

The Data Thread Conference started with an introduction of the event by Marlene Mhagami, the Developer Advocate and a Python programmer for Voltron Data. The purpose of the event was to introduce Apache Arrow and its applications in High Performance Computing. The speakers included in the live session included: Wes McKinney, Jacques Nadeu, Carlos Maltzahn, Jim Pivarski, Fernanda Foertter, Peter Wang, and Sebastian Extevez.

The first session featured a presentation by Wes McKinney, a co-founder and the Chief Technical Officer of Voltron Data, and Jacques Nadeu, a co-founder and the Chief Executive Officer of Sundeck, on the origins and the foundation of Apache Arrow. The two are co-creators of Apache Arrow. The most important point in the presentation was the community basis of the Apache Arrow project, especially its development with a diverse community of engineers and the project's willingness to include other communities, which allowed for further growth and the increased sense of community ownership with the project. Nadeu also commented on the success of an open-source project such as Apache Arrow as "a willingness to share ownership of the project." Being an open source project, Apache Arrow was aided by many individual contributions that allowed the project to span through a variety of programming languages and formats like Python and Rust.

The second presentation was an introduction to High Performance Computing(HPC) and its applications moderated by Jing Brewer, Vice-President of Market Strategy at Voltron Data, and featured an interview with Carlos Maltzhan, a founder and director of the University of California, Santa Cruz, Jim Pivarski, a computational physicist at Princeton University, and Fernanda Foertter, Director of DevRel and HPC Business Development at Voltron Data. The main topic of this interview was on what HPC is, current issues with HPC, what kinds of organizations have access to HPC and current trends in the development of HPC. The fundamentals of HPC discussed defined it as a conglomeration of computing cores that are able to process massive amounts of data. The main issues facing HPC and its development discussed in the interview was the need for more efficient methods to transfer data into the ever growing processing capabilities of new processing cores. The interview then shifted towards the availability of HPC resources. This question was addressed by mentioning open HPC resources available to smaller-scale organizations and groups. The interview concluded with the addressing of a desired development in HPC, where Maltzhan and Pivarski expressed their desire for more collaborations and ease of access to the field in order to expand the field. Foertter, on the other

hand, had expressed the need for more efficient data transfer rather than simply moving large amounts of data.

The third panel was a fireside chat with Peter Wang, the Chief Executive Officer of Anaconda Inc. The focus of the panel was to expand the perspective of the audience through the fireside chat where Wang discussed a wide variety of topics ranging from the advent of HPC to the positives of the rise of open-source resources at the global scale to the future development of data processing.

The final panel was a presentation by Sebastian Extevez, an engineer at DataStax. on the future prospects of Apache Arrow and its uses when combined with other technologies. Through the panel, Extevez described the uses of Cassandra and Apache Arrow in machine learning and how the two can be combined. For example, Cassandra and Apache Arrow can be combined to create a program that is capable of both research analytics and marketplace functionalities. This would allow for a reduction in time between the consumer data and the data analytics.

Overall, this event has enlightened me to the possibilities of Apache Arrow and the potential that the project has to impact the world of data analytics. The thing I was the most amazed about throughout the entire conference was the impact of the community on many of these projects. The community would always respond to the effort put into an open-source project and expand upon it to become better than any single team could.