Trends in Big Data

Reza Zadeh

What is "Big Data"?

Don't trust anyone who tells you they know
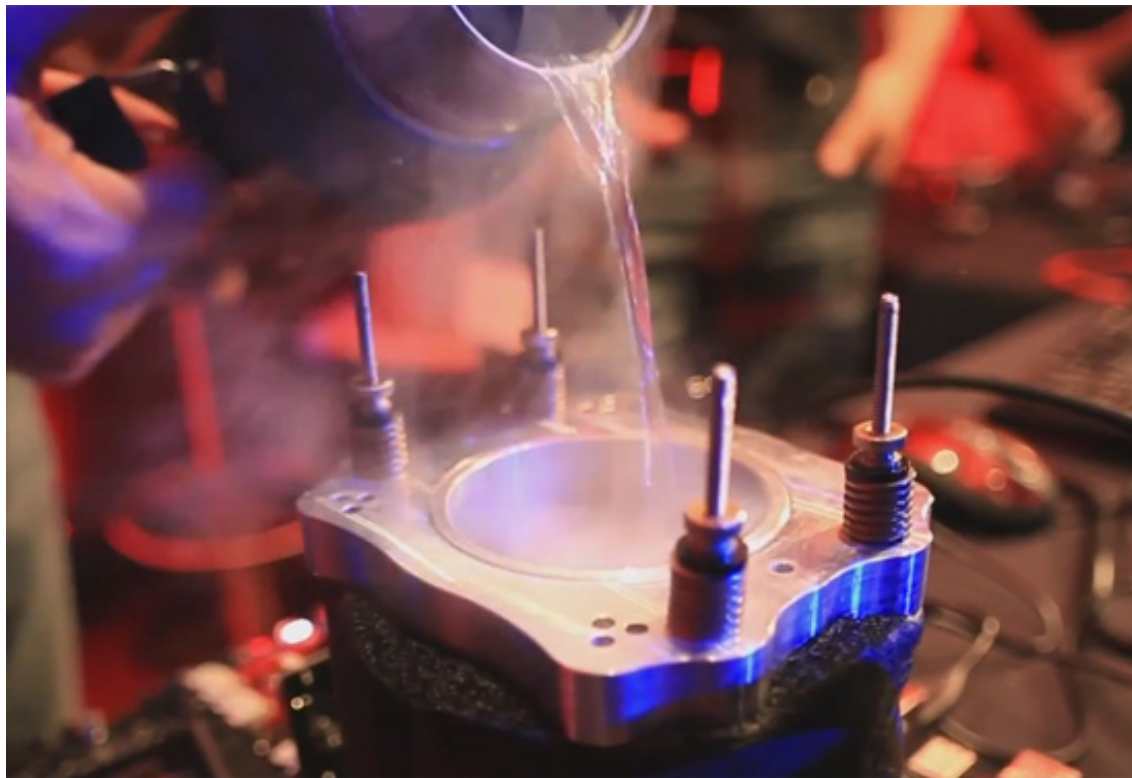
Three Trends

Computation moving from sequential to parallel

Supercomputers being replaced by large clusters of commodity machines

Data moving from disk to memory (RAM)

Parallel Computing

Cooling *fast* CPUs is hitting physical limits

Parallel Computing

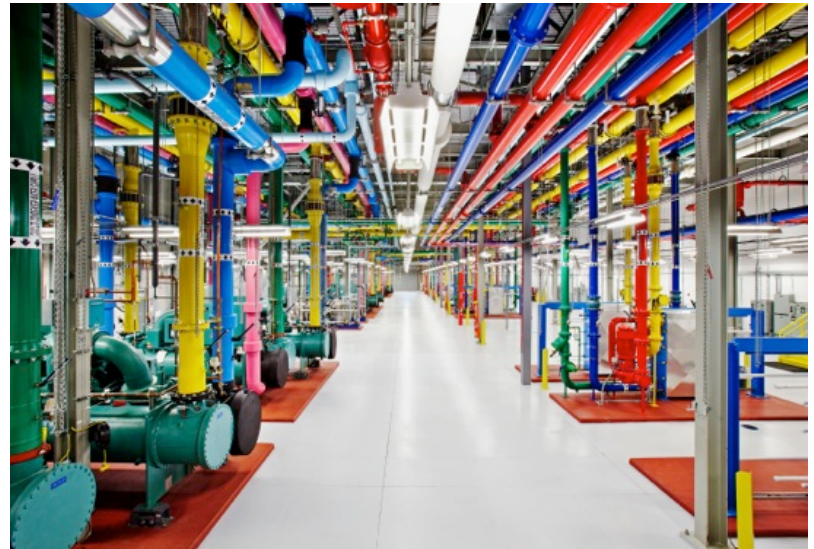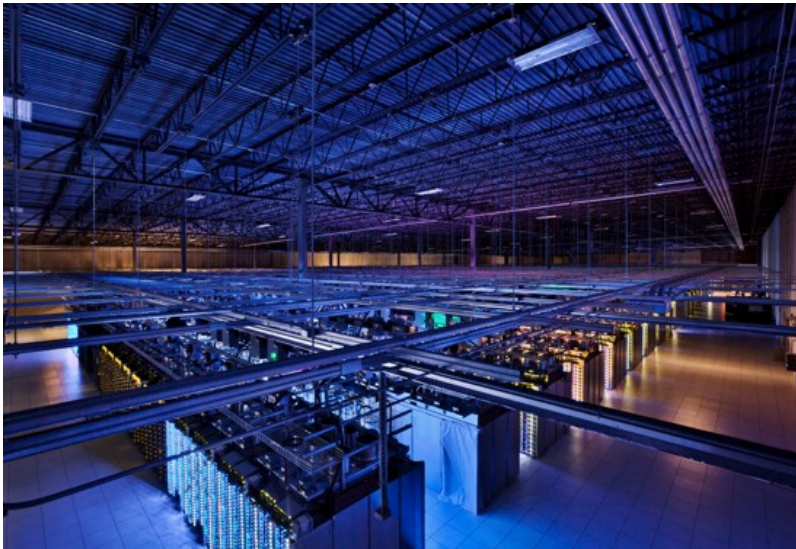Cooling *fast* CPUs is hitting physical limits

Solution: use many CPUs instead of one

Algorithms must be fundamentally changed

Distributed Computing

Supercomputers are expensive

Solution: distribute on large clusters (for rent)

The "Cloud"

Rent large clusters temporarily

Roll your own (commodity hardware after all)

In-memory Computing

Memory size is still doubling every ~2 years

More datasets fit into memory

Use all memory across a cluster:

Spark

Conclusion

"Big Data" is *at least* these three trends.