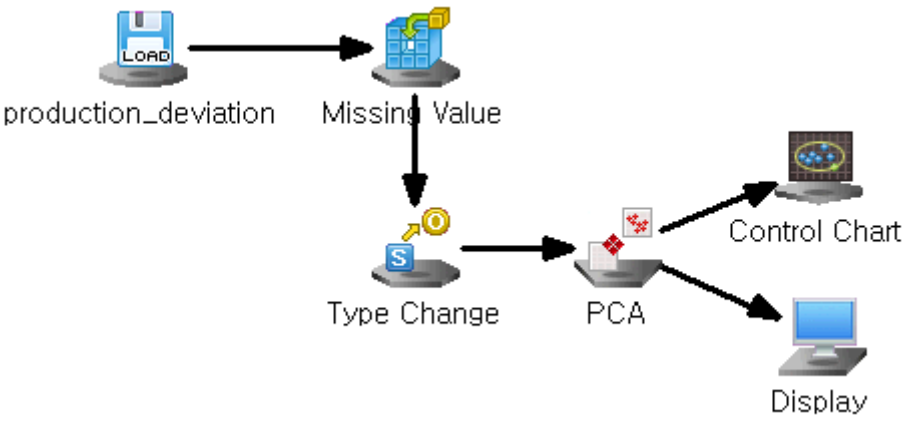
Chapter 2 Stream

ECMiner™ allows for easy and convenient analytical model by connecting **Nodes** into a **Stream**.

Stream is the series of connected nodes, from data loading to the output node. This stream represents a workflow map for data analytics, allowing for easy modifications during execution.

A brief example for Principal component analysis (ex.PCA):

1. Load data (Input data: production data)
2. Data preprocessing (Missing Value, Type Change, etc.)
3. Principal component analysis (PCA)
4. Control Charts for the principal component



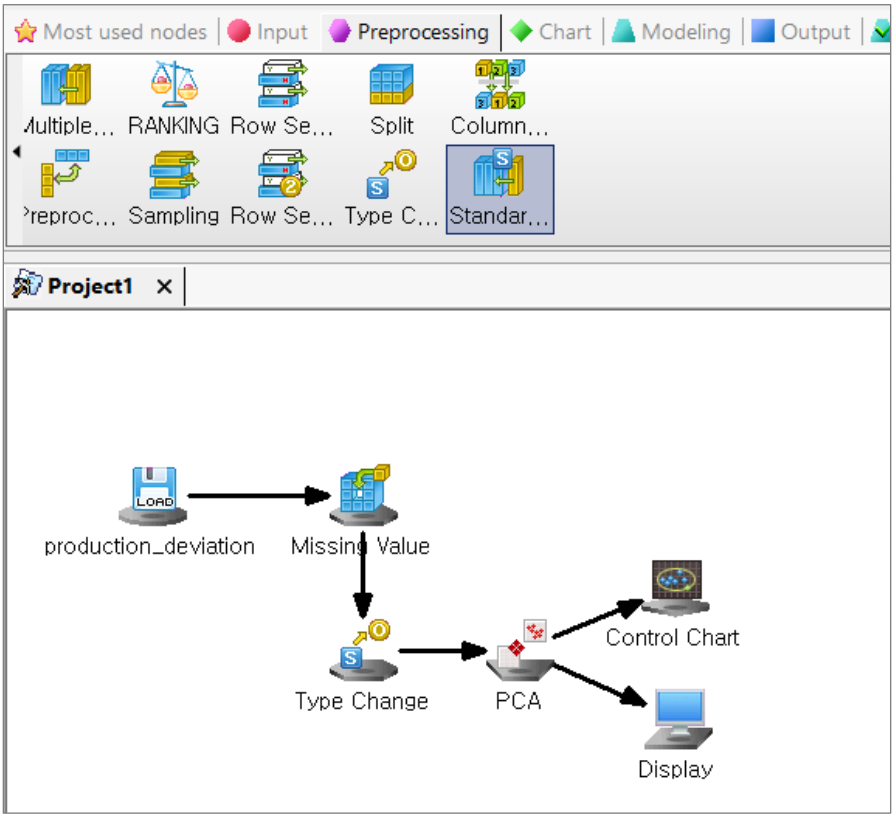
2.1 Stream Configuration

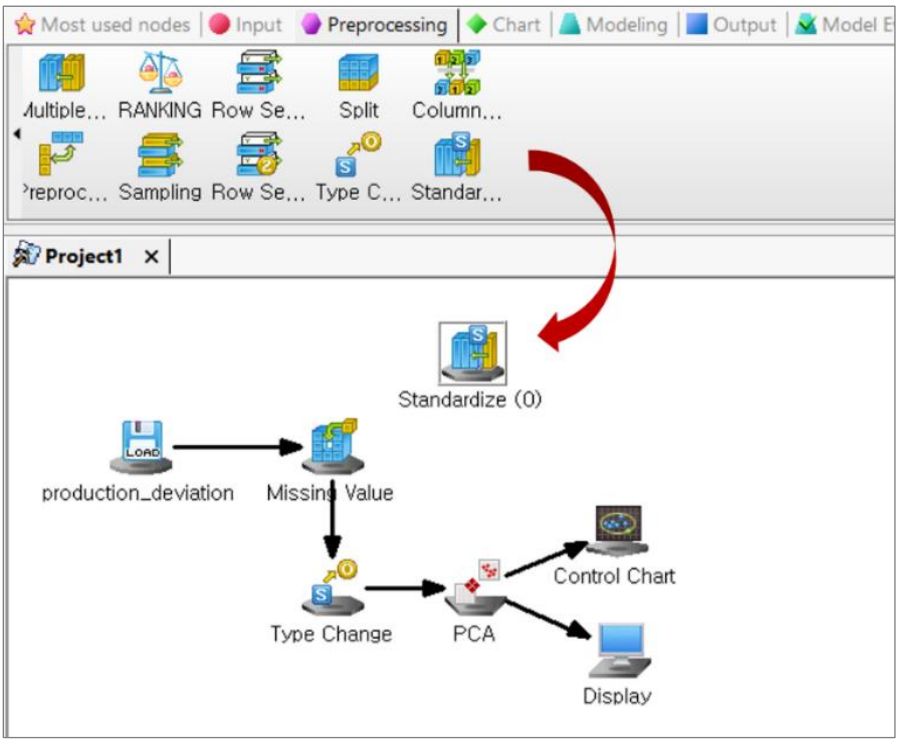
A stream is set up to connect nodes which you choose in each step of data analysis.

1. Pull down nodes (Input data, Preprocessing, Chart, Modeling, Output)
2. Edit node properties
3. Connect node (right-click the mouse and choose 'Connect' to link to the next node)

Create nodes

Pull down nodes in the working project window.

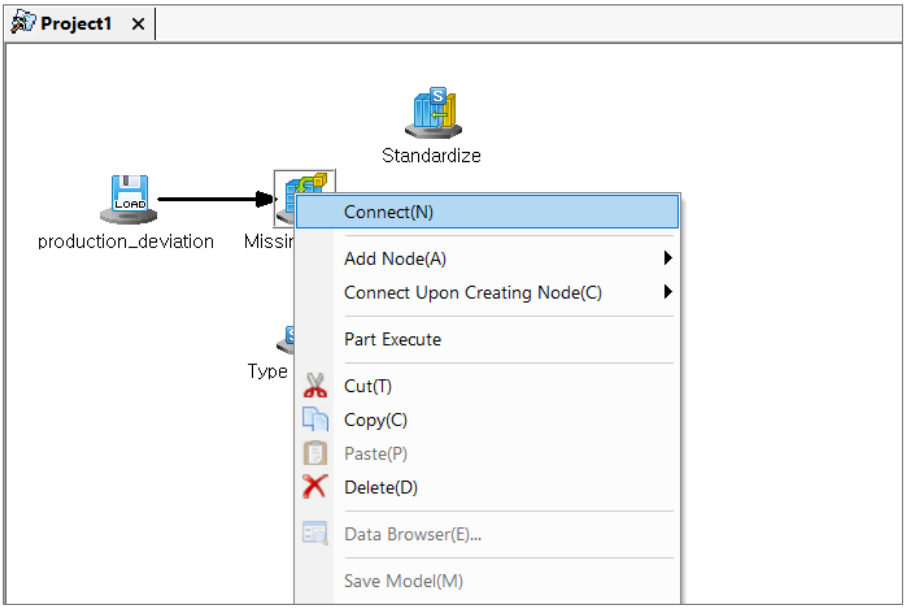




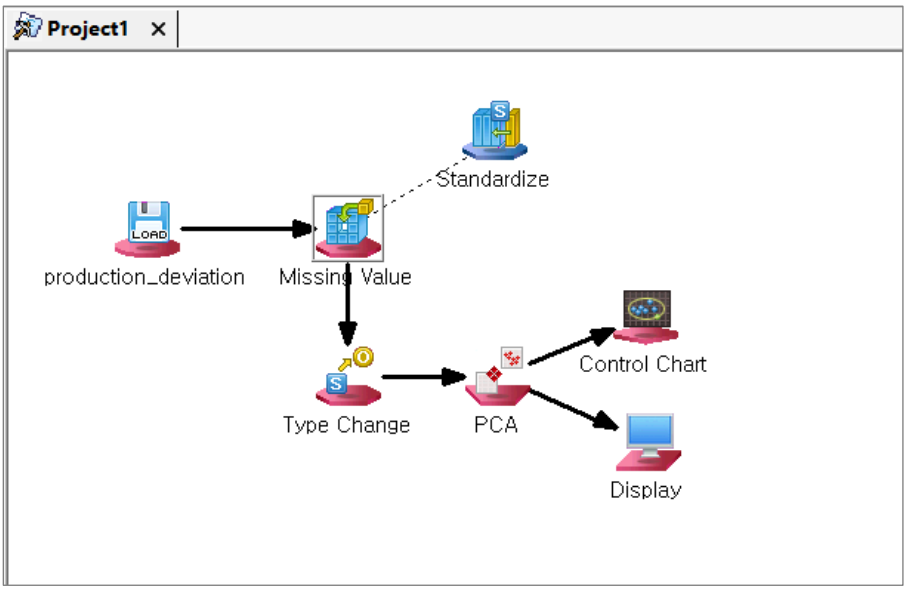
User can add or drop a node in the current project.

Connect nodes

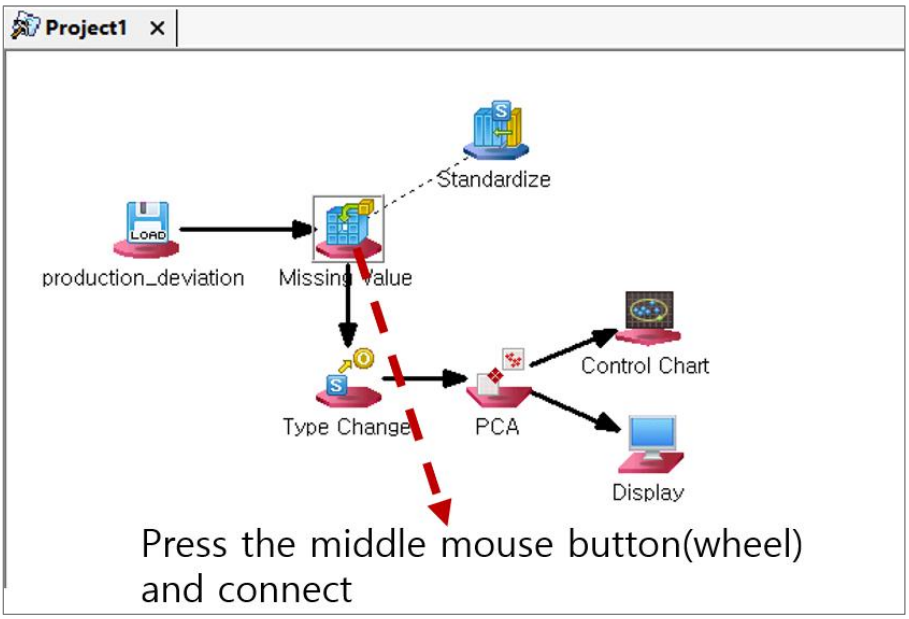
Right-click the mouse, choose 'Connect' and drag the arrow to the next node.



In Project window, you can manage node connections.



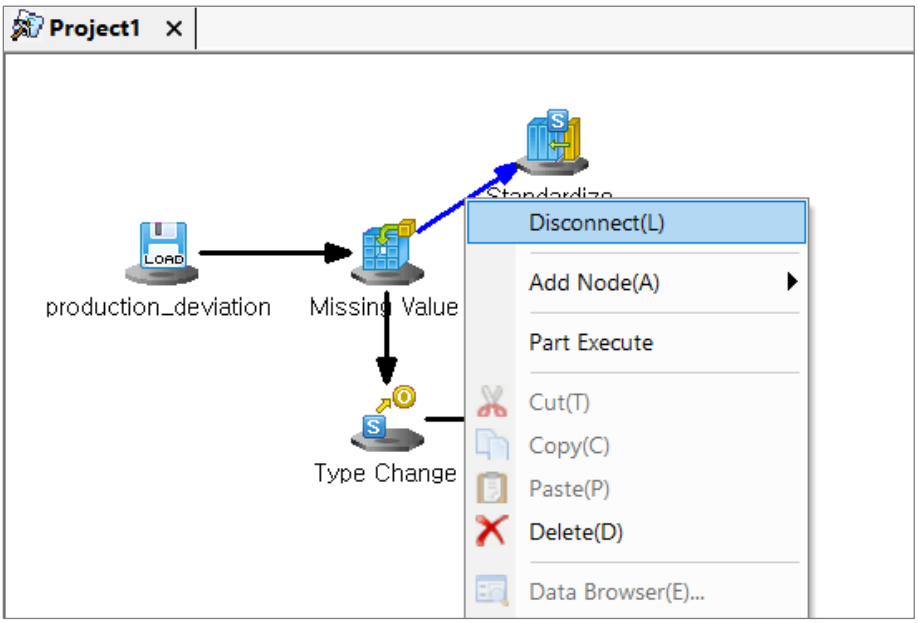
Alternatively, select the node and press the middle mouse button (wheel).



Holding the middle button, move the mouse pointer to the node you want to connect to, and then release the button.

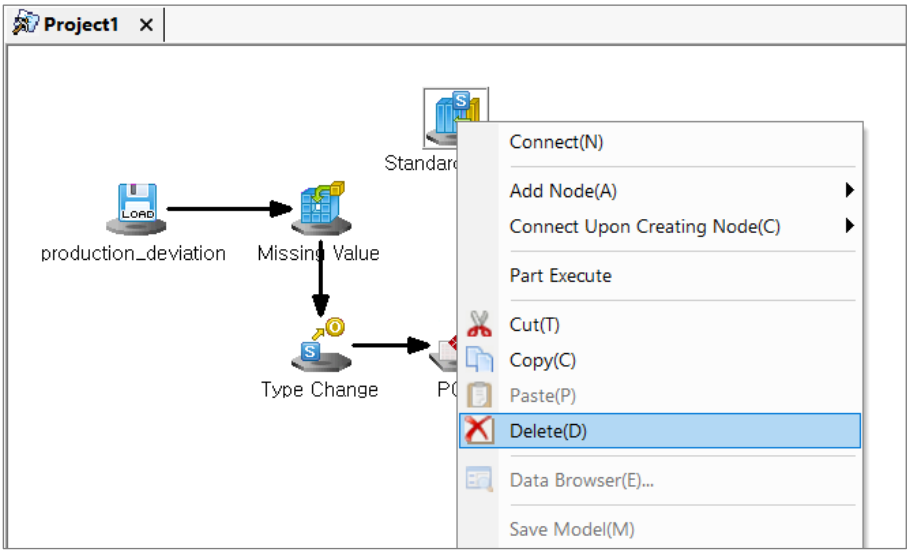
Disconnect nodes

To disconnect nodes, right-click the mouse and select "Disconnect".



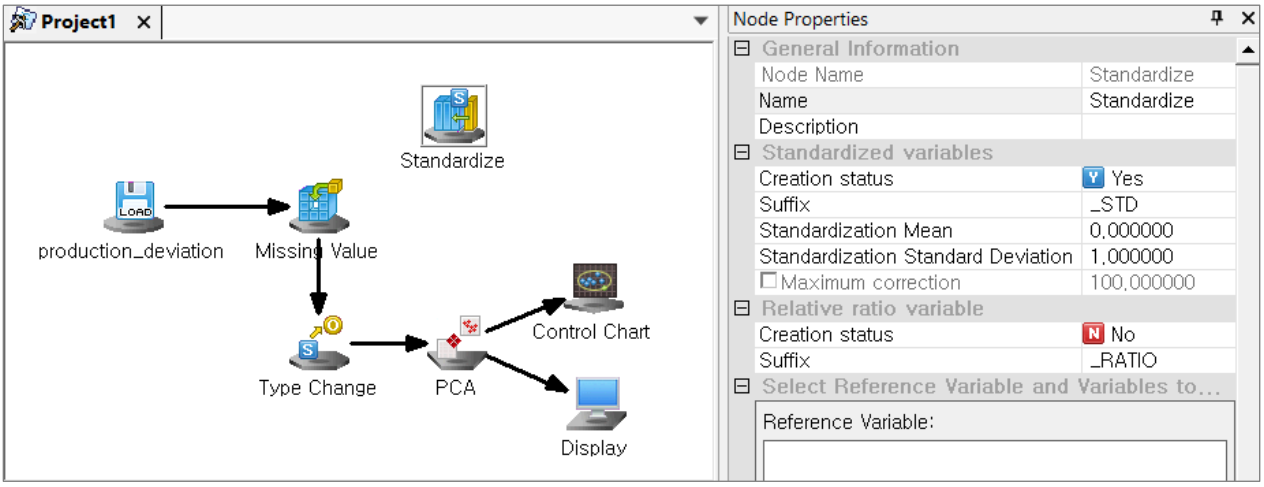
Remove nodes

To remove a node, click "Delete".



Edit node properties

Double-click the node that you want to revise or modify, and change the option of the node property.



Connectivity of nodes

Consider the following guidelines when configuring your stream:

1) Nodes generally have one input and can have multiple outputs.

2) However, Merge, Add, and Column Combiner nodes are exceptions, as they can accept multiple inputs.

3) Chart and Output nodes are output-only nodes that generate results, so they cannot serve as inputs to other nodes.

ECMiner™ provides visual indicators to show how it works

	Input Node	Preprocessing Node	Chart Node	Modeling Node	Output Node	Model Node
Origin						
Connectable						
Unconnectable						

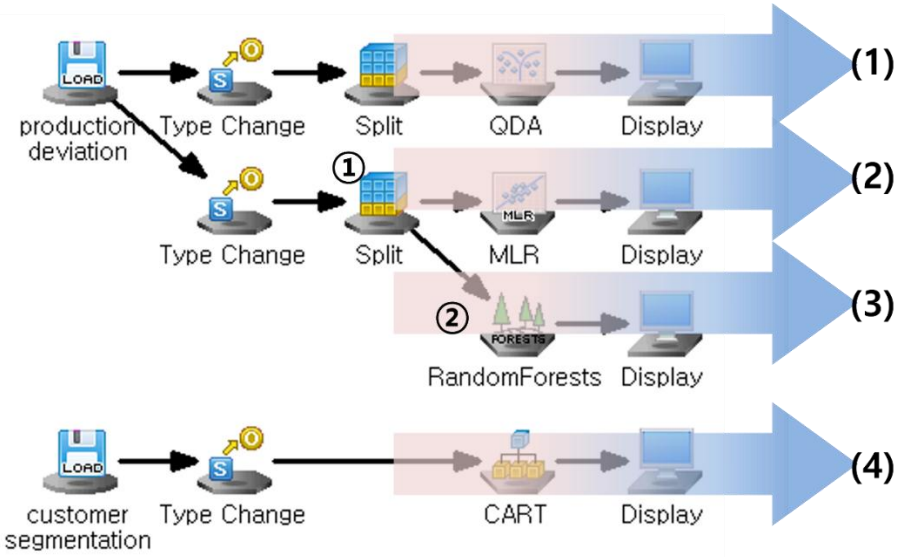
2.2 Stream Configuration Rules

The stream set-up rules:

- A stream includes input, preprocessing, modeling, and output node. Each node is connected to the next, constructing a stream. You may add or remove a node.
- Import data from a file, make some modifications, and then save the file. You may modify configuration stream in "File Reader -> Derived Column -> File Writer”
- If there are multiple streams in the Project window, ECMiner™ determines the order of their execution. Streams are executed sequentially, beginning with the first one in order.
- If node property is not correctly specified, it will not execute. If you have error message, then correct the node property.
- In a stream, nodes can have one-to-many connections. This means a node can receive data from a single source but can send data to multiple nodes.

2.3 Stream Execution

To run a stream, select "Stream > Execute" or "Stream > Part Execute" from the main menu.

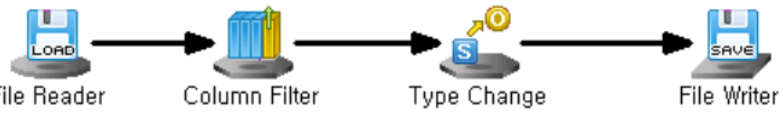


- Select **[Stream(S)] > [Execute]** from the main menu or press F5 to execute all streams (1) through (4).
- Select **[Stream(S)] > [Execute]** if you want to run a part of the stream, for example, node 1 (Split node). To continue running another part, such as node 2 (RandomForests node), select 'Part Execute' again.

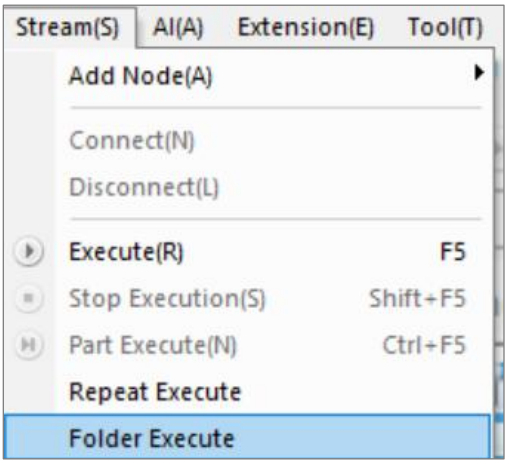
Folder Execute

This feature allows you to repeatedly execute the same project on multiple files of the same format within a specific folder. You can run it as follows:

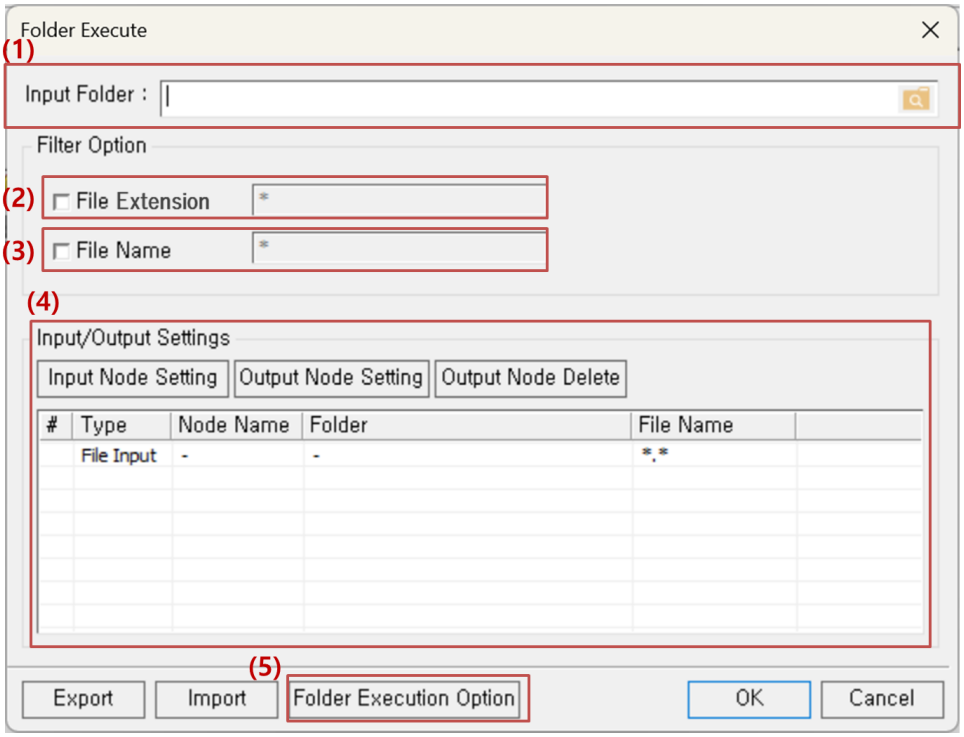
(1) Create the project you want to execute in the workspace.



(2) Select **[Menu] > [Stream(S)] > [Folder Execute].**



(3) Folder Execute Window will appear as shown below.

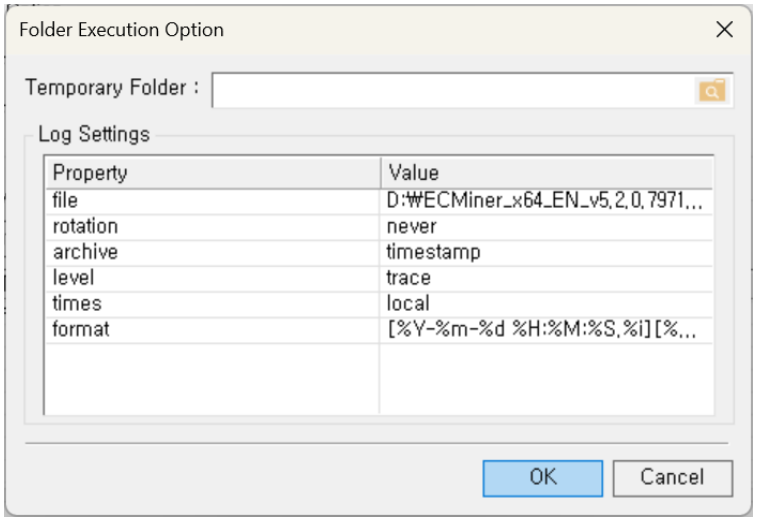


(4) Follow the following steps to set the options for the Folder Execute.

Steps	Setting name		Description
(1)	**Input Folder**		Type in the folder directory or click the right search button to find.
(2)	**Filter Option**	**File Extension**	Type in the file extension of the files within the executing folder.
(3)		**File Name**	Type in the common file name within the executing folder.

(4)	**Input/Output Settings**	Set the Input and Output node within the executing project file by clicking the **Input Node Setting** and **Output Node Setting**. Click the **Output Node Delete** to delete the inserted output node settings.
(5)	**Folder Execution Option**	Log file of the folder execution process can be saved by setting the Folder Execution Options. (Not required)

Note. Folder Execution Option



File	The default log file storage path is the ECMiner™ installation location.
Rotation	Logs can be rotated daily, weekly, or monthly, with the default set to never.
Archive	Log file names can be based on numbers or timestamps, with timestamps as the default.
Level	Only logs at or above the specified level are recorded, with trace capturing the most details.
Times	The reference time for log rotation defaults to local time.
Format	Logs are stored in the format "[%Y-%m-%d %H:%M:%S.%i][%q][%s] %t" by default. * %s : Log record source %t: Log record