



# Flink Data Collector

## Cloud Insights

Tony Lavoie

February 27, 2020

This PDF was generated from [https://docs.netapp.com/us-en/cloudinsights/task\\_config\\_telegraf\\_flink.html](https://docs.netapp.com/us-en/cloudinsights/task_config_telegraf_flink.html) on May 08, 2020. Always check docs.netapp.com for the latest.

# Table of Contents

- Flink Data Collector ..... 1
  - Installation..... 1
  - Setup ..... 3
  - Objects and Counters..... 4
  - Troubleshooting ..... 8

# Flink Data Collector

Cloud Insights uses this data collector to gather metrics from Flink.

## Installation

1. From **Admin > Data Collectors**, click **+Data Collector**. Under **Services**, choose Flink.

Select the Operating System or Platform on which the Telegraf agent is installed.

2. If you haven't already installed an Agent for collection, or you wish to install an Agent for a different Operating System or Platform, click *Show Instructions* to expand the [Agent installation](#) instructions.
3. Select the Agent Access Key for use with this data collector. You can add a new Agent Access Key by clicking the **+ Agent Access Key** button. Best practice: Use a different Agent Access Key only when you want to group data collectors, for example, by OS/Platform.
4. Follow the configuration steps to configure the data collector. The instructions vary depending on the type of Operating System or Platform you are using to collect data.



## Flink Configuration

Gathers Flink metrics.

### What Operating System or Platform Are You Using?

[Need Help?](#)

Windows

### Select existing Agent Access Key or create a new one

Default (405fb5ec-d4cb-4404-977b-71fa931e1ad3)

[+ Agent Access Key](#)

\*Please ensure that you have a Telegraf Agent in you environment before configuring [Show Instructions](#)

### Follow Configuration Steps

[Need Help?](#)

- 1 Install Jolokia on your Flink JobManager(s) and Flink Task Manager(s). For details refer to the following [document](#).
- 2 Copy the contents below into a new .conf file under the C:\Program Files\telegraf\telegraf.d\ folder. For example, copy the contents to the C:\Program Files\telegraf\telegraf.d\cloudinsights-flink.conf file.

```
## #####  
## JobManager  
## #####  
[[inputs.jolokia2_agent]]  
  ## USER-ACTION: Provide address(es) of flink Job Manager(s), port for jolokia, add one URL  
  ## for each Job Manager for name <name>
```

- 3 Replace <INSERT\_FLINK\_JOBMANAGER\_ADDRESS> with the applicable Flink Job Manager address(es). Please specify a real machine address, and refrain from using a loopback address.
- 4 Replace <INSERT\_FLINK\_TASKMANAGER\_ADDRESS> with the applicable Flink Task Manager address(es). Please specify a real machine address, and refrain from using a loopback address.
- 5 Replace <INSERT\_JOLOKIA\_PORT> with the applicable jolokia port.
- 6 Modify 'Namespace' if needed for server disambiguation (to avoid name clashes).
- 7 Modify 'Cluster' if needed for Flink cluster designation.
- 8 Restart the Telegraf service.

```
Stop-Service -Name telegraf -ErrorAction SilentlyContinue; Start-Service -Name telegraf
```

# Setup

A full Flink deployment involves the following components:

**JobManager:** The Flink master. Coordinates a series of TaskManagers (slaves). In a High Availability setup, system will have more than one JobManager.

**TaskManager:** The Flink slaves. This is where Flink operators get executed.

The Flink plugin is based on the telegraf's Jolokia plugin. As such as a requirement to gather info from all Flink components, JMX needs to be configured and exposed via Jolokia on all components.

## Compatibility

Configuration was developed against Flink version 1.7.0.

## Setting Up

### Jolokia Agent Jar

For all individual components, a version the Jolokia agent jar file must be downloaded. The version tested against was [Jolokia agent 1.6.0](#).

Instructions below assume that downloaded jar file (jolokia-jvm-1.6.0-agent.jar) is placed under location '/opt/flink/lib/'.

### JobManager

To configure JobManager to expose the Jolokia API, you can setup the following environment variable on your nodes then restart the JobManager:

```
export FLINK_ENV_JAVA_OPTS="-javaagent:/opt/flink/lib/jolokia-jvm-1.6.0-agent.jar=port=8778,host=0.0.0.0"
```

You can choose a different port for Jolokia (8778). If you have an internal IP to lock Jolokia onto you can replace the "catch all" 0.0.0.0 by your own IP. Notice this IP needs to be accessible from the telegraf plugin.

### TaskManager

To configure TaskManager(s) to expose the Jolokia API, you can setup the following environment variable on your nodes then restart the TaskManager:

```
export FLINK_ENV_JAVA_OPTS="-javaagent:/opt/flink/lib/jolokia-jvm-1.6.0-agent.jar=port=8778,host=0.0.0.0"
```

You can choose a different port for Jolokia (8778). If you have an internal IP to lock Jolokia onto you

can replace the "catch all" 0.0.0.0 by your own IP. Notice this IP needs to be accessible from the telegraf plugin.

## Objects and Counters

The following objects and their counters are collected:

Object:	Identifiers:	Attributes:	Datapoints:
Flink Task Manager	Cluster Namespace Server	Node Name Task Manager ID Node IP	Network Available Memory Segments Network Total Memory Segments Garbage Collection PS MarkSweep Count Garbage Collection PS MarkSweep Time Garbage Collection PS Scavenge Count Garbage Collection PS Scavenge Time Heap Memory Committed Heap Memory Init Heap Memory Max Heap Memory Used Thread Count Daemon Thread Count Peak Thread Count Thread Count Total Started

<b>Object:</b>	<b>Identifiers:</b>	<b>Attributes:</b>	<b>Datapoints:</b>
Flink Job	Cluster Namespace server Job ID	Node Name Job Name Node IP Last Checkpoint External Path Restarting Time	Downtime Full Restarts Last Checkpoint Alignment Buffered Last Checkpoint Duration Last Checkpoint Size Number of Completed Checkpoints Number of Failed Checkpoints Number of in Progress Checkpoints Number of Checkpoints Uptime
Flink Job Manager	Cluster Namespace Server	Node Name Node IP	Garbage Collection PS MarkSweep Count Garbage Collection PS MarkSweep Time Garbage Collection PS Scavenge Count Garbage Collection PS Scavenge Time Heap Memory Committed Heap Memory Init Heap Memory Max Heap Memory Used Number Registered Task Managers Number Running Jobs Task Slots Available Task Slots Total Thread Count Daemon Thread Count Peak Thread Count Thread Count Total Started

Object:	Identifiers:	Attributes:	Datapoints:
Flink Task	Cluster Namespace Job ID Task ID	Server Node Name Job Name Sub Task Index Task Attempt ID Task Attempt Number Task Name Task Manager ID Node IP Current Input Watermark	Buffers In Pool Usage Buffers In Queue Length Buffers Out Pool Usage Buffers Out Queue Length Number Buffers In Local Number Buffers In Local Per Second Count Number Buffers in Local Per Second Rate Number Buffers In Remote Number Buffers In Remote Per Second Count Number Buffers In Remote Per Second Rate Number Buffers Out Number Buffers Out Per Second Count Number Buffers Out Per Second Rate Number Bytes In Local Number Bytes In Local Per Second Count Number Bytes In Local Per Second Rate Number Bytes In Remote Number Bytes In Remote Per Second Count Number Bytes In Remote Per Second Rate Number Bytes Out Number Bytes Out Per Second Count Number Bytes Out Per Second Rate Number Records In Number Records In Per Second Count Number Records In Per Second Rate



Object:	Identifiers:	Attributes:	Datapoints:
Flink Task Operator	Cluster Namespace Job ID Operator ID Task ID	Server Node Name Job Name Operator Name Sub Task Index Task Attempt ID Task Attempt Number Task Name Task Manager ID Node IP	Current Input Watermark Current Output Watermark Number Records In Number Records In Per Second Count Number Records In Per Second Rate Number Records Out Number Records Out Per Second Count Number Records Out Per Second Rate Number Late Records Dropped Assigned Partitions Bytes Consumed Rate Commit Latency Avg Commit Latency Max Commit Rate Commits Failed Commits Succeeded Connection Close Rate Connection Count Connection Creation Rate Count Fetch Latency Avg Fetch Latency Max Fetch Rate Fetch Size Avg Fetch Size Max Fetch Throttle Time Avg Fetch Throttle Time Max Heartbeat Rate Incoming Byte Rate IO Ratio IO Time Avg (ns) IO Wait Ratio IO Wait Time Avg (ns) Join Rate Join Time Avg Last Heartbeat Ago

# Troubleshooting

Additional information may be found from the [Support](#) page.

## Copyright Information

Copyright © 2020 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

## Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.