

A BIG DATA CHEAT SHEET: THE BIG PHARMA EDITION

TAMARA DULL, DIRECTOR OF EMERGING TECHNOLOGIES

sas best
practices
THOUGHT PROVOKING BUSINESS

**Big data
is not
new.**



PATIENT RECORDS	HOSPITAL ADMISSIONS	FINANCIAL DATA	INSURANCE DATA	SCHEDULING DATA	20%
EMAIL	PDF FILES	SPREAD-SHEETS	WORD PROCESSING DOCUMENTS	RFID TAGS	
GPS	WEB LOG DATA	PHOTOS	SATELLITE IMAGES	SOCIAL MEDIA DATA	80%
RESEARCH DATA	FORUMS	LAB RESULTS	VIDEOS	CLINICAL TRIALS	
MOBILE DATA	WEBSITE CONTENT	MARKETING DATA	AUDIO FILES	OPEN DATA	

HERE'S OUR 3-4-5 PLAN:



❑ 3 Definitions

❑ 4 Trends

❑ 5 Questions

3 DEFINITIONS

sas best
practices
THOUGHT PROVOKING BUSINESS

THE DEFINITIONS

BIG DATA

*“**Big Data** refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods...*

***Volume, Velocity, and Variety**—often referred to as the three V’s of Big Data—capture the true meaning of Big Data.”*

SOURCE: Frost & Sullivan: “Drowning in Big Data? Reducing Information Technology Complexities and Costs for Healthcare Organizations”



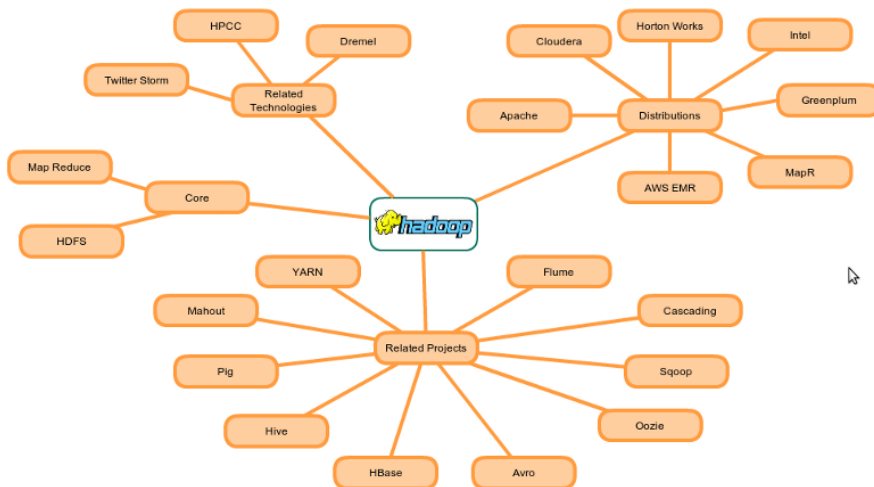
“That amount of data or complexity which puts you out of your comfort zone.”

Paul Kent
VP of Big Data
SAS Institute

Is it a project...



...or an ecosystem?



NOTE:

Hadoop is not
synonymous
with big data



*“A **data lake** is a storage repository that holds a vast amount of raw data in its native format, including structured, semi-structured, and unstructured data. The data structure and requirements are not defined until the data is needed.”*



*“If you think of a datamart as a store of bottled water – cleansed and packaged and structured for easy consumption – the **data lake** is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples.”*

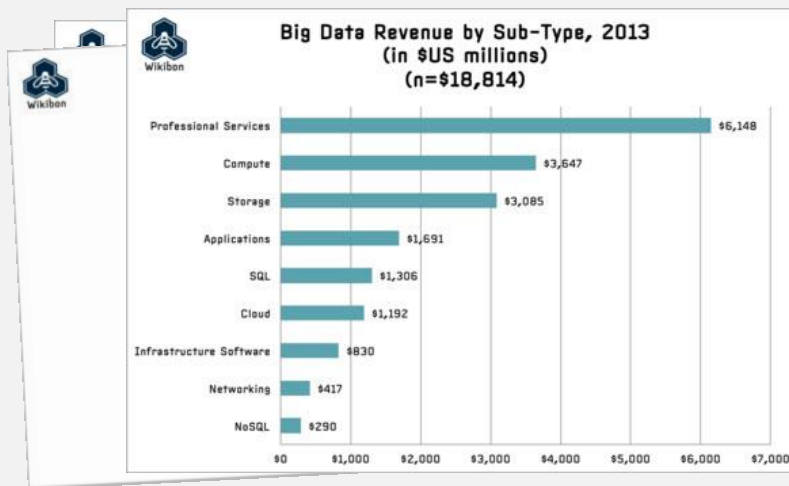
James Dixon
CTO, Founder & Chief Geek
Pentaho

4 TRENDS

sas best
practices
THOUGHT PROVOKING BUSINESS

The market is growing.

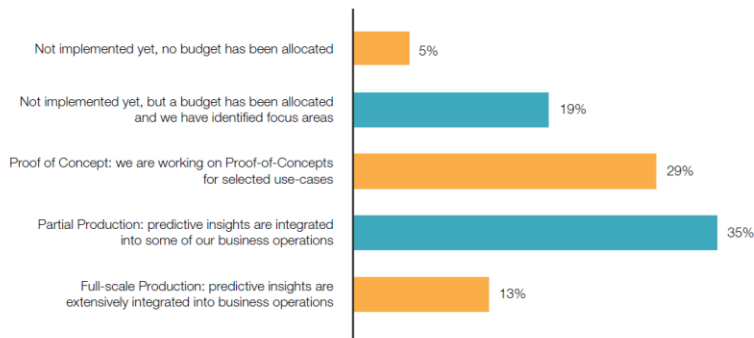
Trend #1



SOURCE: http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017

The success rate is meh.

Figure 2: Status of Big Data Implementations



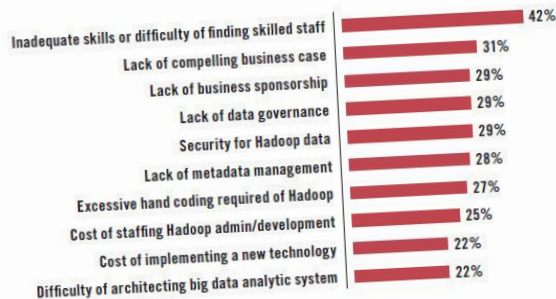
Source: Capgemini Consulting, "Big Data Survey", November 2014

People issues trump technology issues.

Trend #3

Figure 3: Key Challenges for Big Data Implementation

What are the most likely barriers to implementing Hadoop technologies in your organization? Select eight or fewer.

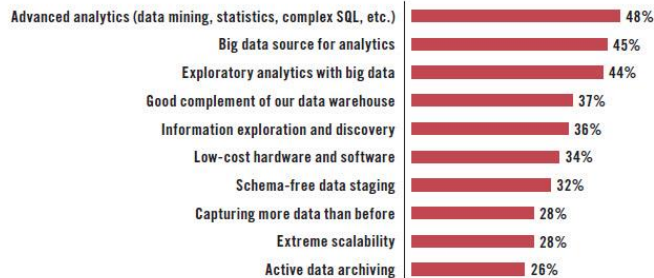


Source: TDWI Best Practices Reports - Hadoop for the Enterprise: Making Data Management Massively Scalable, Agile, Feature-Rich and Cost-Effective

Source: Capgemini Consulting, "Big Data Survey", November 2014

Analytics keeps them coming back.

If your organization were to implement Hadoop technologies, which business processes, data, and applications would most likely benefit? Select eight or fewer.



Source: TDWI Best Practices Reports - Hadoop for the Enterprise: Making Data Management Massively Scalable, Agile, Feature-Rich and Cost-Effective

5 QUESTIONS

sas best
practices
THOUGHT PROVOKING BUSINESS



HERE'S THE 5 QUESTIONS:

1. What can Hadoop do that my data warehouse can't?
2. We're not doing "big" data, so why do we need Hadoop?
3. Is Hadoop enterprise-ready?
4. How is big data impacting Big Pharma today?
5. What are the primary threats to big data adoption?

QUESTION #1

WHAT CAN HADOOP DO THAT MY DATA WAREHOUSE CAN'T?



1. Store data more cheaply.



**2. Process data more quickly
(and cheaply).**

QUESTION #2

WE'RE NOT DOING "BIG" DATA, SO WHY DO WE NEED HADOOP?



Stage structured data.



Process structured data.



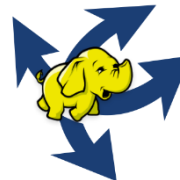
Archive any data.



Process any data.



Access any data.
(via data warehouse)



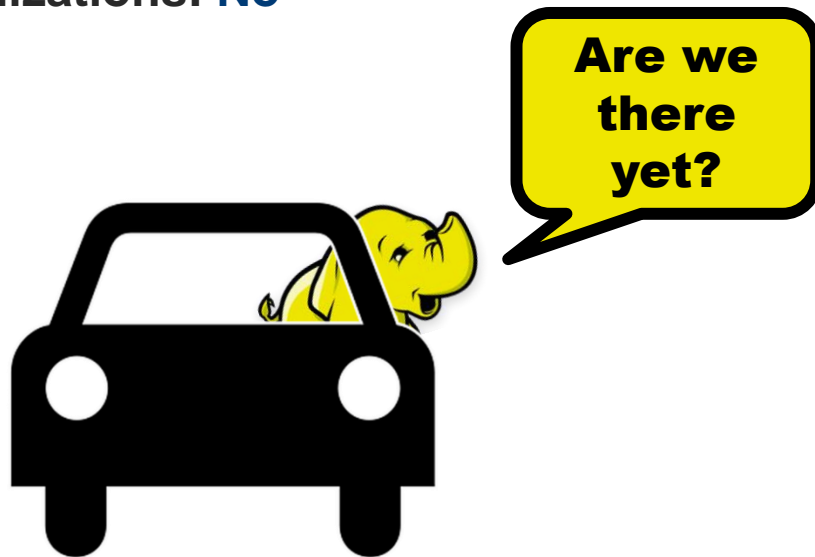
Access any data.
(via Hadoop)

QUESTION #3

IS HADOOP REALLY ENTERPRISE-READY?

For your organization: **Maybe**

For all organizations: **No**



QUESTION #4

HOW IS BIG DATA IMPACTING BIG PHARMA TODAY?



patientslikeme®



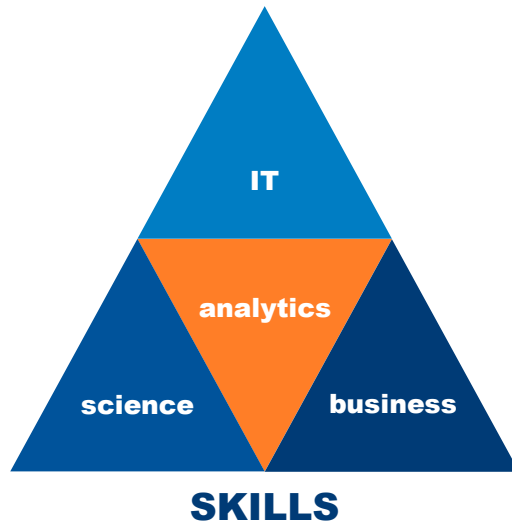
+ AllTrials

ClinicalTrials.gov



QUESTION #5

WHAT ARE THE PRIMARY THREATS TO BIG DATA ADOPTION?



WRAP-UP

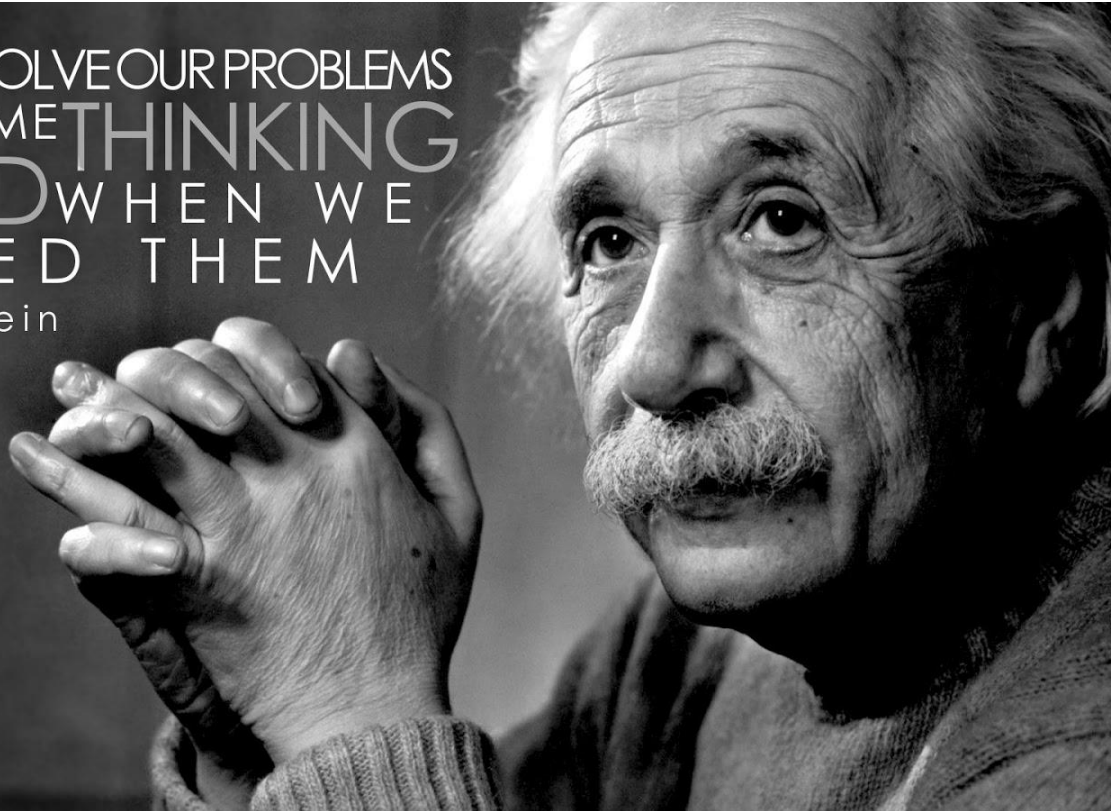
sas best
practices
THOUGHT PROVOKING BUSINESS

HERE ARE YOUR KEY TAKEAWAYS:



- ✓ It's the big data technologies – not the data itself – that's new
- ✓ Understand the context when talking about Hadoop
- ✓ If you're doing big data without analytics, you're wasting your time
- ✓ Approach big data smartly and learn from other...industries, mistakes, etc.

WE CANNOT SOLVE OUR PROBLEMS
WITH THE SAME THINKING
WE USED WHEN WE
CREATED THEM
- Albert Einstein



**IT'S A BIG DATA WORLD OUT THERE.
NOW LET'S BE SAFE.**

sas best
practices
THOUGHT PROVOKING BUSINESS



Tamara.Dull@sas.com



@tamaradull