**What is dbt (data build tool)**

**A gentle introduction to dbt that is taking over the data world**



Photo by [Robin Pierre](https://unsplash.com/@robinpierre?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on [Unsplash](https://unsplash.com/s/photos/data?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText)

dbt, or **d**ata **b**uild **t**ool, is an open-source command-line tool that helps organizations build, test, and maintain their data infrastructure. The tool is designed to make it easier for data analysts and engineers to work with data, by providing a consistent and standardized approach to data transformation and…

dbt allows users to define their data models using SQL, and then uses these models to generate optimized SQL code that can be run against a data warehouse or other data storage system. This allows users to build a maintainable and scalable data infrastructure that can be easily updated and extended over time.

In addition to generating SQL code, dbt also provides a number of features that make it easier to work with data. These features include the ability to manage dependencies between data models, run tests to ensure data integrity, and track the lineage of data to understand how it has been transformed over time.

**Use-Cases for dbt**

dbt can be used in a variety of ways. Some common use cases for the technology include:

* **Building and maintaining data pipelines**: dbt can be used to define data models using SQL, and then generate optimized SQL code that can be run against a data warehouse or other data storage system. This allows users to build and maintain a scalable data infrastructure.
* **Ensuring data quality and integrity**: dbt provides a number of features that make it easier to ensure the quality and integrity of data. This includes the ability to run tests to validate data, as well as track the lineage of data to understand how it has been transformed over time.
* **Standardizing data transformation processes**: dbt provides a consistent and standardized approach to data transformation and analysis, making it easier for data analysts and engineers to work with data. This can help organizations improve the quality and reliability of their data, and make it easier to extract insights and drive business decisions.
* **Providing a collaborative environment for data teams**: dbt allows data analysts and engineers to work together on the same data models and transformations, providing a collaborative environment for data teams. This can help improve communication and collaboration within data teams, and make it easier to work on complex data projects

**dbt-core vs dbt Cloud**

dbt-core and dbt Cloud are two different products offered by Fishtown Analytics, the creators of data build tool.

dbt-core is an open-source command-line tool that allows users to define their data models using SQL, and then uses these models to generate optimized SQL code that can be run against a data warehouse or other data storage system.

dbt Cloud, on the other hand, is a cloud-based platform that provides additional features and functionality on top of dbt-core. dbt Cloud provides a web-based interface for managing data models, as well as additional features such as scheduling, collaboration tools, and integrations with other data tools.

In summary, dbt-core is the underlying open-source tool that powers dbt, while dbt Cloud is a cloud-based platform that provides additional features and functionality. dbt-core can be used on its own, or in conjunction with dbt Cloud to provide a more comprehensive data infrastructure solution.

**What dbt isn’t?**

dbt is **not a data warehouse** or a database itself, but rather a tool that can be used in conjunction with a data warehouse to make it easier to work with and manage data. Additionally, dbt is **not a programming language**, but it does use a programming-like syntax to specify how data should be transformed and loaded into a data warehouse. It is also **not a visualization tool**, although it can be used in conjunction with visualization tools like Tableau or Looker to help users understand and analyze their data

**Getting started with dbt**

If you are new to dbt and would like to kick start new dbt projects, you can refer to some of my recent articles that will help you [install dbt for your specific data warehouse or storage](https://towardsdatascience.com/install-dbt-1bd6a4259a14) and [structure your dbt projects](https://towardsdatascience.com/dbt-models-structure-c31c8977b5fc) in a meaningful and scalable way.

**Final Thoughts**

Overall, dbt is a powerful tool that can help organizations improve their data infrastructure and make it easier for data analysts and engineers to work with data. By providing a consistent and standardized approach to data transformation and analysis, dbt can help organizations improve the quality and reliability of their data, and make it easier to extract insights and drive business decisions.