

# Apache Spark Through Email

Markus Dale

Nov 2018

# Slides And Code

- ▶ Slides: <https://github.com/medale/spark-mail/blob/master/presentation/ApacheSparkThroughEmail.pdf>
- ▶ Spark Code Examples:  
<https://github.com/medale/spark-mail/>

# Data Science for Small Dataset



Figure 1: Laptop

## Data Science for Larger Dataset



Figure 2: Standalone Server

# Data Science for Larger Dataset (Vertical Scaling)

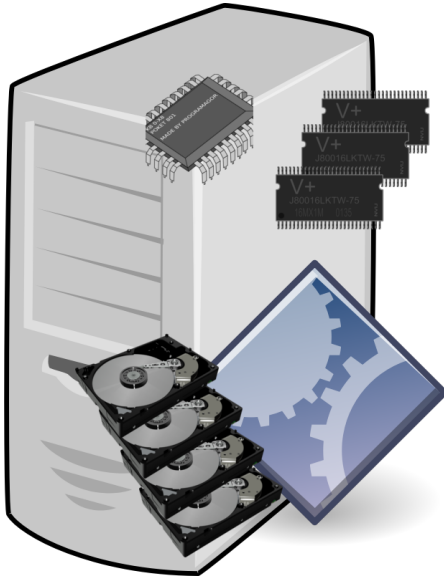


Figure 3: Beefed-up Server

# Data Science for Large Datasets (Horizontal Scaling)



Figure 4: Multiple cooperating Servers

# Big Data Framework - Apache Hadoop



Figure 5: HDFS, MapReduce

# Hadoop Ecosystem

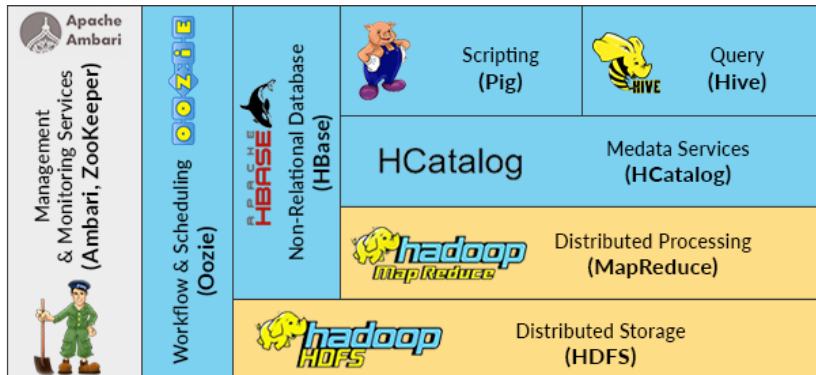


Figure 6: Some Frameworks Around Hadoop