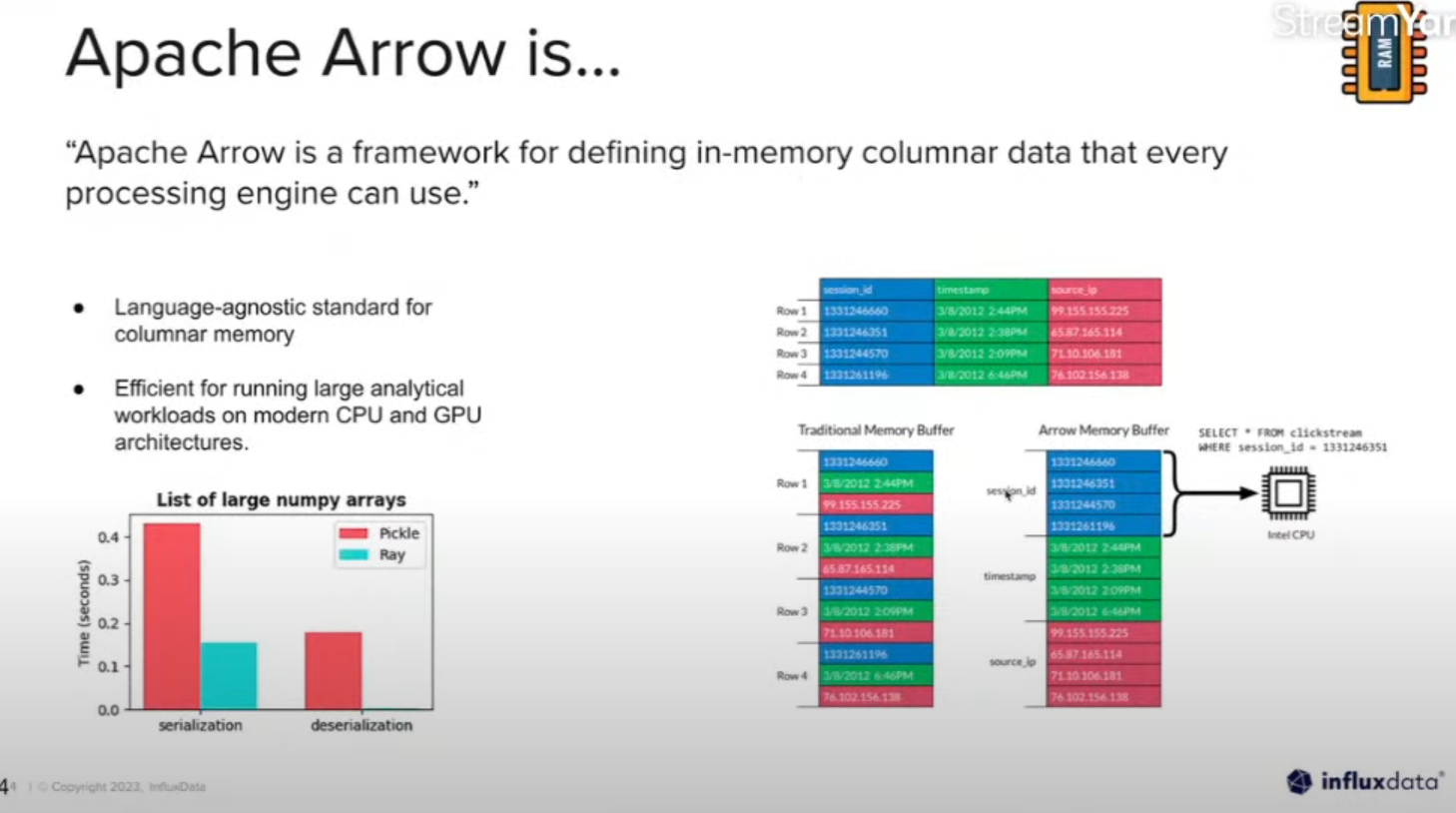


1. Apache arrow is in memory data format
2. Parquet is a file based format
3. Apache arrow flight is a transportation protocol for columnar data



Apache arrow is a framework that provides a cross language in memory data format designed to improve the performance and efficiency of big data processing.

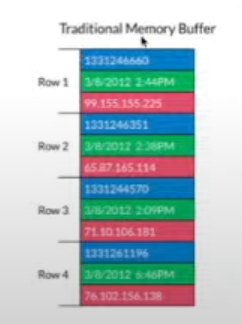
Basically allows data to be easily passed between different frameworks and languages (programming). Also provide a standard way of representing complex data structures in flat memory format.

For example:

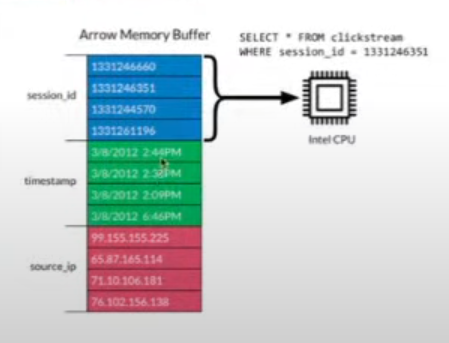
We have this simple monitoring data. And if we are going to save this in our memory:



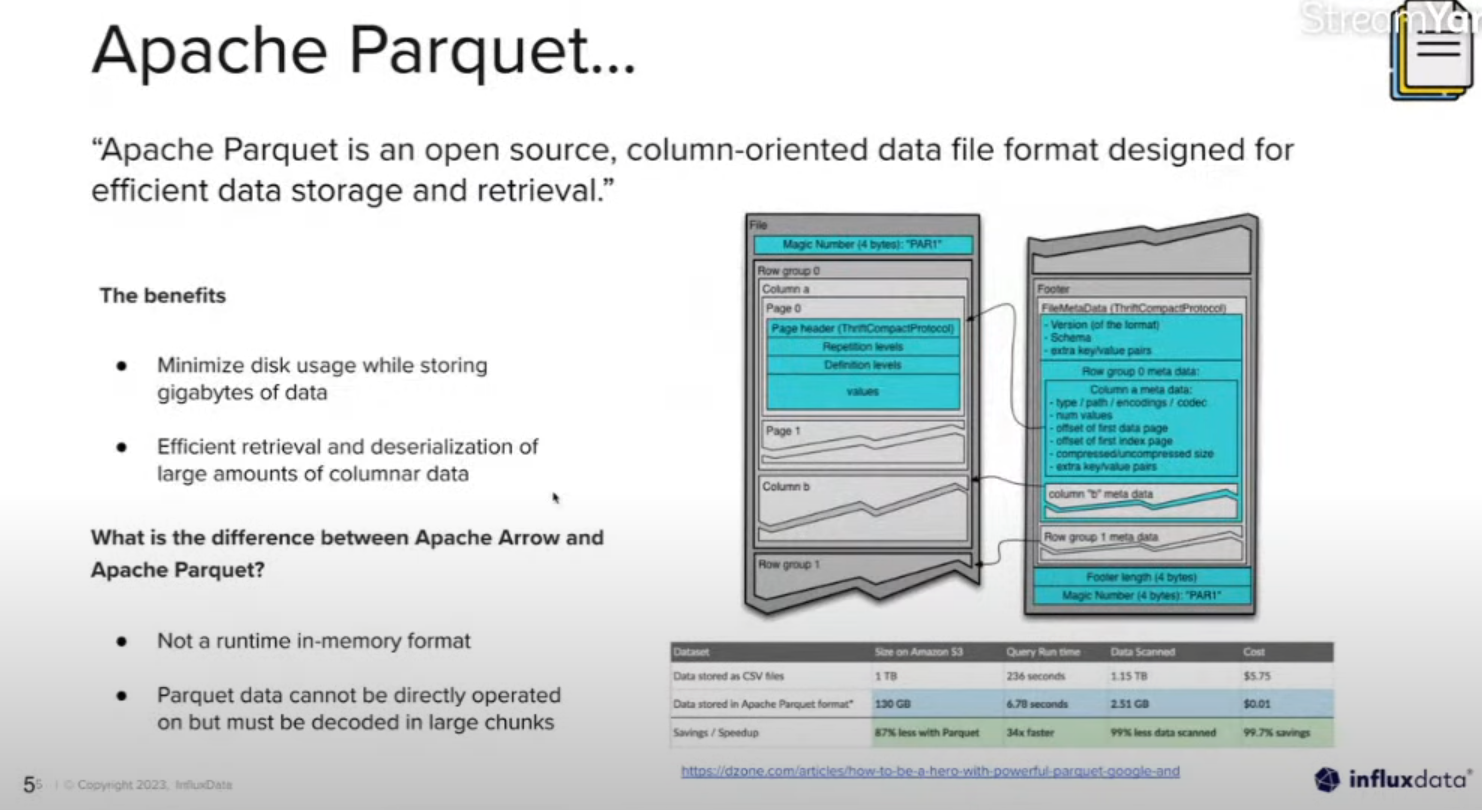
This would look like in our memory which is not efficient:



But with pyarrow in-memory format it was designed to store each column individually and its more efficient for doing big data processing.



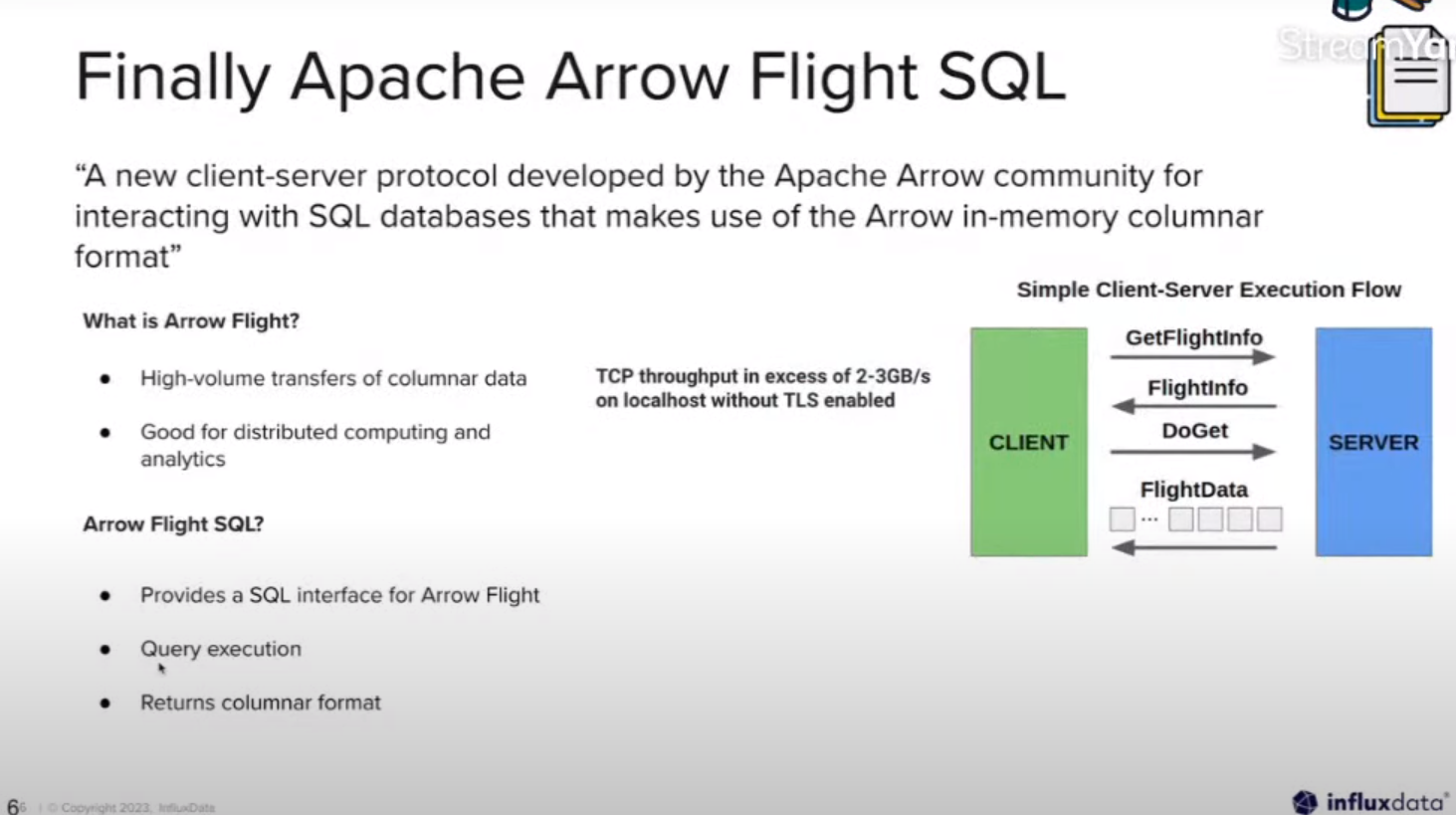
Basically using apache arrow was designed to be data efficient across many languages where we can implement it in java, python, c++ and many more that we can interact between different programming languages.



Basically, Apache Parquet is all about how data is organized, compressed, and stored efficiently on physical storage disks.

Apache arrow vs apache parquet:

Arrow focuses on how data is structured in memory and facilitates efficient data interchange between languages, while Parquet focuses on how data is stored on disk to enable efficient storage and retrieval in distributed data



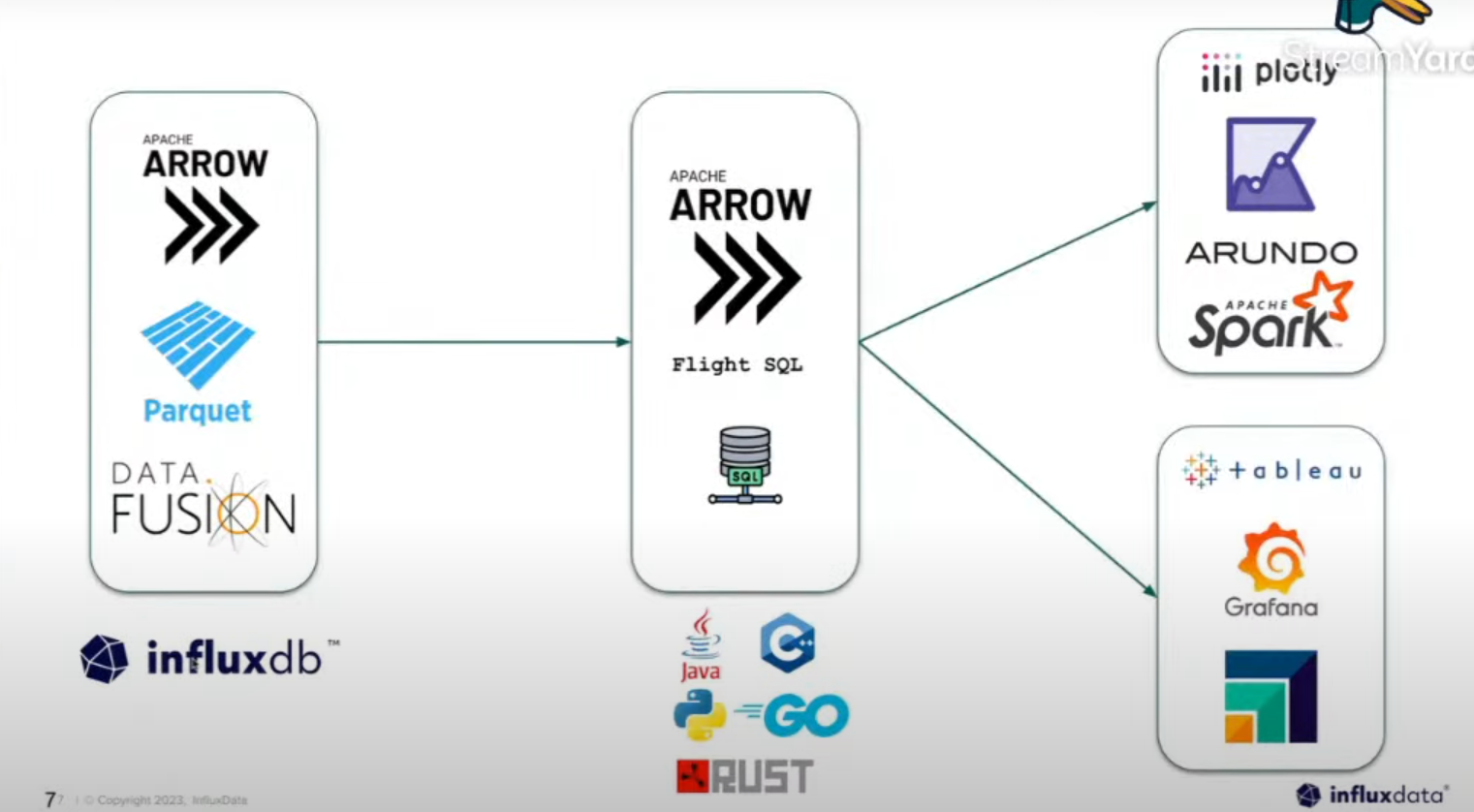
Arrow flight - Apache Arrow Flight focuses on efficient, low-latency data transport and remote procedure call (RPC) between different components or systems.

1. **Apache Arrow Flight**: Focuses on efficient, low-latency data transport and remote procedure calls (RPC) between different systems or components. It optimizes data exchange over networks and emphasizes using the Arrow data format for efficient data sharing.
2. **Apache Parquet**: Primarily concerned with how data is structured and stored on physical storage disks. It defines a columnar storage file format optimized for efficient storage and retrieval of data on disk.
3. **Apache Arrow**: Defines a standardized, language-agnostic, and efficient in-memory data structure for columnar data. It facilitates efficient data interchange between different programming languages and is optimized for data processing tasks.

When we query data from influxdb, influxdb will send back that data within columnar format to the client.

Why?

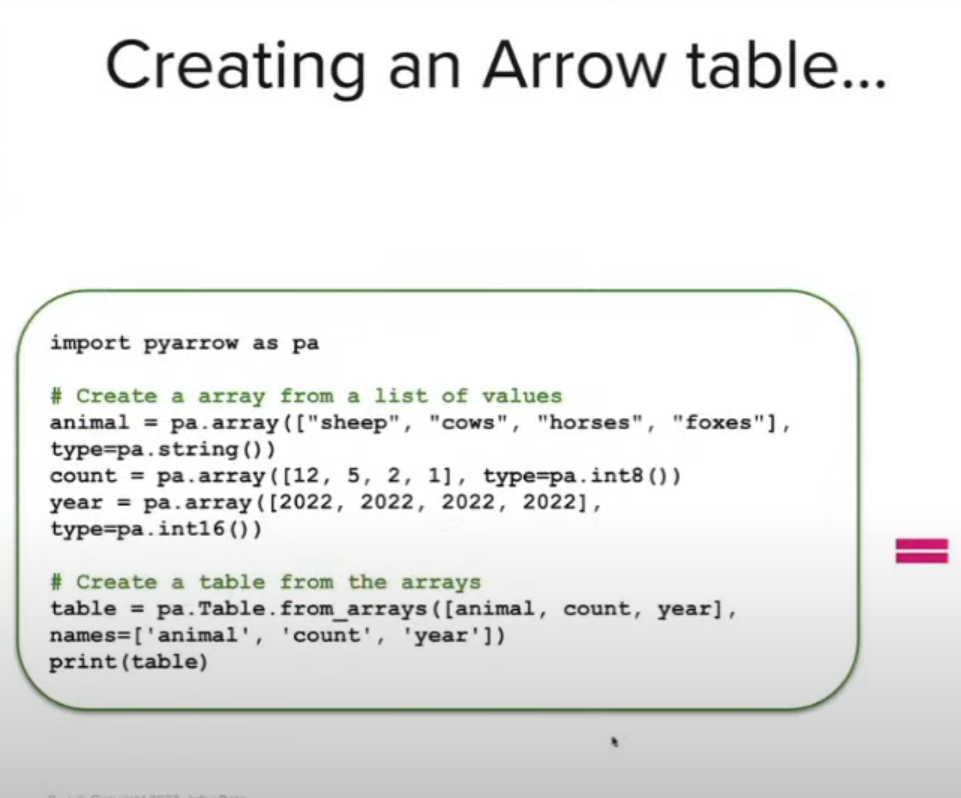
Because influxdb is built on all three of this ecosystems.



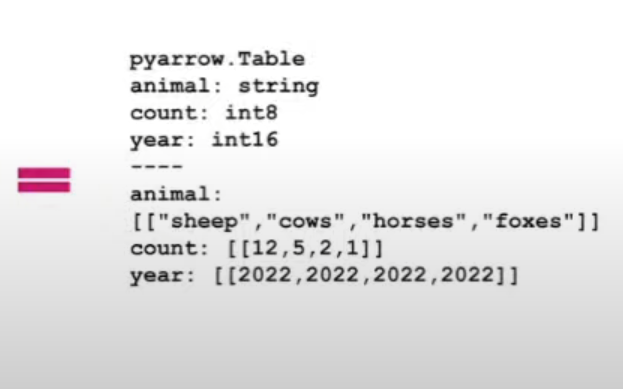


So in interacting directly with influxdb database, we should understand first what is the common data structure which is the **Arrow Table.**

Arrow table represents the columns of data in a series of arrays in which you can see in the code snippet below:



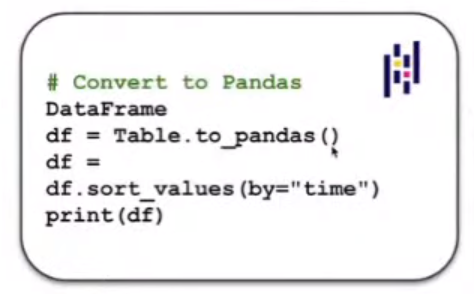
If we run it the output will look like this:



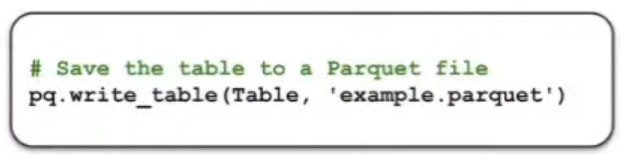


One of the primary methods that most people use with arrow tables is to convert it to a format such as pandas.

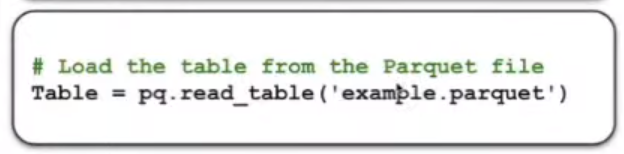
Arrow has an amazingly efficient way of converting to a pandas data frame and it’s as simple as:



We can also save our table to a parquet file



We can also load the table from parquet file to our influx database



Pyarrow also provide a series of aggregation as well.

