

# Spark Native Accelerators and Associated Technologies

## Landscape

There are a few budding technologies that are looking to accelerate Spark, generally by replacing the execution engine with a more efficient one. As of 2024-05-30, many of these technologies are not yet stable.

Feel free to email me at [trent@trenthauck.com](mailto:trent@trenthauck.com) if you have any suggestions or corrections.

Description	Execution Engine	Language	Related Companies	Open Source	TPC-H Link <sup>1</sup>
Gluten + Velox	Velox	C++	Meta wrote Velox, MSFT is using Gluten + Velox in Fabric	Yes	Velox TPC-H
DataFusion Comet	DataFusion	Rust	Apple released DataFusion Comet	Yes	DataFusion Comet TPC-H
Photon	Photon	C++	Databricks	No	N/A
Blaze	DataFusion	Rust	Kwai	Yes	Blaze TPC-H
RAPIDS	RAPIDS	C++/Cuda	NVIDIA	Yes	N/A

## Component Notes

- Gluten dubs itself a middle layer for offloading computation to native engines. In practice, it seems to be used primarily with Velox.
- Velox is a C++ execution engine developed by Meta. It is used Fabric from Microsoft.
- Apache DataFusion Comet is a native execution engine for Spark written in Rust. It is based of Apache DataFusion and was released by Apple.
- Photon is a C++ execution engine developed by Databricks available on their platform.
- RAPIDS is a suite of libraries for data science and analytics that uses GPUs. It includes a Spark plugin. <https://github.com/NVIDIA/spark-rapids-jni/> contains the bindings implemented in Cuda and C++.

---

<sup>1</sup>These are links, if available, to TPC-H benchmarks for the given technology. They may not use the same scale factor or configuration as other technologies, so please be cautious when comparing them. Also, if you have better links, please let me know.