



# HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytics Workload





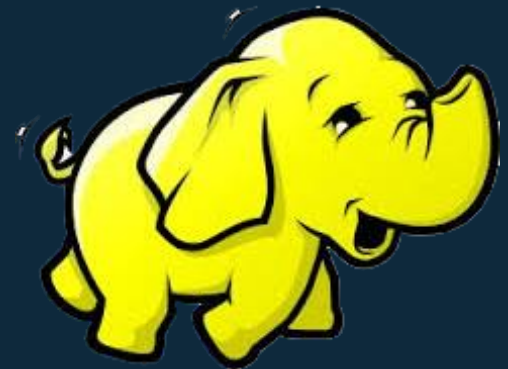
# Introduction

- What is Big Data?
- Limitations of Current Solutions
- Introduction of Parallel DBMS
- MapReduce

# What is Hadoop?

- Opensource Parallel computing model for Big data.
- Fault-tolerant
- Scalable framework
- Java-based

“How to process big data with reasonable cost and time?”



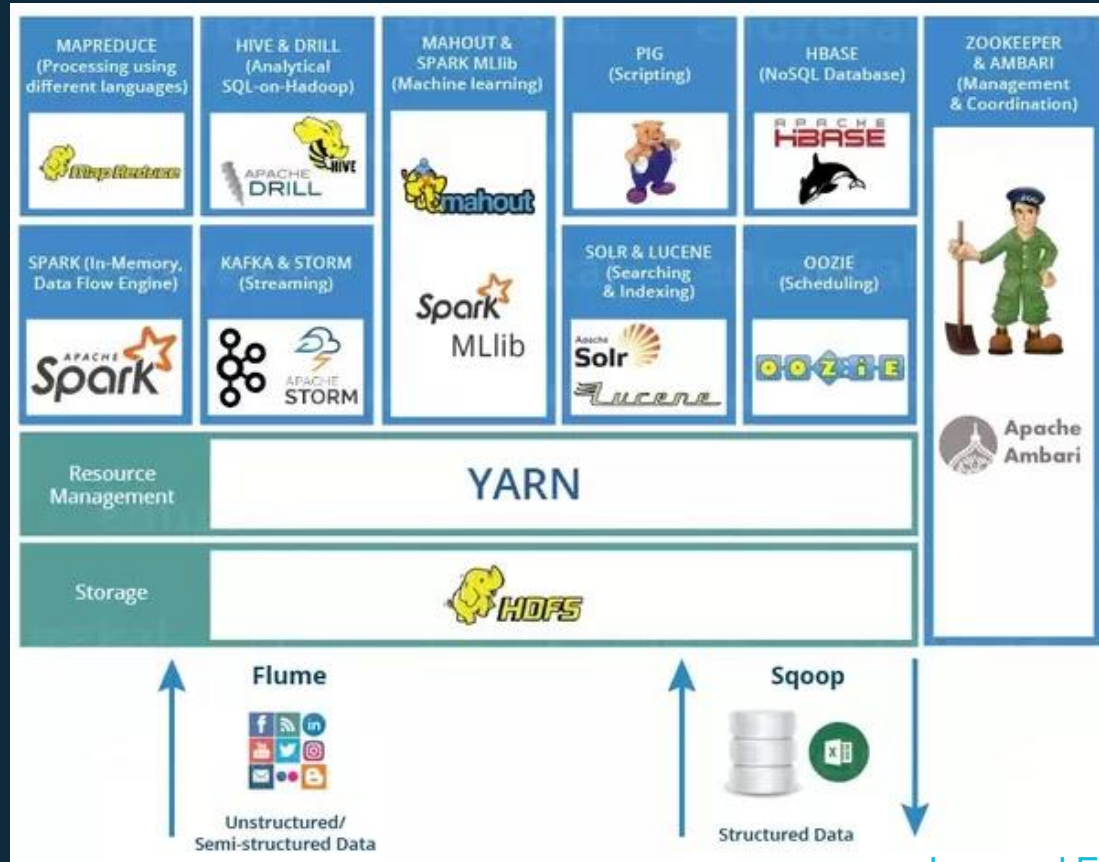
**hadoop**



# Hadoop Architecture

- Distributed DBMS with some centralization.
- Main nodes
- Data Node
- Central control node/Name Node

# The Hadoop Ecosystem



# Uses of Hadoop:

- **Yahoo!**
- **Amazon**
- **Facebook**
- **Netflix**



# Benefits of Hadoop:



- **Speed**
- **Flexible**
- **Fault - resistant**
- **Cost-effective**
- **Scalable**

# Challenges of Hadoop

- ◇ Talent Gap
- ◇ Data Security
- ◇ Data Management and Governance.





# Conclusion





# Thanks!

## Any questions?

You can find me at:

◇ [@lemanuel@ur.rochester.edu](mailto:@lemanuel@ur.rochester.edu)





# Credits

Special thanks to all the people who made and released these awesome resources for free:

- ◇ Presentation template by [SlidesCarnival](#)
- ◇ Photographs by [Unsplash](#)

