Cloud computing - ch - 6

Cloud Security and introduction to Hadoop

TOPICS

- Tools and technologies to secure the data in Private and Public Cloud Architecture
- Security Concerns
- Legal issues and Aspects
- Multi-tenancy issues
- introduction to cloud simulator
- Hadoop
- Advantages & disadvantage
- Hadoop environment

Tools and technologies to secure the data in Private and Public Cloud Architecture

Securing data in both private and public cloud architectures is crucial to protect sensitive information and ensure compliance with data protection regulations.

Public Cloud Security

Identity and Access Management (IAM):

AWS IAM, Azure Active Directory, or Google Cloud IAM: These services help you manage and control user access to cloud resources.

Encryption:

Cloud-native Encryption: Most cloud providers offer encryption for data at rest and in transit. Use services like AWS KMS, Azure Key Vault, or Google Cloud KMS to manage encryption keys.

Application-Level Encryption: Encrypt sensitive data within your applications using libraries like OpenSSL or libraries provided by your cloud provider.

Public Cloud Security

Network Security:

Virtual Private Cloud (VPC) or Virtual Network (VN): Segregate your cloud resources into isolated networks.

Network Security Groups (NSGs), Security Groups, or Firewalls: Implement network-level access controls to restrict traffic to and from your resources.

Logging and Monitoring:

Cloud-specific Monitoring Tools: Use services like AWS CloudWatch, Azure Monitor, or Google Cloud Monitoring to monitor resource activities.

SIEM Solutions: Implement Security Information and Event Management (SIEM) tools like Splunk, Sumo Logic, or Elasticsearch to centralize and analyze logs.

Private Cloud Security

Hypervisor Security:

Ensure that the hypervisor layer is secure by regularly patching and updating it.

Network Segmentation:

Implement network segmentation within your private cloud infrastructure to isolate different parts of your network.

Firewalls and Intrusion Detection/Prevention:

Use firewalls and IDPS systems to protect your private cloud's network and monitor for suspicious activities.

Private Cloud Security

Encryption:

Encrypt data at rest and in transit using encryption tools and protocols such as VPNs and SSL/TLS.

Access Controls and Privilege Management:

Implement strong access controls and privilege management to restrict user access to only what is necessary.

Logging and Auditing:

Set up comprehensive logging and auditing mechanisms to track activities within your private cloud.

Security Concerns

Security Concerns

Amazon Web Services (AWS) is one of the most widely used cloud computing platforms in the world, and like any other cloud service, it comes with its own set of security concerns.

Security in AWS is a shared responsibility model, where AWS manages the security of the cloud infrastructure, and customers are responsible for securing their data and applications running on AWS

Security Concerns

Data Breaches:

- Data Encryption: Use encryption for data at rest (using services like AWS KMS) and data in transit (using SSL/TLS).
- Access Controls: Implement strict IAM (Identity and Access Management) policies to control who has access to your resources.
- Data Classification: Classify your data and use appropriate access controls based on sensitivity.

Identity and Access Management (IAM):

- Least Privilege Principle: Grant only the minimum necessary permissions to users, roles, and services.
- Multi-Factor Authentication (MFA): Enable MFA for IAM users to add an extra layer of security.
- Regular Review: Continuously review and audit IAM policies to ensure they align with your security requirements.

Security Concerns

Network Security:

- Virtual Private Cloud (VPC): Use VPCs to isolate and segment your network resources.
- Security Groups and Network ACLs: Implement security groups and network ACLs to control inbound and outbound traffic.
- DDoS Mitigation: Use AWS Shield and AWS WAF to protect against Distributed Denial of Service (DDoS) attacks.

Data Loss Prevention:

- Regular Backups: Implement automated backups and snapshot policies for critical data.
- Versioning: Enable versioning for Amazon S3 buckets to recover from accidental deletions or modifications.
- Cross-Region Replication: Replicate data across multiple regions for disaster recovery.

Legal issues and Aspects

Legal issues and Aspects

When using Amazon Web Services (AWS), there are several legal issues and aspects that organizations and users should be aware of.

AWS provides a robust and secure cloud infrastructure, but the responsibility for legal compliance often falls on the customer.

Data Privacy and Compliance

Data Protection Laws: Depending on your location and the location of your customers, you may need to comply with data protection regulations such as the General Data Protection Regulation (GDPR) in Europe or the California Consumer Privacy Act (CCPA) in the United States.

Data Processing Agreement (DPA): AWS offers a DPA that outlines the responsibilities of both AWS and the customer regarding data protection. It's important to understand and agree to the terms of this document when handling personal data.

Data Ownership and Responsibility

Data Ownership: Clarify who owns the data you store on AWS and understand AWS's policies regarding data ownership.

Data Classification: Properly classify and label data to ensure that sensitive data is adequately protected.

Intellectual Property (IP) Rights

License Compliance: Ensure that you have the necessary rights and licenses to use any software or intellectual property when running applications on AWS.

AWS Marketplace: When using AWS Marketplace software, review and comply with the licensing terms specified by the software vendors.

Compliance and Regulatory Requirements

Industry-Specific Regulations: Many industries have specific regulatory requirements (e.g., healthcare, finance). Ensure that your AWS environment complies with any applicable industry regulations.

Third-Party Audits: Understand that AWS undergoes third-party audits for various compliance standards (e.g., SOC 2, ISO 27001). These audits can be leveraged to demonstrate compliance to customers and regulators.

Multi-tenancy issues

Multi-tenancy issues

Multi-tenancy is a common architectural model in cloud computing where multiple customers or tenants share the same physical infrastructure and resources, such as servers, storage, and network infrastructure.

AWS, as a leading cloud provider, supports multi-tenancy, and while it offers strong isolation mechanisms to protect customer data and resources, there are still potential issues and considerations to be aware of:-

Multi-tenancy issues

Security Isolation:

- Data Segregation
- Virtual Private Cloud (VPC)

Data Privacy and Compliance:

- Data Residency
- Data Encryption

Resource Allocation and Performance:

- Resource Contention
- Auto Scaling

Access Control and Identity Management:

- IAM Policies
- Role-Based Access Control (RBAC)

INTRODUCTION TO CLOUD SIMULATOR

- CloudSim is a framework for modeling and simulation of cloud computing infrastructures and services.
- CloudSim is completely written in Java
- CloudSim Plus is a totally re-engineered CloudSim fork providing general-purpose cloud computing simulation and exclusive features such as: multi-cloud simulations, vertical and horizontal VM scaling, host fault injection and recovery, joint power- and network-aware simulations and more.
- Though CloudSim itself does not have a graphical user interface, extensions such as CloudReports offer a GUI for CloudSim simulations.
- CloudSimEx extends CloudSim by adding MapReduce simulation capabilities and parallel simulations.
- Cloud2Sim extends CloudSim to execute on multiple distributed servers, by leveraging Hazelcast distributed execution framework.

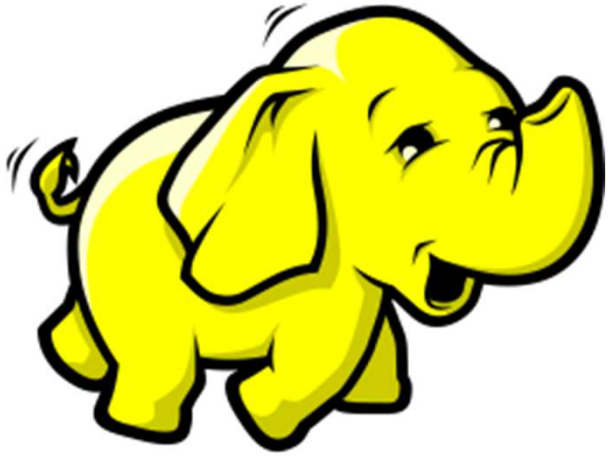

Cloud Sim setup
process in eclipse
Java

HADOOP

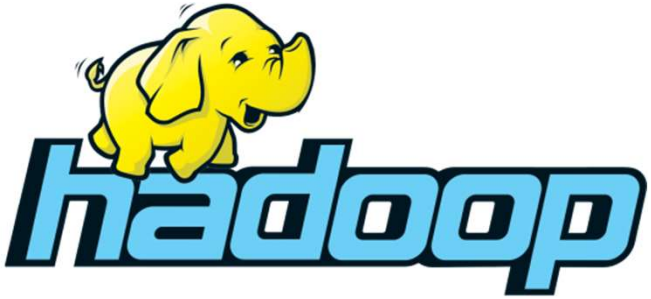
Introduction to Hadoop Framework

Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation.

It provides a software framework for distributed storage and processing of big data using the MapReduce programming model



What is Hadoop?



An Open Source framework that allows distributed processing of large data-sets across the cluster of commodity hardware

Open Source:

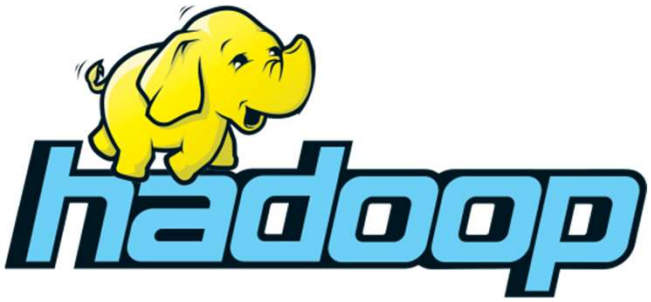
Source code is freely available,It may be redistributed and modified

Distributed Processing:

Data is processed distributedly on multiple nodes / servers

Multiple machines processes the data independently

What is hadoop ? ..II



An Open Source framework that allows distributed processing of large data-sets across the cluster of commodity hardware

Cluster:

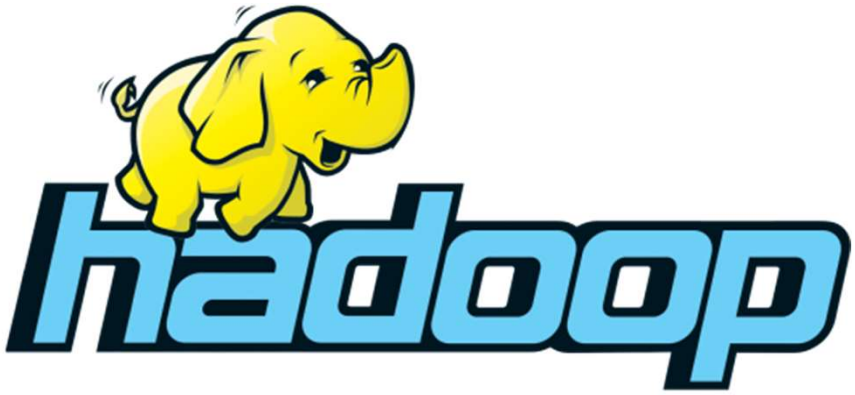
Multiple machines connected together,Nodes are connected via LAN

Commodity Hardware:

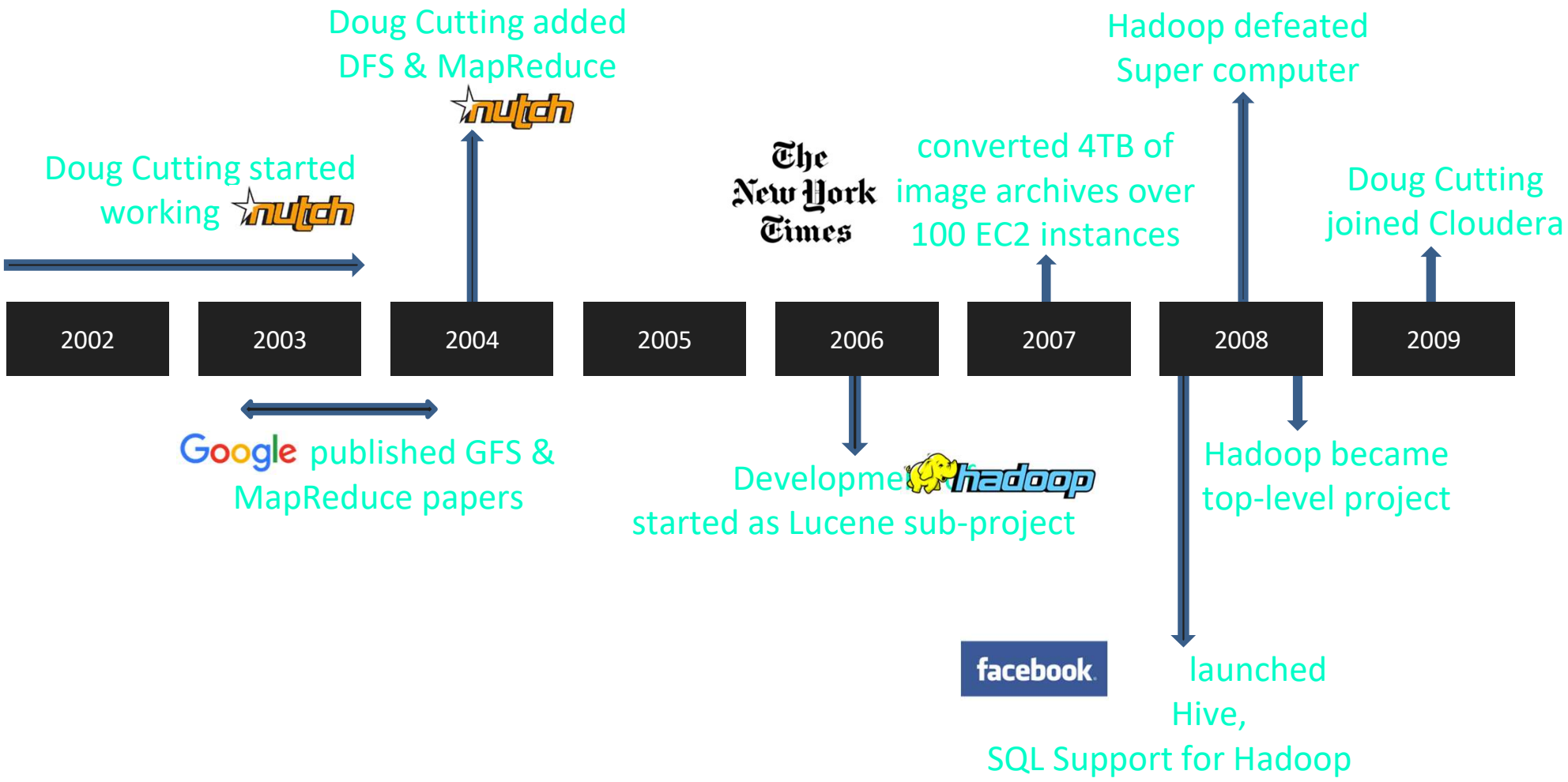
Economic / affordable machines ,Typically low performance hardware

What is Hadoop?

- Open source framework written in Java
- Inspired by Google's Map-Reduce programming model as well as its file system (GFS)

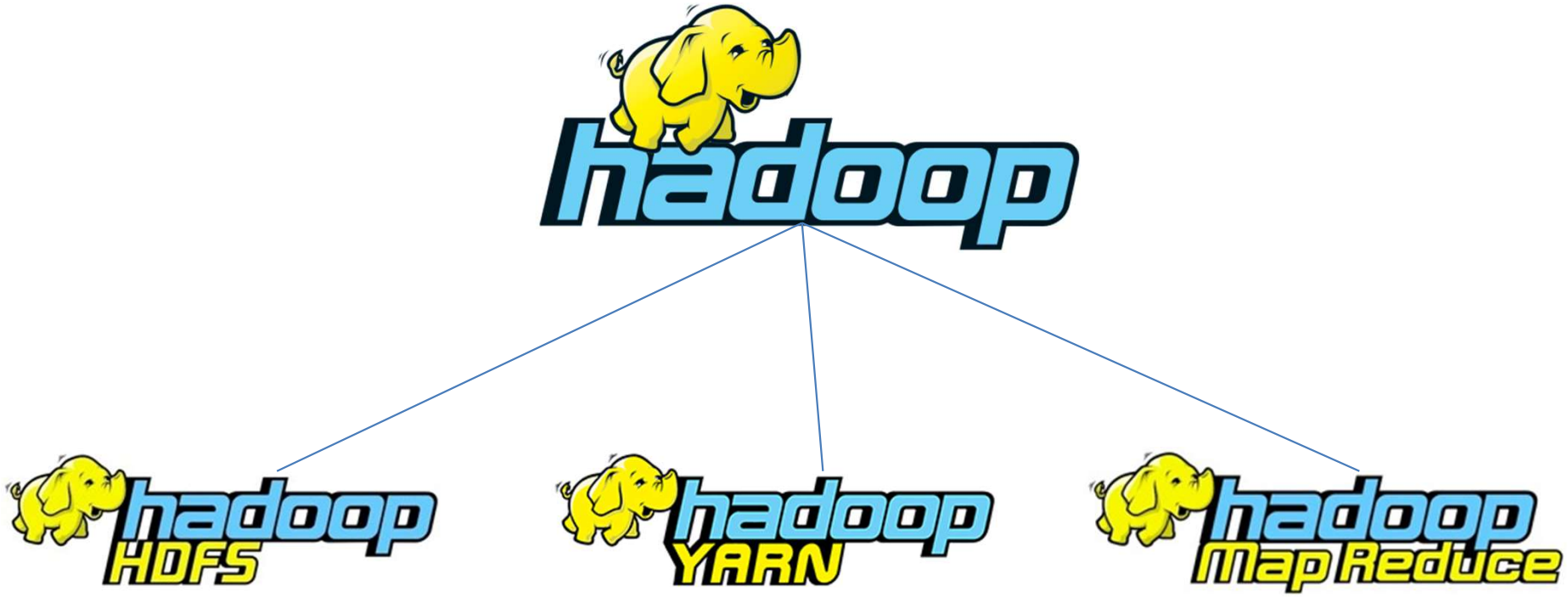


Hadoop History

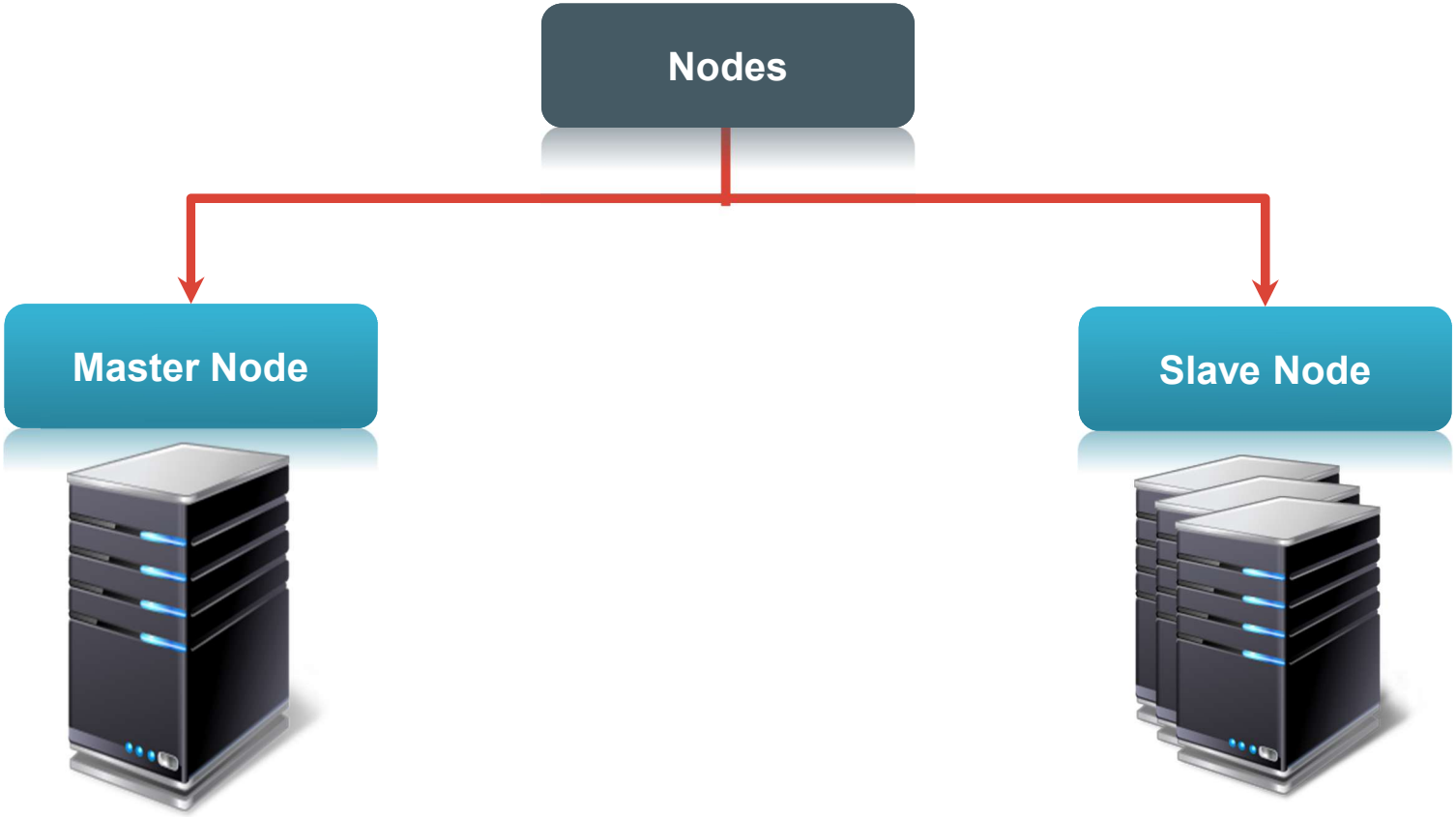


Hadoop Components

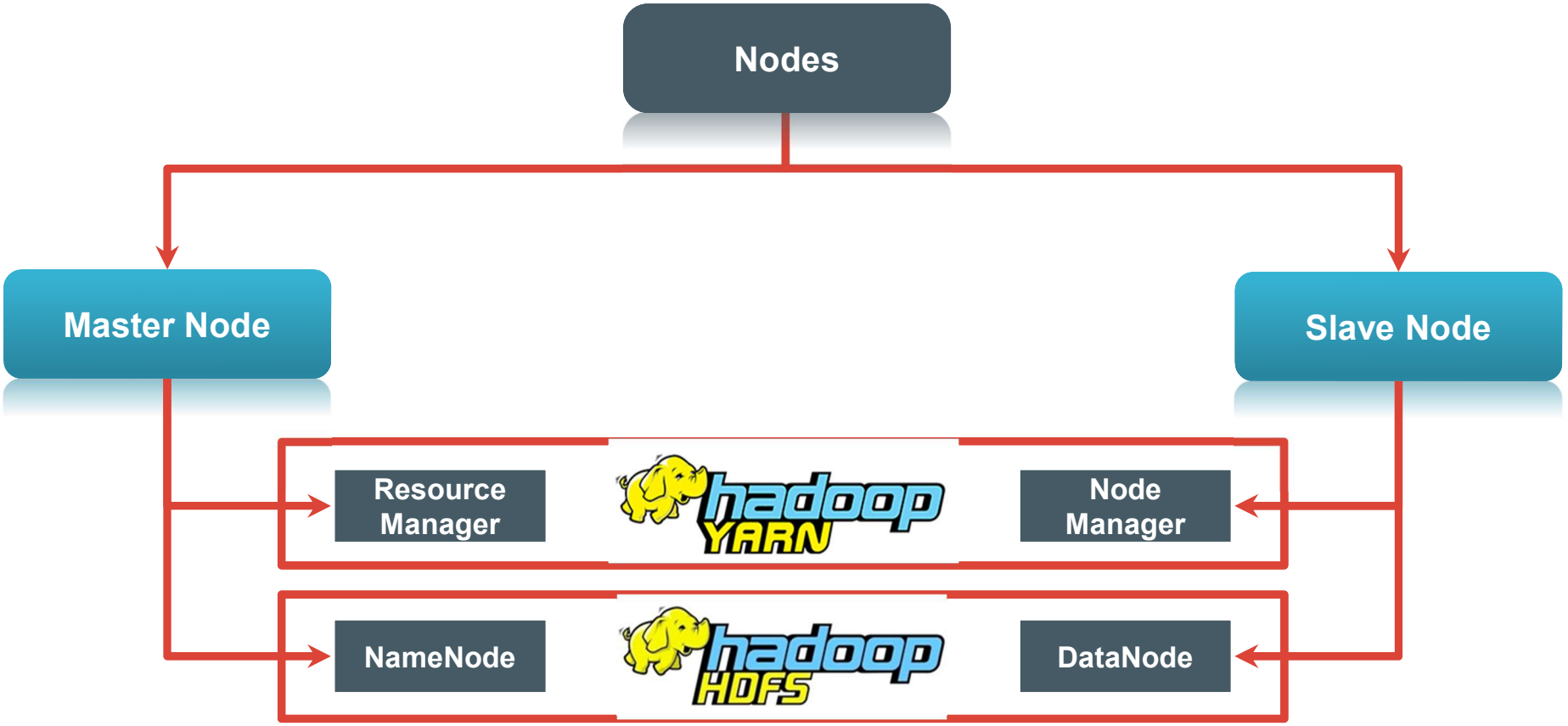
Hadoop consists of three key parts



Hadoop Nodes

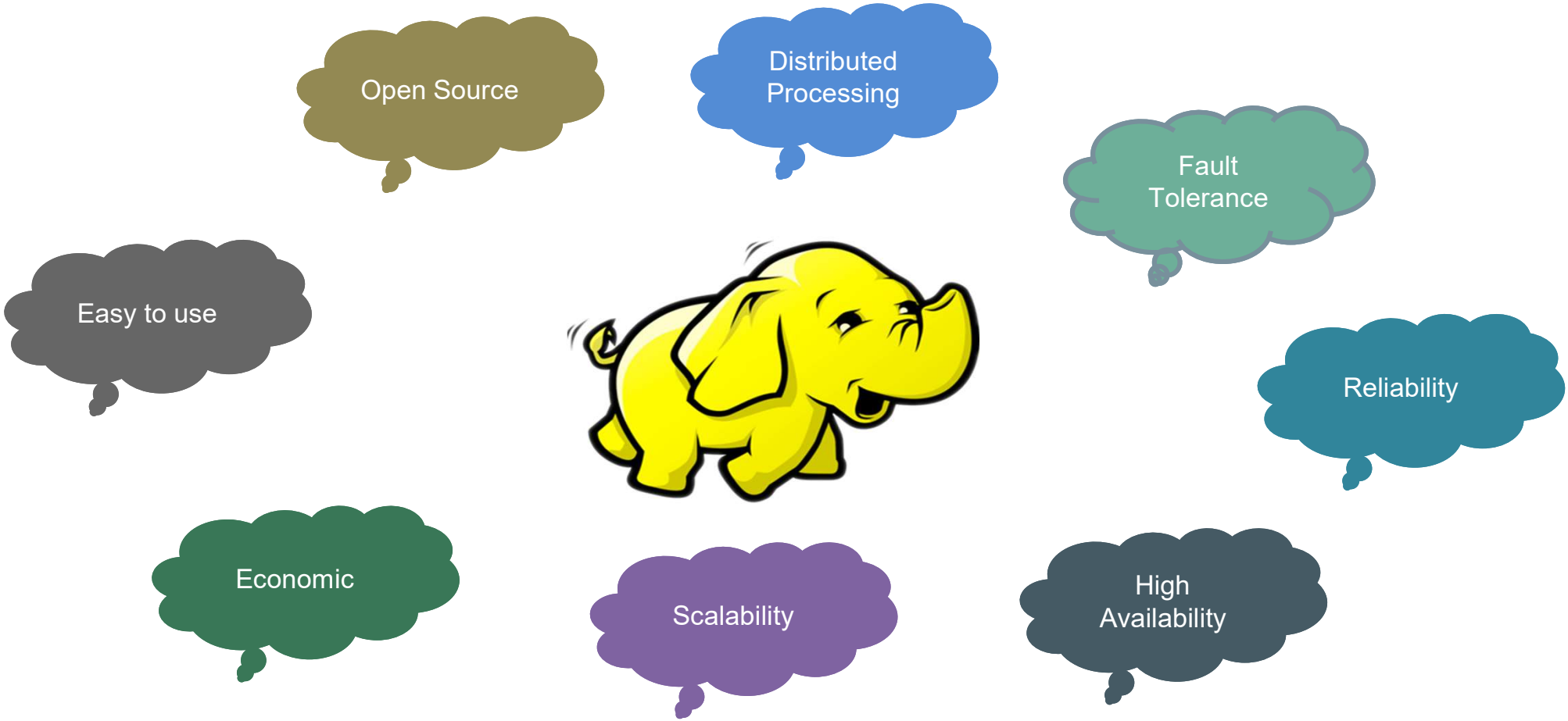


Hadoop Daemons



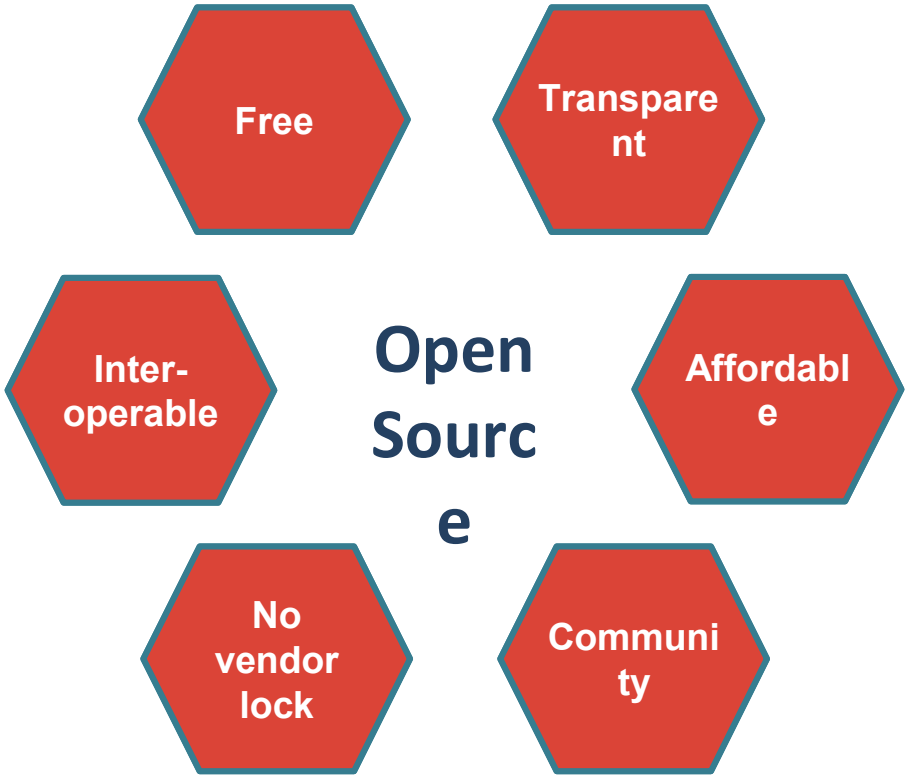
Basic Hadoop Architecture

Hadoop Characteristics



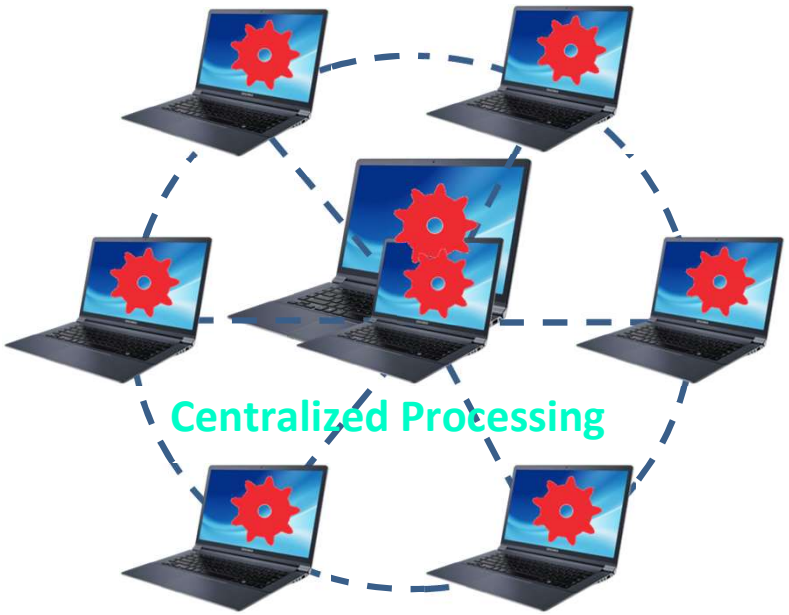
Open Source

- Source code is freely available
- Can be redistributed
- Can be modified



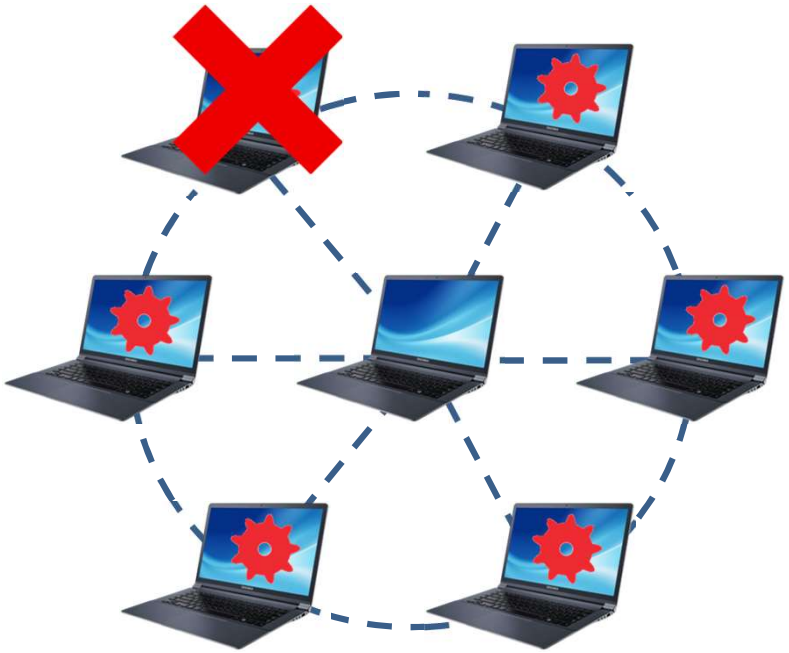
Distributed Processing

- Data is processed distributedly on cluster
- Multiple nodes in the cluster process data independently



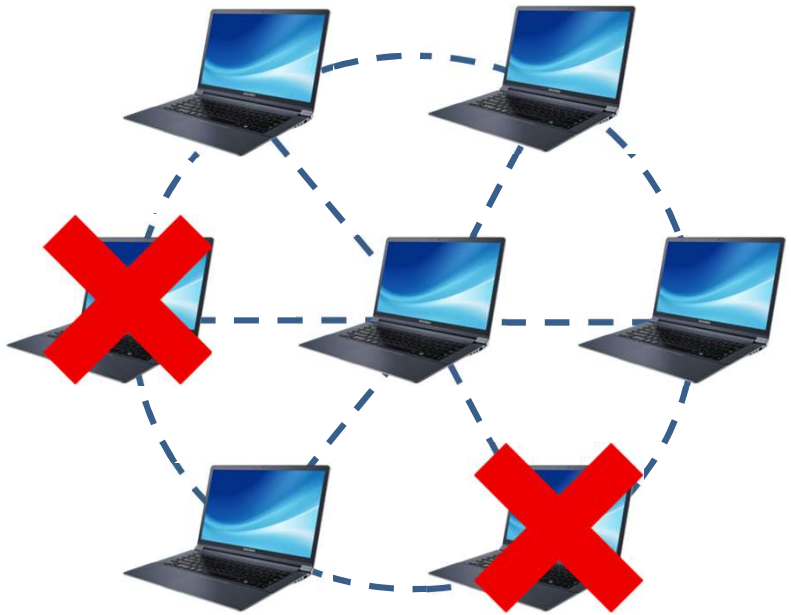
Fault Tolerance

- Failure of nodes are recovered automatically
- Framework takes care of failure of hardware as well tasks



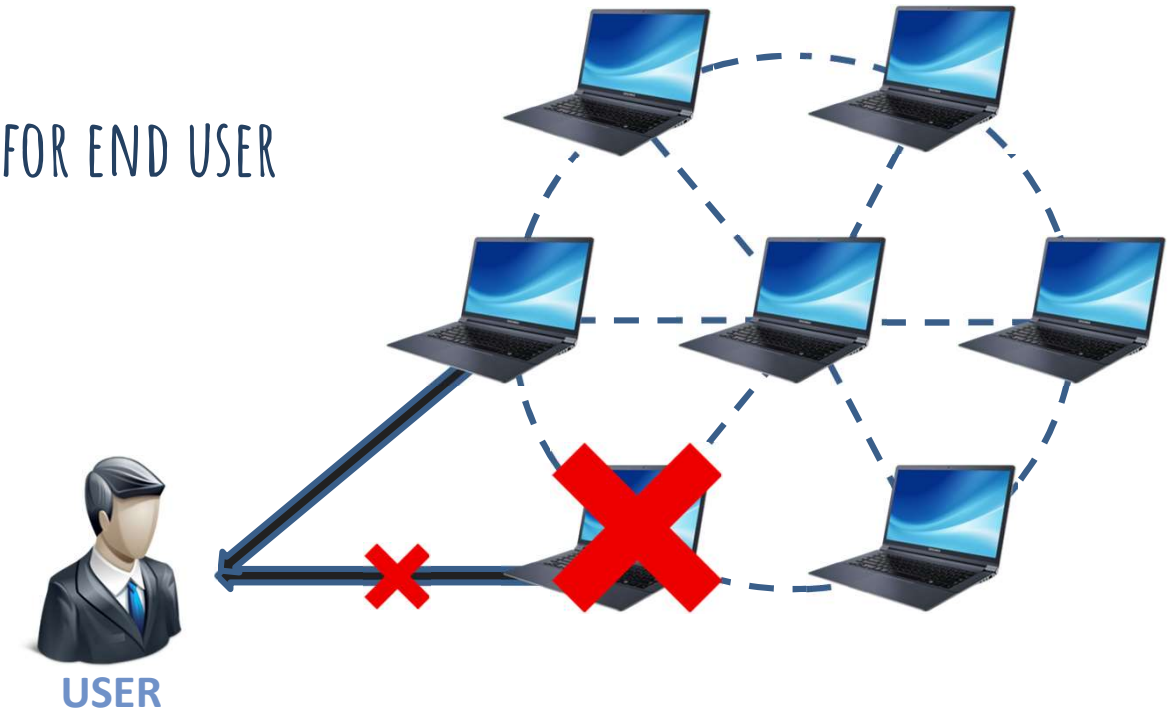
Reliability

- Data is reliably stored on the cluster of machines despite machine failures
- Failure of nodes doesn't cause data loss



High Availability

- Data is highly available and accessible despite hardware failure
- There will be no downtime for end user application due to data



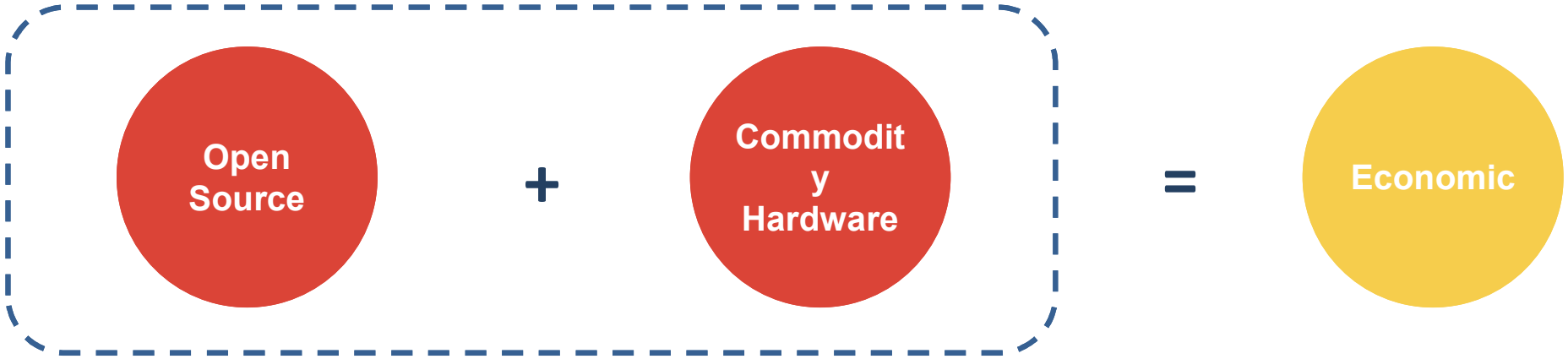
Scalability

- Vertical Scalability – New hardware can be added to the nodes

- Horizontal Scalability – New nodes can be added on the fly

Economic

- No need to purchase costly license
- No need to purchase costly hardware

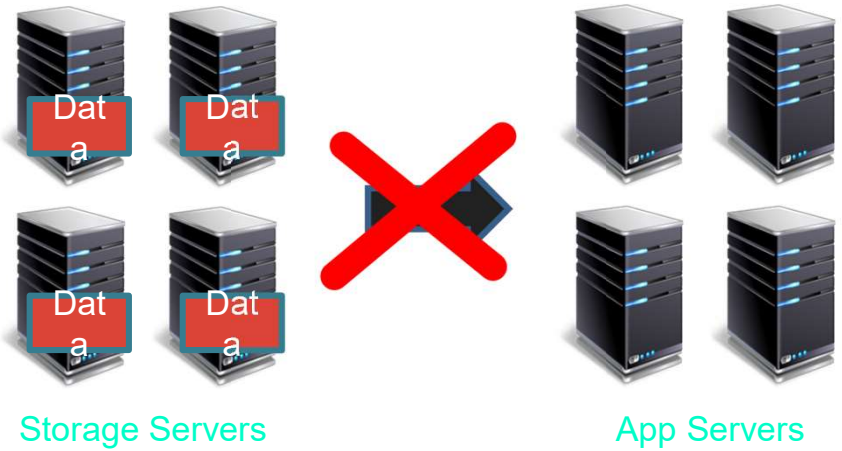


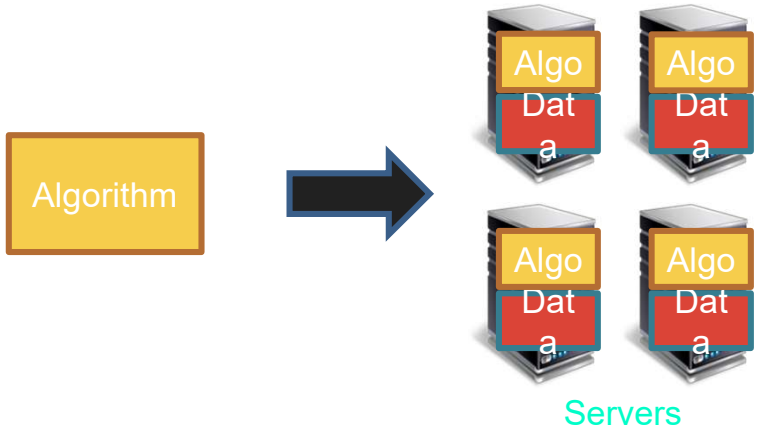
Easy to Use

- Distributed computing challenges are handled by framework
- Client just need to concentrate on business logic

Data Locality

- Move computation to data instead of data to computation
- Data is processed on the nodes where it is stored

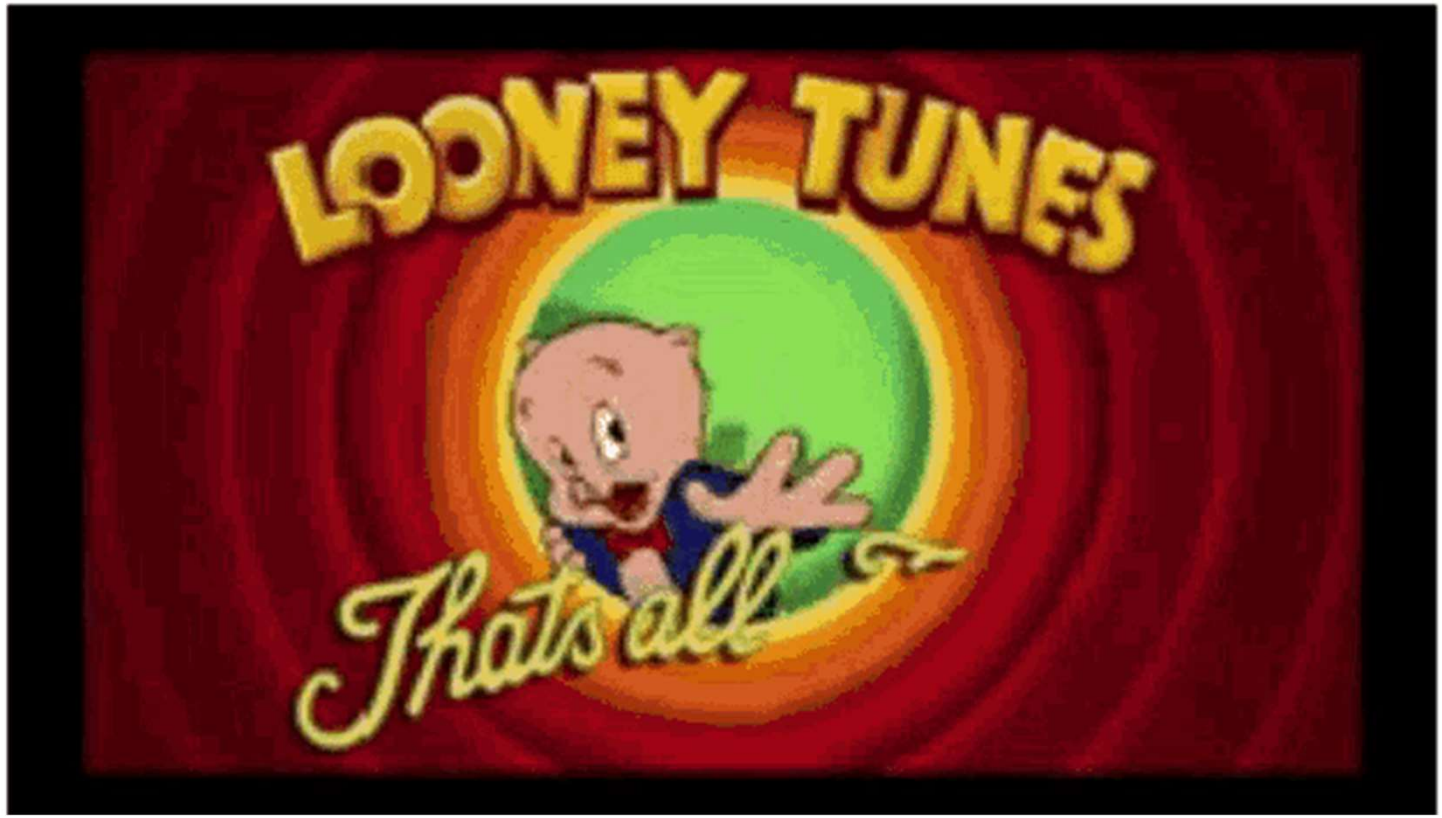




Pros and cons of using hadoop

- Cost
- Scalability
- Flexibility
- Speed
- Fault Tolerance
- High Throughput
- Minimum Network Traffic

- Problem with Small files
- Vulnerability
- Low Performance In Small Data Surrounding
- Lack of Security
- High Up Processing
- Supports Only Batch Processing



Thats all folks for this chapter !!!!