1. How write a python code that runs as fast as C++?

By using cython. Cython is a programming language that is a superset of Python. It allows developers to write python code which is compiled in C language. So, it is much faster than python and allows low level access.

2. What is the fastest data type in python?

The fastest data structure in Python that has an O (1) time complexity for most operations is a dictionary. A dictionary is an unordered collection of key-value pairs, where each key is unique and has a certain value associated with it. Dictionaries are implemented using hash tables, which are a data structure that allows for efficient lookup, insertion, and deletion of key-value pairs. The hash function takes the key and returns an index into an array, where the corresponding value is stored.

3. What is Hadoop? and what type of data does it use?

Hadoop is an open-source software framework that is used to store and process large datasets across clusters of computers. It is designed to run on commodity hardware, which makes it a cost-effective solution for processing and storing large datasets. It is widely used in industry for tasks such as data warehousing, data mining, and machine learning. Hadoop does not use a traditional relational database to store data. Instead, it uses the Hadoop Distributed File System (HDFS) to store large amounts of data across multiple machines in a distributed environment. HDFS is a distributed file system that is designed to store large files and provide high throughput access to those files across a cluster of computers.

4. What are the ways to compensate for missing values in data?

- Deletion of rows with missing values.
- Imputation like mean, median, mode, multiple and regression imputation.
- Prediction of missing values.
- Weighting the correct data with higher values than the missing data.