Distributed and parallel. They are not the same. They focus on different parts. Distributed means not on an individual system. Parallelism means a big job is divided into small parts which are done at the same time to finish a big job in order to decrease processing time.

They are always used together.

The goal is to handle larger data set and increase performance. Note the primary goal is to handle larger data set, which is different from traditional parallel algorithm.

3. challenge

But it also brings problems in storage, I/O, communication, etc.

Communication happens at least between each node and the scheduler. It is much harder than usual communication, since the individual system is more complex. When talking about communication, we have to choose synch or asynch.

4. How

4.1 MapReduce

4.2 GPU

What GPU dose is exactly the same as MapReduce but on a single chip. A GPU is divided into several blocks and blocks are divided into threads which do smaller job.

4.3 Stochastic GD

通常大规模机器学习都是深度学习

Why Deep learning?

When it comes to image or natural language processing, normal ML algorithms have bad performance. While going deep into neural network, we can extract the more abstract information contained in the data.

GPU stands for graphic processing unit. It is designed for processing images using a parallel architecture. CPU is designed for calculating. When processing, the gpu is divided into blocks and blocks are divided into threads. Every thread dose a part of the big job.

TensorFlow is a data flow chart developed by google. In the chart nodes represent computation, while edges represent the multi-dimensional vectors between the nodes, also called tensor. It is a large scale machine learning architecture that can work on heterogeneous systems.