Voltron Conference

The Voltron Conference was very interesting and made me feel so optimistic and ambitious to be a part of a conference like it one day, being around successful people who have achieved a lot of their own dreams and who are willing to share their experiences and the obstacles they have faced is always a great experience and will always be ambitious and optimistic to be like them one day.

Speakers were Wes McKinney CTO/Co-Founder at Voltron Data Apache Arrow PMC Creator Python pandas, Author python for Data Analysis, and also Jacques Nadeau CEO/Co-founder at Sundeck.io, they were the first speakers at the conference, and they focused on a lot of different points and mentioned a lot of different areas that they wanted to focus on during their Voltron data journey.

Other speakers were Carlos Malzahn who is the founder & Director University of California Santa Cruz (CROSS), Jim Pilarski who is a Computational physicist, Princeton University, Fernanda Foerster Director of DevRel and HPC business Development, Voltron Data, Peter Wang Chief Executive officer Anaconda INC., Sebastian Estevez Engineer DataStax

Both Jacques Nadeau and Wes Mckinney started the talk about the Voltron data journey, and they mentioned that they wanted to grow a community KATAMI STYLE in a way that means they wanted people who are not afraid to experiment.

Arrow is the go-to project for data analytics systems that need to move and process data fast. Starting in 2016, Arrow has since become the de facto standard for efficient in-memory columnar analytics and low-overhead data transport, with support for more than 12 programming languages. Beyond providing an in-memory data format and protocols for interoperability, Arrow offers an expansive toolbox of modular computing libraries to provide a strong foundation for the next generation of analytical computing systems.

When they started hiring people, they mainly focused on making room at the table for more leaders and listening to everybody else ideas and thoughts and which lead them to have a big part in ambitious projects, also Wes McKinney also mentioned” when hiring someone, hire someone who can help be the replacement”.

Jim Pivarskia mentioned HPC is very hard to maintain and that it requires newer developers and good trends and with these resources, it will be good to see it continue to collaborate and contribute.

Carlos Malzahn who specialized in storage systems mentioned that production data is a challenge and if something goes wrong you have to restart the simulation and do checkpoints and get data from the storage system.

Fernanda mentioned that in the future she wishes that there will be performance extract, and find ways that are more efficient for computer resources that take advantages without bad impacts.

Peter Wang mentioned that he loved python as a language, executable sudo code, ease of use turned to an easier way to accessibility, he also stressed “Code is Law” in a way that coding became more political when making it easy to use which was very odd to him and that people with agency and data want to have the most power out there, making the code hard and finding a harder way to use isn’t the code problem it is the people problem more than being a technical problem, and that having python as an easy to use source was the whole plan of working together to build standards that work for everyone.

Sebastian Estevez talked about Apache Cassandra and mentioned that it is a distributed NoSQL database optimized for transactional workloads (OLTP) which are designed for performance, and availability, and uses a log-structured merge for persistence, and drives some of the biggest transactional applications in the world.

In the memory format and ecosystem of standards designed for analytics like language-independent Columnar and nesting, zero-copy data transfer, and an optimized for vectorized processing like SIMD, GPU, and Coprocessors.

AstraDB is a DataStax serverless multi-tenant database as a service that has bills per request/storage/data transfer, support for REST, graphql, and APIs via stargate.io, and it also splits Cassandra into services that scale independently through coordination, data, streaming data to data nodes.

Rodrigo Aramburu was also a speaker in the Voltron as well and he talked about Ibis as he said it is an open-source Python framework for accessing data and performing analytical computations using multiple backends, including SQL databases like PostgreSQL and MySQL and big data systems like Google’s BigQuery. Use this lightweight interface to future-proof possible migrations, unify SQL calls for diverse backends, reduce compute load on the frontend, and provide users with a powerful expression language that is familiar and cohesive to read.

Resources:

<https://www.youtube.com/watch?v=QN_NVseO-NA>