



@ . ' O . . .  
O . ' °

Thank you for downloading this Voltron Data resource. Carahsoft is the distributor for Voltron Data's AI and ML solutions available via NASA SEWP V, ITES-SW2, NASPO ValuePoint, and many more contract vehicles.

To learn how to take the next step toward acquiring Voltron Data's solutions, please check out the following resources and information:



For Voltron Data overview:  
[carah.io/Voltron-Data](https://carah.io/Voltron-Data)



For upcoming events:  
[carah.io/Voltron-Events](https://carah.io/Voltron-Events)



For additional resources:  
[carah.io/Voltron-Resources](https://carah.io/Voltron-Resources)



For additional Artificial Intelligence:  
[carah.io/ai-solutions](https://carah.io/ai-solutions)



To set up a meeting:  
[VoltronData@carahsoft.com](mailto:VoltronData@carahsoft.com)



To purchase, check out the contract vehicles available for procurement:  
[carah.io/procurement](https://carah.io/procurement)

# Ibis and Snowflake at the Speed of Arrow

Kae Suarez, Phillip Cloud



[Arrow](#) is the open source standard for columnar data in memory. The project helps enterprises accelerate their stacks due to its benefits for parallelization and access patterns that enable faster execution for almost any task in data analytics.

Recently, the community at the [Ibis Project](#), the Python interface for data analytics, looked into the gains from using Arrow when interfacing with Snowflake and saw 3x - 5x speed gains when accessing large datasets. Today, we'll explore what this looks like and what it means for Snowflake — and what it could mean for you.

## Snowflake

[Snowflake](#) is a powerful cloud database platform with impressive scaling capabilities that enable analytics of any size. With an emphasis on using the best technologies for the task, Snowflake uses columnar data to enhance performance and access patterns. This is one of many decisions that enables high performance in Snowflake, and it performs at its best when other applications in the stack match this design, as it reduces conversions between row and column-based memory structures.

## Ibis

[Ibis is the Python-native interface for data analytics](#) and connects to a variety of backends, including Snowflake. Each backend has considerations for the shape and traits of its output. By default, Ibis outputs all final results to pandas Dataframes for maximum compatibility. However, columnar data is powerful, and the community at the Ibis Project knows this, so any output can instead be directed to a PyArrow table. Furthermore, when retrieving data, Ibis can accept Arrow data from queries if the backend has support for such output.

## Snowflake + Ibis

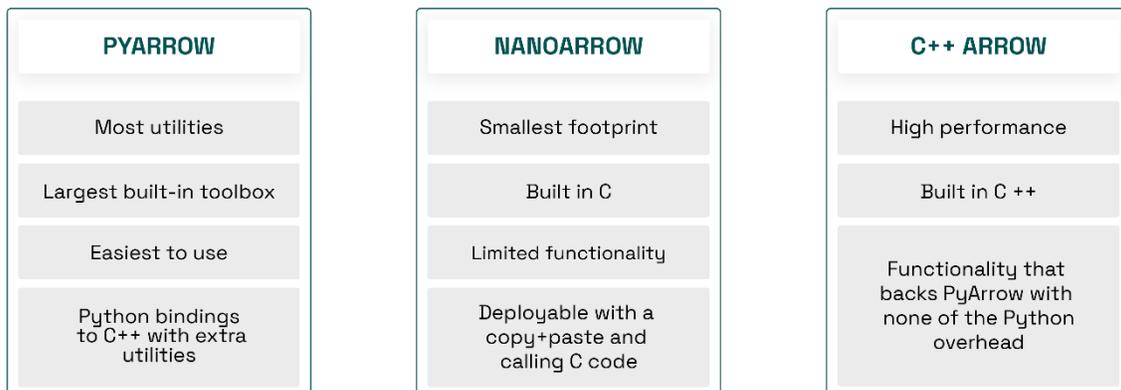
Since Snowflake uses Arrow, and Ibis supports Arrow, Ibis accepts data from Snowflake in Arrow format. However, this support was optional, and performance differences were untested. New tests show a 3x - 5x speedup when retrieving datasets in Arrow format. Thus, the Ibis Project community has now made Arrow the primary format for data fetch from Snowflake — there is no reason to leave that much performance on the table. It was a small, easily accessible change, and ensures no one accidentally holds back their applications anymore.

Other backends, given Arrow support, will get the same treatment, as the Arrow in-memory format fuels performance within and between tools. In Snowflake, columnar data backs the high performance of cloud data computing, and in Ibis, Arrow provides the fastest way to get data from backend to user. A fully Arrow-based stack is more and more feasible with each tool that supports it, and Ibis + Snowflake is simply one of the newest combinations to profit.

## Conclusion

This is just another in a long line of use cases where Arrow enables interoperability and computation easily and efficiently. There are [several more success stories](#), including Velox, from Meta.

If you want to join in, there are plenty of routes. As a starting place, here are three common paths to Arrow that we recommend — they all fulfill different roles, but every single one gets you to the columnar format that we love to use.



Voltron Data designs and builds composable data systems using standards like Arrow, Ibis, Substrait, and more. Check out our [Product page](#) to learn more about our approach.