Cluster Monitoring: Consolidation -- how can I see it is working?

Well, usually you don't. Consolidation is an efficient internal storage method that normally just works quietly in the background. So, it is not something that is intended to be viewed directly.

But it can be, from cmsh if we really want to. This is shown later on in this text, after the background explanation.

Background explanation

In the first part of this KB, we will give a background explanation:

Consider the following example, with the following settings:

- We sample raw data every 2 minutes.
- And we consolidate data every 10 minutes.

So there are always two types of data to bear in mind. Raw data, and consolidated data.

We will illustrate this with some ASCII graphics.

Every "|" indicates a data point.

Note that we have 5 times as many raw data points.

(just in case the formatting is lost: switch to a fixed width text)

```
                --- time --->

raw:         | | | | | | | | | | | | | | | |

consolidated: |     |     |     |     |
```

In the preceding example it makes no sense to use consolidated data. That is because for the entire period we also have raw data that is 5 times more accurate.

As time passes we will start dropping old raw data, to save space on the disk.

So, in the following, for the first 20 minutes we no longer have raw data.

Cluster Monitoring: Consolidation -- how can I see it is working?

--- time --->

raw: | | | | | | | | | | | | |

consolidated: | | | | | |

But we do still have the consolidated data points for this period.

When we plot data, we automatically switch to consolidated data for periods without raw data.

So we'll get a combination of both data sources.

--- time --->

plot: | | | | | | | | | | | | | |

So, the preceding gives a background understanding of how consolidation works.

Viewing in cmsh

In this, the second part of this KB, we will show this behavior too.

This assumes that our cluster has been UP for long enough that raw data is being dropped.

Long enough is at least 7 days, which is the minimal raw-data-interval. Because we use run-length-encoding (RLE) to compress monitoring on disk, this raw-data-interval can be (much) longer depending on the metric check.

The forks metric is changes very quickly, and thus can do little run-length-encoding. This makes it an ideal choice for this example.

In the following, we plot the last 7 days for forks. The first, older, part of plot has 1 sample per hour, and this is consolidated data.

But the last (more recent) part has 1 sample per 2 minutes, which is the raw data.

```
[bright8->device[bright8]]% dumpmonitoringdata -7d now forks
```

Timestamp   Value   Info

Cluster Monitoring: Consolidation -- how can I see it is working?

```
------------------------- ------------------- ----------
```

2018/10/17 10:30:00   2.76243 processes/s

2018/10/17 11:30:00   2.52528 processes/s

2018/10/17 12:30:00   2.53972 processes/s

...

2018/10/24 10:42:00   2.66669 processes/s

2018/10/24 10:44:00   2.63333 processes/s

2018/10/24 10:46:00   2.64167 processes/s

We can also query consolidated data directly.

This functionality is mainly for testing purposes, and is not available in Bright View.

Forks was set up with the 3 default (1 hour, 1 day, 1 week) consolidators. By specifying '--consolidationinterval' we can pick which consolidated data we want to see.

[bright8->device[bright8]]% dumpmonitoringdata --consolidationinterval 1h -7d now forks

...

[bright8->device[bright8]]% dumpmonitoringdata --consolidationinterval 1d -7d now forks

...

[bright8->device[bright8]]% dumpmonitoringdata --consolidationinterval 1w -7d now forks

...

Unique solution ID: #1440
Author: Frank Furter
Last update: 2018-10-26 17:54