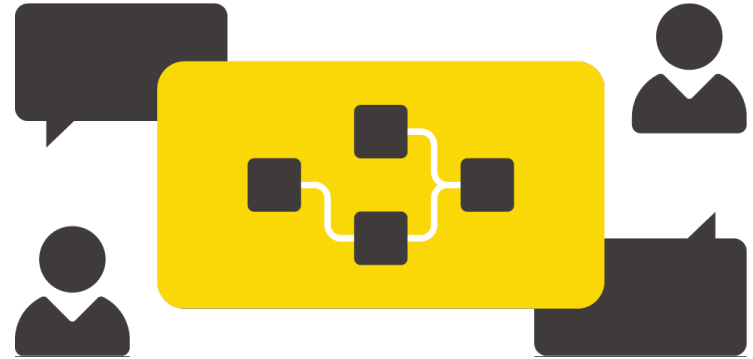


Open for Innovation

**KNIME**

# Working with the new File Handling Framework

KNIME Spring Data Talks



**Tobias Kötter, Björn Lohrmann & Mark Ortman**

March 24, 2021

# Framework Parts

---



File Systems



File Handling  
Nodes



Path Type

# Framework Parts

---



File Systems



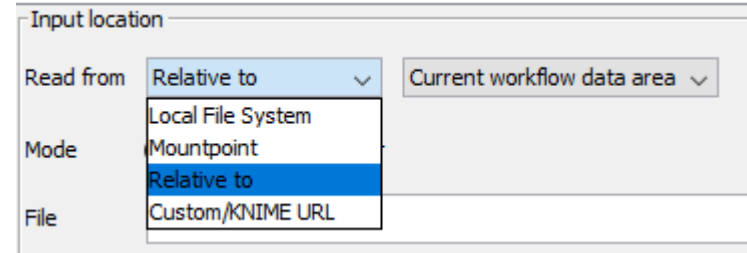
File Handling  
Nodes



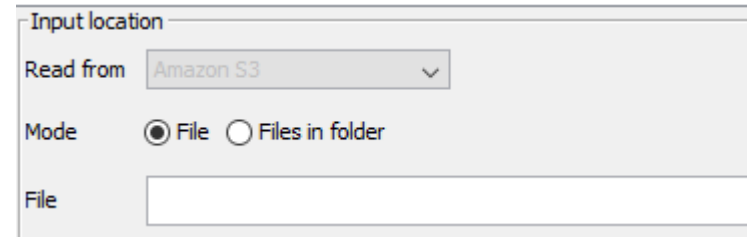
Path Type

# File Systems

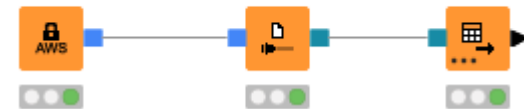
- Standard File Systems
  - Local File System
  - Mountpoint
  - Relative to (workflow, mountpoint, workflow data area)
  - Custom/KNIME URL
- Connected File Systems
  - Microsoft Azure
  - Google
  - Amazon
  - Databricks
  - BigData file systems (hdfs, httpFS, ...)
  - On-premise (e.g. ssh, ftp, ...)
- Planned for Summer Release
  - Azure Data Lake Storage Gen2 (available in [nightly build](#))
  - SMB (Samba)



CSV Reader



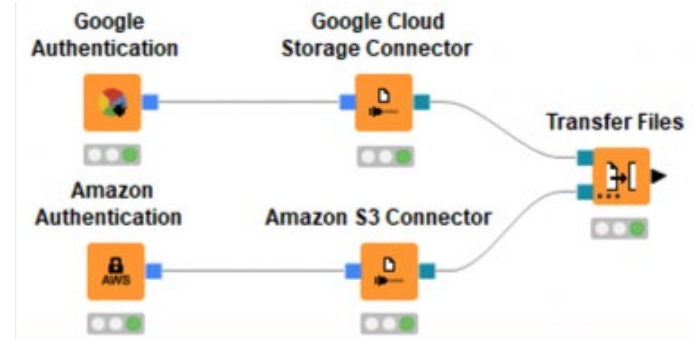
Amazon Authentication Amazon S3 Connector CSV Reader



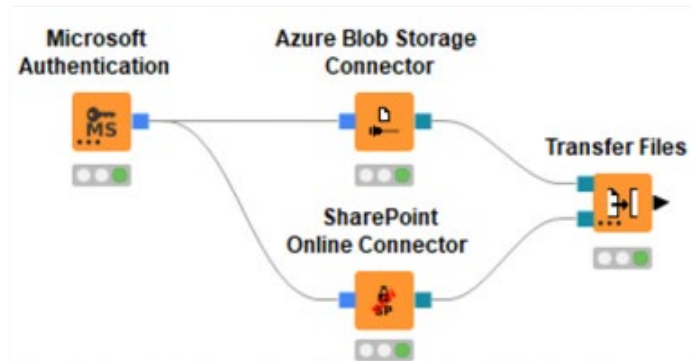
# Full Flexibility



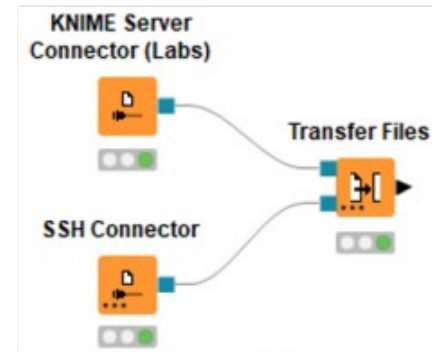
Local file system and KNIME mountpoints



Cross cloud environments

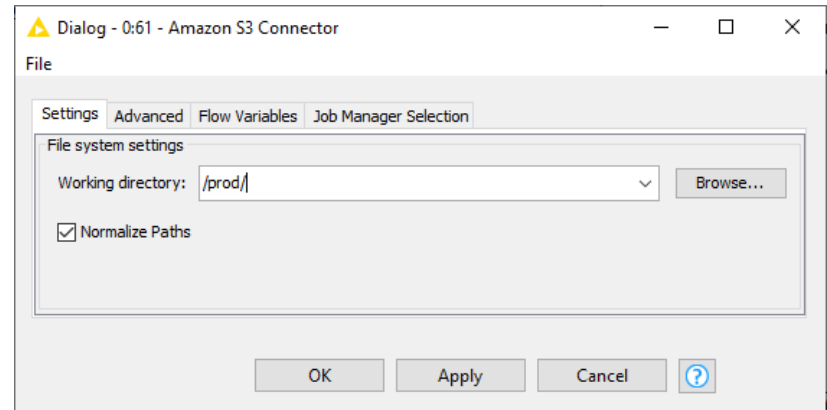
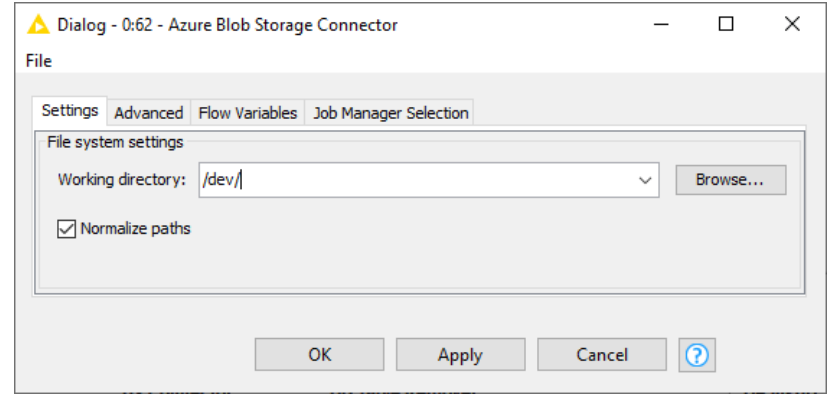
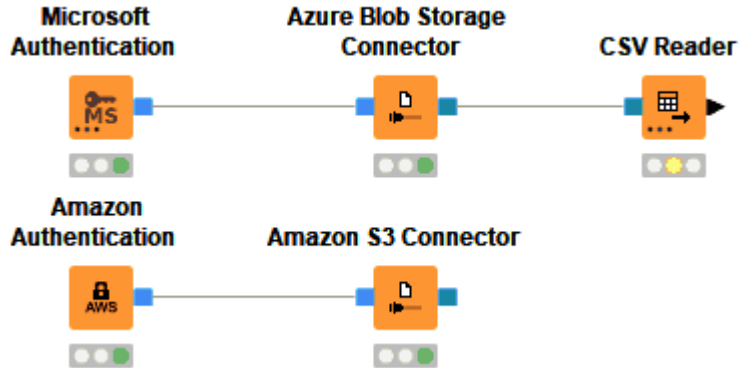


Same cloud environment



On-premise

# Seamless Migration



# Framework Parts

---



File Systems



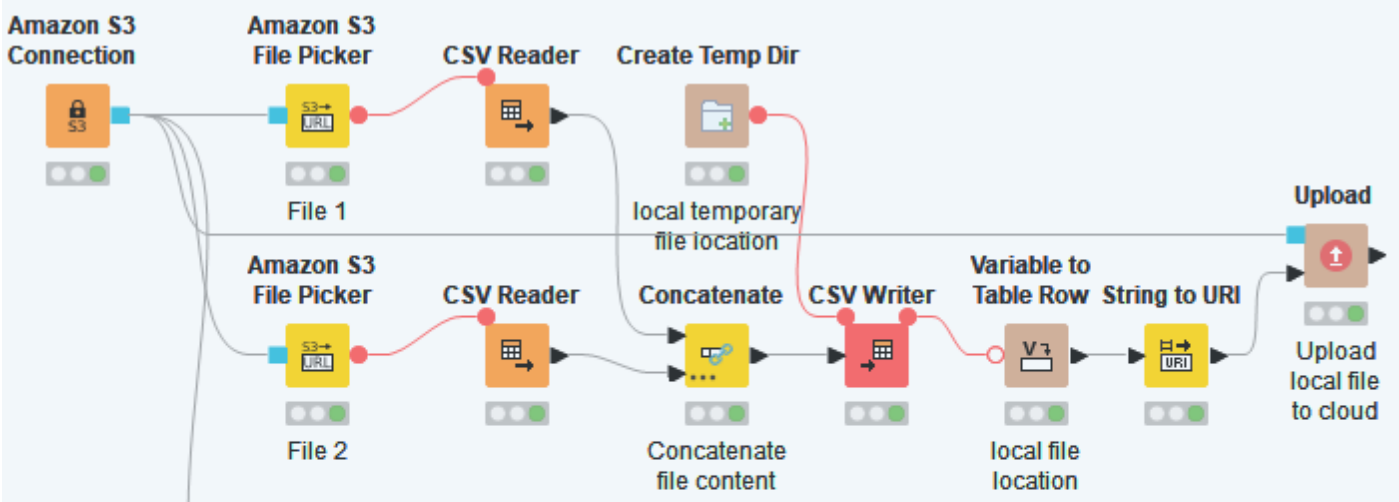
File Handling  
Nodes



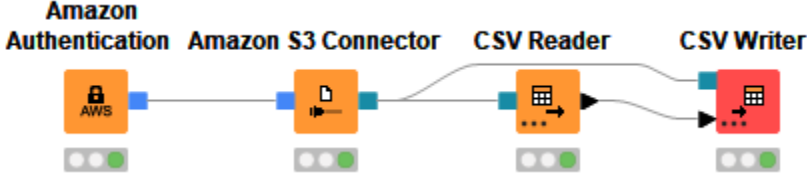
Path Type

# Working with Remote Files

Old



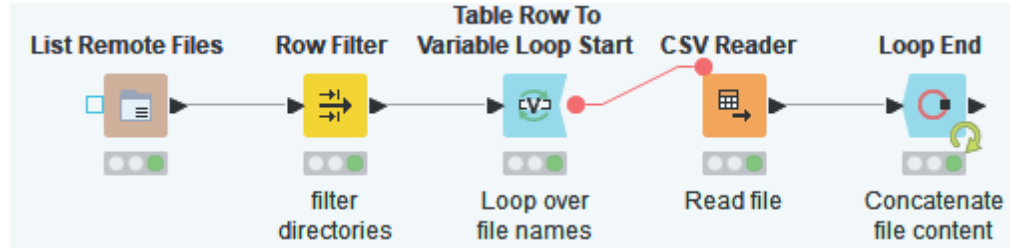
New






# Files in Folder

Old



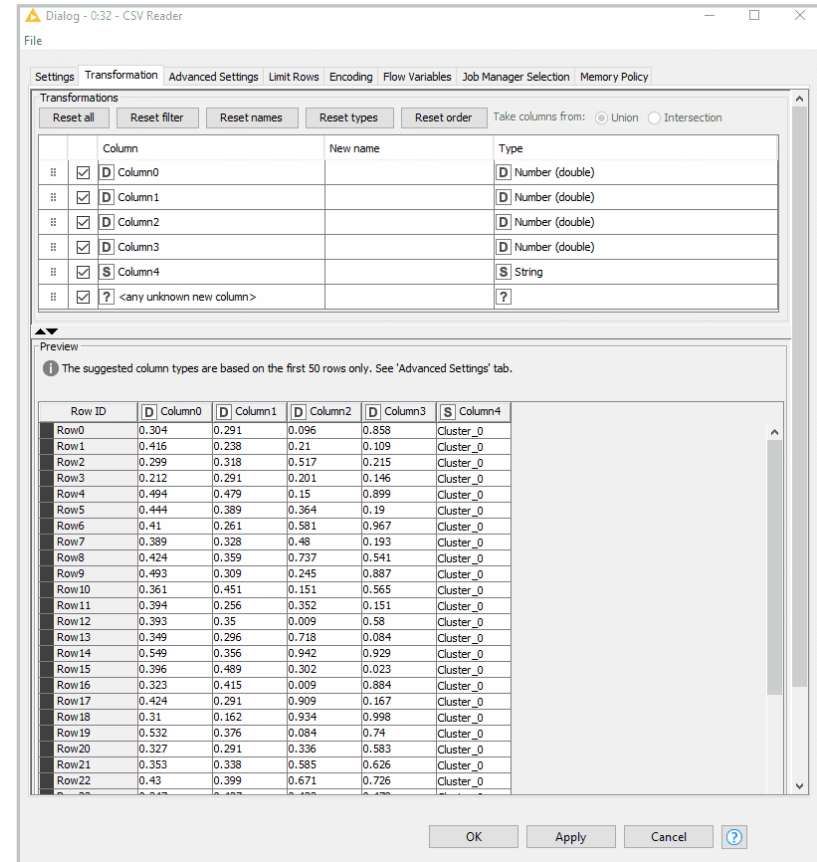
New

**CSV Reader**  
  
**Files in folder**

The screenshot shows the **Dialog - 3:30 - CSV Reader** window. The **File** tab is active. Under **Input location**, **Read from** is set to **Local File System**. Under **Mode**, **Files in folder** is selected, and the **Filter options** button is highlighted. The **Folder** is set to **c:\temp**, and **Include subfolders** is checked. Below, the **Reader options** section shows **Format** set to **Autodetect format**, **Column delimiter** as **,**, **Row delimiter** as **\r\n**, **Quote char** as **"**, and **Quote escape char** as **"**. A **Filter options** dialog box is open, showing **File filter options** with **File extension(s)** set to **csv**, **Case sensitive** unchecked, **File name** set to **\***, **Case sensitive** unchecked, **Wildcard** selected, **Regular expression** unselected, and **Include hidden files** unchecked. **Folder filter options** are also shown with **Folder name** set to **\***, **Case sensitive** unchecked, **Wildcard** selected, **Regular expression** unselected, and **Include hidden folders** unchecked. **OK** and **Cancel** buttons are at the bottom.

# Easy to Use Data Transformation

- Available
  - via dedicated node with table inputs
  - and contained in every new reader node that has a table output
- Supported operations
  - Column sorting
  - Column filtering
  - Column renaming
  - Column type mapping
  - Concatenation of multiple files/tables
- Live preview provides instant feedback of the changes



# Framework Parts

---



File Systems



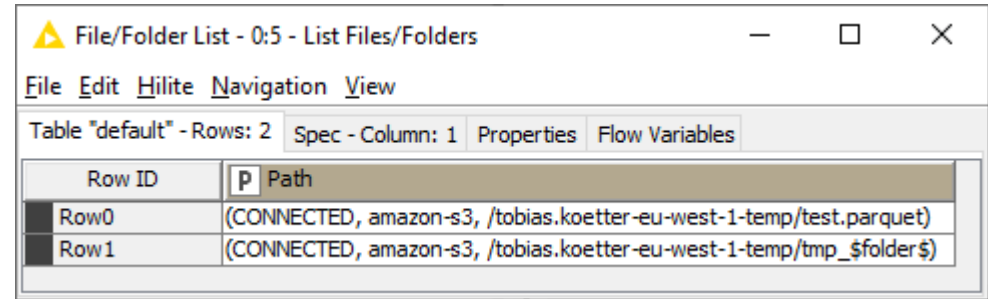
File Handling  
Nodes



Path Type

# Path Data Cell and Flow Variable

- Consists of three parts:
  - Type
  - Specifier
  - File path

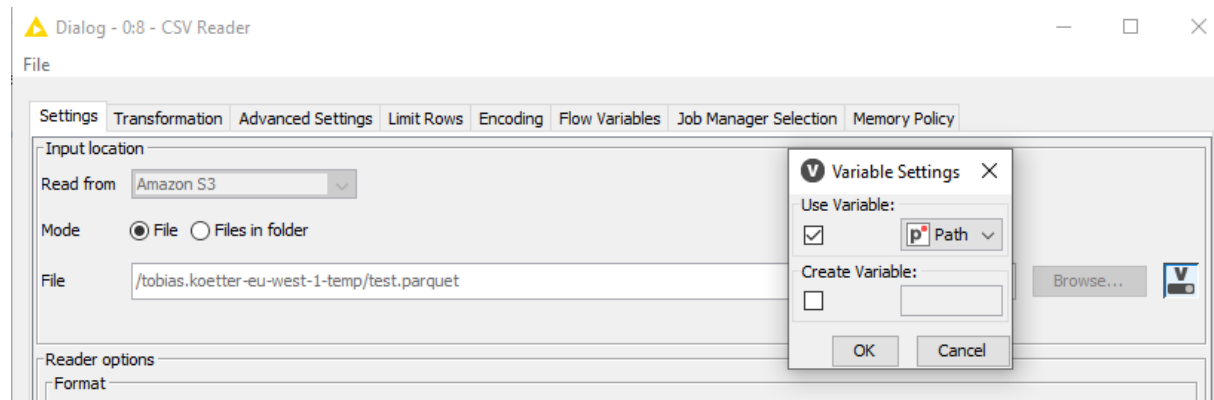


File/Folder List - 0:5 - List Files/Folders

File Edit Hilite Navigation View

Table "default" - Rows: 2 Spec - Column: 1 Properties Flow Variables

Row ID	P Path
Row0	(CONNECTED, amazon-s3, /tobias.koetter-eu-west-1-temp/test.parquet)
Row1	(CONNECTED, amazon-s3, /tobias.koetter-eu-west-1-temp/tmp_\$folder\$)



Dialog - 0:8 - CSV Reader

File

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables Job Manager Selection Memory Policy

Input location

Read from: Amazon S3

Mode:  File  Files in folder

File: /tobias.koetter-eu-west-1-temp/test.parquet

Reader options

Format

Variable Settings

Use Variable:  Path

Create Variable:  [ ]

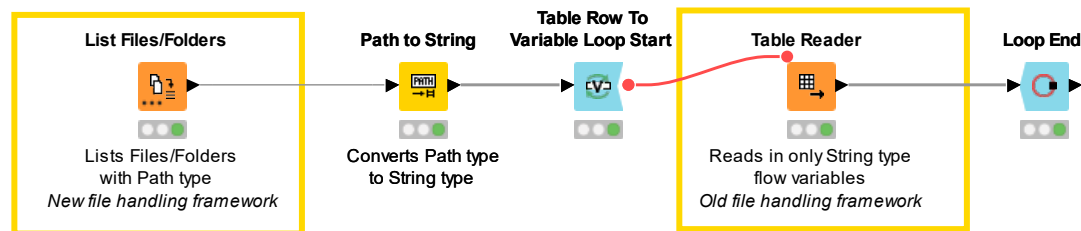
OK Cancel

# Compatibility

- New path type not compatible with old nodes!
- Not all KNIME and community nodes have been converted yet
- Identify converted nodes via their dynamic input port node icon



- [Path to String](#) and [Path to String \(Variable\)](#) node convert new path type to old String representation



- Check out the [Compatibility and migration](#) section in the [File Handling Guide](#)

# More Information

1. Documentation:  
[KNIME File Handling Guide](#)
2. Blog:  
[New File Handling Out of Labs and Into Production](#)
3. Video:  
[Watch Bernd talk about the new File Handling Framework](#)
4. Example Workflows:  
[Working with Utility nodes](#)  
[Data transfer between clouds](#)
5. Questions:  
<https://forum.knime.com/>

