



# What Is Hadoop?

School of Information Studies  
Syracuse University

# What Is Hadoop?

A system for:

- Storing,
- Processing, and
- Managing

data.

*Sound familiar?*

Yes. Change “data”  
to “Big Data.”



*“It does look similar—but this one  
is powered by Hadoop”*

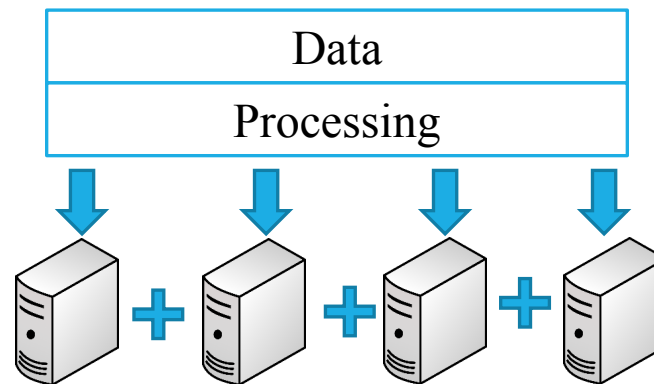
# Hadoop Handles Big Data by *Scaling Out*

**Problem:** File too large to fit on a single storage platform?

**Solution:** Distribute the file over several computers.

**Problem:** Server not “fast” enough to perform ETL over your data?

**Solution:** Distribute data processing over several computers.





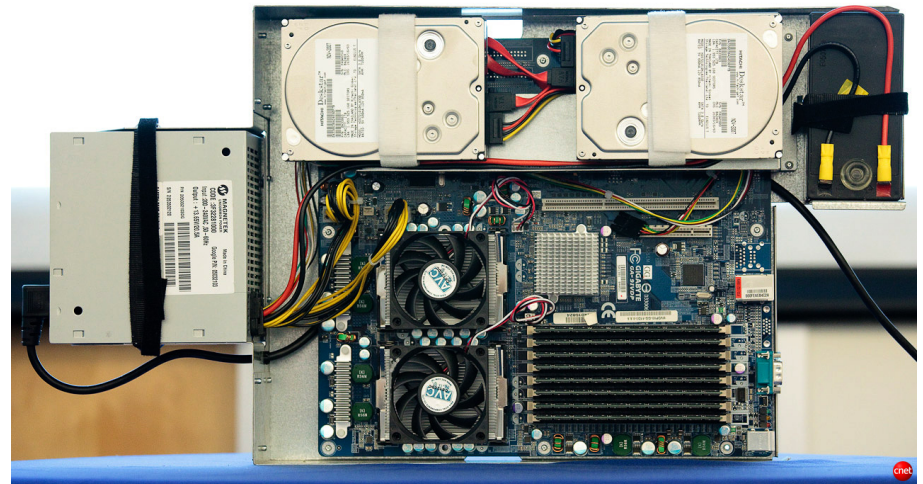
# Hadoop Is Designed to Use *Commodity Hardware*

Hadoop hardware:

- Modular.
- Easy to add and remove nodes.
- Failure is not only acceptable but *expected*.

This is contrary to enterprise hardware.

- High-redundancy/fault-tolerant
- Vertical scaling
- Storage arrays



\* Google Hardware spec for server source: C|NET