# Energy AnalytiX
# Backfill of Hyper Historian Data

APPLICATION NOTE

January 2016

**Description:** Guide to backfilling Energy AnalytiX with Hyper Historian data

**General Requirement:** ICONICS AnalytiX version 10.90 or later, Microsoft SQL Server 2008 R2 or later, understanding of OPC UA data access.

## Introduction

In previous versions of Energy AnalytiX, users were only able to trigger the import of Hyper Historian data using periodic triggers. Version 10.9 of Energy AnalytiX and Hyper Historian now support both Manual and On-Demand historical data backfill.

Hyper Historian Data Backfill task allow to meet the following scenarios:

### Scenario 1

1. Energy AnalytiX is set up to get data from Hyper Historian from 1/1/2014.
2. Data is imported into Hyper Historian from 1/1/2013 to 12/31/2013.
3. Energy AnalytiX should backfill all the imported data.
4. Energy AnalytiX should also recalculate any data that pertains to that time period.

### Scenario 2

1. Energy AnalytiX is set up to get data from Hyper Historian.
2. The machine that hosts Energy AnalytiX is taken down for maintenance for four (4) hours.
3. Energy AnalytiX is started back up.
4. Energy AnalytiX should backfill data from Hyper Historian for the four (4) hours.
5. Energy AnalytiX should also recalculate any data that pertains to that time period.

## Procedure

Users can now configure, schedule and monitor historian data import tasks.

- The first step is to configure an on-demand manual trigger task within Workbench for an active period trigger.
- Next, the task itself can be scheduled.
- Energy AnalytiX then detects that a historian data task has been scheduled and processes the task. First, it will validate the task. Then, if the task fails the validation step, it will be marked as failed and it will no longer be processed. If the task passes validation, Energy AnalytiX will post historian data requests based on the task's settings:
  - Single Request
  - Multiple Requests
- On-demand data requests will execute in parallel with the periodic data requests for the same trigger. Internally, the Runtime distinguishes the requests and separates the data so that they are not mixed. For an On-Demand data backfill request, Energy AnalytiX will first delete any existing data for the assets or meters within the interval of the request.
- At any point in time, users can cancel an executing or scheduled task or monitor the status of a task.

## Management Support

A new **Manage Data Backfill Tasks** button has been added to the Trigger configuration form, as shown in Figure 1. This button launches the 'Manage Data Backfill Tasks' historian data backfill form, as shown in Figure 2.

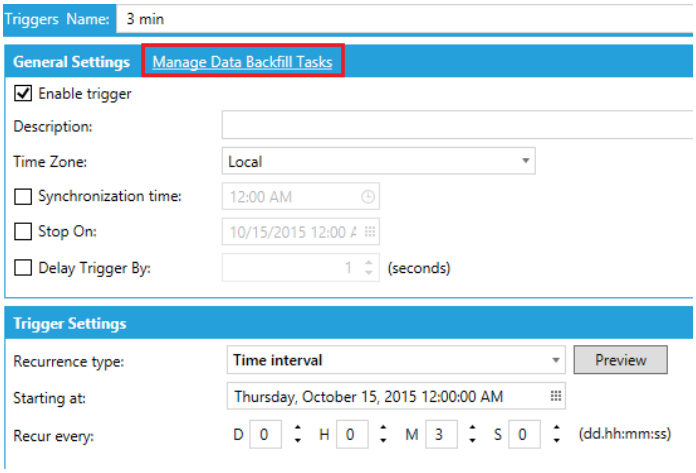


**Figure 1 - 'Manage Data Backfill Tasks' Button in Triggers Settings**

The new historical data backfill management settings allow users to:

- Configure a New Task
- Schedule (Add) the New Task
- Monitor New and/or Existing Tasks
- Cancel a Data Backfill Task

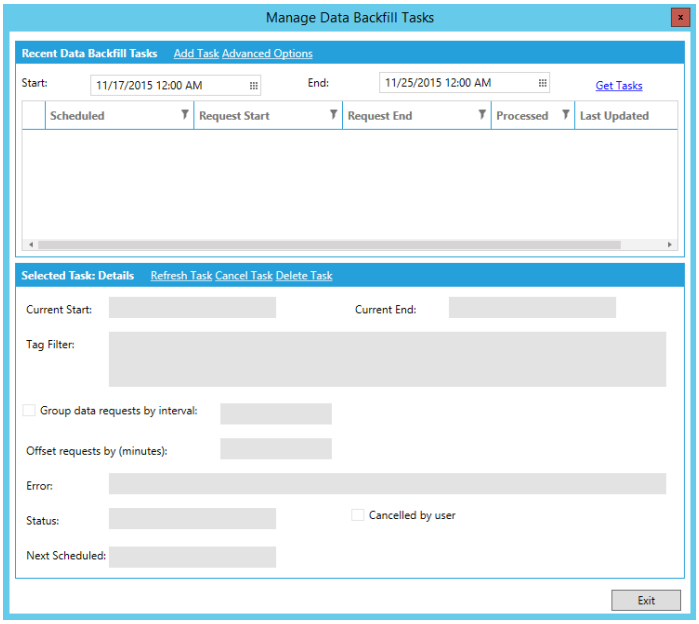


**Figure 2 - 'Manage Data Backfill Tasks' Window**

- Users can click on the **Add Task** button to configure a new data backfill task. This button will be disabled if the specific trigger is not enabled.

- Once a new task has been added, the grid will refresh to show the new task.

- In the **Selected Task Details** section, the **Current Start** and **Current End** times represent the start and end times of the data request for next execution. Imagine requiring a data import from 12:00 AM to 2:00 AM within the same day. If the user chooses to post a single request (i.e., not split the data request into smaller intervals), then the Current Start and Current End times will always be 12:00 AM and 2:00 AM. However, if the users chooses to post grouped requests (i.e., group data requests in smaller intervals of a time period [for example, in 15 minute increments]), then the Current Start and Current End times will be 12:00 AM and 12:15 AM, then 12:15 AM and 12:30 AM, then 12:30 AM and 12:45 AM, etc.

- The **Next Scheduled** time is the time that the next data request will execute based on the data backfill task settings.

- If the user selects an existing task from the grid, then the user can cancel the task by clicking on the **Cancel task** button.

- Cancelling a task sets the **Cancelled by user** field to true (checked), so that when the Energy AnalytiX service detects the change, it will then finalize the task cancellation, in which case, the status field of the task will be 'Cancelled'.

- Note that all the displayed times are in the local time zone, while they are stored as UTC inside the newly created EA-TriggerDataImport table.

## Configuration Support

When users click on the **Add Task** button, the **Configure Data Backfill Tasks** historian data backfill configuration form, as shown in Figure 3, will appear.

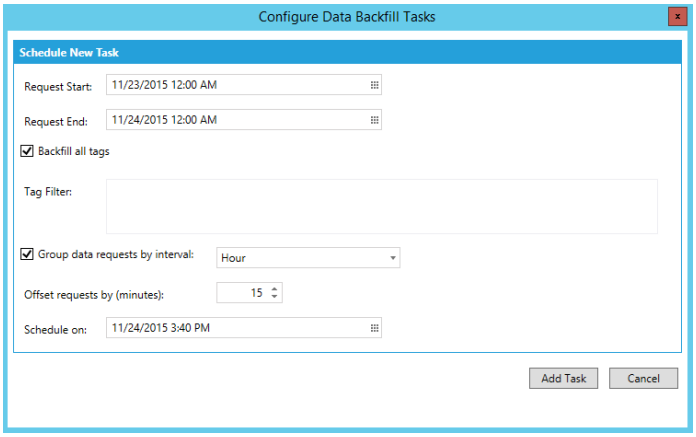


**Figure 3 - 'Configure Data Backfill Tasks' Window**

This form provides the following options:

- Data backfill **Request Start** and **Request End** times
- An option to **Backfill all tags** - backfill all the historical tags associated with the trigger or some of them (by using a tag filter)
- **Tag Filter** – For example: Company\Department\Location AAA\meter *
- **Group data requests by interval** – This provides an option to request all data in one execution or to group data request by base summary period (typically 15 minutes), hour, day, or week. Ideally, the bigger the data request is, the more that grouping of requests should be utilized.

- **Offset requests by (minutes)** – This provides options to offset (delay) data request by x minutes (rather than starting at the scheduled date). It will start at the scheduled date + Delay interval (min).

The form also provides validation. If the validation fails, the task will not be added. The validation is applied when the user clicks on the **Add Task** button and checks that:

- The data request start and end dates are earlier than "now"
- The data request start is earlier than the data request end

## Advanced Options

The **Advanced Options** button in the **Manage Data Backfill Tasks** window opens the **Configure Advanced Data Backfill Tasks** window, as shown in the figure below.

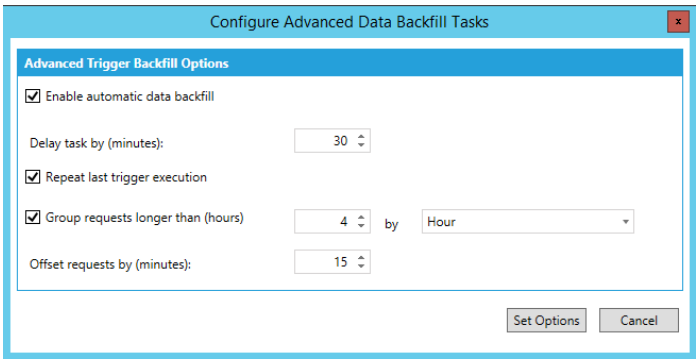


**Figure 4 - 'Configure Advanced Data Backfill Tasks' Window**

**NOTE**: Automatic data backfill support is a one-time operation, which is activated when the service restarts.

This form provides the following options:

- **Enable automatic data backfill** – Enables or disables automatic historical trigger data synchronization
- **Delay task by (minutes)** – Option to allow for a delay interval on posting the automatic historian data backfill request
- **Repeat last trigger execution** - option to schedule the trigger to repeat the last trigger request on startup. This will allow for smoother data backfills. This way, if the Energy AnalytiX service was shut down during a historical trigger processing request, the request will be posted again and data will be retrieved.
- **Group requests longer than (hours)** – Users can set a threshold of data duration (in hours) and then group by a selected time increment (Hour, etc.).
- **Offset requests by (minutes)** – This provides options to offset (delay) data request by x minutes (rather than starting at the scheduled date). It will start at the scheduled date + Delay interval (min).

## Monitoring Support

These configuration forms can be used to manage and monitor historical data tasks. At a more detailed level, the Energy AnalytiX provider's **Monitor View** branch's **Trigger Information** tab can be used to identify data requests on a per trigger basis for on-demand historian data requests. Depending on the set options, the name of the trigger that appears in the Monitor View could be: Trigger Name (Manual) or Trigger Name (Autosync).

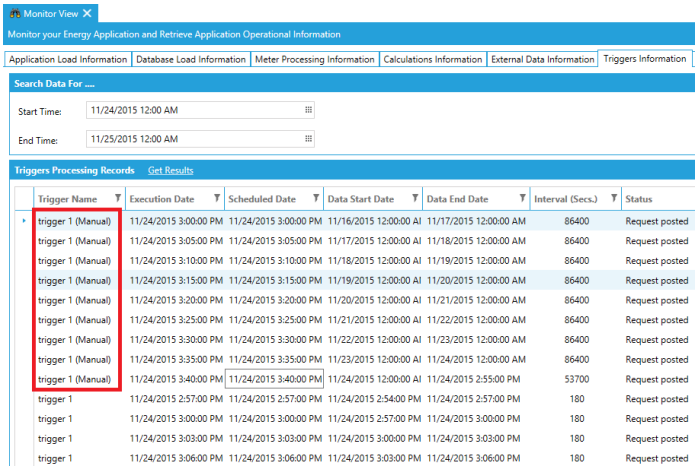


**Figure 5 - Trigger Information Tab Selected Within the Monitor View Branch of the Energy AnalytiX Provider in Workbench with (Manual) Trigger**

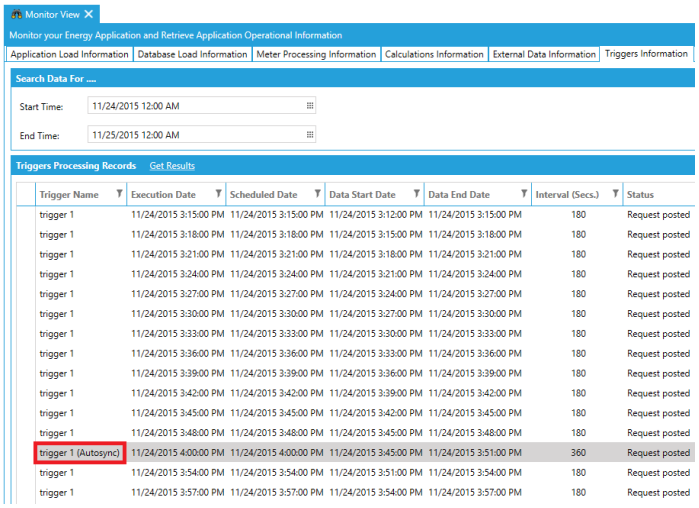


**Figure 6 - Trigger Information Tab Selected within the Monitor View Branch of the Energy AnalytiX Provider in Workbench with (Autosync) Trigger**

## Runtime Support

Energy AnalytiX can detect if a historian data task has been scheduled and can then process that task.

- First, it will validate the task. It will check that:
  - The scheduled date must be later than "now"
  - The data request start and end dates are earlier than "now"
  - The data request start is earlier than the data request end
  - The related (parent) periodic trigger is enabled and has started
  - The related (parent) periodic trigger has been associated with meters or metadata
    - If the task fails validation, it will mark the task as failed and it will no longer be processed.
    - If the task passes validation, Energy AnalytiX will update the task and:
      - Set the status as 'processing'
      - Set the Current Start and Current End dates based on user settings for the data backfill task
      - Set the "Next" scheduled date for the task, based on the delay interval for the task
      - Schedule the task to start executing
      - Start posting historian data requests based on the task settings:
        - Single Request (no grouping)
        - Multiple Requests (grouping enabled)

- On-demand data requests will execute in parallel with the periodic data requests for the same trigger
  - Internally, the Runtime distinguishes the requests and separates the data so that they are not mixed.
  - For an on-demand data backfill request, Energy AnalytiX will first delete any existing data for the assets or meters within the interval of request.
  - If there are multiple on-demand tasks for a specific trigger, Energy AnalytiX will execute them in sequential manner, to better load-balance the system.
  - There will be only one data backfill task executing per trigger. All additional tasks will stay in "default" status until the currently executing task is complete or gets cancelled by the user.
  - Also, if the Energy AnalytiX server detects that there is a heavy flow of new meter summary data, it may temporarily delay calculation evaluation, depending on the flow rate of meter data, to give more priority to importing meter data.

**Energy AnalytiX Service Restart**

If the service restarts or the server reboots, Energy AnalytiX will try to continue with the manual data backfill task and it may also move the next scheduled time within the closest 15 minute interval after the time it restarts.

A typical example of a historian data backfill for a specific trigger may be:

- **Task Configuration**
  - Scheduled Date: 20:00:00 PM
  - Delay: 15 minutes
  - Request Start: 10:00:00 AM
  - Request End: 18:00:00 PM
  - Request Group Interval: 1 hour

- **First Run:**
  - Next Scheduled date: 20:15:00 PM
  - Current Start: 10:00:00 AM
  - Current End: 11:00:00 AM

- **Second Run:**
  - Next Scheduled date: 20:30:00 PM
  - Current Start: 11:00:00 AM
  - Current End: 12:00:00 PM

## Additional Considerations

- On-demand historian data backfill can potentially process large volumes of data.
- When using Hyper Historian Data Backfill within Energy AnalytiX, the hardware requirements for Microsoft SQL Server, Energy AnalytiX, FrameWorX server, and Hyper-Historian nodes should be considered.
  - Especially for Microsoft SQL Server, the processor count, available memory and disk access speed are fundamental.
  - The more memory that is available to Microsoft SQL Server, the more data will be cached into memory and the faster the overall performance will be.
- For large data requests, users should always group the requests by smaller chunks, depending on historical tag count and available hardware.
- In cases of importing past historical data, users should consider that:
  - Energy AnalytiX considers staging meter data as temporal data.

- Energy AnalytiX considers meter summary data the "long term" data to keep.

  Imagine that you wish to import meter data for last three (3) years but, for the staging meter data, you only want to keep them for three (3) months. You would set the table maintenance options for the meter staging data to three (3) years. You would then import your historical data and let Energy AnalytiX process it and create meter summaries, calculations, etc. Then, you would reset the table maintenance options for meter staging data to three (3) months (see the Table Management Example and Microsoft SQL Database Statistics Plan near the end of this document).

- Setting the initial Microsoft SQL Server database size to a large value improves performance by minimizing the database expansion operations while data is backfilled.
- Users should consider the impact of importing or updating a large volume of data on the performance of Microsoft SQL Server and Energy AnalytiX.
  - Microsoft SQL Server will always try to process data as fast as possible.
  - To accomplish this, it utilizes all the indexes in related tables as well as the statistics for each index.
  - The statistics look at the data of each table and get a histogram of their distribution.
  - This enables Microsoft SQL Server to get an estimate for the volume of data for a certain query and try to service it in the best possible way.
  - When a lot of data is imported or updated, the statistics may get out of synch by not reflecting a current estimate for the data distribution.
  - Microsoft SQL Server automatically updates the statistics when it detects a change of about 20% on the table records. In some cases, this may take a bit of time.
  - Users can define a management task on Microsoft SQL Server to update statistics (for example, once a day or every few days), therefore keeping the database performance to each peak.
  - If you do commit a periodic statistics plan, keep in mind the following (see the figures below).
    - Try to set the database to update statistics asynchronously.
    - Never use FULL sampling option.
    - Try to use a reasonable sampling rate – say 5% of rows in a given table.
    - Run the scheduled plan at off-peak times, say 1:00 AM.

## Table Management Example

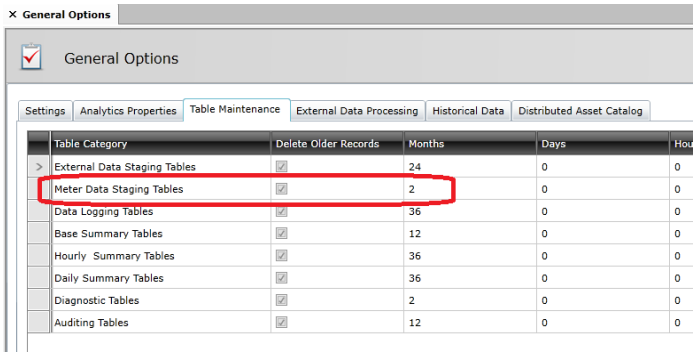


**Figure 7 - Table Maintenance Tab in the General Options Branch of the Energy AnalytiX Provider in Workbench-SL**

## Microsoft SQL Database Statistics Plan

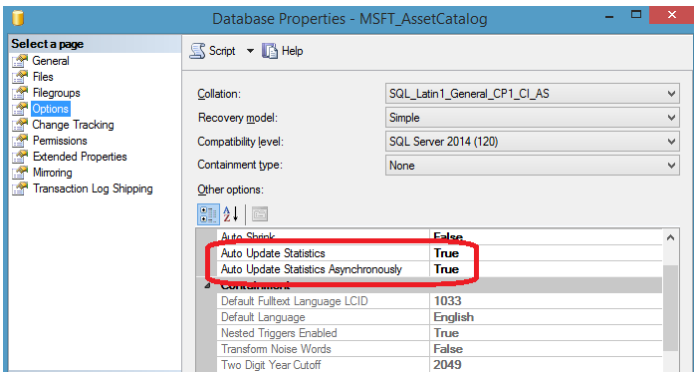


**Figure 8 - Database Properties – MSFT_Asset Catalog**

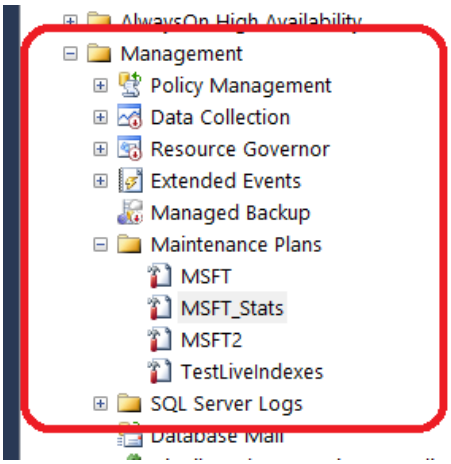


**Figure 9 - Selecting a Maintenance Plan to Update**

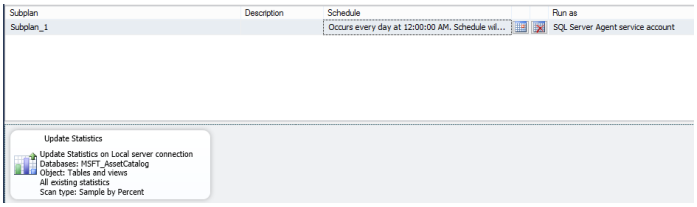


**Figure 10 - Setting a Schedule and Updating Statistics**

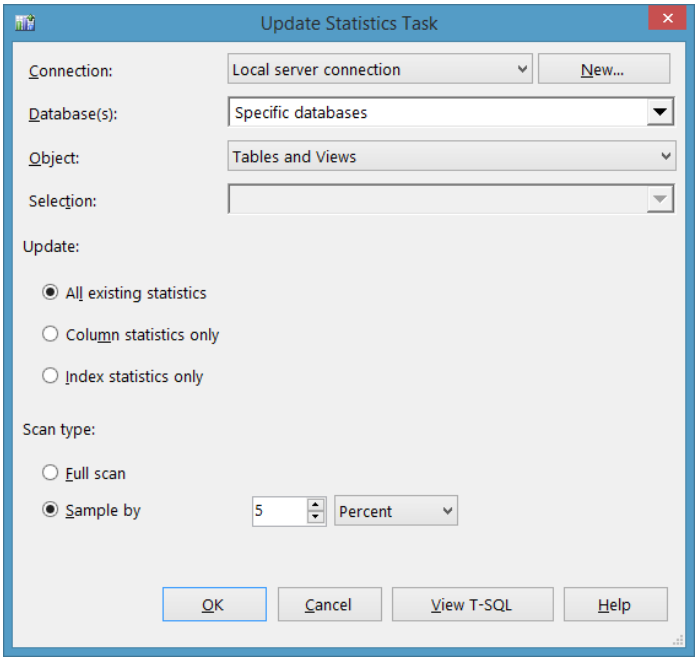


**Figure 11 - Update Statistics Task Window**