Getting Started:

Why Hadoop?

“The Hadoop Ecosystem is great for Big Data”

The 4 W’s (and H):

***What’s** in the ecosystem?*

***Why** is it beneficial?*

***Where** is it used?*

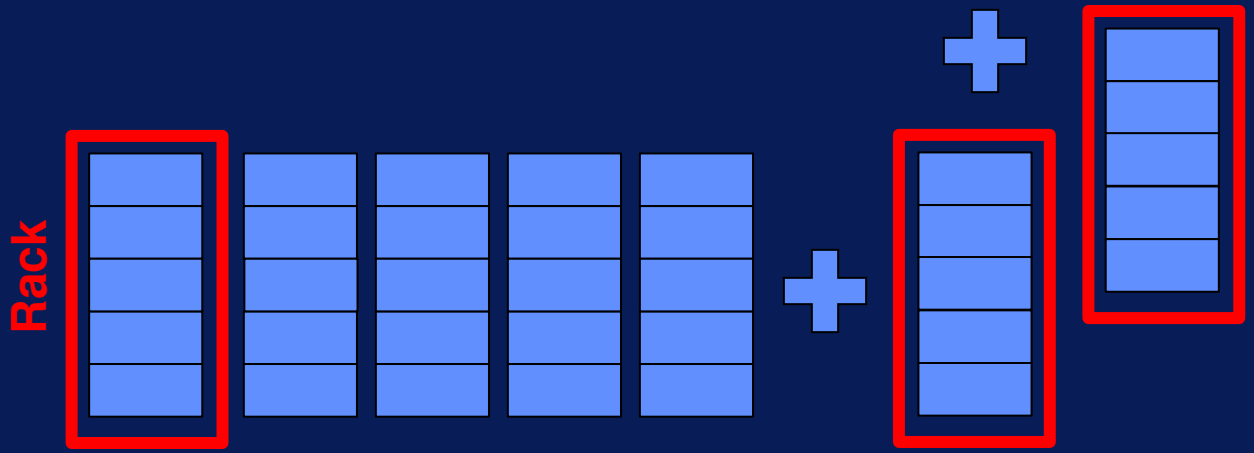
***Who** uses it?*

***How** do these tools work?*

Major Goals

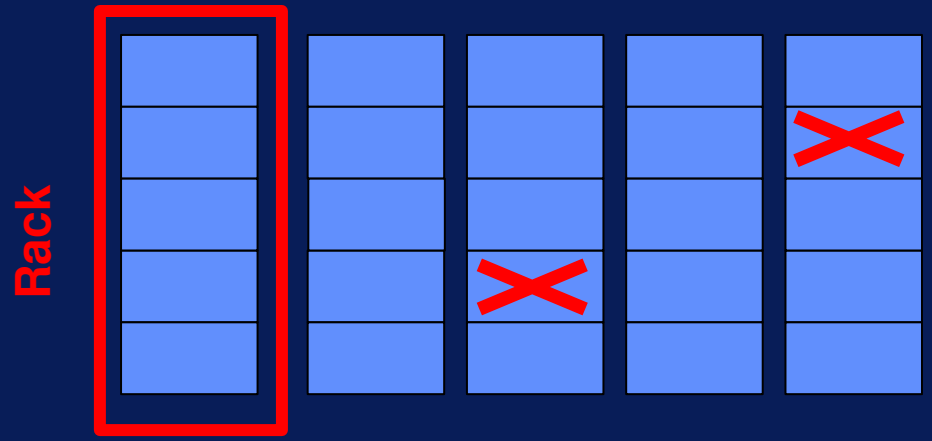
1. Enable Scalability

Commodity hardware is cheap

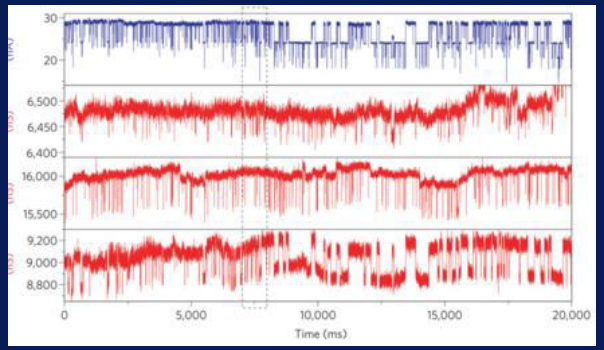


2. Handle Fault Tolerance

Be ready: crashes happen

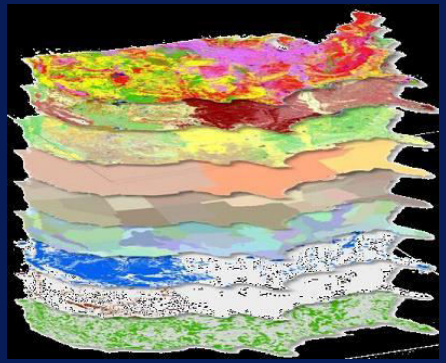


3. Optimized for a Variety Data Types

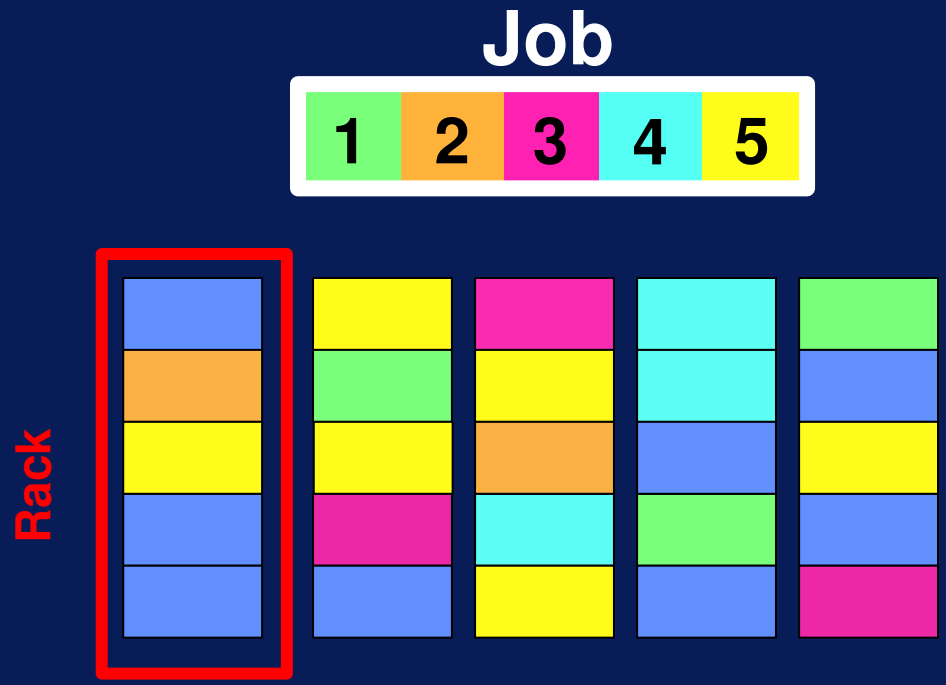


Cars marketplace

vendor	Model	Price	Mileage	VIN Code
Chevrolet	Corvette	17226	25965.0	ILLAKAWAZDZ
Chevrolet	Corvette	34229	46429.0	RCPNSRYGXOI
Chevrolet	Corvette	27982	50209.0	NWLGCEVEHGI
Chevrolet	Corvette	51825	72998.0	NGVZSCIZGSM
Chevrolet	Corvette	52845	34364.0	PSDRUYYOIJG.
Chevrolet	Malibu	37874	37273.0	VLFPQPWNEFD
Chevrolet	Malibu	15600	71441.0	EXLJGDWOZSA
Chevrolet	Malibu	52447	46700.0	NLMGJZAKBPD
Chevrolet	Malibu	27129	36254.0	OIPFUIENLEHSX
Chevrolet	Malibu	28846	77162.0	WRCOOFREZLI
Chevrolet	Malibu	46165	60590.0	HUFTTHQHSFJI
Chevrolet	Malibu	18263	37790.0	JLMHNAFSHVD



4. Facilitate a Shared Environment



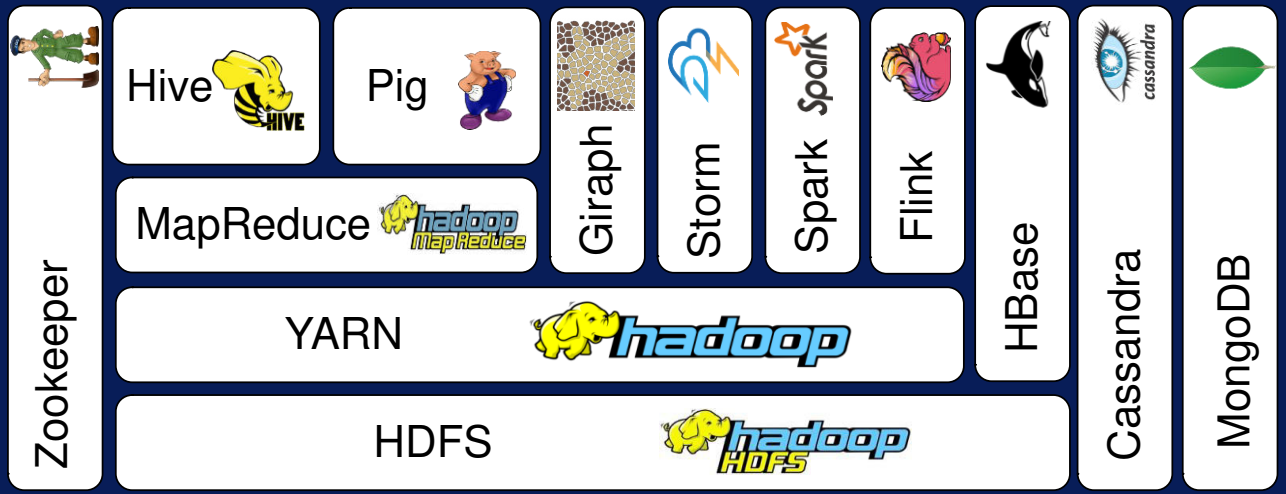
5. Provide Value

Community-supported

Wide range of applications

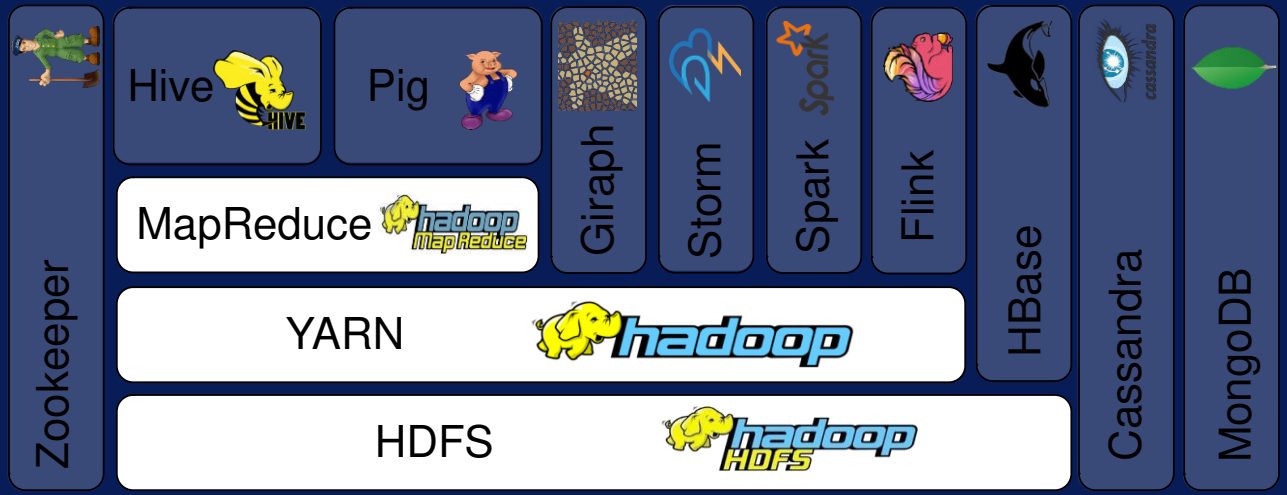
The rest of this Module…

The Hadoop Ecosystem



The rest of this Module…

Main Hadoop Components



The rest of this Module…

Cloud Computing

PaaS

SaaS

IaaS

The rest of this Module…

When to use Hadoop?

Hadoop

~~**Hadoop**~~

The rest of this Module…

Exercises