5 Facts About Hadoop and Open Source You Need to Know
A Panel Discussion with Deloitte, Hortonworks and JPMorgan Chase

Hadoop is in the house.
Hadoop is...

- An Apache open source project
- A stuffed, yellow elephant
- A big data ecosystem of Hadoop-related projects
- Not a database
Meet the Panelists

Hello
my name is
ASHISH VERMA
Director, Deloitte Consulting
Technology Information Management

Hello
my name is
SHAUN CONNOLLY
VP Corporate Strategy

Hello
my name is
BILL WHEATLEY
VP of Digital Analytics
Here’s the Game Plan

The panelists will discuss five topics.

You get to play along.

Let’s have fun learning together.
Big Data and Hadoop
It’s Poll Time!

Text CODE to: 22333
(message and data rates may apply)

OR

Submit responses at:
PollEv.com/bestpractices
Which came first – big data or Hadoop?

Text a **KEYWORD** to **22333**

Submit responses at **PollEv.com/bestpractices**

- Big data
- Hadoop
- They came at roughly the same time

![Graph showing the comparison between big data and Hadoop.](Image)
I dream of a world where chickens can cross the road without having their motives questioned.
Let’s Discuss: Big Data and Hadoop

How do you define big data?

Will the term “big data” go away?

Are big data and Hadoop synonymous?

Hi, my name is Hadoop.
Topic 2 of 5

Hadoop and the Enterprise Data Warehouse
Which statement best describes how your organization is using Hadoop in your enterprise data ecosystem?

Text a **KEYWORD** to 22333

Submit responses at **PollEv.com/bestpractices**

- Hadoop is fully integrated
- We’re in the process of getting Hadoop integrated
- We’re using Hadoop in standalone mode, but we plan to integrate
- We’re using Hadoop in standalone mode, and there’s no current plans to integrate
- We don’t have Hadoop installed
- I’m not sure
In your perception, what would be useful applications of HDFS in your organization? Select all that apply.

<table>
<thead>
<tr>
<th>Application</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complement to a data warehouse, specifically for advanced analytics</td>
<td>78%</td>
</tr>
<tr>
<td>Live archive for Web, machine, sensor, social, and other non-traditional data</td>
<td>52%</td>
</tr>
<tr>
<td>Data staging area for data warehousing</td>
<td>41%</td>
</tr>
<tr>
<td>Sandbox for ad hoc analytics and other data-related development</td>
<td>41%</td>
</tr>
<tr>
<td>Repository for content, document, and records management</td>
<td>35%</td>
</tr>
<tr>
<td>Live archive for traditional enterprise data</td>
<td>23%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
</tr>
</tbody>
</table>

*Figure 7. Based on 712 responses from 263 respondents; 2.7 responses per respondent, on average.*
Let’s Discuss: Hadoop and the EDW

Why do I need Hadoop if I have a data warehouse?

Can Hadoop provide me with greater insight into my data than my data warehouse?

Are companies successfully integrating Hadoop into their data ecosystem?

I play well with others.
Open Source Software
Does open source software deserve a seat at the enterprise table?

- Text a KEYWORD to 22333
- Submit responses at PollEv.com/bestpractices

1. Yes, absolutely
2. Yes, but only if it’s properly governed and managed
3. I’m not sure. It seems less stable than proprietary software
4. Are you kidding me? Absolutely not
5. What is open source software?
“Nucleus…found that there is a place for both open source and SAS in many enterprise environments where SAS has been successfully used, sometimes for decades, to analyze data.”
Let’s Discuss: Open Source Software

Open source software is free, but what does it really cost?

How secure is open source?

Is open source ready for the enterprise?

I look like an elephant on the outside, but I’m open source software through and through.
Topic 4 of 5

Successes and Failures
Which statement best describes your organization's big data project(s) track record?

We've had several successful projects so far.

We're making good progress, but it's been a slow, bumpy ride.

We've had a mixed bag of successes and failures.

We ask ourselves regularly, “Why are we doing this again?”

We haven't done a big data project yet.

Text a **KEYWORD** to 22333

Submit responses at PollEv.com/bestpractices

**SUCCESS1**

**SUCCESS2**

**SUCCESS3**

**SUCCESS4**

**SUCCESS5**

0% 20% 40% 60% 80% 100%

#pbls14

Presented by SAS Institute Inc.
Management resistance
Selecting the wrong uses
Asking wrong questions
Lacking the right skills
Unanticipated problems beyond big data technologies
Disagreement on enterprise strategy
Big data silos
Problem avoidance
Let’s Discuss: Successes and Failures

Why do some projects succeed and others fail?

What business problems does Hadoop solve well?

What are the barriers to adoption?

Will you adopt me? I’m housebroken.
Topic 5 of 5

Tips for the Journey
How long has your organization been on its big data journey?

- Text a **KEYWORD** to 22333
- Submit responses at PollEv.com/bestpractices

<table>
<thead>
<tr>
<th>JOURNEY1</th>
<th>JOURNEY2</th>
<th>JOURNEY3</th>
<th>JOURNEY4</th>
<th>JOURNEY5</th>
</tr>
</thead>
<tbody>
<tr>
<td>4+ years</td>
<td>2-3 years</td>
<td>We started this past year</td>
<td>We’re planning to start in the next year or two</td>
<td>We have no big data plans</td>
</tr>
</tbody>
</table>
“One does not discover new lands without consenting to lose sight of the shore for a very long time.”

–André Gide
Let’s Discuss: Tips for the Journey

What’s a common mistake companies make when starting out?

What are some lessons learned along the way?

“If I only knew then what I know now…”

Are we there yet?
Wrapping It Up
5 Facts About Hadoop and Open Source You Need to Know

#1 Big Data and Hadoop

#2 Hadoop and the Enterprise Data Warehouse

#3 Open Source Software

#4 Successes and Failures

#5 Tips for the Journey
The Panelist Sound Off
It’s a big data world out there. Now let’s be safe.

Hello
my name is
ASHISH VERMA
asverma@deloitte.com

Hello
my name is
SHAUN CONNOLLY
shaun@hortonworks.com

Hello
my name is
BILL WHEATLEY
william.wheatley@chase.com

Hello
my name is
TAMARA DULL
tamara.dull@sas.com

#pbls14