

Best practices for develop efficient and scalable Power BI DataFlows

Follow the following

- ❖ Use multiple dataflows and link entities
 - To improve performance of large size dataflows, first create a separate dataflow for connecting to the data sources. Perform all transformations needed on the entities as part of the next dataflows by linking the entities. This will ensure to use the enhanced compute engine of Power BI premium tenant and provide a better performance.
- ❖ Limit number of data files to load from SharePoint
 - When working with SharePoint and connecting to multiple files, you may notice sporadic failures. SharePoint throttles multiple requests for data files coming at the same time to ensure it remains reliable and responsive. Splitting the SharePoint data files into multiple dataflows will by a better response.
- ❖ Use multiple workspaces to schedule dataflow for refresh
 - Avoid setting a refresh schedule for a linked dataflow in the same workspace as the source dataflow to avoid the locking behavior
- ❖ Separate data loading and data transformation dataflows by workspaces
 - Permissions on dataflow access can be better managed by separate dataflows on different workspaces
- ❖ Co-locate data lake and Power BI tenant in the same region
 - Dataflows do not currently support multi-geo regions. The Premium capacity must be in the same region as your Power BI tenant.
- ❖ Separate on-premises sources from cloud sources
 - create a separate dataflow for each type of source, such as on-premises, cloud, SQL Server, Oracle, SharePoint etc..
- ❖ Separate dataflows into a separate capacity
 - Separate your dataflows, or scaling up your Premium capacity for additional memory.