanced user usability, and single sign-on configuration. These cookies cannot be disabled.

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at http://www.ibm.com/privacy and IBM's Online Privacy Statement at http://www.ibm.com/privacy/details in the section entitled “Cookies, Web Beacons and Other Technologies” and the “IBM Software Products and Software-as-a-Service Privacy Statement” at http://www.ibm.com/software/info/product-privacy.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Adobe, Acrobat, PostScript and all Adobe-based trademarks are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other product and service names might be trademarks of IBM or other companies.

Product Number:

Printed in USA