[Azure Synapse Analytics] Polybase reject rows are not output if reject rows oversize reject value

Last updated by | Veena Pachauri | Mar 9, 2023 at 12:00 AM PST
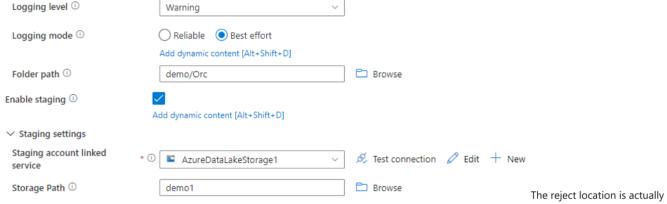

Contents
- Issue
- Resolution

Issue

If customer using polybase to load data into Azure Synapse Analytics, if they set the reject value and logging location in order to output the error rows and the job failed because the # of reject rows is more than reject value, the error row files may not be output to the specified location as expected.

Resolution

As the requirment from polybase public doc ⧉, the reject row location should be a directory within the External Data Source. So if the logging storage and source storage location (staging storage location if enable staging) are in the same container, the error rows can be output for sure.

If you set the two locations as different containers, in fact the error rows will be created within the source data container as a wrong folder path. For example,

The reject location is actually demo1/demo/Orc rather than demo/Orc.

However if the reject row location and source storage location (staging storage location if enable staging) are in different containers, we have a post logic to move the reject files from a wrong folder to the target folder only if the job run successfully. If the job failed, the logic will not be applied. Currently we don't have a plan to add this post logic for failed run, please suggest customer to change the two location into the same storage and same container (the rest folder path can be different).

How good have you found this content?