

Libraries that used to treat with huge data:

1. **pandas:**

- **Description:** pandas is a powerful data manipulation library that provides data structures like DataFrames, which are particularly useful for working with tabular data. While pandas is generally suitable for large datasets, its performance can be optimized by using features like chunking for reading and processing data in smaller pieces.

2. **Dask:**

- **Description:** Dask is a parallel computing library that integrates with pandas, NumPy, and other Python libraries. It allows for parallel and distributed computing, making it suitable for working with larger-than-memory datasets. Dask enables lazy evaluation and parallel execution of operations.

3. **NumPy:**

- **Description:** NumPy is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays. NumPy arrays are more memory-efficient than standard Python lists, making them suitable for large datasets.

4. **PySpark (Apache Spark with Python):**

- **Description:** PySpark is the Python API for Apache Spark, a fast and general-purpose cluster computing system. Spark is designed for distributed data processing and can handle large-scale data efficiently. PySpark provides high-level APIs for distributed data processing and machine learning.

5. **Vaex:**

- **Description:** Vaex is a Python library for lazy, out-of-core DataFrames. It allows you to work with large datasets that don't fit into memory by reading and processing data on disk in a memory-mapped format. Vaex is particularly fast and memory-efficient for common data manipulations.

6. **Modin:**

- **Description:** Modin is designed to scale pandas workflows by using parallel and distributed computing. It allows you to switch between pandas and Modin seamlessly, providing a performance boost for data manipulation operations on larger datasets.

7. **HDF5 (h5py):**

- **Description:** HDF5 is a file format and set of tools for managing complex data. The h5py library allows Python programs to interact with HDF5 files efficiently. It is suitable for storing and retrieving large numerical datasets.