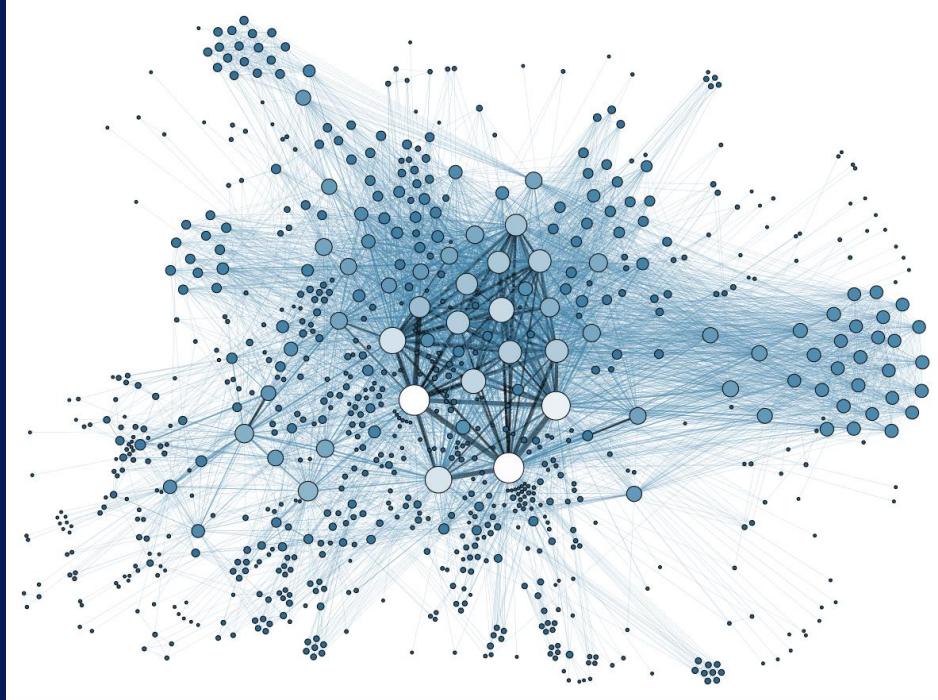


Why is Big Data Processing Different?

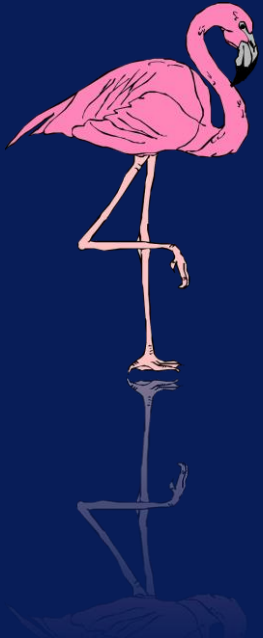


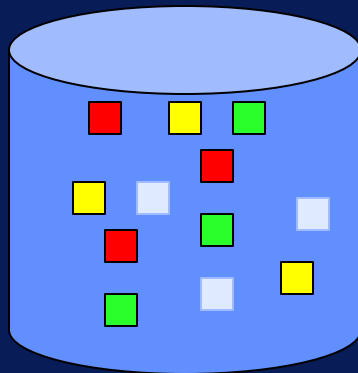
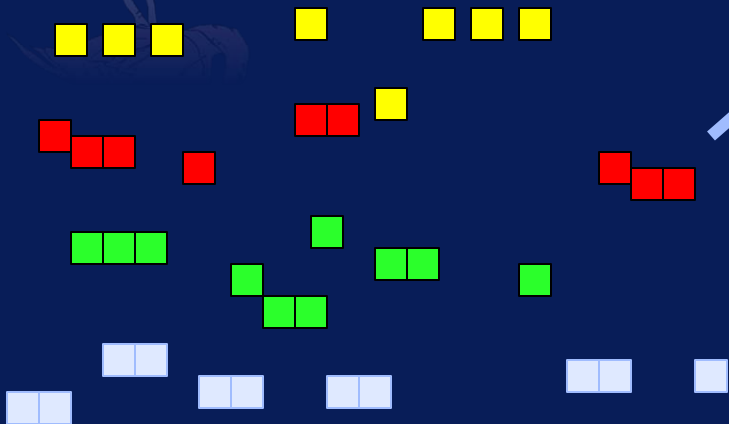
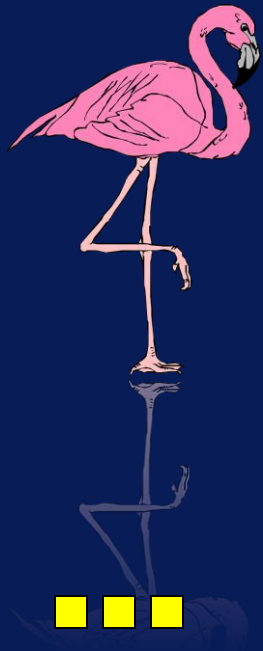
After this video you will be able to..

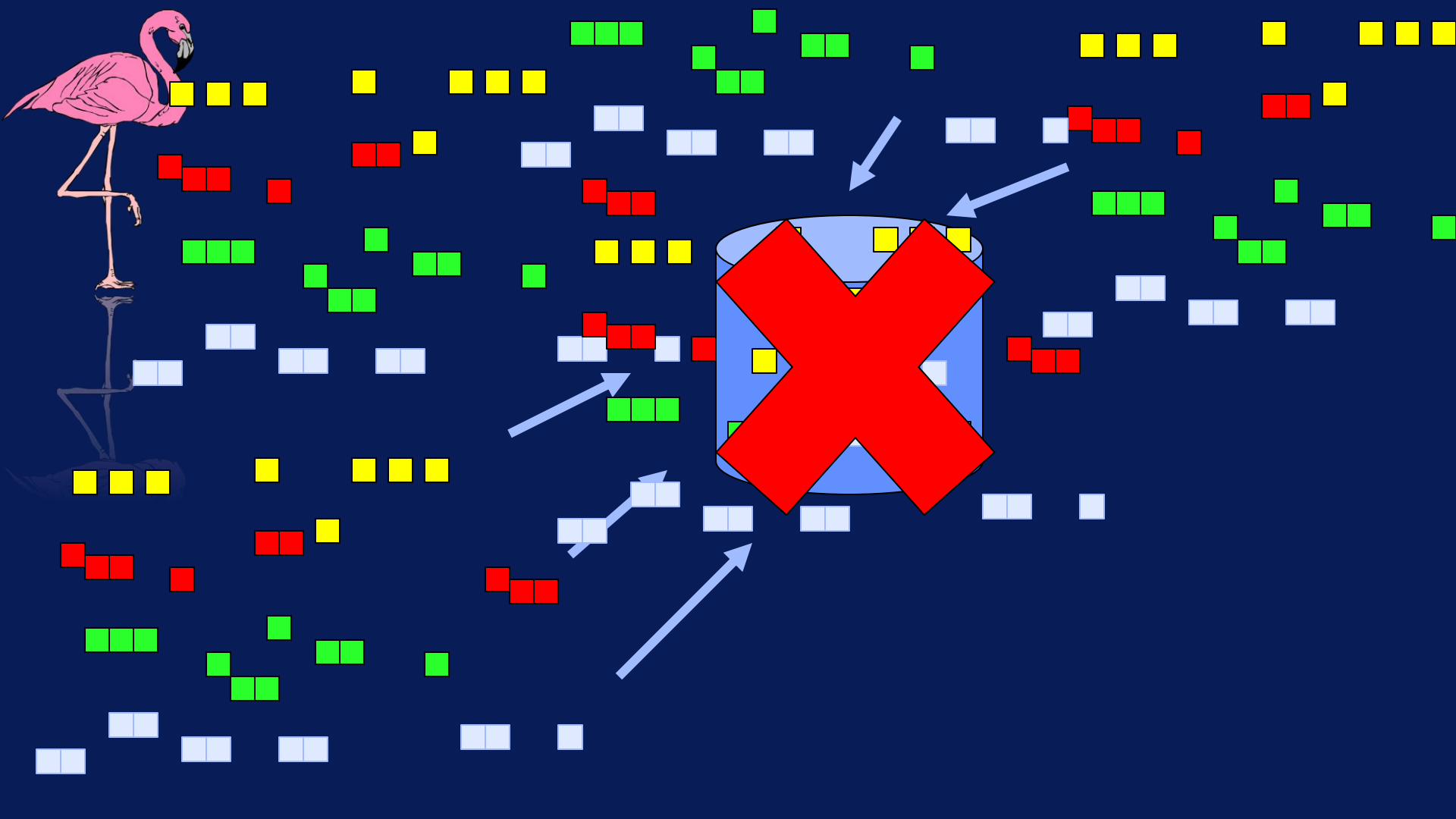
- Summarize the requirements of programming models for big data and why you should care about them
- Explain how the challenges of big data related to its variety, volume and velocity affects its processing

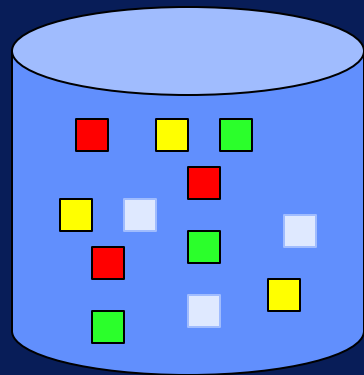
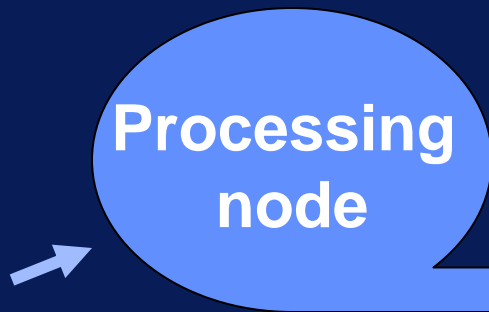
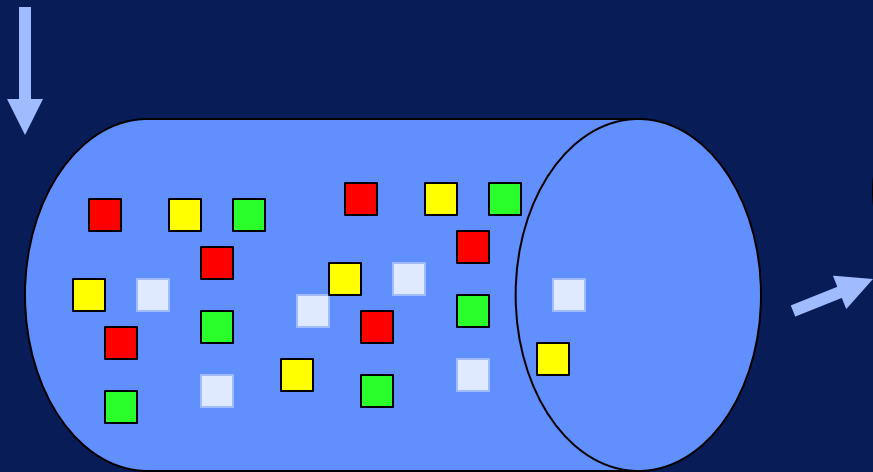
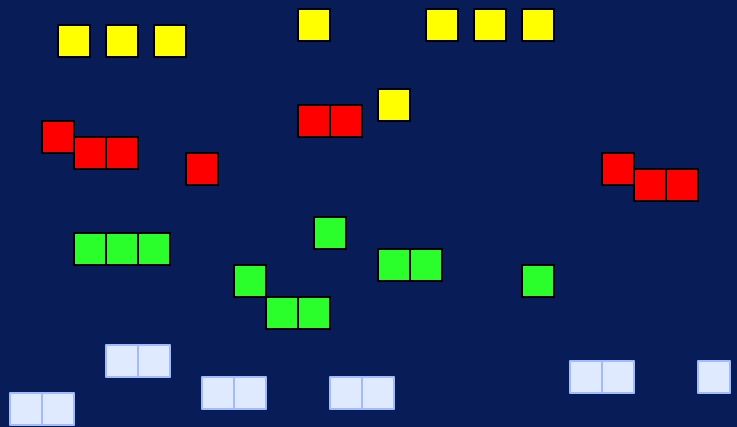
Requirements for Big Data Systems

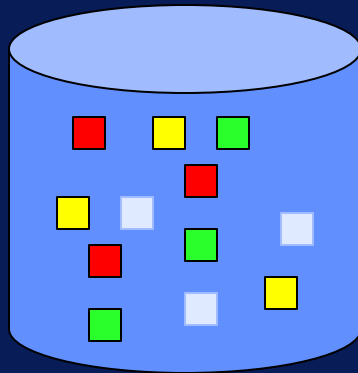
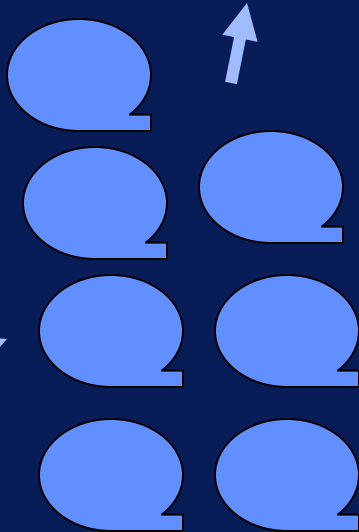
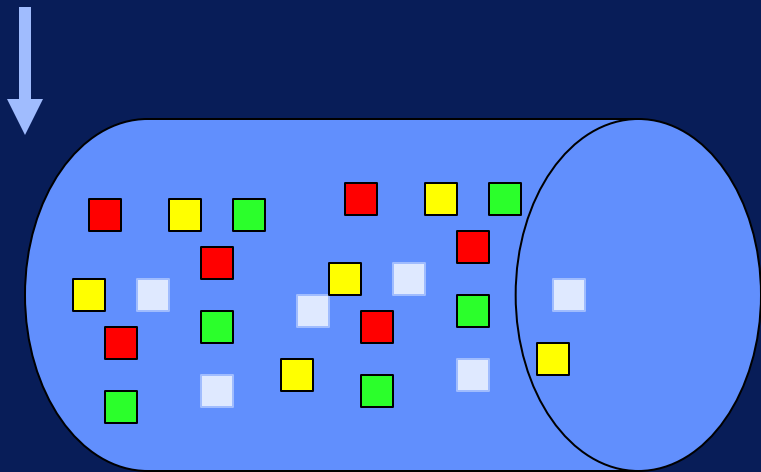
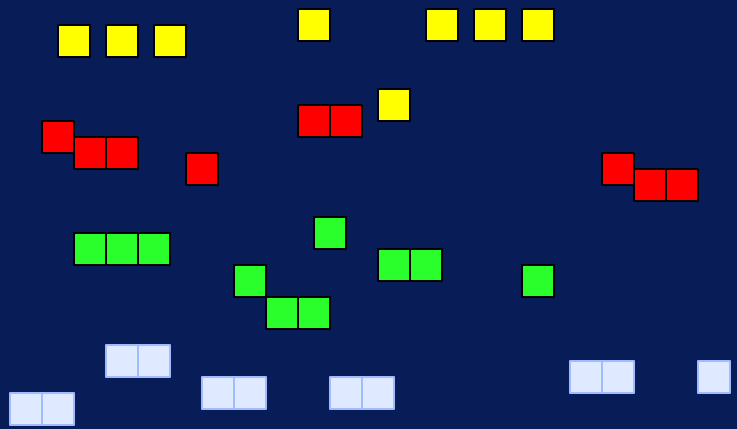
A Big Data System for an Online Game



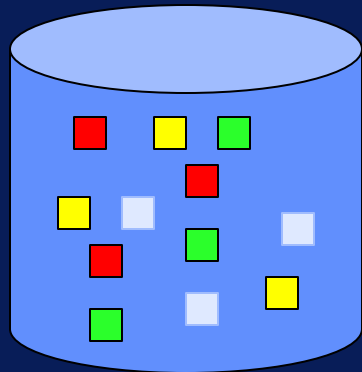


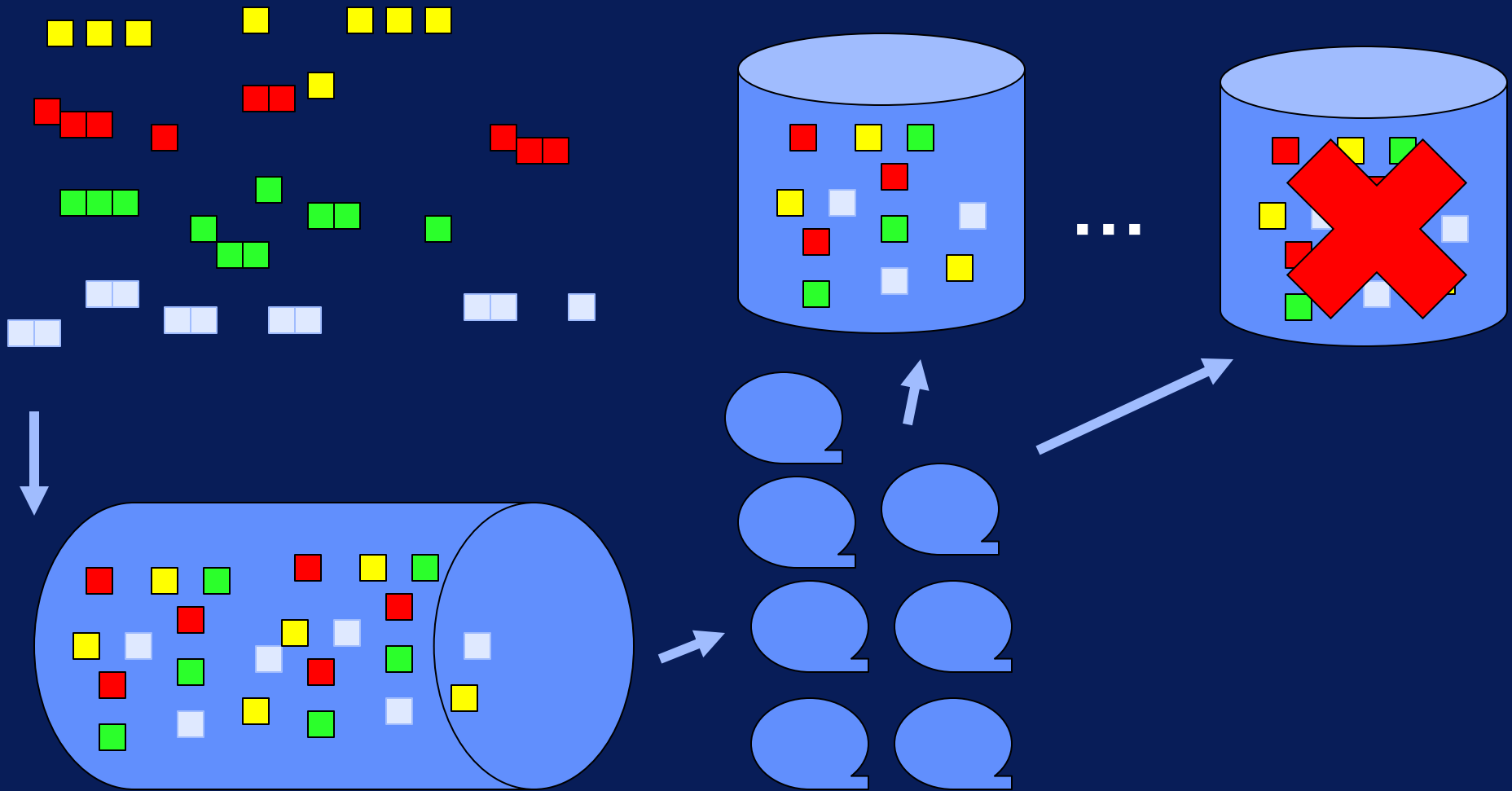


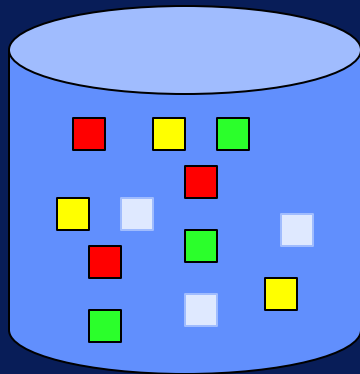
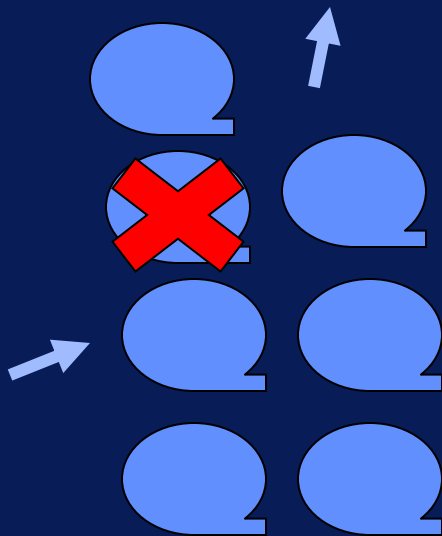
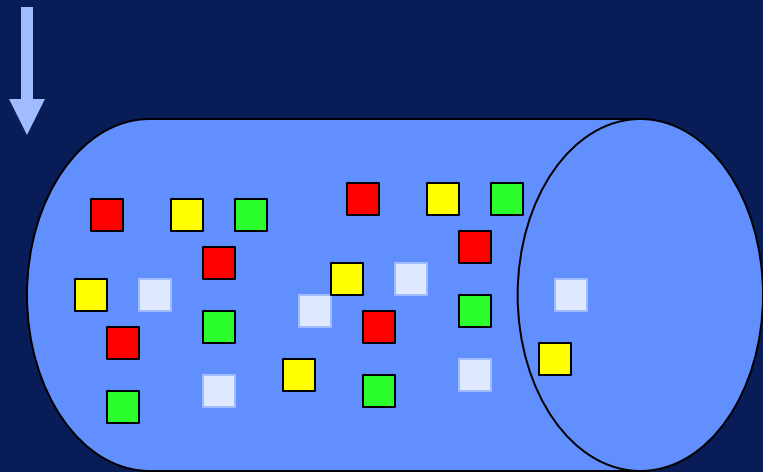
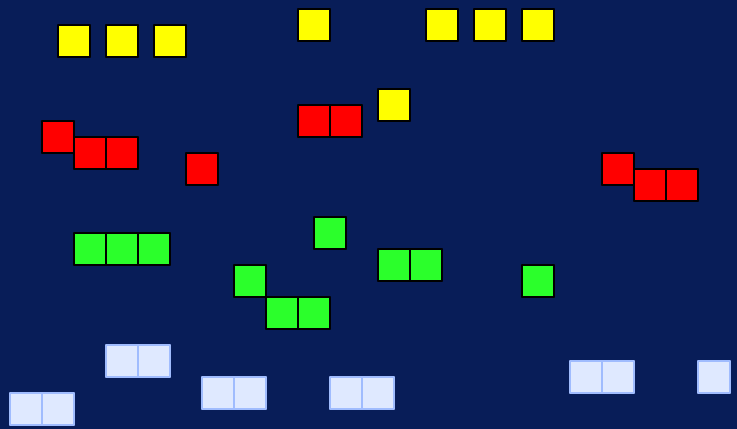




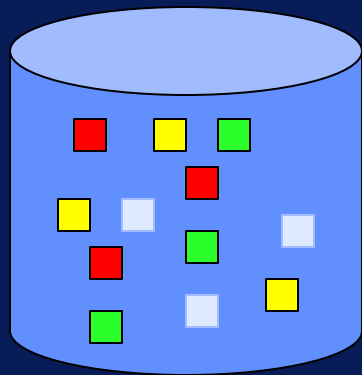
...

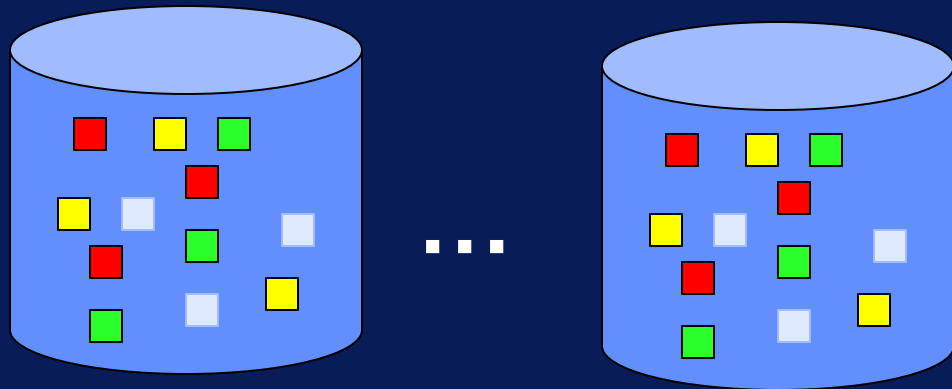






...





**Batch
Processing**

A large, light blue arrow pointing upwards, with a black outline, positioned on the left side of the image.

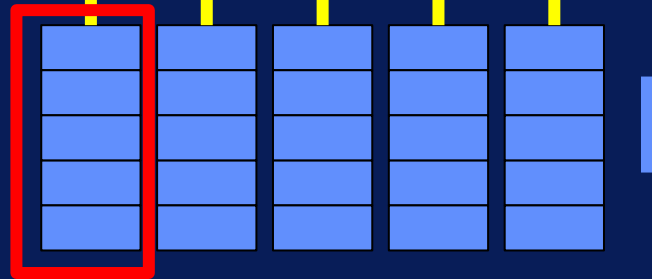
Scalability

A large, light blue arrow pointing downwards, with a black outline, positioned on the right side of the image.

Complexity

Network

Rack



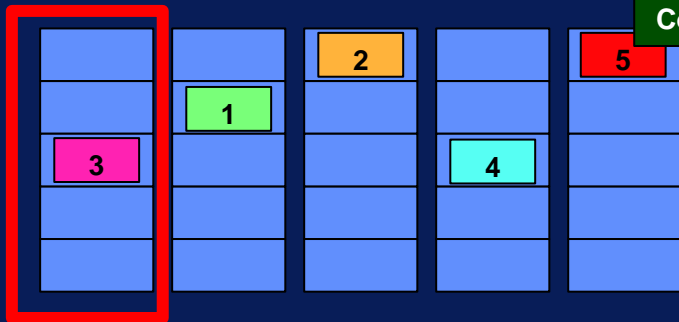
Data-parallel
scalability



Data



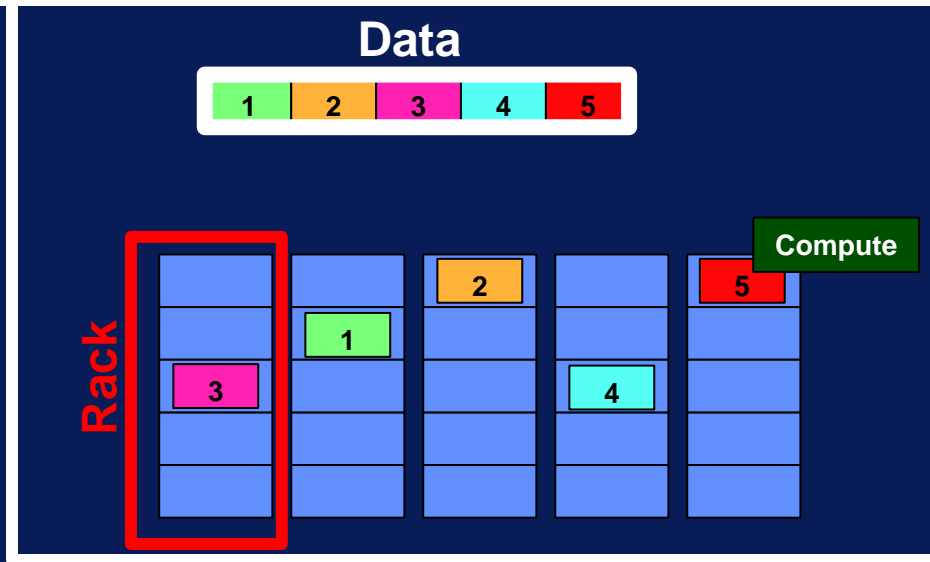
Rack



Compute

Programming Model = abstractions

Runtime Libraries + Programming Languages



Requirements for Big Data Systems

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

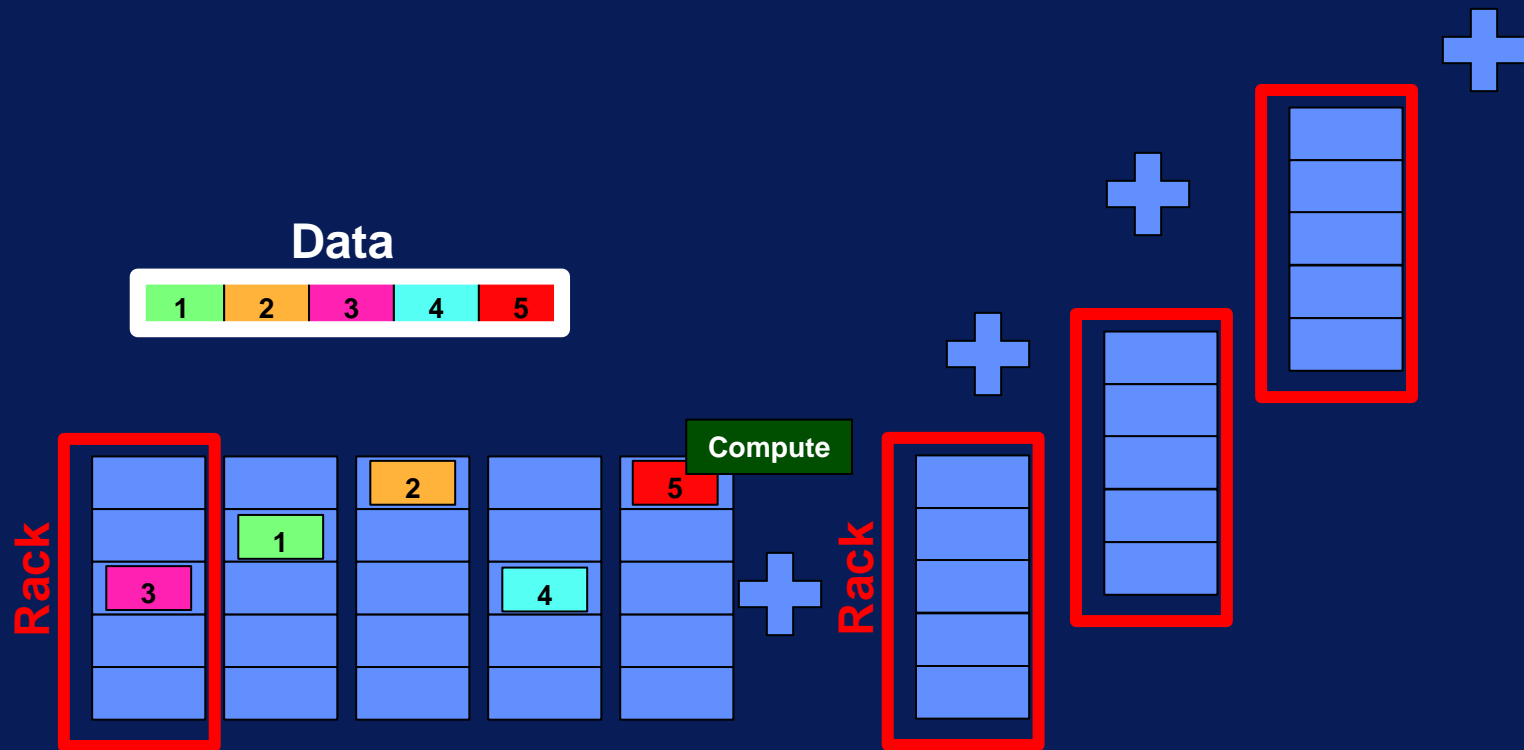
Distribute computations to nodes

2. Handle Fault Tolerance

Replicate data partitions

Recover files when needed

3. Enable Adding More Racks



4. Optimized and extensible for many data types

Document

Table

Key-value

Graph

Multimedia

Stream

5. Enable both streaming and batch processing

Low latency processing
of streaming data

Accurate processing
of all available data

Volume



**Scalable batch
processing**

Velocity



Stream processing

Variety



**Extensible data storage,
access and integration**