# Kepler - Bug #5319

## The workflow which archive sensor data into metacat can upload incorrect data set when new data is coming

02/22/2011 04:20 PM - Jing Tao

| | | | |
|---|---|---|---|
| **Status:** | Resolved | **Start date:** | 02/22/2011 |
| **Priority:** | Normal | **Due date:** | |
| **Assignee:** | Jing Tao | **% Done:** | 0% |
| **Category:** | sensor-view | **Estimated time:** | 0.00 hour |
| **Target version:** | sensor-view-0.9.0 | **Spent time:** | 0.00 hour |
| **Bugzilla-Id:** | 5319 | | |

**Description**

I used sensor simulator to create data set, then killed the simulator. It created data with timestamp from 2011-02-22 03:16:54 to 2011-02-22 03:17:52.

I ran the workflow and got the eml has the title:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-22 03:16:54" and "2011-02-22 03:17:52"

I checked the data file in metacat and it has the data
from 2011-02-22 03:16:54 to 2011-02-22 03:17:52

Everything looks good.

Then I ran the sensor simulator again and created some new data. Then I killed the simulator. It created data with timestamp from 2011-02-22 03:57:35 to 2011-02-22 03:58:33

I ran the workflow again and two eml documents were uploaded:
1. Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-22 03:16:54" and "2011-02-22 03:58:33"

2. Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-22 03:57:35" and "2011-02-22 03:58:33"

I double checked the data file in metacat and found the data files comply with the metadata.

From the title, we can see the dataset 1 combines both dataset 2 and the previous dataset which was uploaded in the last time.

So we have the duplicated data. The second run should only create the dataset 2.

### History

#### #1 - 02/22/2011 04:54 PM - Jing Tao

I use sensor simulator created data with timestamp from 2011-02-22 04:32:02 to2011-02-22 04:33:00.

Then I ran the workflow and three eml documents (data files) were uploaded to metacat.

The time intervals are:
1. from 2011-02-22 03:16:54 to 2011-02-22 04:33:00
2. from 2011-02-22 03:57:35 to 2011-02-22 04:33:00
3. from 2011-02-22 04:32:02 to 2011-02-22 04:33:00

documents 1 and 2 are duplicated data.

#### #2 - 02/23/2011 05:02 PM - Jing Tao

Today, i fixed the issue that "last update" wasn't persistent.
I ran the workflow again and found:
First time:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:56:15" and "2011-02-23 03:58:13"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:57:15" and "2011-02-23 03:58:13"

Second time:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:56:15" and "2011-02-23 04:23:44"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:57:15" and "2011-02-23 03:58:13"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:21:46" and "2011-02-23 04:23:44"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:22:46" and "2011-02-23 04:23:44"

Third time:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:56:15" and "2011-02-23 04:38:35"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 03:57:15" and "2011-02-23 03:58:13"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:21:46" and "2011-02-23 04:38:35"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:22:46" and "2011-02-23 04:23:44"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:36:37" and "2011-02-23 04:38:35"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-23 04:37:37" and "2011-02-23 04:38:35"

#### #3 - 02/25/2011 11:22 AM - Jing Tao

This logic issue on two inputs, start time and interval, of DataTurbineActor 3.

The start time is using the current time, but the interval is using current time - previous time. We change the start time is previous time and interval is current time - previous time.

It works. Here is the result after running 3 times of the workflow:
1
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:05:24" and "2011-02-25 10:06:22"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:06:24" and "2011-02-25 10:07:22"

2
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:07:22" and "2011-02-25 10:07:22"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:07:24" and "2011-02-25 10:32:22"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:32:24" and "2011-02-25 10:33:22"

3
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:33:22" and "2011-02-25 10:33:22"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:33:24" and "2011-02-25 10:47:06"
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:47:08" and "2011-02-25 10:48:06"

But it still has issue on the boundary:
data with timestamp 2011-02-25 10:33:22 has been upload in the second time run:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:32:24" and "2011-02-25 10:33:22

However, it was uploaded again in the third run:
Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:33:22" and "2011-02-25 10:33:22"

You can see the document Dataset for sensor:"sensor0" at site:"gpp" for time period "2011-02-25 10:33:22" and "2011-02-25 10:33:22" in the third run only has a single data.

I dug around and found:

When lastUploadedTime 2011-02-25 10:33:22 was passed to DataTurbineActor 2, three output came out:
2011-02-25 10:33:24
2011-02-25 10:47:07
2011-02-25 10:48:07

Actually there is not data or metadata with the timestamp 2011-02-25 10:33:24.
Is it a bug of DataTurbineActor?

#### #4 - 03/22/2011 04:30 PM - Jing Tao

This is bug on TimeDifference class. I wrote another class MetadataRangesDeterminer class to replace it. This issue was fixed by stored the timeformat into database.

#### #5 - 03/27/2013 02:30 PM - Redmine Admin

Original Bugzilla ID was 5319