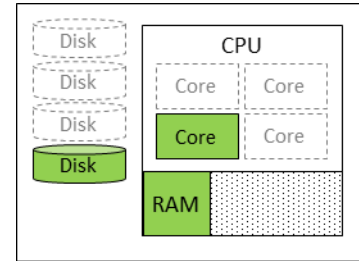


Core-to-Cloud Parallel Computing – The Key to Big Data Machine Learning

Conventional machine learning tools were not built for Big Data. They were designed to run on desktop workstations and departmental servers and operate on small data sets.

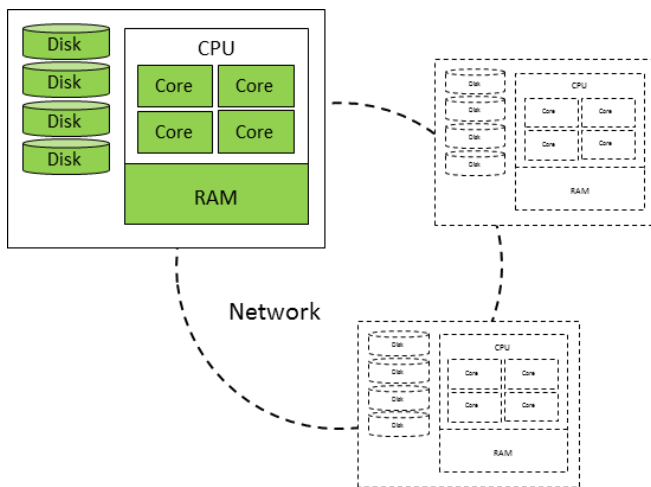
As a result, most machine learning programs are single-threaded and can only execute one instruction on one core of one CPU of one computer at a time, and they cannot handle very much data in memory. These programs have no *data parallelism*, making them very slow and unable to build big models.



So, while desktop computers and data center servers have gotten more powerful, adding more, faster cores, CPUs, RAM and disks, most machine learning software hasn't been able to take advantage of those advances.

If you have system programming skills, you can re-program some open source machine learning software to use the whole machine; or, if you have a generous machine learning budget, you can buy from a handful of expensive enterprise server offerings to get some single-system speedup.

But, what if the models you need to build are bigger than a single machine? Even if you have more servers sitting on your network, few conventional machine learning programs can use them.



And, even if you can find software that runs on all of your networked computers, what if your model is bigger than your network?

Few companies can just buy more computers, but everyone can "rent" them in the cloud by the hour.

The cloud enables economical elastic compute provisioning up to any scale.

If you can fully utilize cloud-based compute clusters, the scale of your models is unlimited.

But, just because a machine learning program runs in the cloud doesn't mean it automatically runs on multiple machines – far from it. Most cloud software, like most downloaded software, is still single threaded and only uses a fraction of a single machine to operate, as in the first picture above.

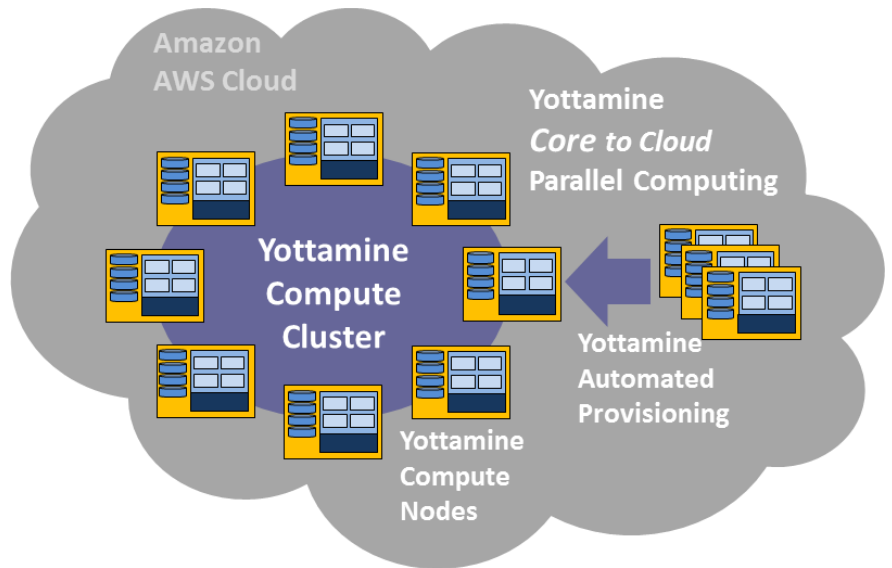
The only practical, affordable way to get the scale and speed needed to do Big Data Machine Learning is with software that uses parallel programming from core to cloud. Such a hyper-parallel software architecture enables far greater speed, scale, and accuracy, at much lower cost than any other approach to machine learning.

Core-to-Cloud Parallel Computing – The Key to Big Data Machine Learning

Yottamine Analytics is pioneering Big Data Machine Learning using *Core-to-Cloud* parallel programming. Yottamine’s cloud-based model building software uses every core of every CPU, and all the memory you need to build predictive models of any size.

Yottamine analyzes your data and automatically selects and configures the resources needed to build your model, no matter how large it may be.

Then under the control of Yottamine’s unique highly parallel machine learning algorithms, all the nodes in a cluster work together to pool their resources and share the workload. This produces large models faster and enables you to produce very large models that would not otherwise be possible. Here’s proof.



	LIBSVM Standard Production Version	Yottamine Predictive Services
Architecture		
Test 1: Large Data Set (Covtype) – 580,000 examples, 2 classes, 54 features		
Power	1 core out of 4 in Xeon Server CPU 16 GB RAM	80 Xeon cores (10 networked 8-core CPUs) 70GB RAM
Speed	Evaluate 1 parameter in 6.3 hours	Evaluate 210 parameters in 1.5 hours
Accuracy	77.11%	95.75%
Test 2: Very Large Data Set (MNIST8M) – 8.1M data points, 10 classes, 784 features		
Power	1 core out of 4 in Xeon Server CPU 16 GB RAM	640 cores (20 networked dual 16-core CPUs) 1.2TB RAM
Speed	Evaluate 1 parameter in 40 days (est.)	Evaluate 12 parameter in 6 hours
Accuracy	Unknown – terminated after 3 days	97.82%

Yottamine is scalable, fast, easy to use and very affordable. See for yourself at Yottamine.com.

- A laptop supercomputer experience for the professional Big Data scientist.

