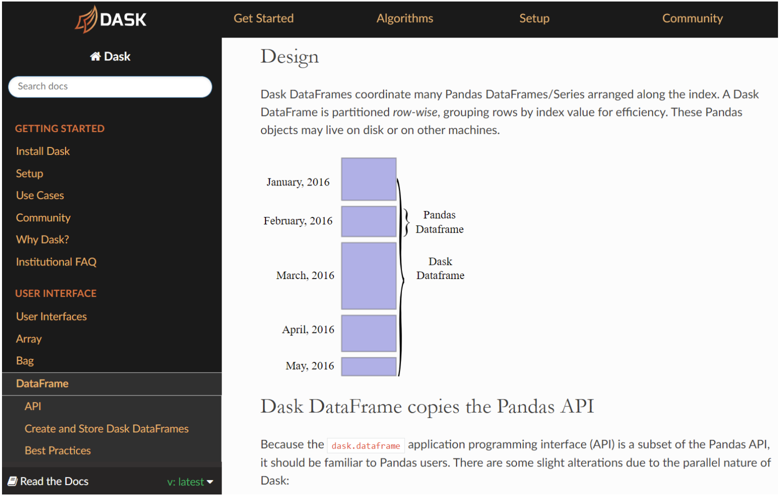
Cannlytics

Cannabis Data Science

Meetup

May 5, 2021

Working with big data



DASK

Get Started | Algorithms | Setup | Community

Dask

Search docs

GETTING STARTED

- Install Dask
- Setup
- Use Cases
- Community
- Why Dask?
- Institutional FAQ

USER INTERFACE

- User Interfaces
- Array
- Bag
- DataFrame
  - API
  - Create and Store Dask DataFrames
  - Best Practices

Read the Docs v: latest

Design

Dask DataFrames coordinate many Pandas DataFrames/Series arranged along the index. A Dask DataFrame is partitioned *row-wise*, grouping rows by index value for efficiency. These Pandas objects may live on disk or on other machines.

Dask DataFrame copies the Pandas API

Because the `dask.dataframe` application programming interface (API) is a subset of the Pandas API, it should be familiar to Pandas users. There are some slight alterations due to the parallel nature of Dask:

ARMA processes

It is helpful to introduced the autoregressive moving average (ARMA) representation.

Definition: An autoregressive moving average (ARMA) model is a linear representation of a stochastic process with constant coefficients.

An ARMA(p, q) representation of x_t is

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q},$$

Until next time

Read in some big data and make some forecasts. Next week we can check our forecasts.