

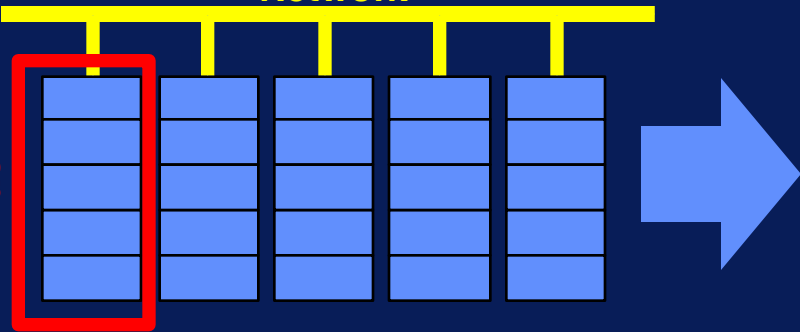
Programming Models for Big Data

After this video you will be able to..

- Explain the requirements of programming models for big data and why you should care about them
- Tell your friends how you can scale the speed of pasta sauce generation in your kitchen by applying big data programming models

Rack

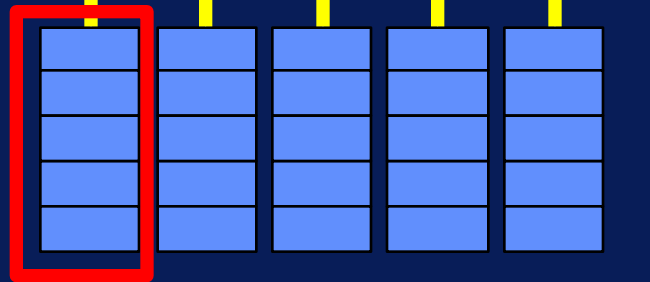
Network



**Data-parallel
scalability**

Network

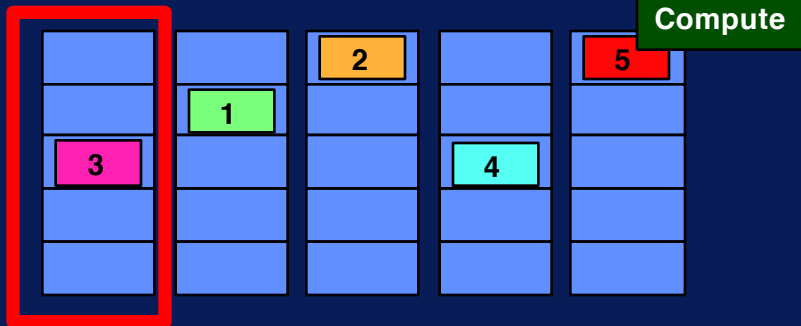
Rack



Data

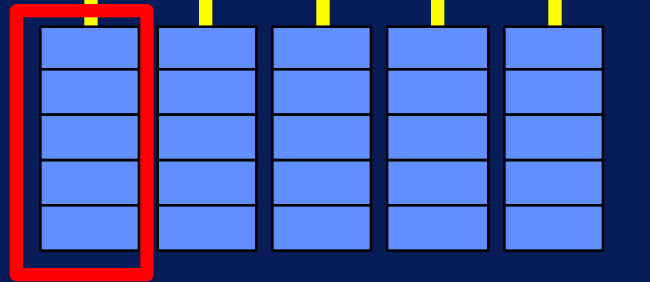


Rack



Network

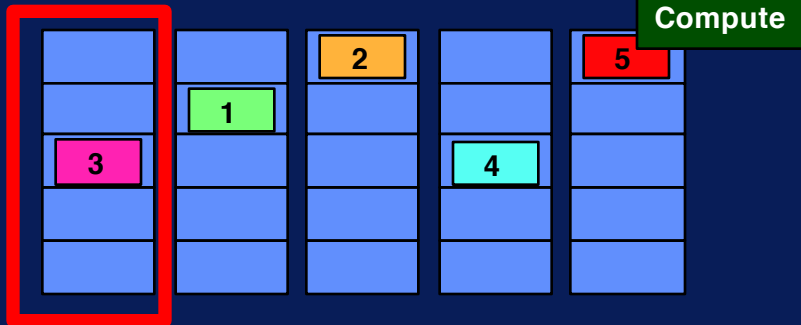
Rack



Data

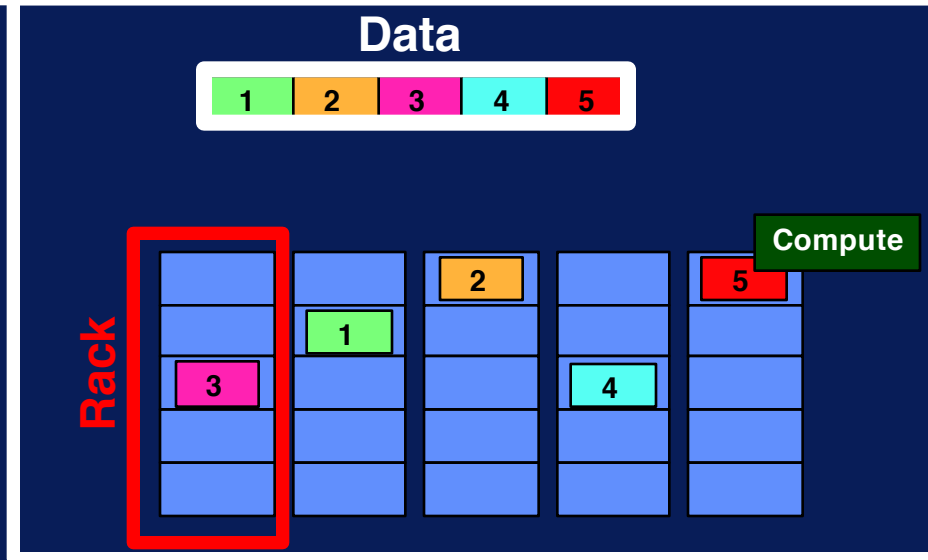


Rack

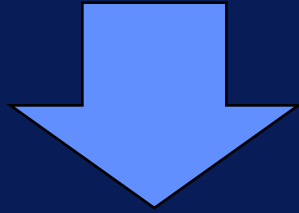


Programming Model = abstractions

Runtime Libraries + Programming Languages



Programming Model for Big Data



Programmability
on top of
Distributed File Systems

Requirements for Big Data Programming Models

1. Support Big Data Operations

Split volumes of data

1. Support Big Data Operations

Split volumes of data

Access data fast

1. Support Big Data Operations

Split volumes of data

Access data fast

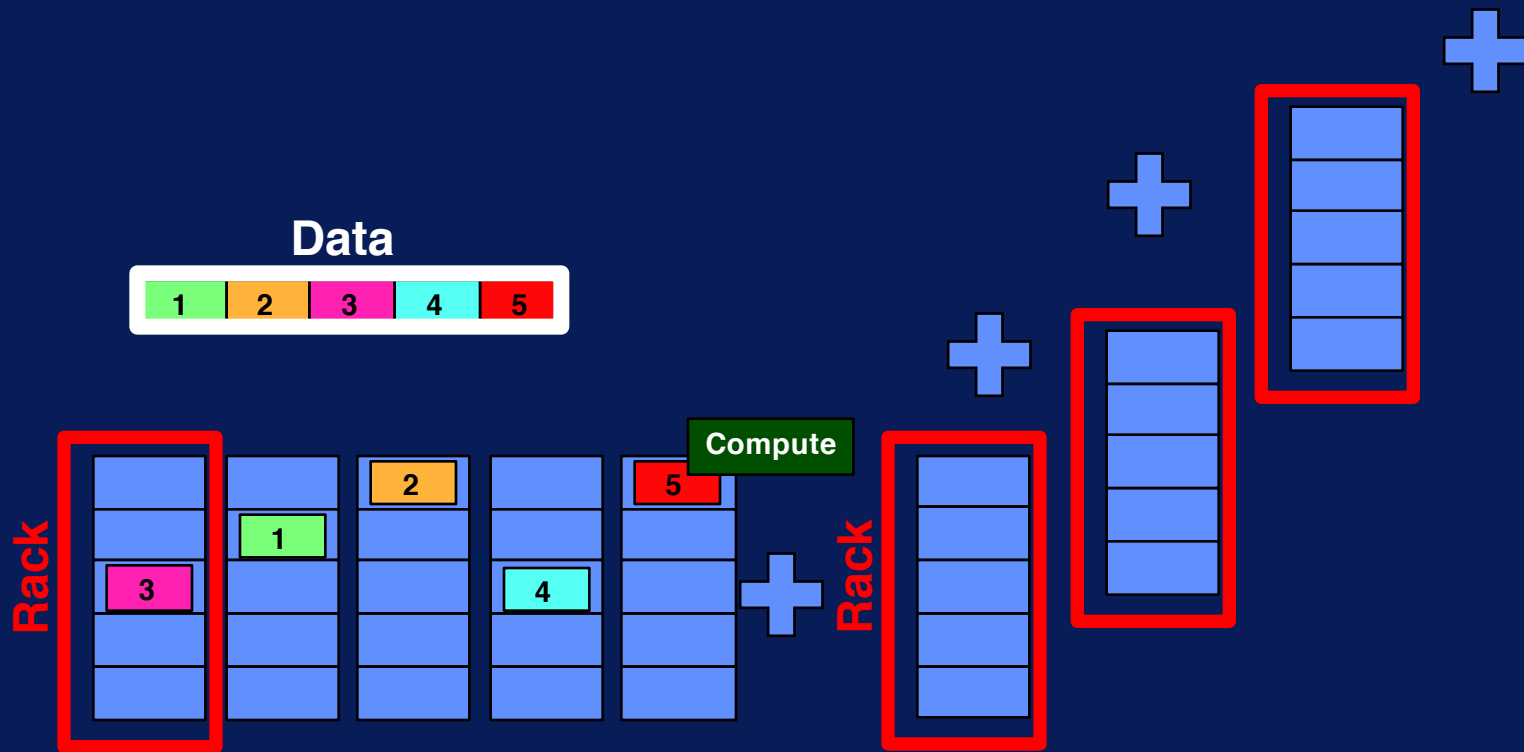
Distribute computations to nodes

2. Handle Fault Tolerance

Replicate data partitions

Recover files when needed

3. Enable Adding More Racks



4. Optimized for specific data types

Document

Table

Key-value

Graph

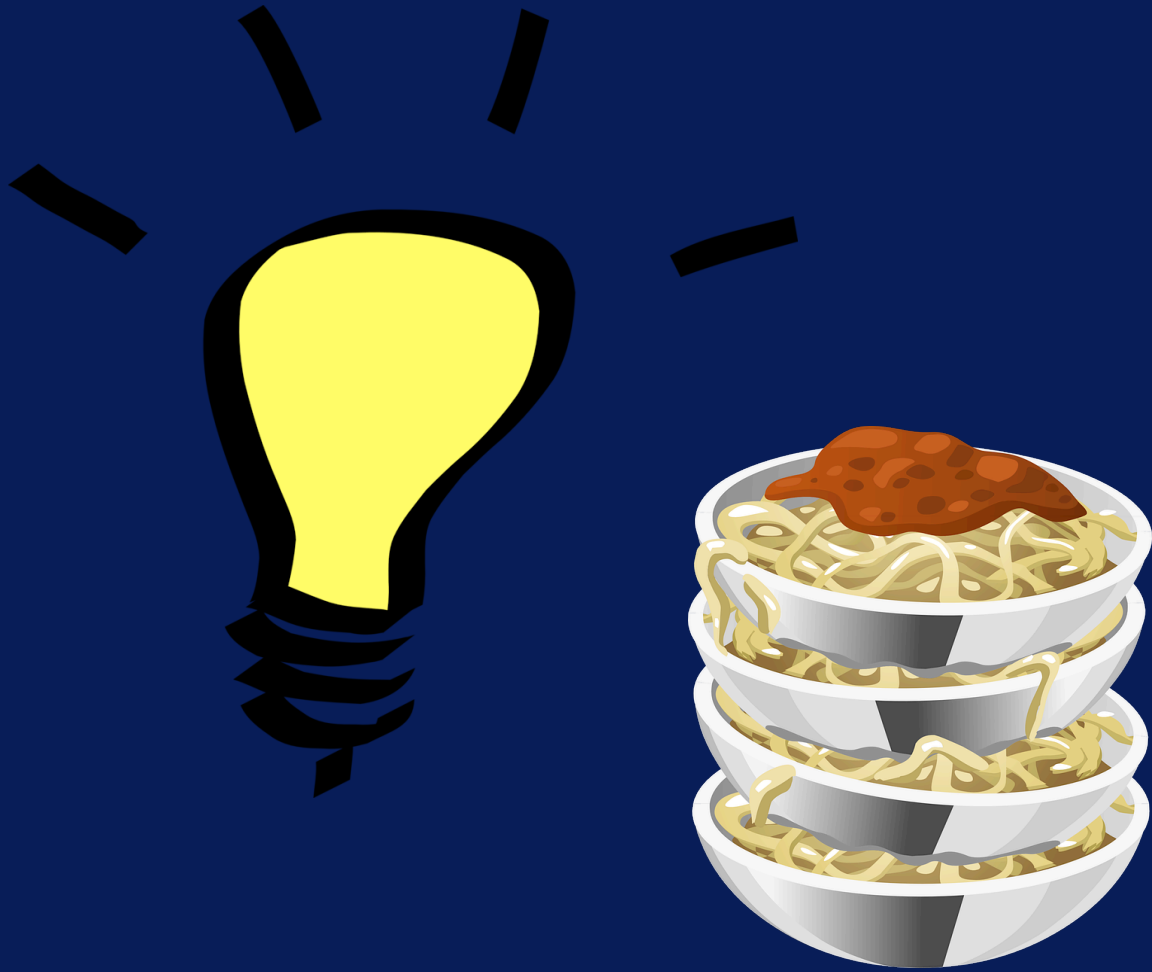
Stream

Multimedia

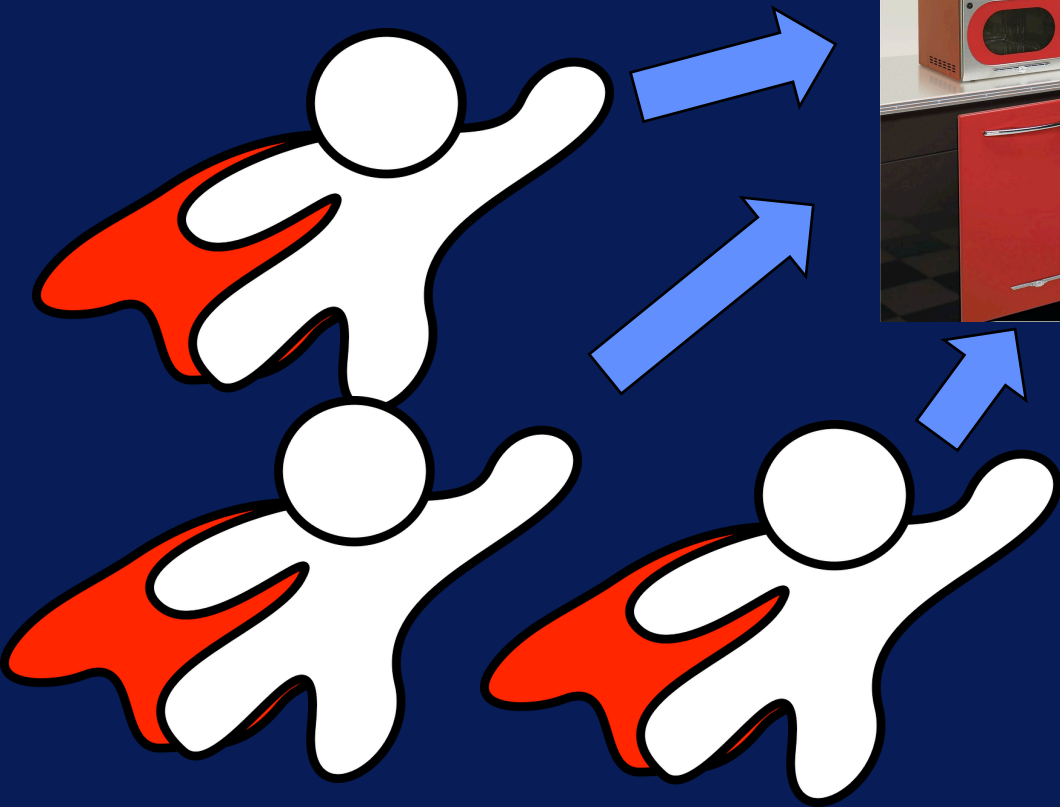
Natural model for independent
parallel tasks over multiple resources!



Coming over
for dinner in half
an hour...



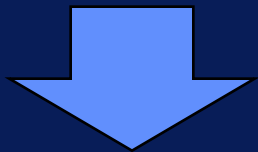
Helpers!







MapReduce



A programming model for Big Data



Many implementations

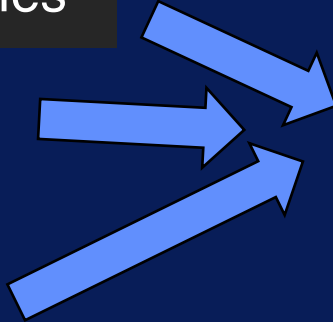
Programming Model = abstractions

Runtime Libraries + Programming Languages

Support large data volumes

Provide fault tolerance

Enable scale out



MapReduce