

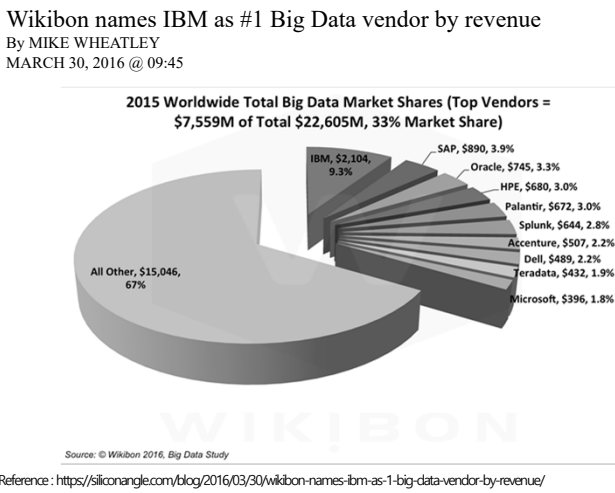
# Big Data Emerging Technologies

## Introduction

Prof. Jong-Moon Chung

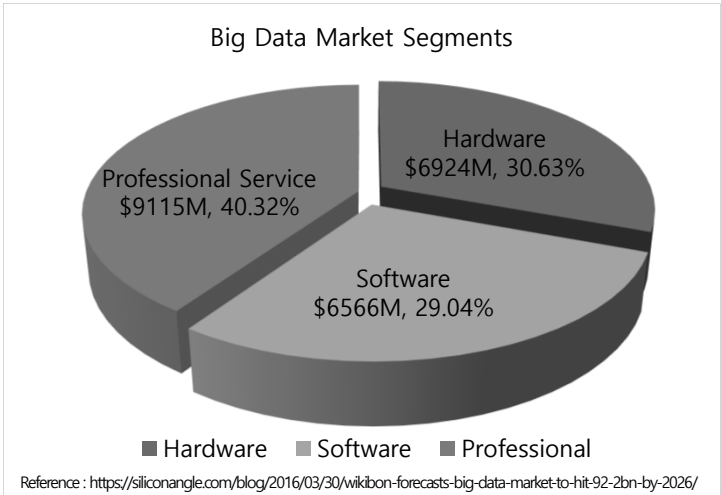
### Big Data Emerging Technologies

#### ❖ Big Data Corporate Rankings



## Big Data Emerging Technologies

### ❖ Big Data Market Shares

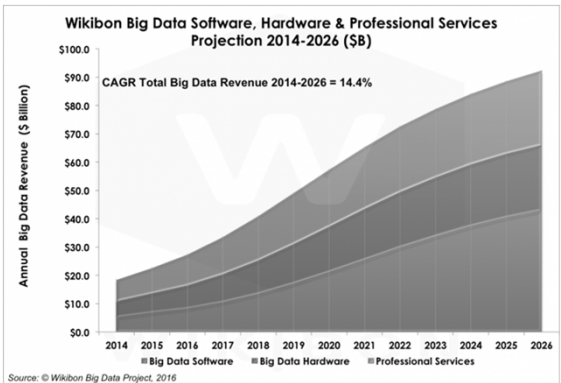


## Big Data Emerging Technologies

### ❖ Software vs. Hardware vs. Services

Wikibon forecasts Big Data market to hit \$92.2B by 2026

By MIKE WHEATLEY  
MARCH 30, 2016 @ 09:44



Reference : <https://siliconangle.com/blog/2016/03/30/wikibon-forecasts-big-data-market-to-hit-92-2bn-by-2026/>

## Big Data Emerging Technologies

### ❖ Top Big Data Corporate Products



## Big Data Emerging Technologies

### ❖ H1N1 Flu Virus Starts in the U.S.

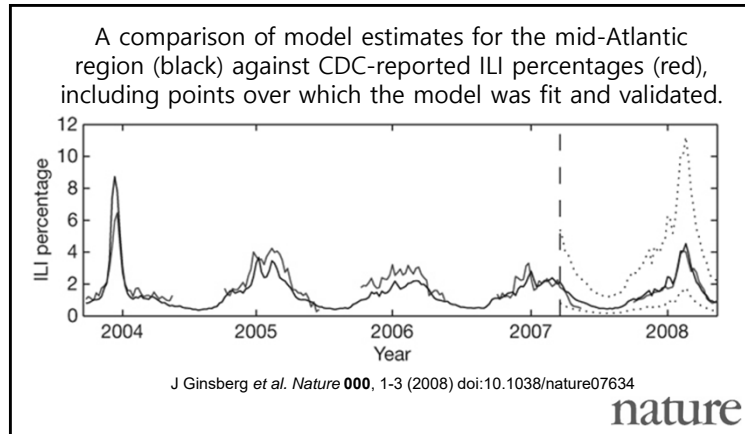
- U.S. CDC (Centers for Disease Control and Prevention) diagnostic data
- What vaccine was needed?
- How much vaccine was needed?
- Where was the vaccine needed?

The Google logo, consisting of the word "Google" in its characteristic multi-colored font.

## Big Data Emerging Technologies

### Detecting Influenza Epidemics using Search Engine Query Data

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. *Nature*, 457, pp. 1012–1014, February 19, 2009.



Reference: <https://www.nature.com/articles/nature07634t>

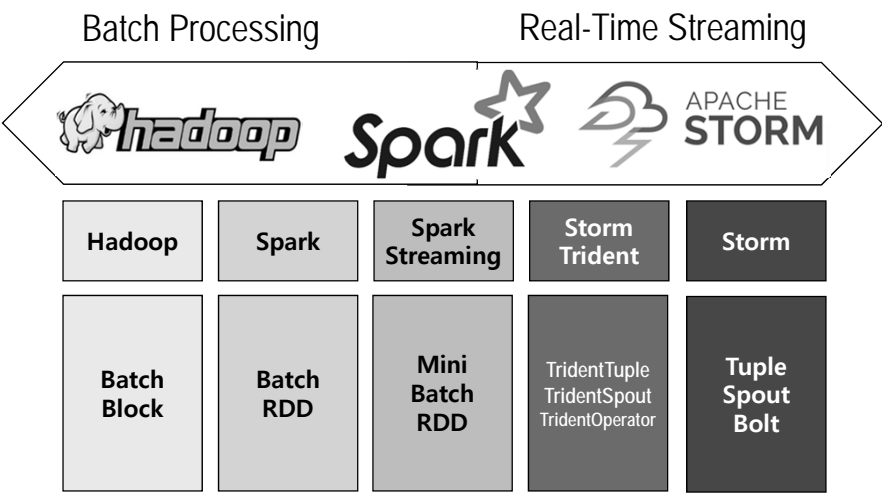
## Big Data Emerging Technologies

### ❖ 4V Big Data Challenges

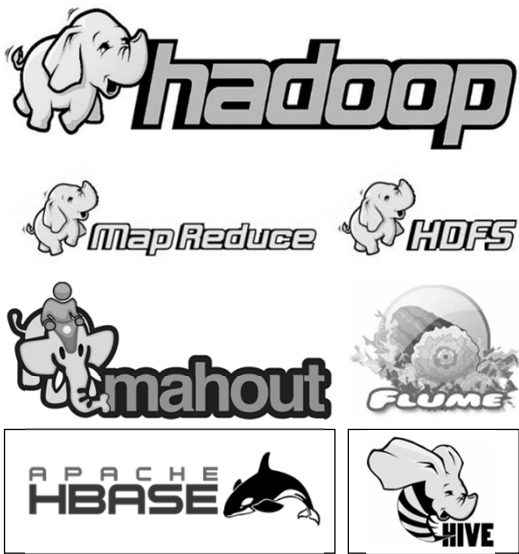
- Volume - Data Size
- Variety - Data Formats
- Velocity - Data Streaming Speeds
- Veracity - Data Trustworthiness



Big Data Emerging Technologies

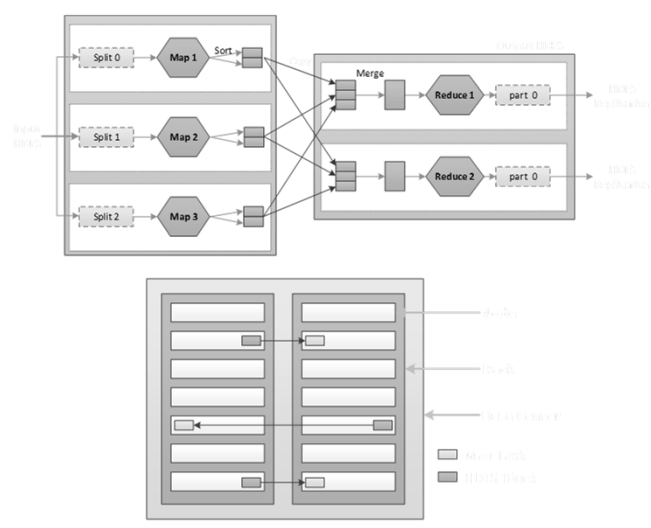


Big Data Emerging Technologies



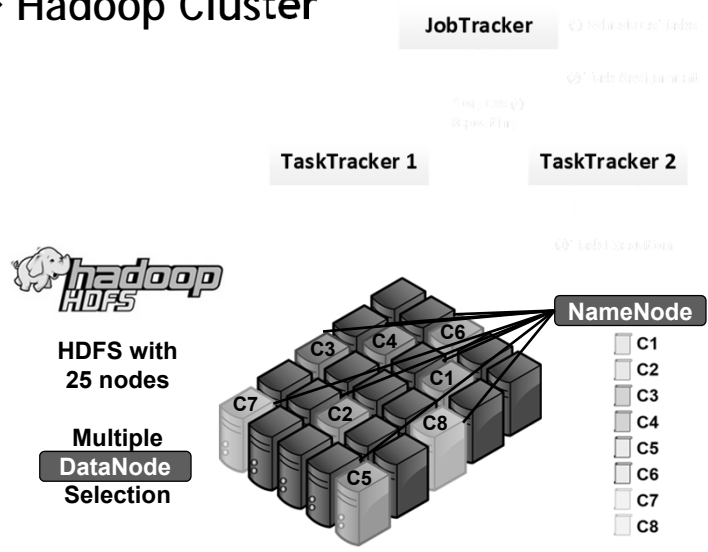
# Big Data Emerging Technologies

## ❖ Hadoop MapReduce & HDFS



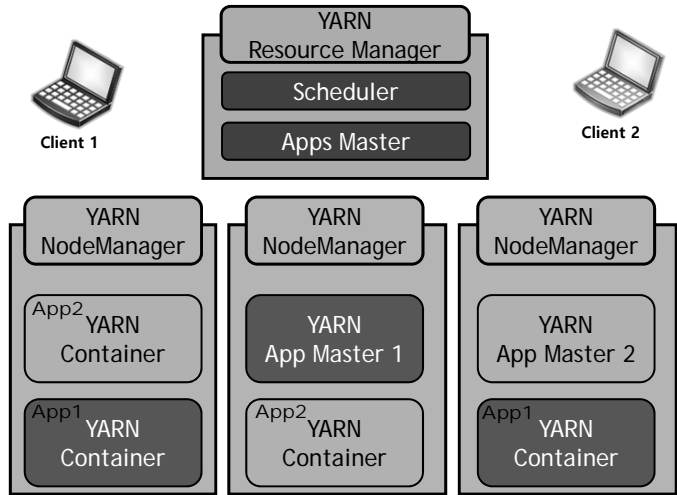
# Big Data Emerging Technologies

## ❖ Hadoop Cluster



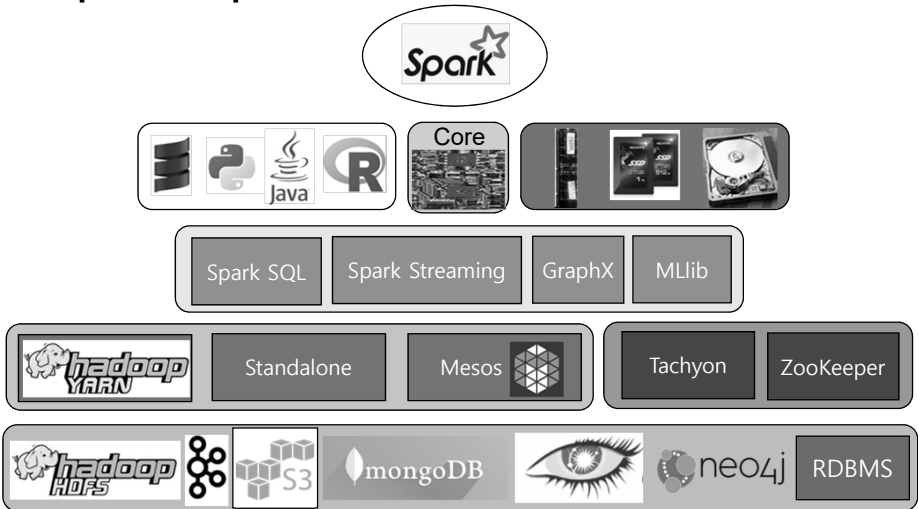
# Big Data Emerging Technologies

## ❖ Hadoop YARN



# Big Data Emerging Technologies

## ❖ Apache Spark



## Big Data Emerging Technologies

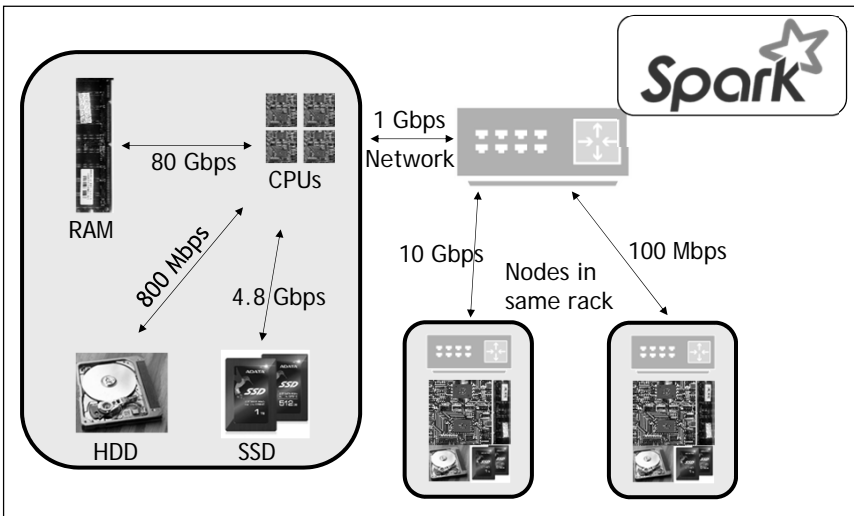
### ❖ ML (Machine Learning)



- Basic statistics
  - Correlation, Hypothesis testing (P-value)
- Classification and Regression
  - Linear models (SVM, Linear & Logistic Regression), Naive Bayes, Decision tree
- Others
  - Collaborative Filtering, Clustering (*k*-means), Dimensionality Reduction, etc.

## Big Data Emerging Technologies

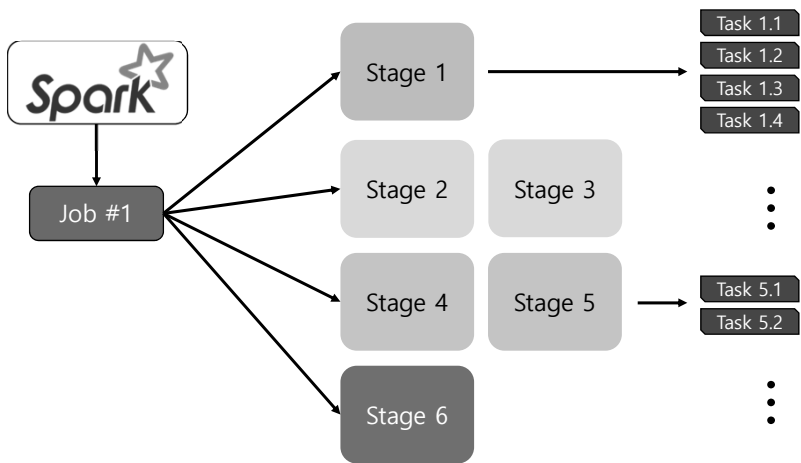
### ❖ Speed Optimized Big Data Processes





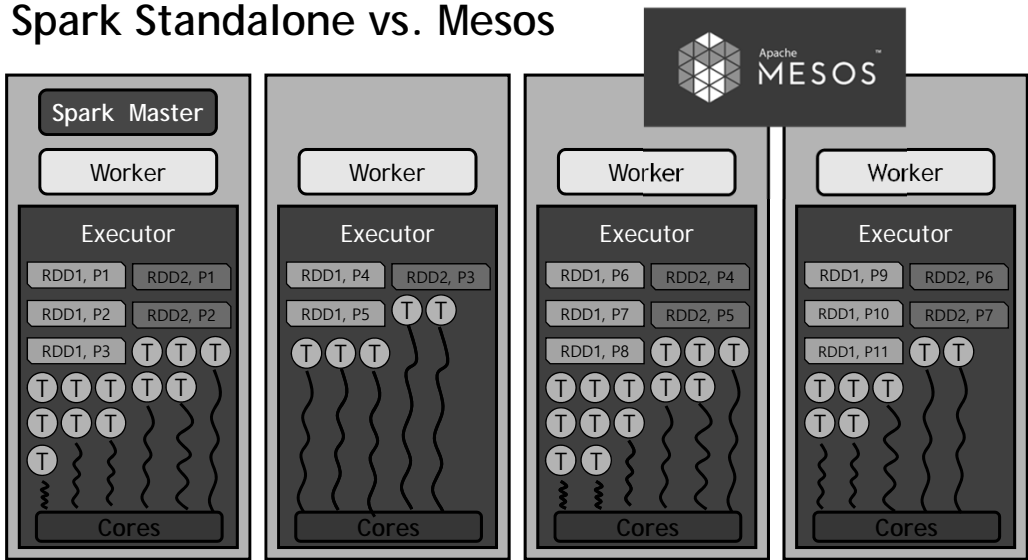
## Big Data Emerging Technologies

❖ Spark Job → Stage → Task → Thread → Core



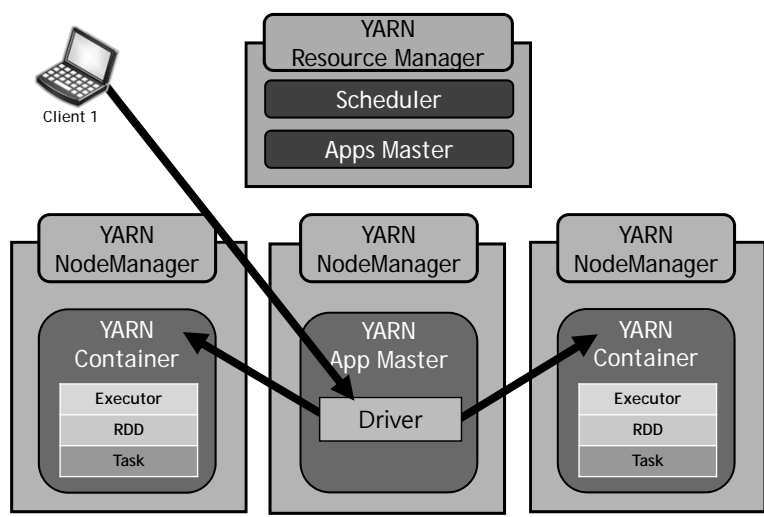
## Big Data Emerging Technologies

❖ Spark Standalone vs. Mesos



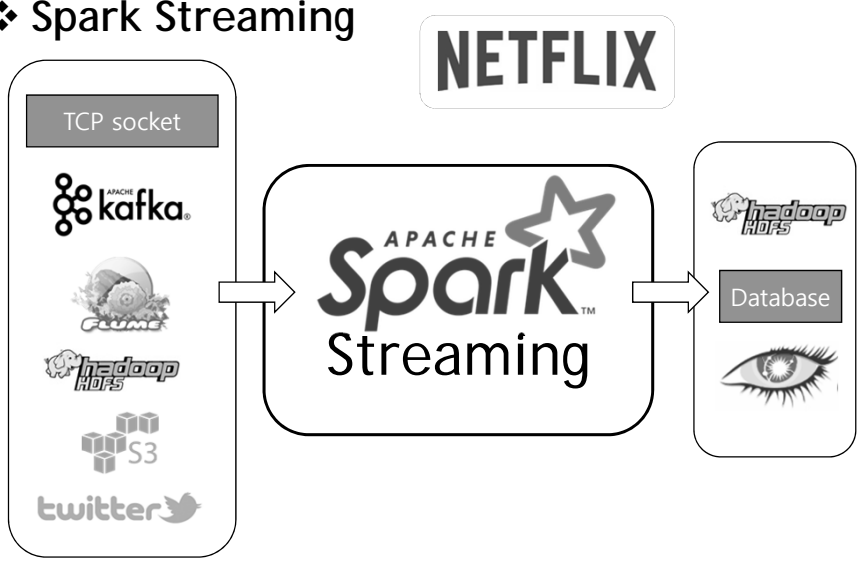
## Big Data Emerging Technologies

### ❖ Spark YARN Cluster Mode



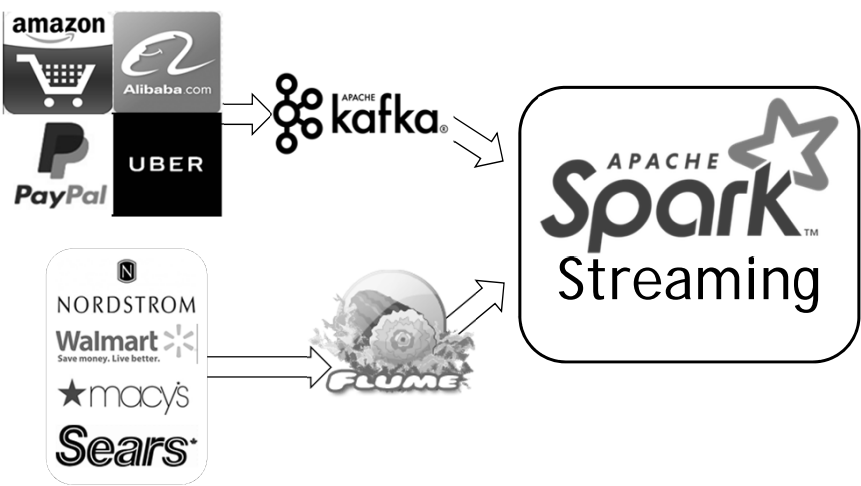
## Big Data Emerging Technologies

### ❖ Spark Streaming

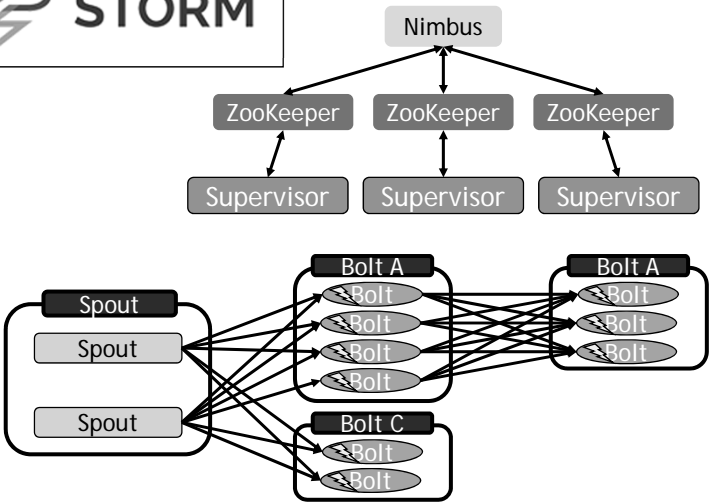


# Big Data Emerging Technologies

## ❖ Spark Streaming

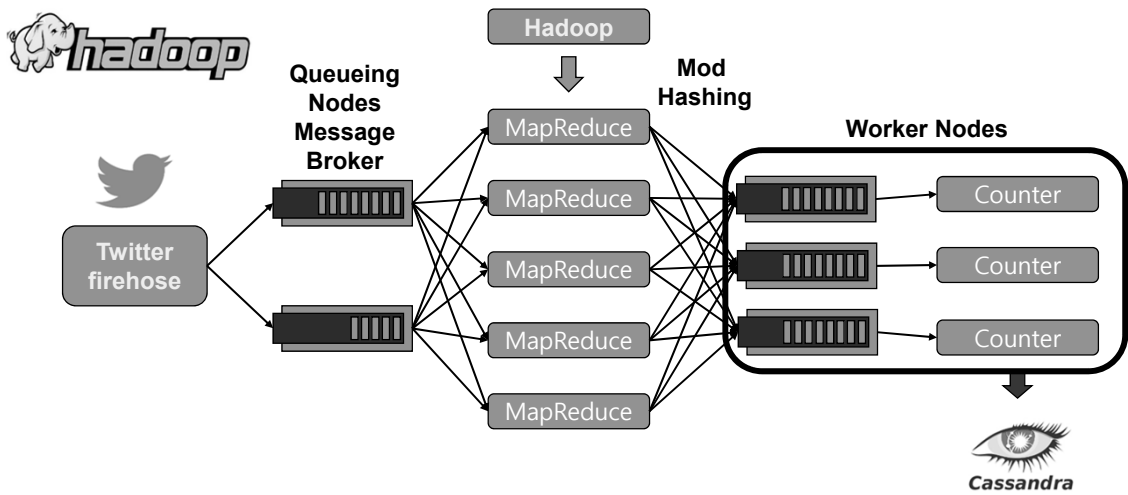


# Big Data Emerging Technologies



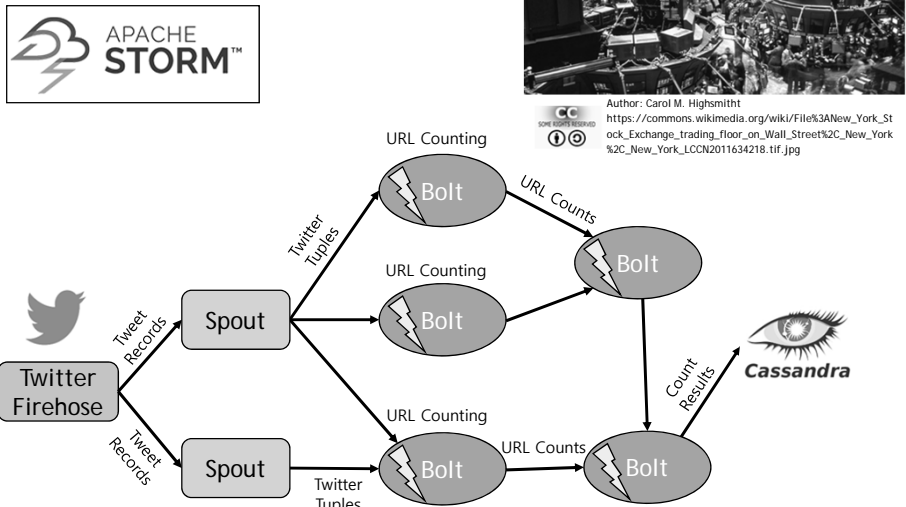
Big Data Emerging Technologies

❖ Twitter using Hadoop



Big Data Emerging Techn

❖ Storm Applications



# Big Data Emerging Technologies

## ❖ IBM SPSS Statistics Project



- IBM’s Big Data Corporate Rankings
  - 1<sup>st</sup> in Big Data Corporations
  - 4<sup>th</sup> in Big Data Software
  - 2<sup>nd</sup> in Big Data Hardware
  - 1<sup>st</sup> in Big Data Professional Services

# Big Data Emerging Technologies

## ❖ IBM SPSS Statistics Project 1 & 2

