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.

Source: IDC 2016
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 than 1000 employees across all industries

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

Source: IDC 2016
Key Research Findings

Source: IDC 2016
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?

- Less than 100 instances
- 100 to 199 instances
- 200 to 499 instances (77%)
- 500 or more instances

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

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

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?

- 5 or fewer
- 6 to 10 copies
- 11 to 15 copies
- 16 to 20 copies
- More than 20 copies

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
More that \(\frac{3}{4}\) 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?

- Microsoft SQL Server
- Oracle Application
- Oracle Database
- Open Source Databases (e.g. MySQL)
- SAP
- Microsoft Exchange
- SAS
- Microsoft SharePoint
- IBM DB2
- Epic (Healthcare Application)
- Other

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?

- I wrote my own scripts to mask away sensitive data and prepare copies: 30%
- I am using off-the-shelf product for masking: 20%
- I do not know, my DBA runs the process: 10%
- I do not know, my storage admin runs the process: 20%

More than 80% are home-grown.
For 62% of organizations, copies take \( \frac{1}{2} \) 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

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

- A few days: 32%
- A week: 42%
- A few weeks: 23%
- A month: 15%
- A few months: 6%

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.

Source: IDC 2016

Copy Data Impact

101011
01101
10101
01001
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?

- Cost of additional silo of storage for copy data
- Explosive growth in data copies
- Necessity to support separate infrastructure for secondary data copies
- Ineffective management tools
- Lack of visibility into copy infrastructure

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
Major concerns about managing copy data

Performance (QoS)  
Security  
Cost

Source: IDC 2016
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**

- General business processing applications (e.g. OLTP, ERP, CRM) (N=509)
- Collaborative applications (e.g. email, SharePoint) (N=489)
- Virtual Infrastructure (Virtual desktop and/or virtual server) (N=484)
- Web Infrastructure (N=495)
- Fixed content (audio, video, images) and general file storage (N=486)
- Business analytics (N=486)
- Industry-specific applications (e.g. PACS or EMR in Healthcare, CAD/CAM in Manufacturing, etc.) (N=28)
- Video surveillance (N=445)
- Media Streaming (N=454)
- Provide storage as a service to third party outside your organization (N=405)
- Other (N=5)

Every app is 50% or higher

**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
Copy Data Management Benefits

Source: IDC 2016
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?

Increase

Decrease

Remain about the same

Correlates to copy data growth expectations

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
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?

Fast Recovery
High availability no downtime at all during unplanned event
Save time for administrators
Reduced storage or other infrastructure costs
Faster time to revenue
Reduced risk of noncompliance
Faster/improved electronic discovery
Greater customer satisfaction
Faster development time
Faster insights and decision making
Overall faster time to market
Faster bug fixing by development
Greater collaboration between test and development
Quicker triage of problems by QA
No data loss/more granular recovery
Other

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
Conclusions

Source: IDC 2016
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.

Source: IDC 2016