



By M.Saravanan

# Finishing Bits/Drivers

**National Workshop on  
Recent Trends in Big Data Analytics  
February 27, 2016**

# Hadoop Job Trends

## Hadoop Developer

A Hadoop developer is responsible for the actual coding/programming of Hadoop applications. Ideally the candidate should have at-least 2 years of experience as a programmer. This role is synonymous to a software developer but in the Big Data domain.

## Hadoop Architect

A Hadoop architect is responsible for planning and designing next-generation "big-data" system architectures. He/She is also responsible for managing the development and deployment of Hadoop applications.

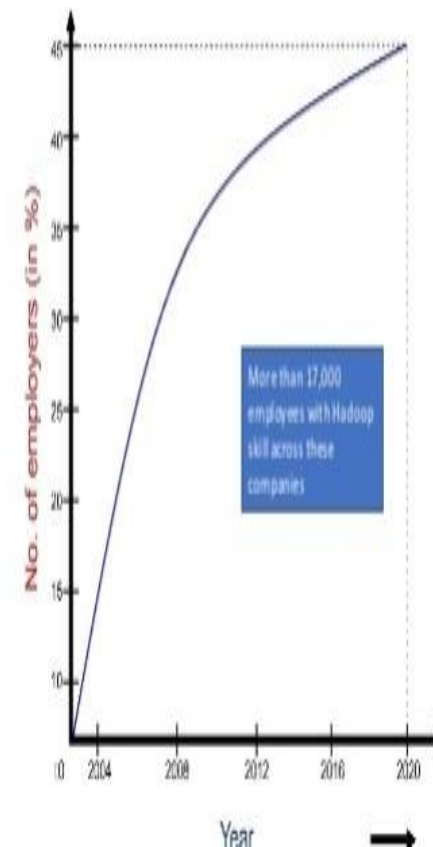
## Data Scientist

For lack of a better term - Data Scientist is believed to be the "Sexiest" job title of the 21st century. Data scientists thrive on solving real world problems with real data. They are very good at using different techniques for analyzing data from different sources to help business make intelligent decisions.

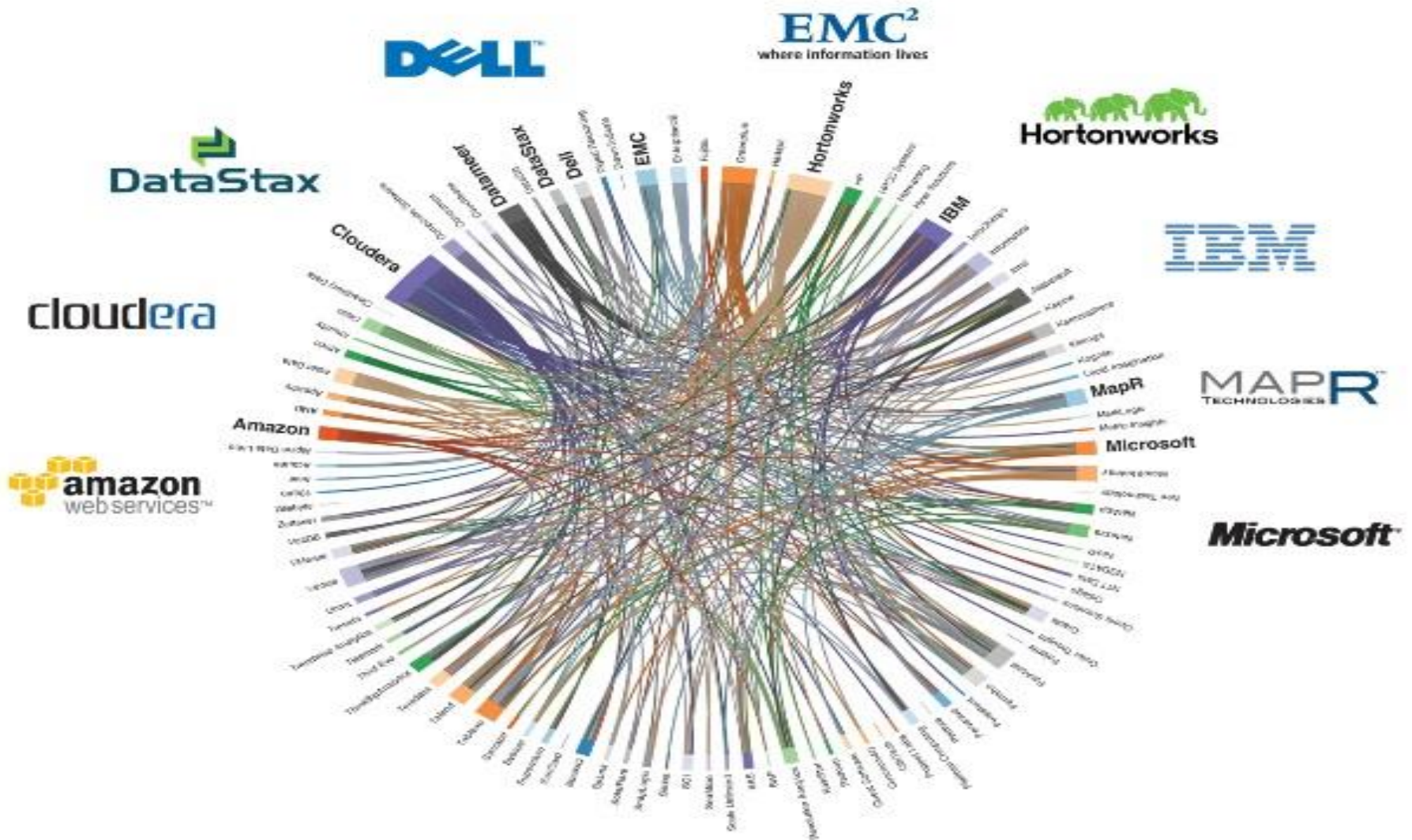
## Hadoop Global Job Trends



Top Hadoop Technology Companies



# Major Players





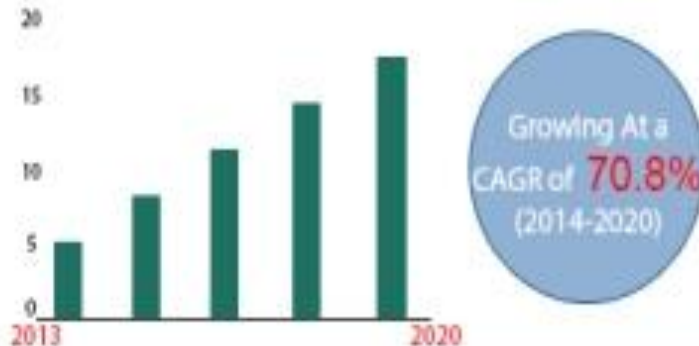
# HaaS

## GLOBAL HADOOP-AS-A-SERVICE (HaaS) MARKET

Size & Forecast, (2013-2020)

### GLOBAL HADOOP-AS-A-SERVICE (HaaS) MARKET

Market is Expected to Reach **\$16.1 billion** by 2020



### GLOBAL HaaS MARKET BY DEPLOYMENT OPTIONS

- Run-it-Yourself (RIY)
- Pure Play/Managed HaaS

### GLOBAL HADOOP-AS-A-SERVICE (HaaS) MARKET BY GEOGRAPHY



### GLOBAL HADOOP-AS-A-SERVICE (HaaS) MARKET BY END USER



### GLOBAL HADOOP-AS-A-SERVICE (HaaS) MARKET DYNAMICS

#### Drivers

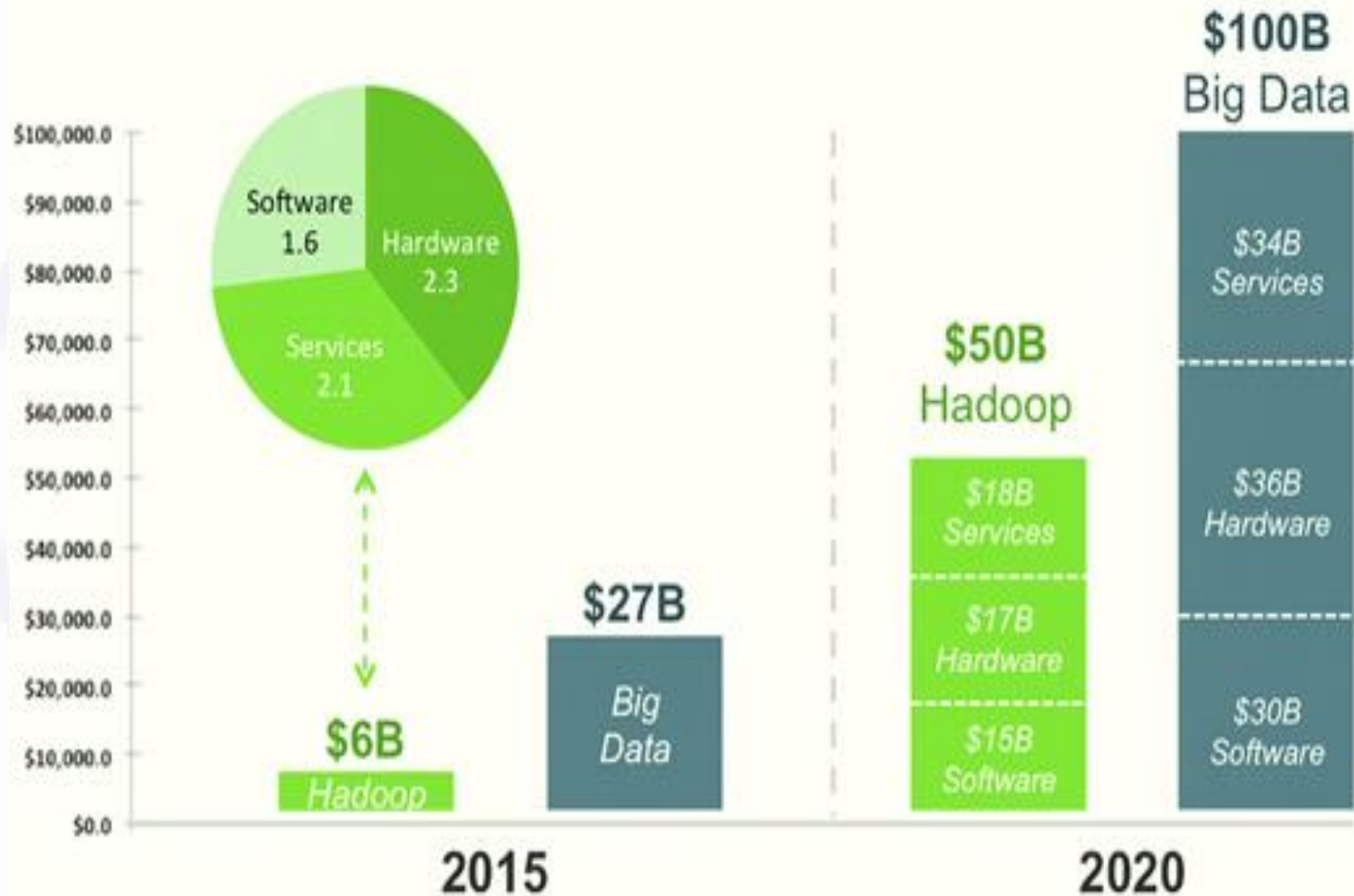
- Increasing competition in the business environment
- Extremely low upfront costs compared to on-premises Hadoop
- Growing demand from SMEs
- Flexibility and Agility for businesses
- Need for technical expertise and complexity reduction
- Dropping price of cloud services

#### Restraints

- Data Privacy
- Lack of maturity of hadoop and awareness among consumers

For More Details See Table Of Contents

# Big Data and Hadoop Markets Growing Sharply

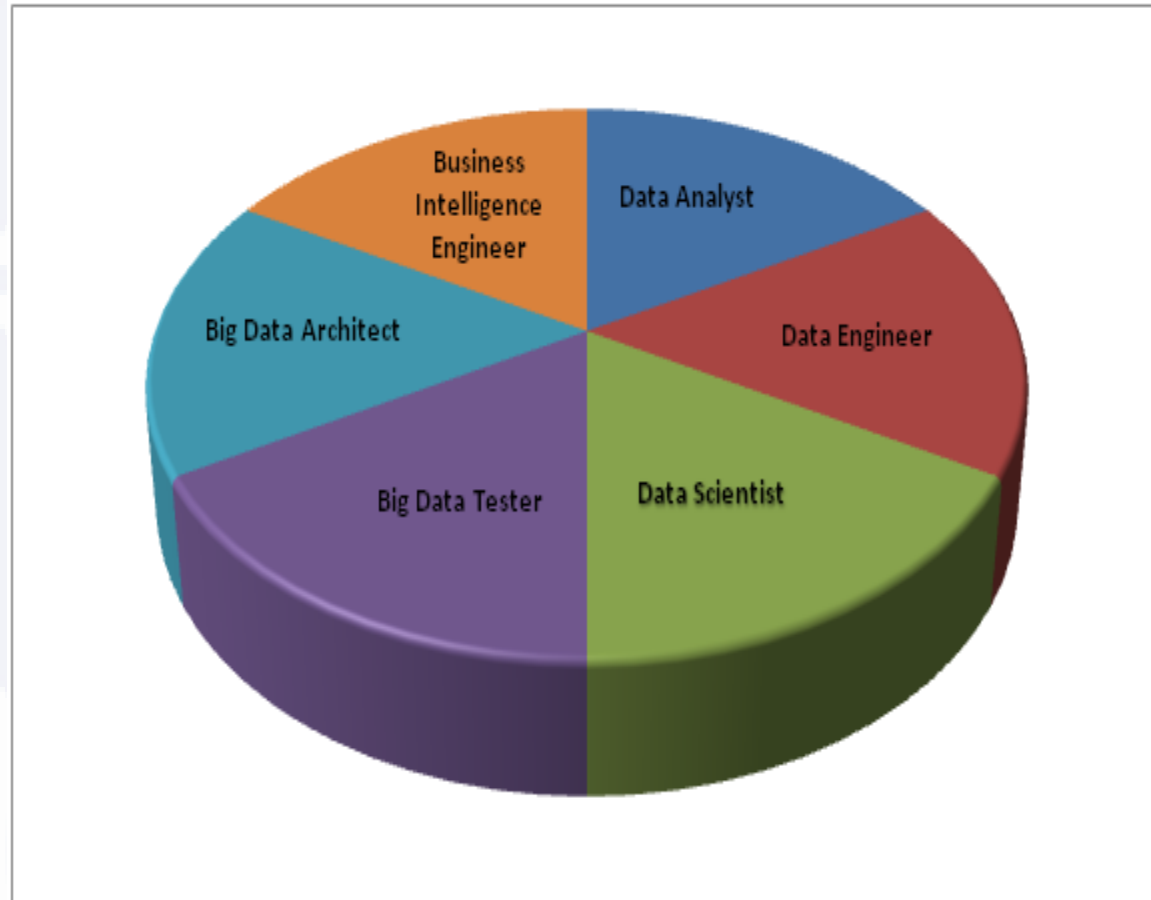




# Virtual Machines

	Hortonworks	Cloudera	MapR
<b>Performance and Scalability</b>			
Data Ingest	Batch	Batch	Batch and streaming writes
Metadata Architecture	Centralized	Centralized	Distributed
HBase Performance	Latency spikes	Latency spikes	Consistent low latency
NoSQL Applications	Mainly batch applications	Mainly batch applications	Batch and online/real-time applications
<b>Dependability</b>			
High Availability	Single failure recovery	Single failure recovery	Self healing across multiple failures
MapReduce HA	Restart jobs	Restart jobs	Continuous without restart
Upgrading	Planned downtime	Rolling upgrades	Rolling upgrades
Replication	Data	Data	Data + metadata
Snapshots	Consistent only for closed files	Consistent only for closed files	Point-in-time consistency for all files and tables
Disaster Recovery	No	File copy scheduling (BDR)	Mirroring
<b>Manageability</b>			
Management Tools	Ambari	Cloudera Manager	MapR Control System
Volume Support	No	No	Yes
Heat map, Alarms, Alerts	Yes	Yes	Yes
Integration with REST API	Yes	Yes	Yes
Data and Job Placement Control	No	No	Yes
<b>Data Access</b>			
File System Access	HDFS, read-only NFS	HDFS, read-only NFS	HDFS, read/write NFS (POSIX)
File I/O	Append only	Append only	Read/write
Security: ACLs	Yes	Yes	Yes
Wire-level Authentication	Kerberos	Kerberos	Kerberos, Native

# Why Java professionals learn Hadoop ?



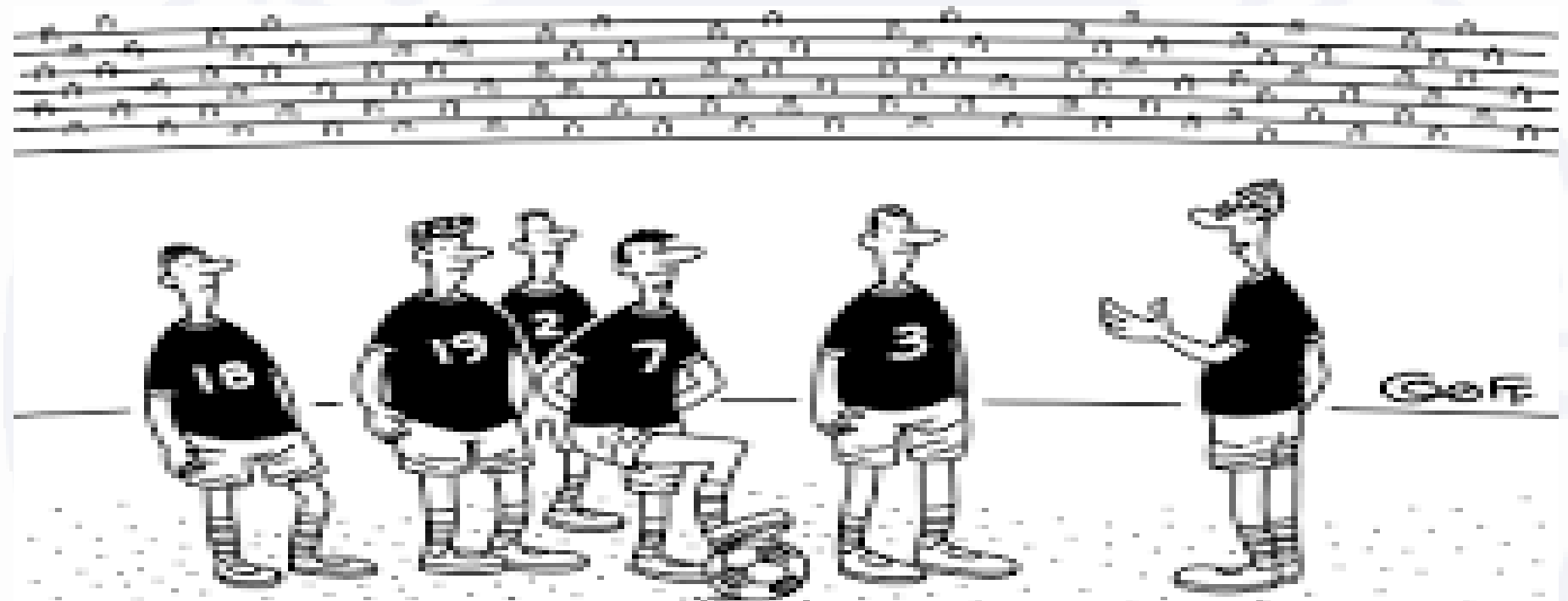
# Welcome



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."



# Big Data Insights



**“Remember, the other team is counting on Big Data insights based on previous games. So, kick the ball with your other foot.”**

# Go out for Dinner

