Big Data Cloud Technology Report

How Organizations Run Big Data Applications and Workloads in the Cloud

pepperdata®

Executive Summary

Key Findings:

This survey was conducted to better understand how organizations run their big data applications and workloads in the cloud. When IT doesn't possess visibility into their big data system performance, overspending is almost guaranteed.

We wanted to learn how enterprises are thinking about their big data applications in the cloud, whether they are staying within budget, and what their strategies are for optimizing operations.

This survey was conducted online, in November 2020, among 750 participants from a range of industries.

For 64% of respondents, **"cost management and containment"** is their biggest concern with running cloud big data technologies and applications.

For a majority of respondents, a desire to "better optimize current cloud resources" was their **highest priority big data cloud initiative.**

20-40%

In 2020, for 1 in 3 respondents, cloud spend was projected to be over budget by **between 20% and 40%.**

40%+

For 1 in 12 respondents, cloud spend was expected to be over budget by **40% or more.**

On Big Data in the Cloud Strategy

The majority of respondents prefered the **private cloud** over any other type of cloud. The minority of respondents had no plans to move to the cloud.

What type of cloud strategies are enterprises pursuing?

A. Public

B. Private

C. Hybrid

D. No plans to move to the cloud

E. Don't know
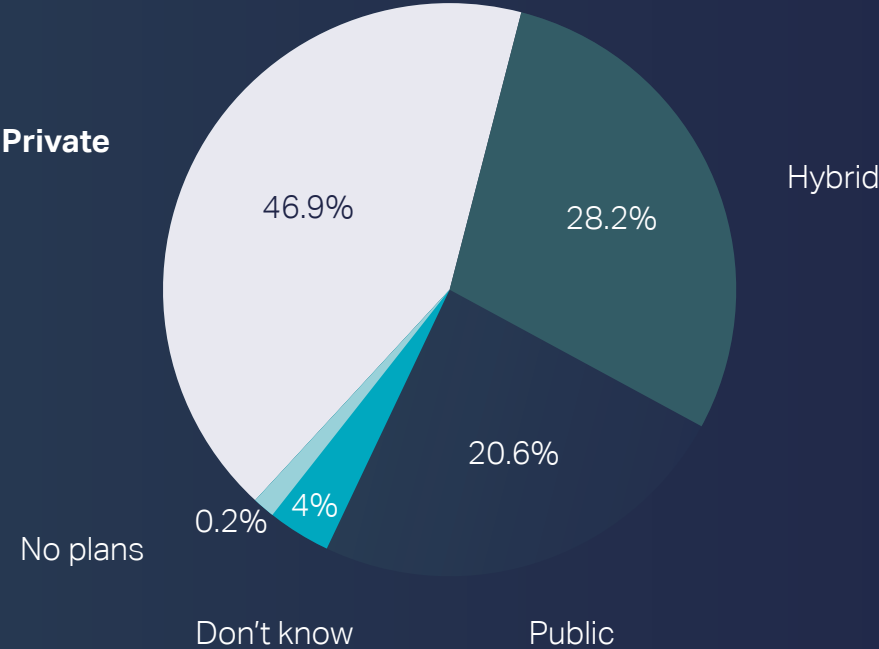
On Big Data in the Cloud Strategy

The majority of respondents cite **cost management and containment** and increased complexity as their biggest concern about running big data applications in the cloud.

What do enterprises cite as their biggest concern about running applications in the cloud?

A. Shifting from CapEx to OpEx for cloud expenditures

B. Cost management and containment

C. Increased complexity and lack of expertise

D. Lack of control as the business units are spinning up their own instances

CapEX to OpEx

Lack of Control

14.3%

13.4%

39.3%

33%

Cost Management & Containment

Increased Complexity
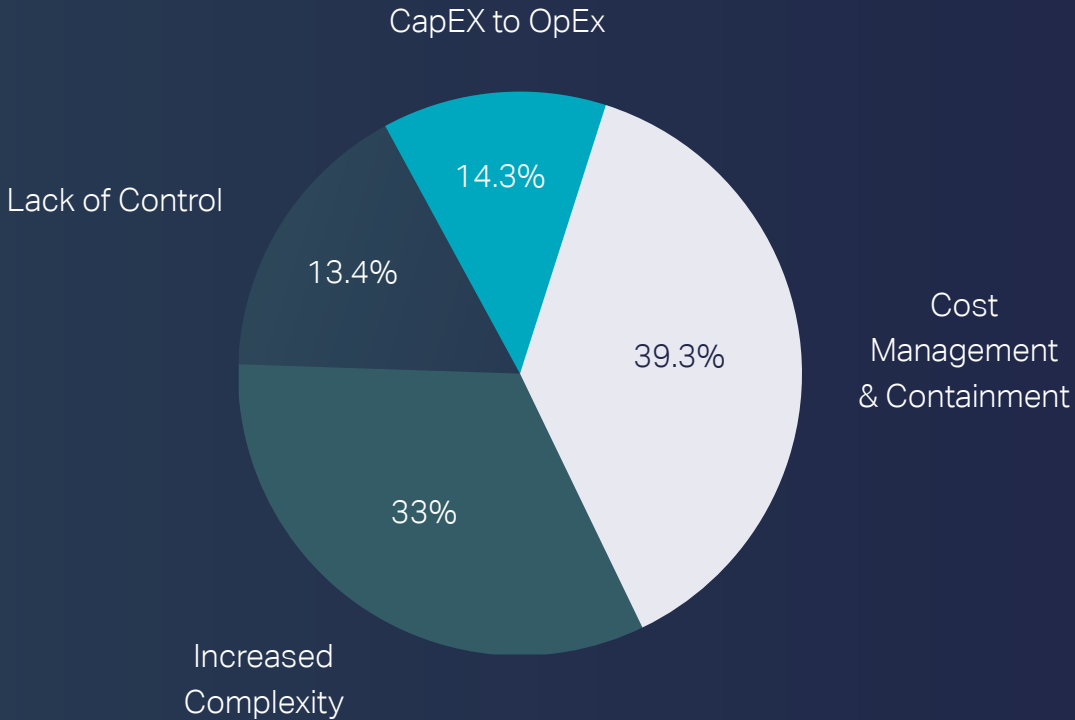
On Big Data in the Cloud Strategy

The majority of respondents measure application/workload performance using an **application performance monitoring solution.** Observability and recommendations provide the visibility needed to get the most from your big data solution.

**How are enterprises measuring application/workload performance based on their cloud instance?**

A. Manual monitoring tools or a homegrown solution

B. **An application performance monitoring solution**

C. An observability solution: insights, alerts and automation

D. Cloud provider tools

E. We are not monitoring workload performance at this time

B. 29.8%
D. 28.1%
A. 19.3%
C. 17.4%
E. 5.5%

0% 10% 20% 30%
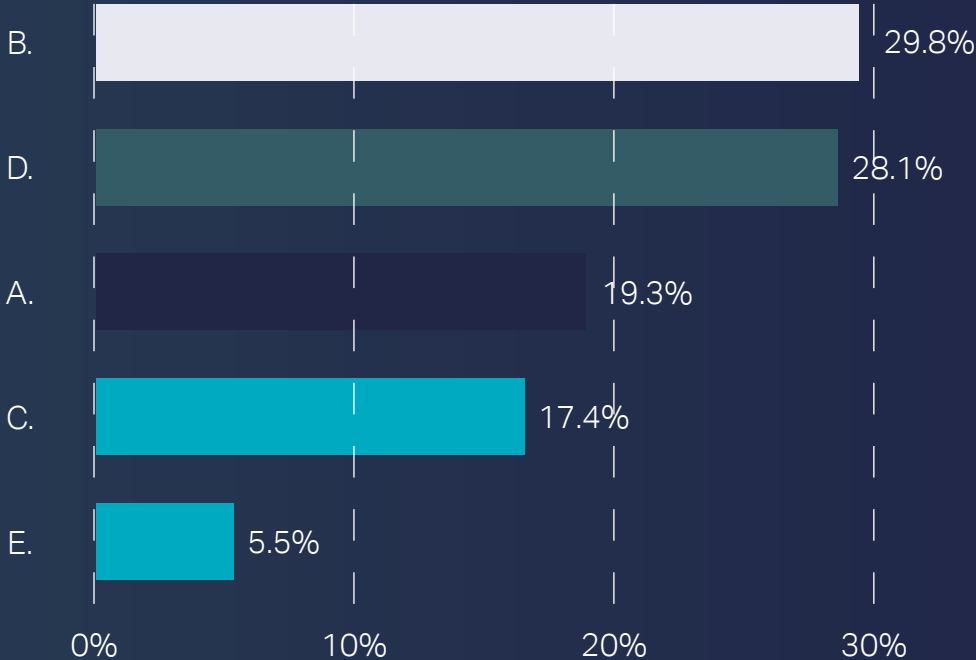
On Big Data in the Cloud Strategy

The majority of respondents manage the support and troubleshooting through **shared support model between ITOps and business units/line-of-business developers.** The ability to monitor and chargeback for the use of big data resources helps ensure proper budget allocation.

With the move to cloud, how are organizations managing the support and troubleshooting for big data applications?

A. Support stays with ITOps

B. Shared support model between ITOps and business units/LOB developers

C. Support is managed by the dev organization within the business units
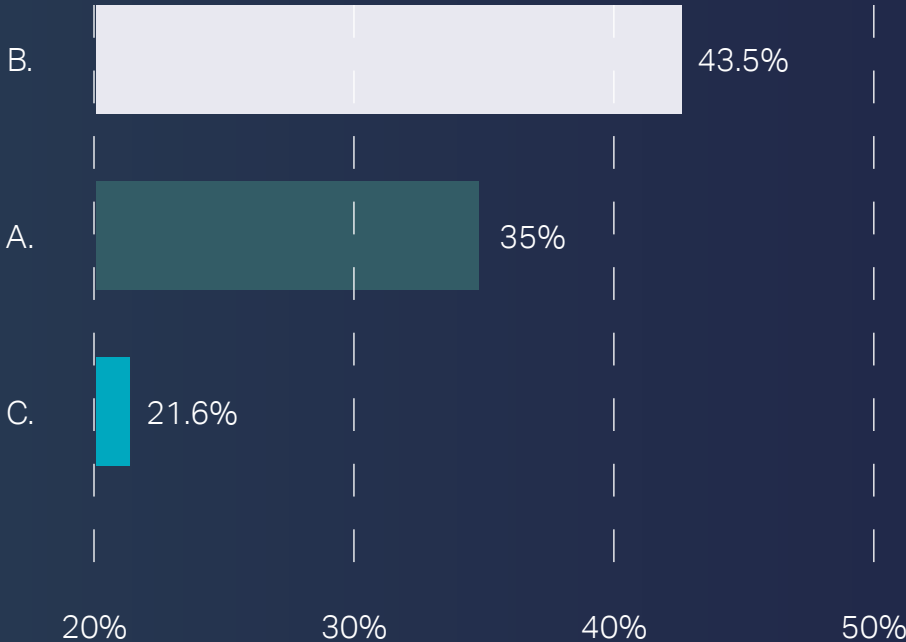
On Cloud Budgets

The majority of respondents say the companies' budgets **stayed the same and reside with IT.** The minority moved into one or multiple business units.

On moving to the cloud, how have companies' budgets been impacted?

A. Stayed the same and reside with IT

B. Moved into one or multiple business units

C. Become shared with IT (responsible for administering)

Become shared with IT

30.5%

Stayed the same and reside with IT

39.6%

29.8%

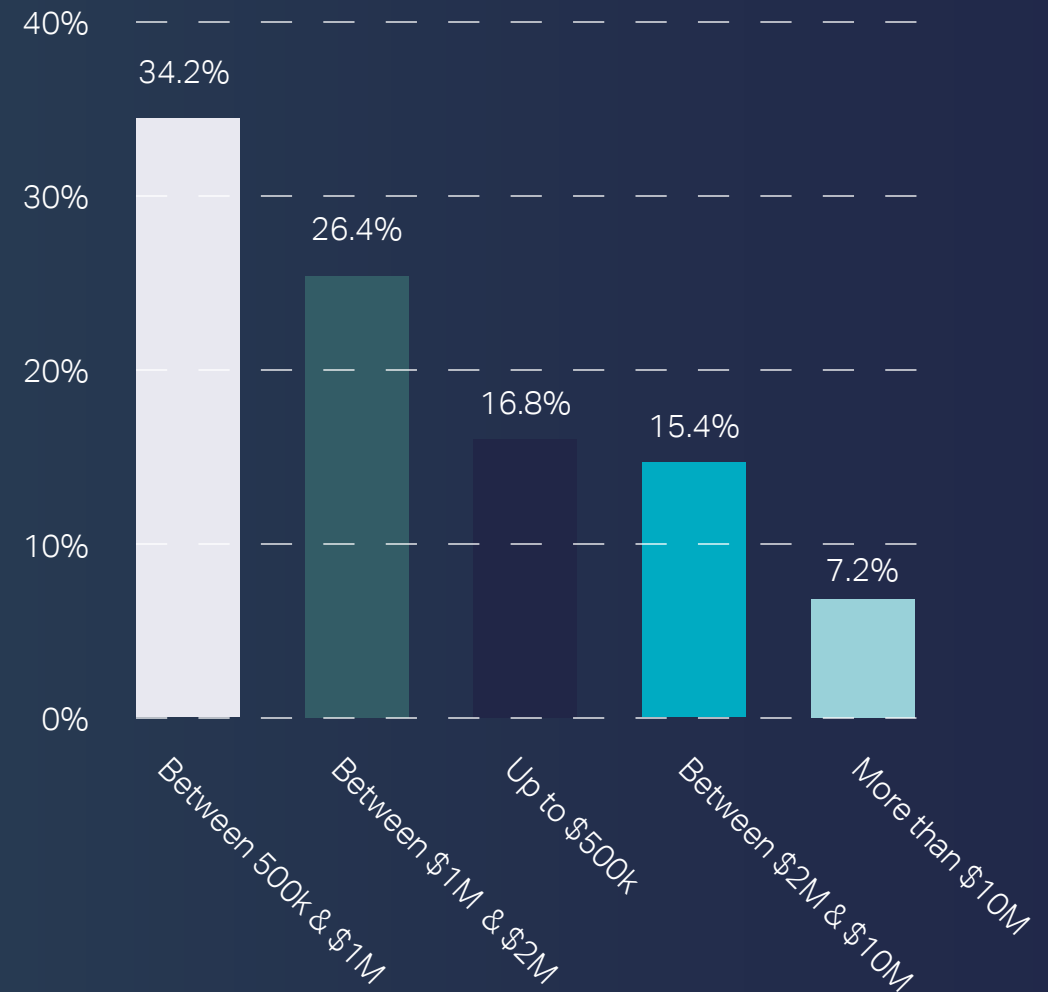Moved into one or multiple business units
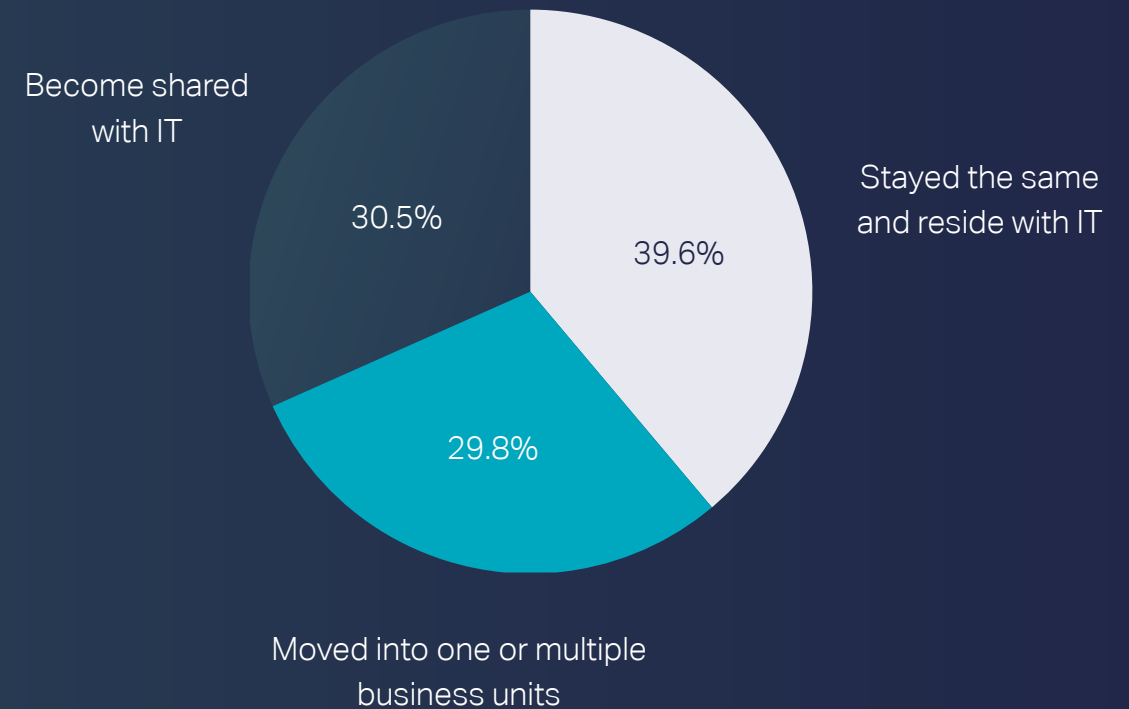
On Cloud Budgets

The majority of respondents estimated they will spend **between $500k and $1M** this year on big data analytics in the cloud.

This is how much respondents estimated they will spend this year on big data analytics in the cloud:

A. Up to $500k

B. Between $500k and $1M

C. Between $1M and $2M

D. Between $2M and $10M

E. More than $10M

On Cloud Budgets

Almost 35% of respondents expected to **exceed budgets between 20% and 40%** and over 8% expected to exceed budgets by 40% or more.
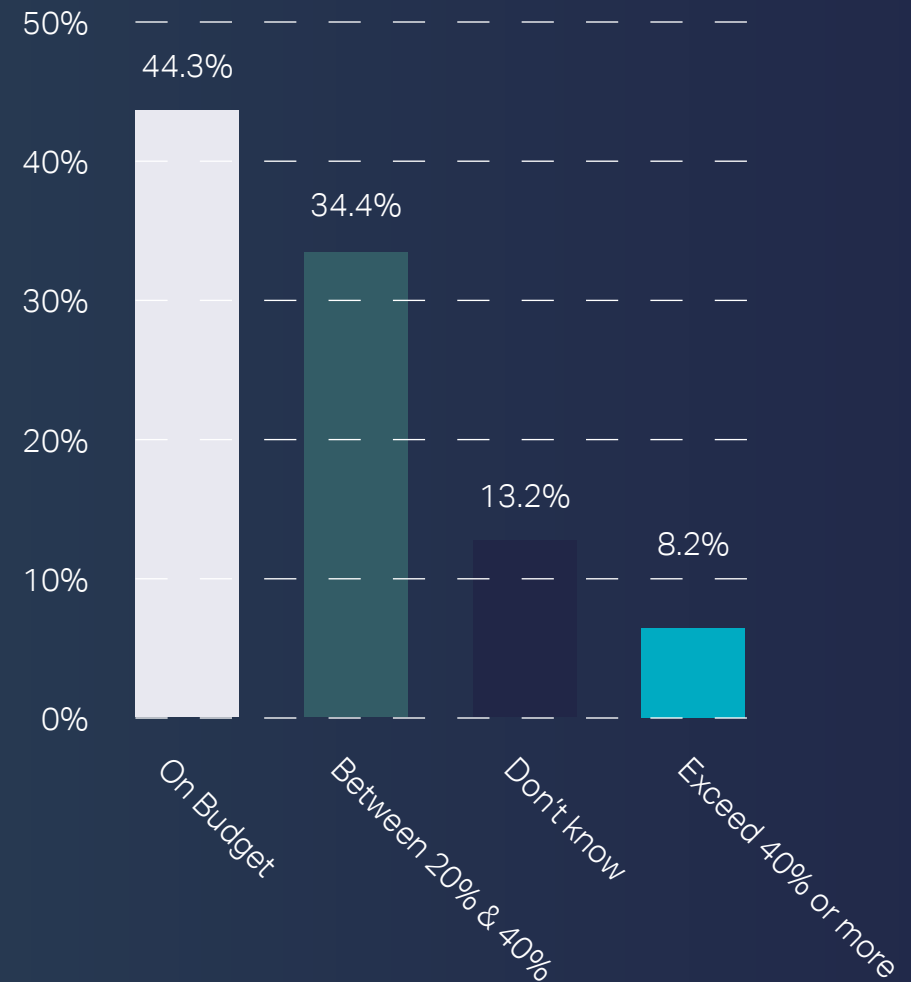
For the year 2020, how within-budget were enterprises?

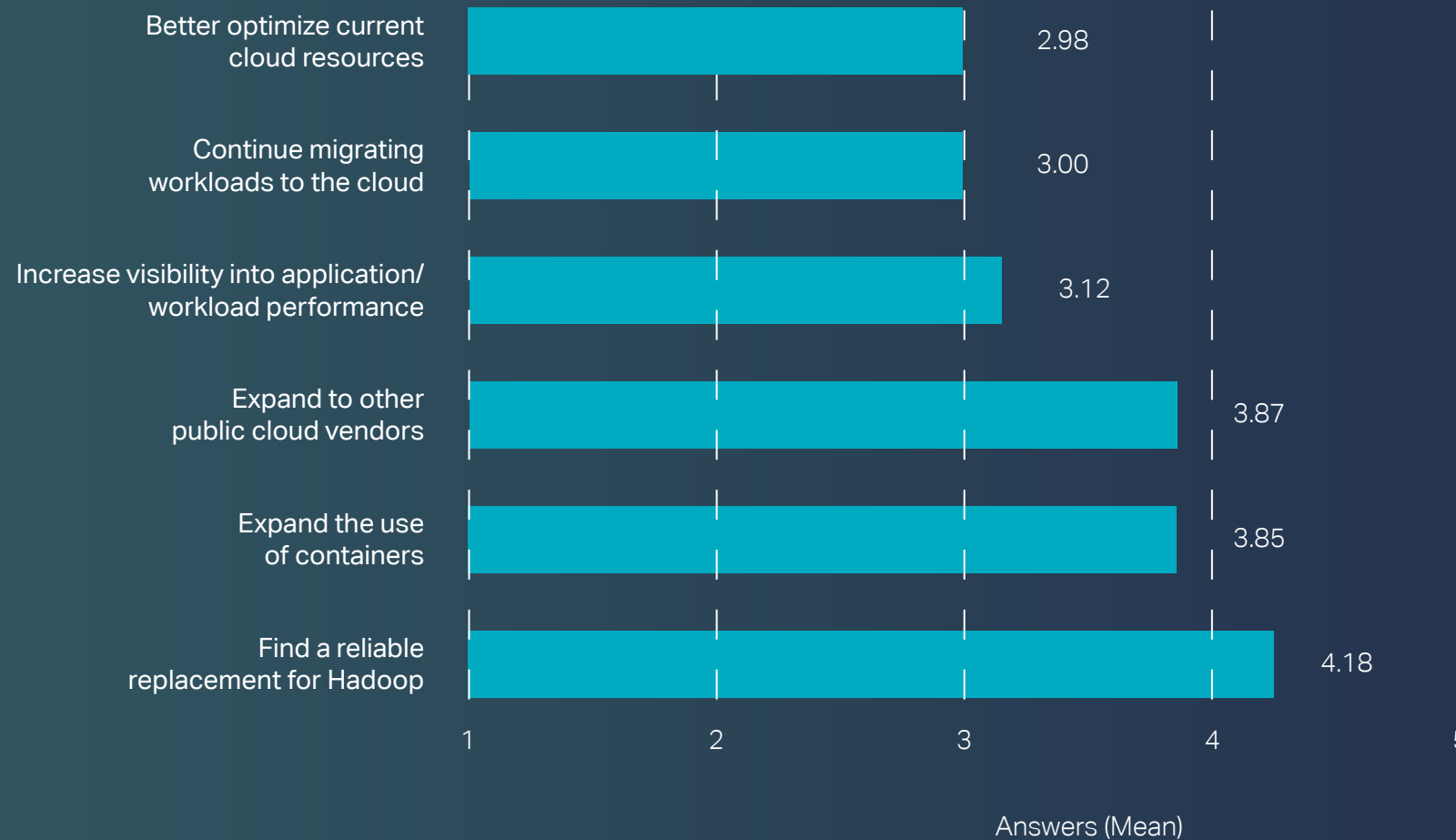**A. On budget**

B. Expected to go over between 20% and 40%

C. Expected to exceed 40% or more

D. Don't know

On Big Data in the Cloud Initiatives

What Are Enterprises' Top Big Data Cloud Initiatives?

| Initiative	Mean
Better optimize current cloud resources	2.98
Continue migrating workloads to the cloud	3.00
Increase visibility into application/ workload performance	3.12
Expand to other public cloud vendors	3.87
Expand the use of containers	3.85
Find a reliable replacement for Hadoop	4.18

Answers (Mean)

Using the mean score, the largest number of respondents identified **optimizing current cloud resources** as their top initiative. Their second priority was migrating workloads to the cloud.

On Big Data in the Cloud Workloads

The majority of respondents say **Hive workloads** consumed the most resources in their big data cloud instances, followed by Spark workloads.

These are the types of applications or workloads that consumed the most resources in respondents' big data cloud instances:

    A. Spark

    B. MapReduce

    C. Tez

    D. Hive

    E. Others
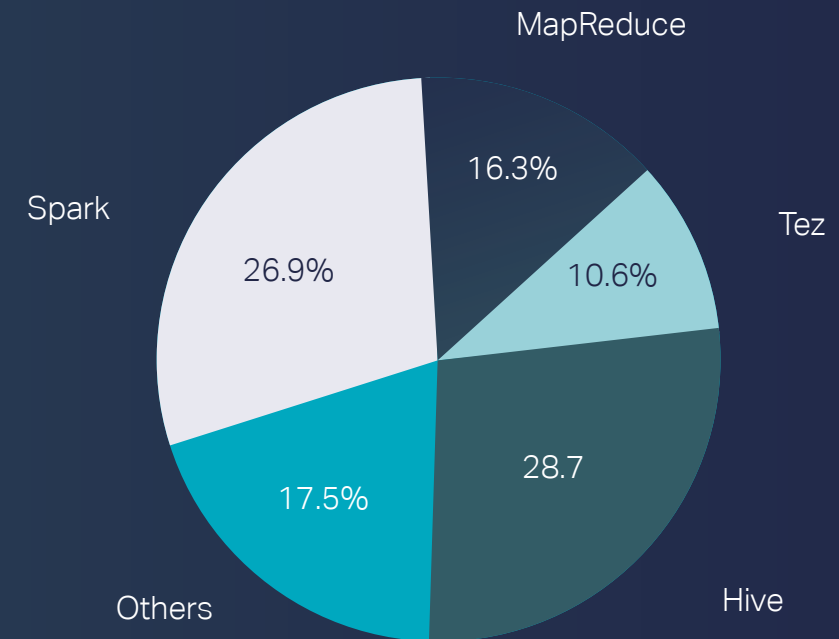
The Pepperdata Performance Suite provides a holistic view that reveals usage patterns, peak computing demand, storage utilization, and processor requirements. The ability to access this information, understand it, and take action on it is key to optimization and ROI.

Learn how you can start optimizing today. Cut waste and win back task hours, resulting in true cloud optimization and savings.

GET THE FACTS ON PEPPERDATA NOW

pepperdata.

www.pepperdata.com