**Shared dataset or local dataset**

**Introduction**

When working with multiple data sources, keeping track of the sources can become burdensome. To complicate matters further, as you begin collaborating with other data analysts and more departments, data can be duplicated, copied, and shared across multiple devices. This has the potential to lead to inaccuracies and inconsistencies in reports. There is no one source of truth to the data. Fortunately, Microsoft Power BI has a solution to this problem—datasets.

In this reading, you'll explore how to set up a shared dataset, the difference between local and shared datasets, and when shared datasets are suitable for different business scenarios.

**What are datasets?**

Power BI datasets represent a source of data ready for reporting and visualization. You can create Power BI datasets in the following ways:

* Connect to an existing data model not hosted in Power BI, such as a SQL database.
* Upload a Power BI Desktop file that contains a model to Power BI Service.
* Upload an Excel workbook that contains one or more Excel tables and/or a workbook data model, or upload a comma-separated values (CSV) file to Power BI Service.
* Use the Power BI service to create a dataset.

Datasets allow you to group data into a single repository. The dataset can then act as a single source of truth for reporting.

Now, let's explore the differences between local and shared datasets.

**Local datasets**

Many businesses operate local datasets, sometimes in the form of local files but more often in the form of SQL databases running on in-house servers. There are many advantages and disadvantages to consider when working with local datasets.

**Data control**

As the owner, local datasets offer the business total control of the data. This makes the dataset more secure since it is not accessible to anyone outside the organization.

**Access speed**

Storing a dataset on a local computer or server allows data to be accessed more quickly. Data transfer speed is higher when using local storage. Storing data on local devices eliminates the need for accessing the internet, which can take time, and removes the limitations imposed by the bandwidth constraints of the internet service provider.

**Limited access**

If your team works remotely, accessing a local dataset can be problematic if they are not on your local device. An IT department will often be required to establish a way to access the internal network securely. This presents potential security issues and introduces access speed considerations. For executives who may be traveling, this could present issues in accessing dashboards needed for strategic decisions.

**Scalability issues**

Local datasets offer limited storage capacity based on the purchased computer or server hardware. If the business works with large datasets, this may present scalability issues. For example, if working with historical sales data, the business may be required to reduce the amount of historical data kept, making it more challenging to extract projections using AI forecasting.

**Shared datasets**

As an organization scales, especially with multiple office locations, remote workers, and business executives traveling worldwide, shared datasets become a crucial part of the business. Having one source of truth for reporting allows stakeholders to be aligned in decision-making based on data insights. There are several factors to consider when utilizing shared datasets.

**Data control and governance**

Dataset governance is the ability to control and manage the use of shared datasets in Power BI workspaces. It involves setting permissions and access levels and ensuring that datasets are properly managed and maintained for accuracy and consistency in reporting.

**Access speed**

One consideration to take into account for shared datasets is internet connectivity and data transfer speed, which may slow down access to the reports and dashboards that utilize the shared datasets. Time should be allocated to optimizing visualizations and reports to reduce the amount of data that needs to be transferred from the cloud to the local device.

**Remote access**

Multiple individuals or organizations can access shared datasets, making collaboration and data sharing easier. For remote teams and executives on the go, accessing reports and dashboards while out of the office helps to ensure correct and strategic decision-making.

**Scalability**

If data sources and datasets are cloud-hosted, cloud resources can be used to perform complex calculations without worrying about the limited capacity of local devices. Furthermore, shared datasets in Power BI unlock the ability to build reports on top of existing published datasets within the organization. This eliminates the need to create a separate data model, saving time and effort. Multiple users can access and modify the same dataset, promoting collaboration and efficiency.

**Conclusion**

Datasets are a powerful tool to ensure the consistency, standardization, and accuracy of data. As a business grows, maintaining datasets becomes increasingly important for business intelligence. Without accurate reporting, stakeholders may make incorrect decisions that negatively impact the business. As you establish datasets in an organization, ensure that time is invested into the maintenance and auditing of datasets and what they contain to ensure continued business success.