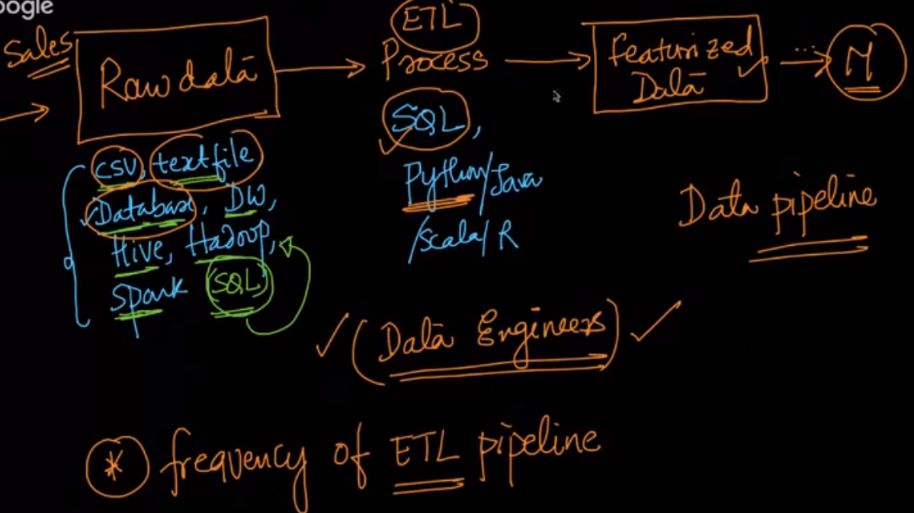


Firstly we’ll perform ETL (extract, transform, load). In this raw data is converted into featurized data.

It’s responsibility of data eng, but in small companies ML eng need to do this all.

Now this ETL process need to be performed at any frequency, like 1 hr, 1 day, 1 week anytime, depending when you are retraining.



There can be 3 types of data volume.

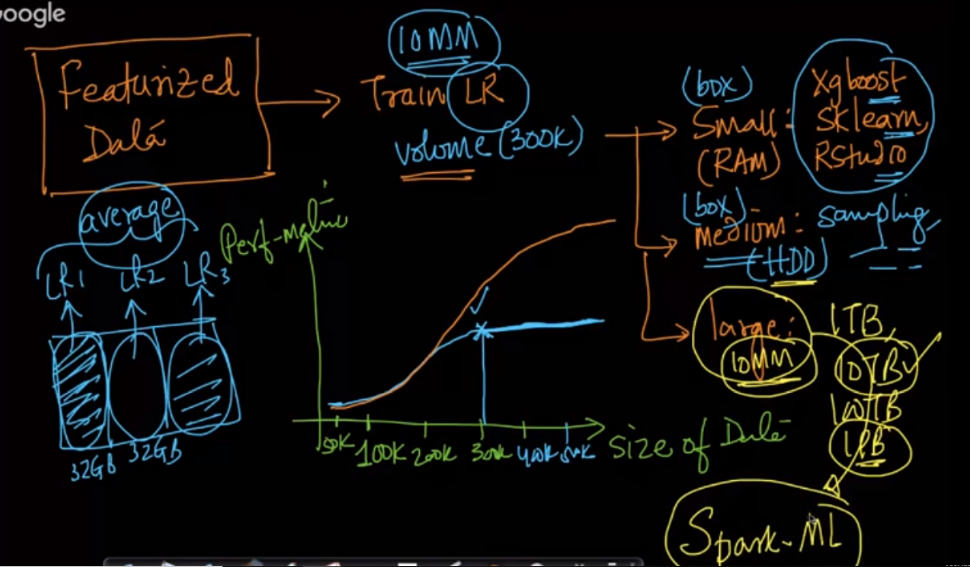
1. Small volume: small data can reside in RAM and therefore we can use Xgboos, sklearn.
2. Medium: It can’t be stored in RAM but can be stored in hard disk, so how do you train,

You pick chunks of data that can be reside in RAM, train that chunk, store that model.

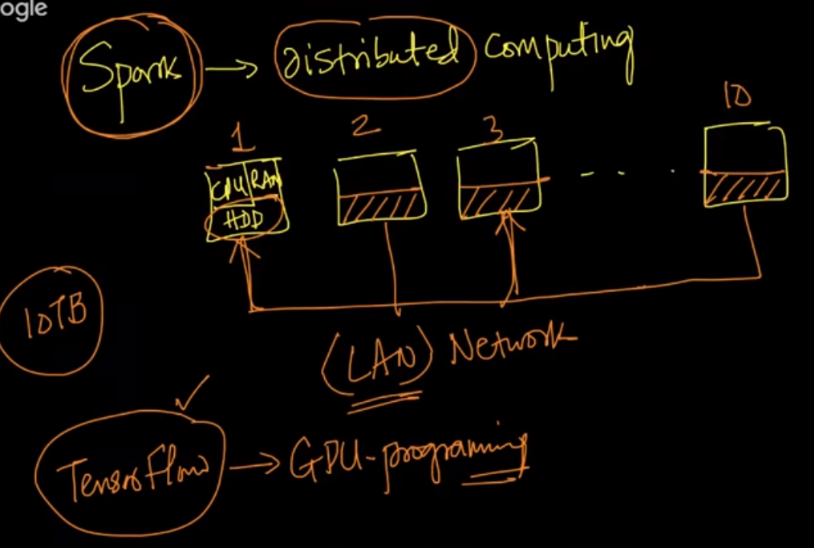
Similarly do with all the chunks, Now do averaging of the weight vectors.

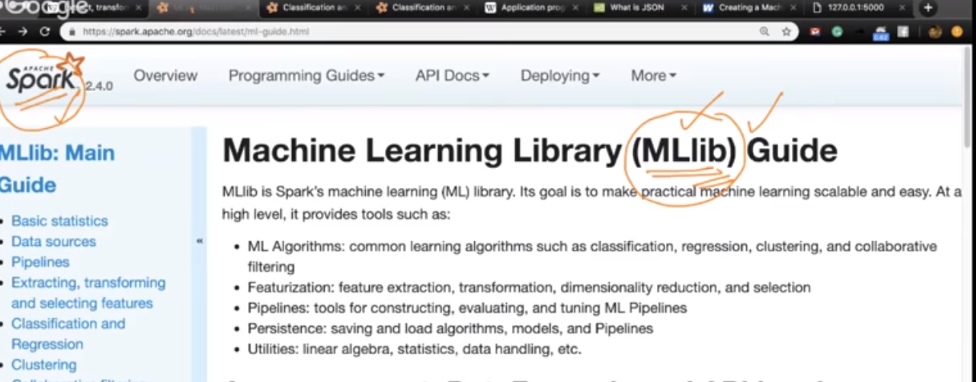
1. Large: If data is very large, that it can’t be reside in hdd, then we use spark ML

Before picking large volume data, check performance against each size of data, and the size after that model performs constantly, that size can be choose.



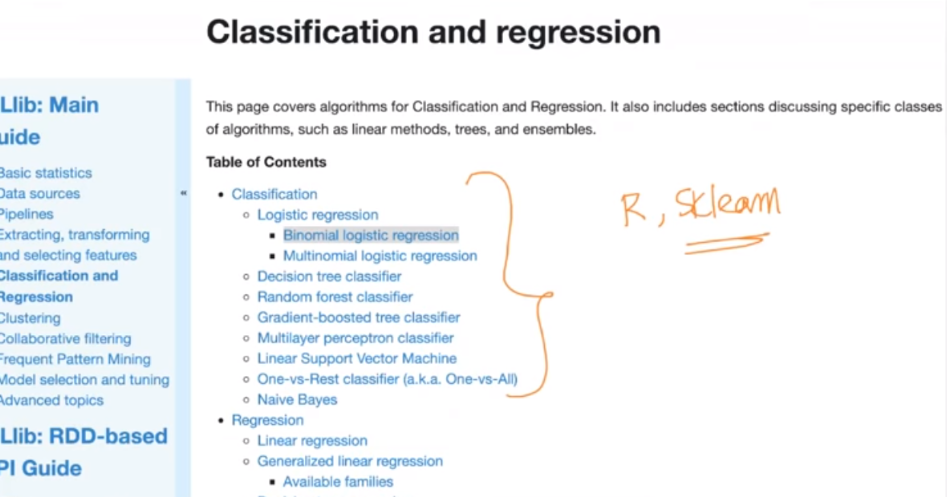
**Spark:** spark is a distributed computing system, It provides a ML library which can be used for large volume data.

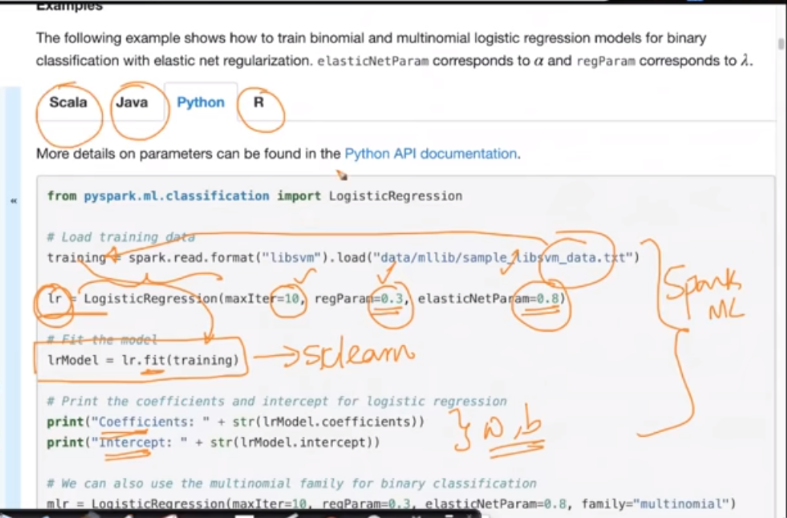


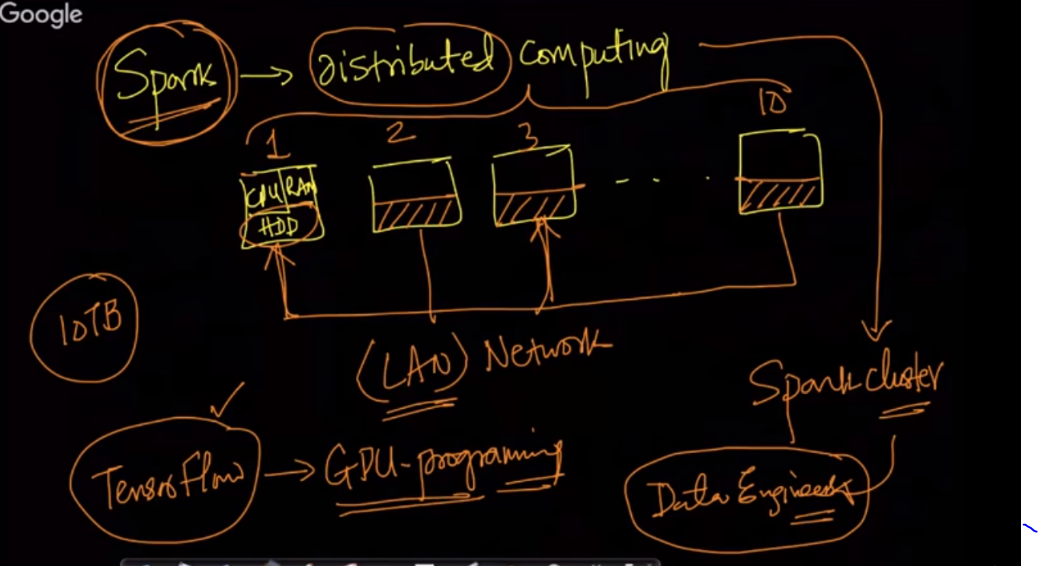


Creating web api using flask for ml

https://www.wintellect.com/creating-machine-learning-web-api-flask/

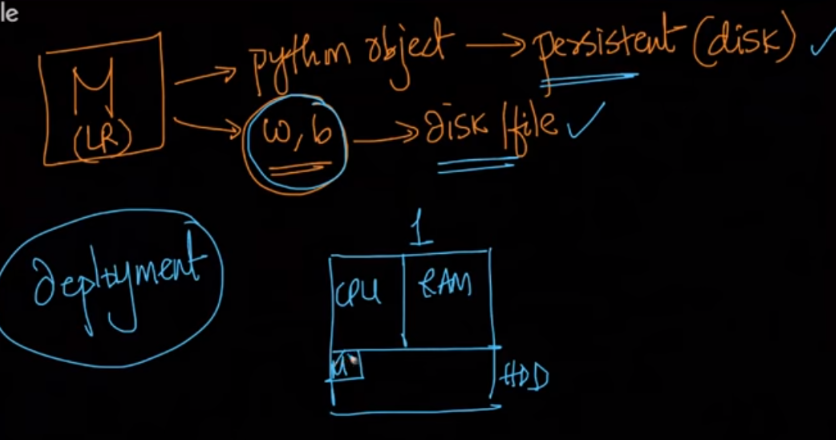






**Deployment of model:**

Once you saved your model as object in hdd, then it can be used with web api. In which new query point will come as request, you apply that to our trained model and get output, and return it via response.



Ideally in large companies, ML team have it’s own server, where main server request for the service.

