

APRIL 2016

# Copy Data Management



COPYRIGHT © 2016

Source: IDC 2016

Layout & Design by:  
Cyclone Interactive

SPONSORED BY

**EMC<sup>2</sup>**

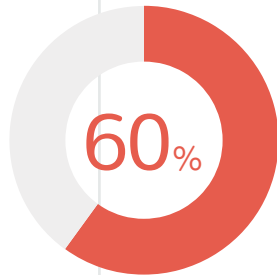
# Introduction



## What is the Copy Data Problem?



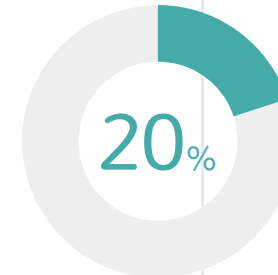
**Copy data is created whenever primary data is copied** for secondary purposes, including protection, operations, test/dev and analytics



According to IDC research, between 45% and 60% of total storage capacity is dedicated to accommodating copy data.



The total cost to IT for copy data storage is estimated to be **\$50.63 billion by 2018**



Fewer than 20% of organizations formally address copy data, but those who do have seen significant decreases in storage capacity growth costs.

# Introduction



## Study Methodology



The goal: to quantify the copy data problem  
and the benefits of addressing it



IDC Survey of more than 500 North American  
organizations with more the 1000 employees across  
all industries



10 in-depth interviews with senior IT leaders regarding  
specify copy data management practices



# Key Research Findings

# 77% of organizations have >200 DB instances



How many Oracle Database or SQL Server instances do you currently have in production in your data center environment?

**77%**  
have more than  
200 instances

Less than 100 instances

100 to 199 instances

200 to 499 instances

500 or more instances

0% 10% 20% 30% 40% 50% 60%

N=513 Base=All Respondents Note: Managed by IDC's Quantitative Research Group.; Weighted by Employee Size; Use caution when interpreting small sample sizes. Source: Data Management Survey for EMC, IDC, November, 2015

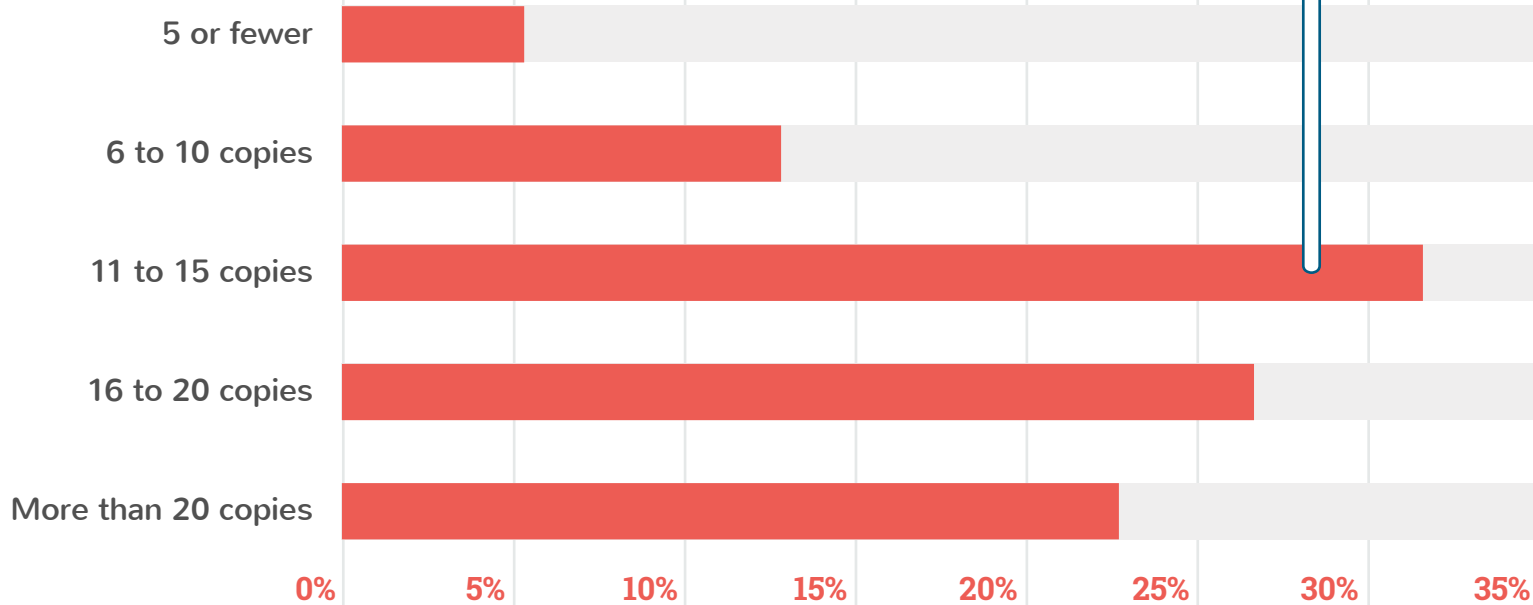
# 82% have more than **10 copies** of each DB

This amounts to **more than 2000 database copies** to manage



What are the typical number of physical copies of a given production instance that are made for the purpose of test/development, QA, DR testing, data warehouse loading, reporting, etc?

**82%**  
82% have more than 10 copies  
Consistent with other IDC research

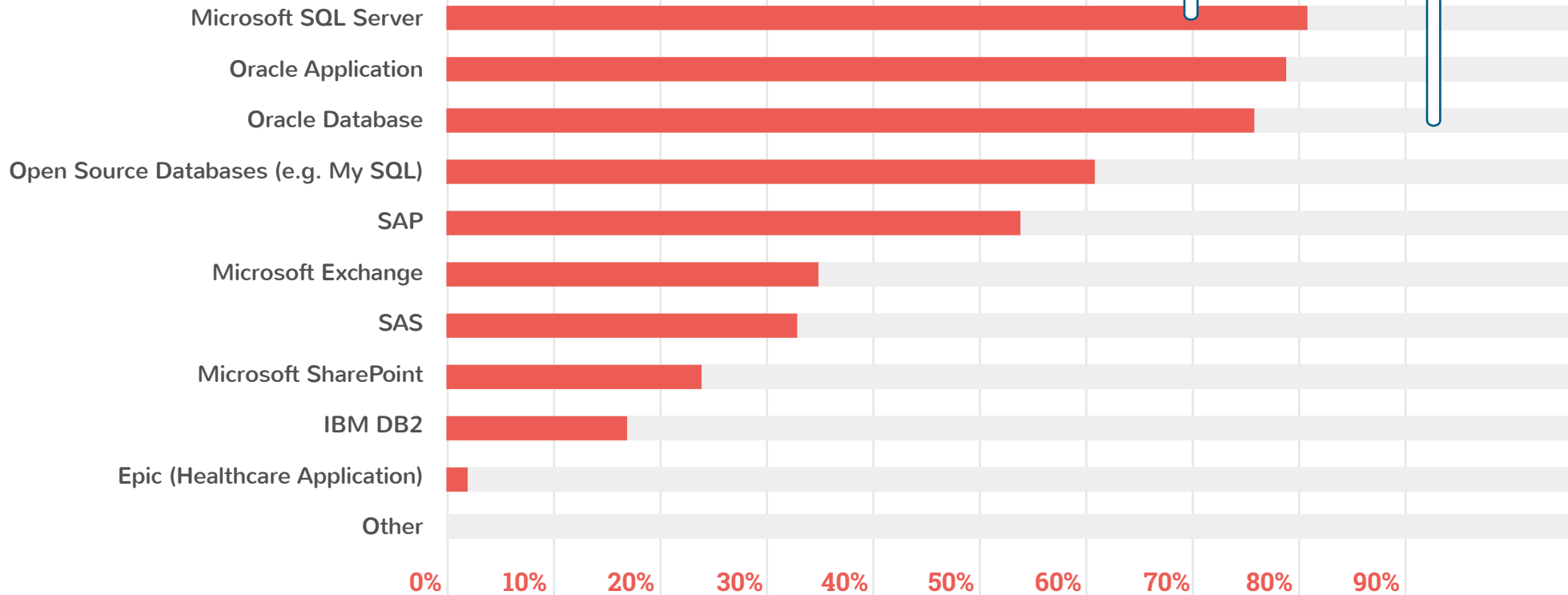


# More than 3/4 of respondents have SQL Server, Oracle apps and Oracle databases and 54% have SAP



Which of the following enterprise applications does your organization use in production?

3/4  
have these three



# More than 80% of organizations use **home-grown masking** software for sensitive data

1 This is **labor-intensive**, may require frequent maintenance and is prone to human error

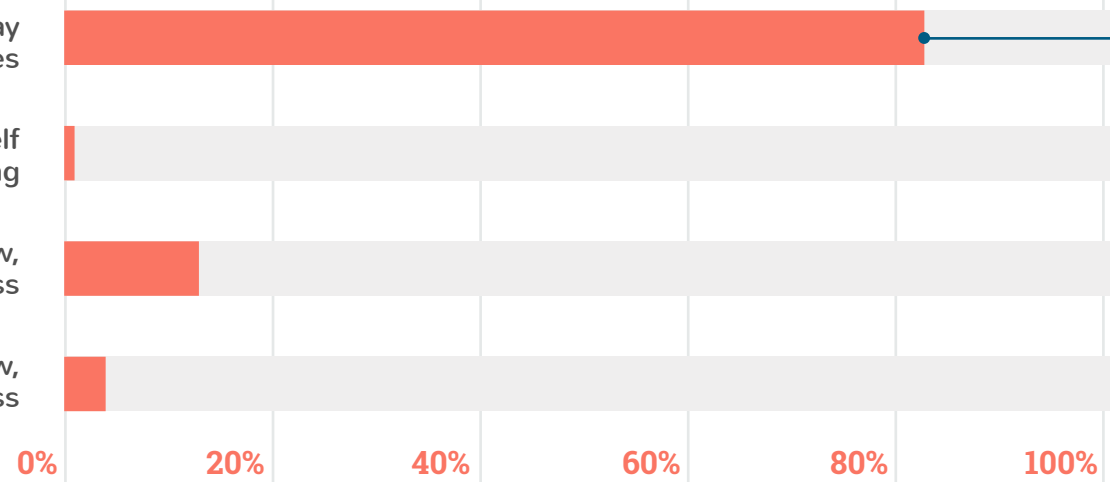
2 Opens a significant risk of sensitive data exposure



What means are you using to prepare your golden image/template for making copies?

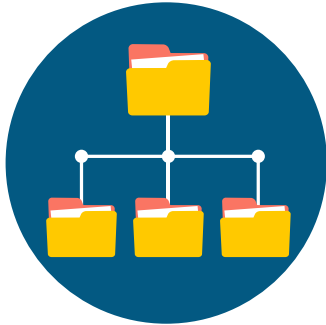
More than **80%** are home-grown

- I wrote my own scripts to mask away sensitive data and prepare copies
- I am using off-the-shelf product for masking
- I do not know, my DBA runs the process
- I do not know, my storage admin runs the process





# For 62% of organizations, copies take 1/2 day or more to create (up to days)



How much time does it take you to refresh copies in an environment (i.e., update based the current system state)?

Up to an hour

A half a day

A day

Days

0%

10%

20%

30%

40%

50%

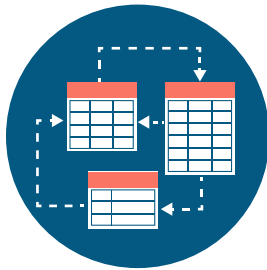
# Frequency of copy refresh



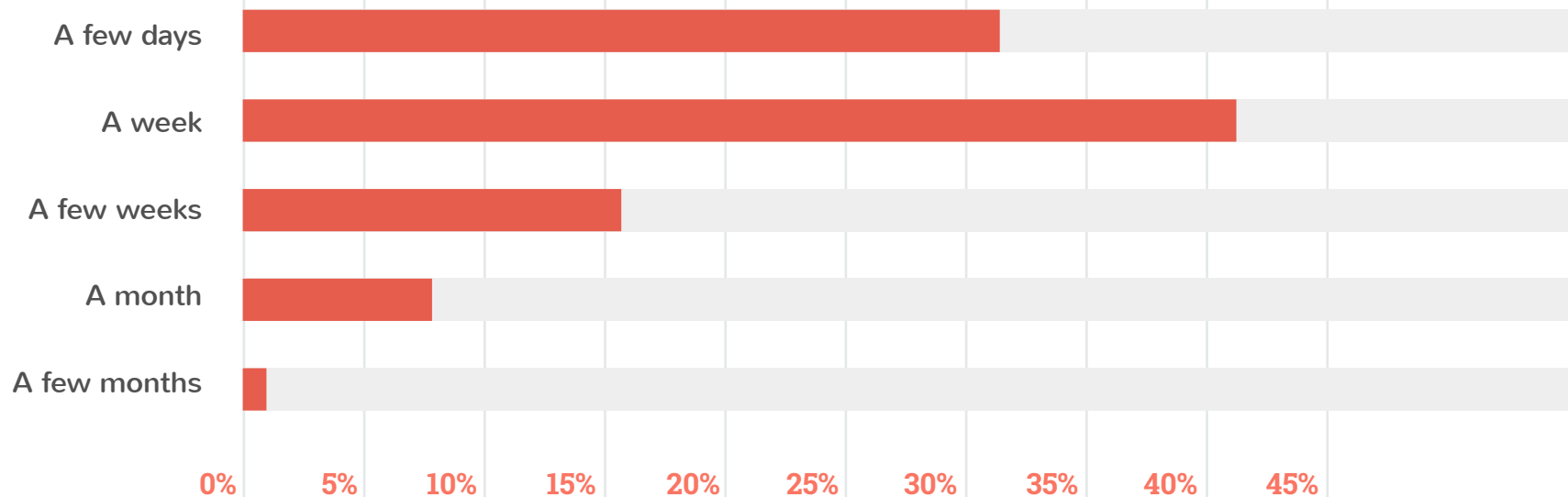
Copies are refreshed every few days by **32%** of organizations, and weekly by **42%**



Time delays limit agility and impact test/dev cycles, slow analytics and may hurt data protection



## What is the frequency to refresh the copies used specifically for dev/test or analytics?





# Copy Data Impact

# Major CDM pain points



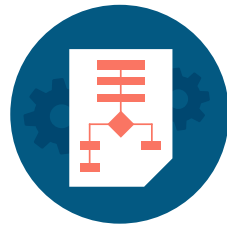
Cost of separate infrastructure for copy data (56%)



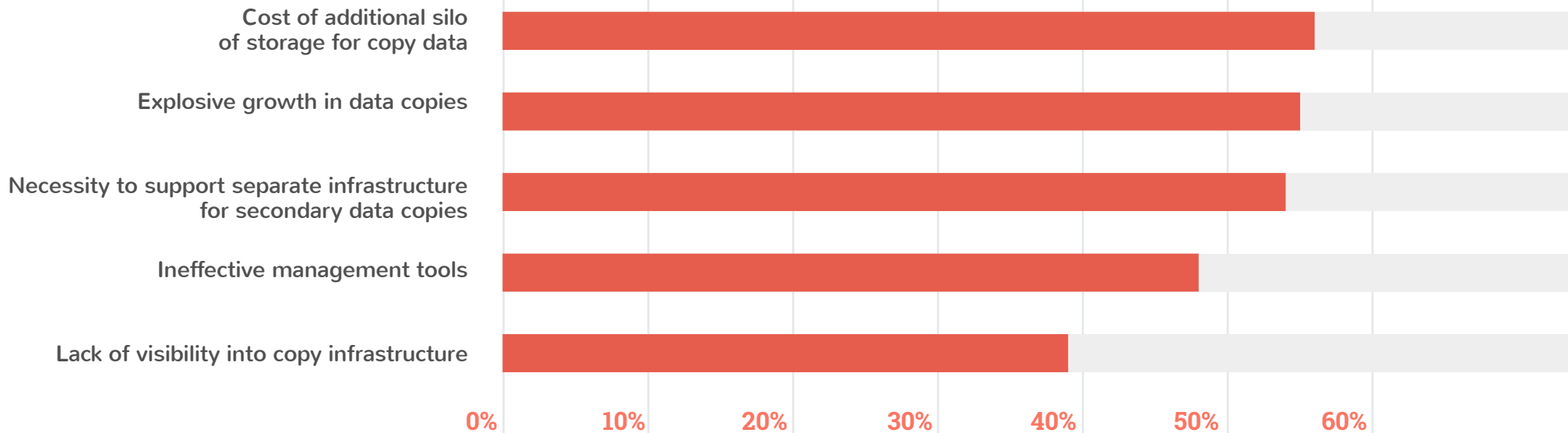
Explosive growth of copy data (55%)



Need to support separate infrastructure for copy data (53%)



What are the major pain points associated with managing secondary data copies?



# Major concerns about managing copy data



**Performance**  
(QoS)



**Security**



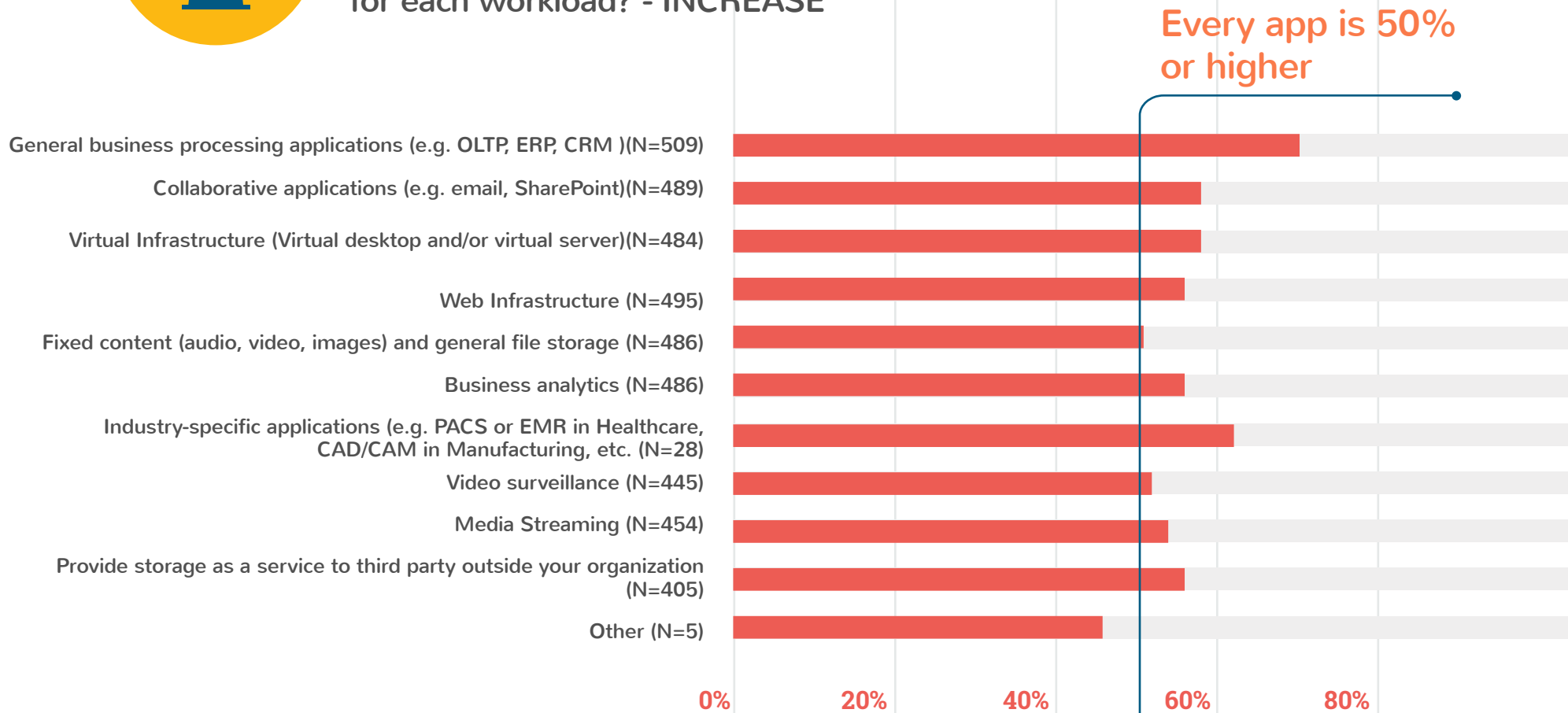
**Cost**

# Major workloads driving copy data

Uniformly distributed across data protection, test/dev, analytics, operations/staging



Do you expect the average number of secondary data copies will increase, decrease or remain about the same for each workload? - INCREASE





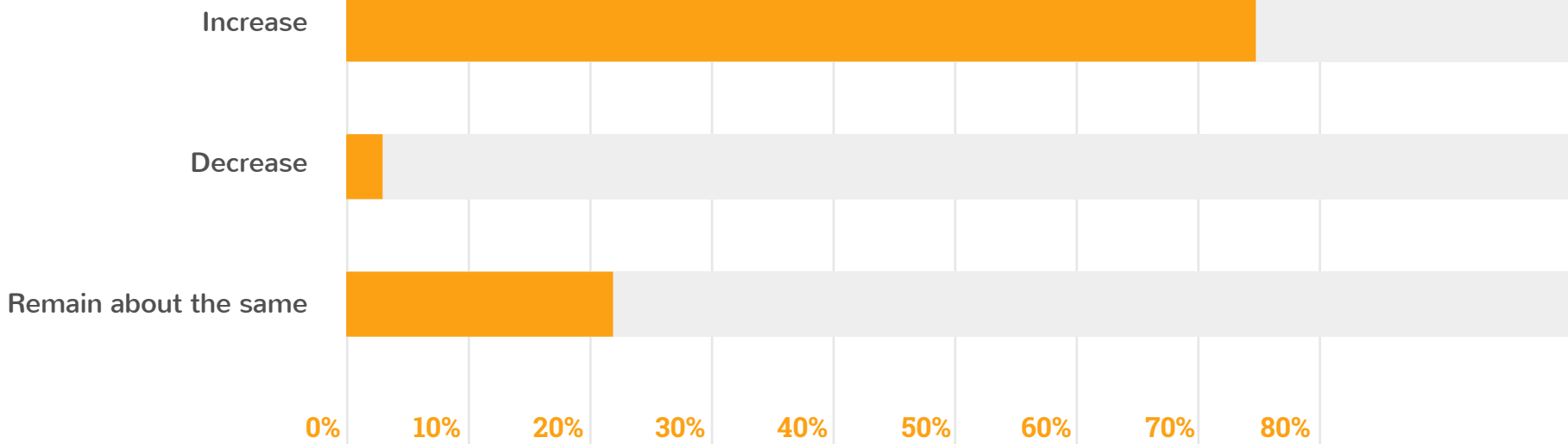
# Copy Data Management **Benefits**

# 74% of organizations expect to spend more on storage in 2016 - Only 4% expect to spend less



Will the budget or planned spending increase, decrease or remain about the same in 2016?

Correlates to copy data growth expectations





# The **#1 Reason** Organizations Expect to Spend Less on Copy Data




**Effective** copy data management


# Top 5 benefits realized from integrated copy data management

 **Faster application recovery**

 **High availability /no downtime**

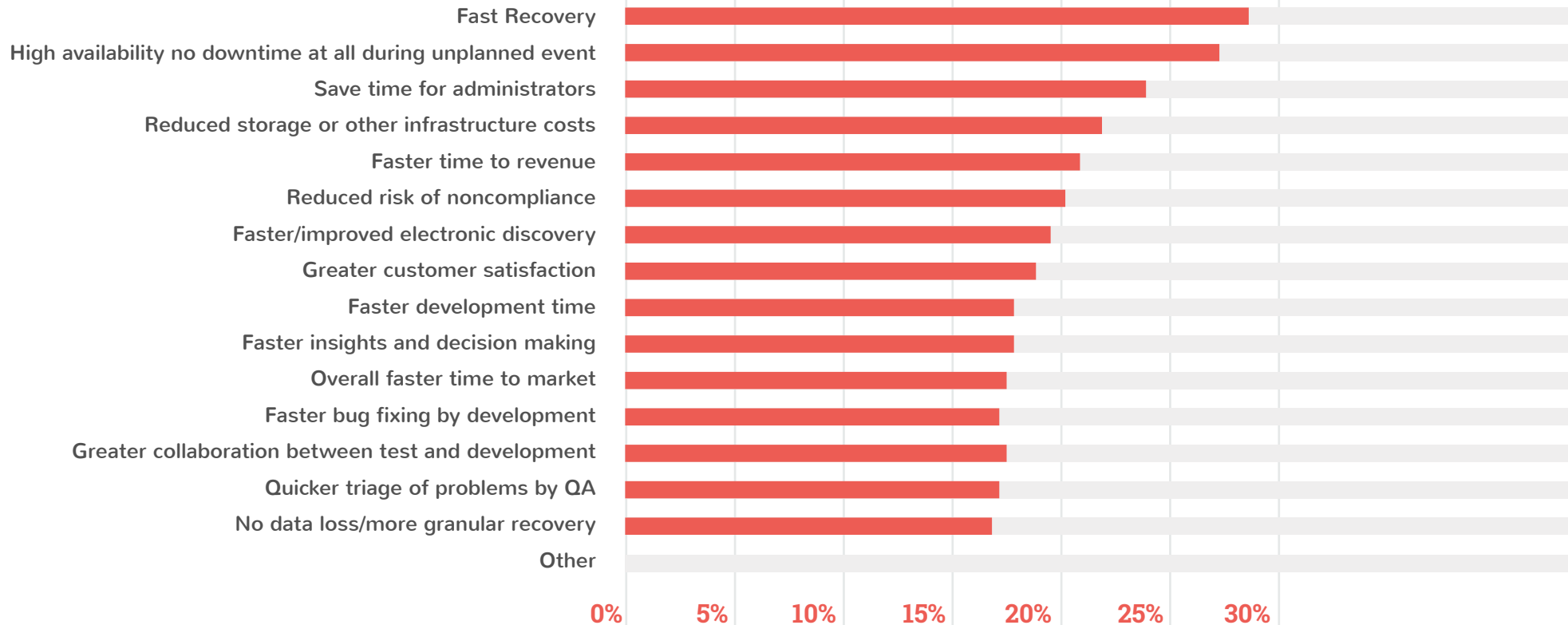
 **Save administrative time**

 **Reduce costs**

 **Faster time to revenue**



What would be the top three benefits to your organization if you could create limitless virtual copies instantly, assuming the overhead cost was negligible?





# Conclusions

# Conclusions



- \* A very **small percentage** of organizations practice copy data management
- \* Those that do:  
Realize both **better operational results** AND **lower costs**
- \* Those that don't:  
Acknowledge the problem/symptoms of **Inefficient** Copy Data problem

- \* Organizations that adopt copy data management can create a relative **competitive advantage** compared to industry peers
- \* An effective copy data management solution will enable **better overall efficiency** of work flows and resources, and therefore will **lower cost** and **streamline operations** – saving time and money.