Research Insight Paper

Mass Data Fragmentation Is Quietly Killing Digital Transformation Efforts

By Steve Duplessie, ESG Founder and Senior Analyst

May 2019

Contents

The Criticality of Data to Modern Organizations

Data is virtually every organization's most valuable asset. Only recently have we made vast advancements in our abilities to mine, organize, and create brand new value from our data sets—enabling companies to "transform" to the digital era— which in turn leads to better productivity, better insight, and bigger profits. Our reliance on data has never been greater and will only grow. Yet, as our data grows and proliferates among and between different application silos, storage silos, geographic silos, operational silos, and various clouds, our ability to see, access, manage, and harness the power of that data weakens.

ESG recently executed research to better understand and quantify the extent to which data is siloed in organizations today and to what degree respondents believe that dynamic creates problems. The research consisted of a survey of 200 IT decision makers influential in their organizations' purchase process for data storage, data protection, and/or data management/analytics. All respondents were based in North America, with 70% employed at enterprises (organizations with 1,000+ employees) and 30% employed at midmarket companies (those with 100-999 employees).

The Problem: Mass Data Fragmentation (MDF)

MDF, global data fragmentation, and distributed data fragmentation all refer to the same thing—by having primary and secondary data sets spread out all over the physical and virtual enterprise, we spend far more time and money trying to manage it than we get in return. *By not having a centralized consistent view of data and easy accessibility to data by any and all of our applications that require it, we will never be able to maximize the value of our data.*

The problem can be broken down into 3 essential elements:

1. **Volume** – our research indicates that secondary data volumes within surveyed organizations will grow by 36% (see Figure 1), on average, by the end of 2019. Data growth eventually destroys all normal IT processes and exposes your shortcomings. From bandwidth to capacity to operations, once you reach your ability to service your peak capacity, the system fails.

2. **Multiple Data Copies** – organizations make copies of data for all the right reasons, whether for backup/data protection, test/development, data mining, QA, or any number of other reasons. While well-intentioned, making copies of data is often the primary culprit of not only immense data growth but the introduction of "inconsistencies" within those copies. People and applications use a copy, modify it in some way, make another copy, and so on. Organizations surveyed report the typical data set is copied and stored an average of six times (see Figure 2). Moreover, these copies are often spread across many locations: 73% of respondents report their organization stores data in multiple public clouds today in addition to their own data centers. Not only are there massive volumes of copied data but it is spread everywhere.

3. **Data Operations** – the term refers to the functional things you do with your data. You protect it by backing it up. You test application rollouts against it. You run analytics on it, etc. On average, companies will have 3 or more test/dev groups accessing copied data. And to make matters worse, the average number of vendors organizations used across all secondary data operations is five!
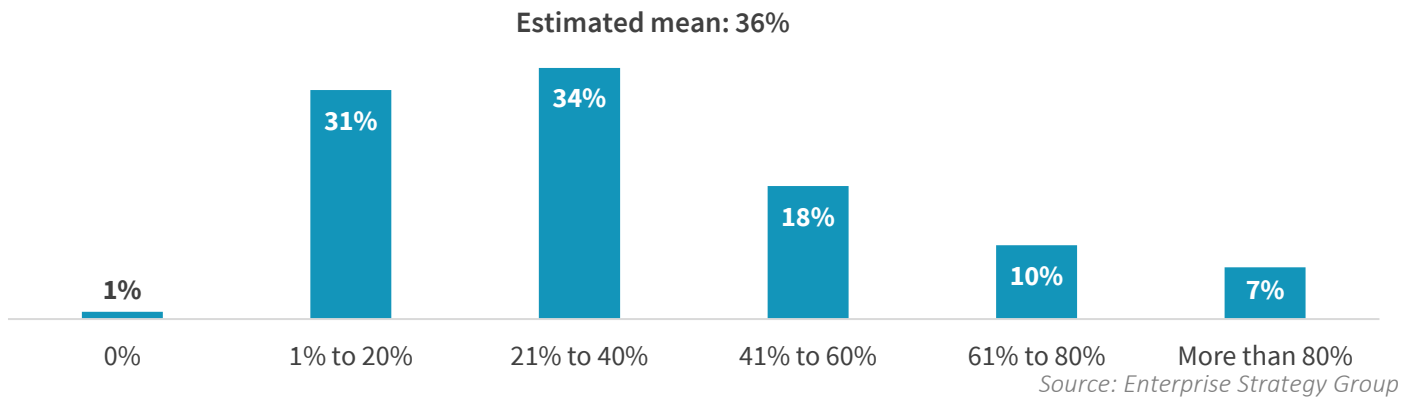
Significant Data Growth Expected

Data growth is the root cause of all data fragmentation, but it is also a universal truth among digitally enabled organizations. Among all organizations, expected growth in secondary storage capacity for the remainder of the calendar year is significant: 36% (see Figure 1). Interestingly, the largest organizations ESG surveyed also expect to experience the fastest growth. Enterprises as a cohort expect a mean growth of 38%, while midmarket organizations expect 29%. This

means organizations with the biggest data management jobs to do today are going to be under the greatest data management pressure as time passes.

Figure 1. Organizations Expect Storage Capacity to Increase in 2019

By approximately how much do you expect your organization's total secondary storage capacity to increase between now and the end of 2019? (Percent of respondents, N=200)
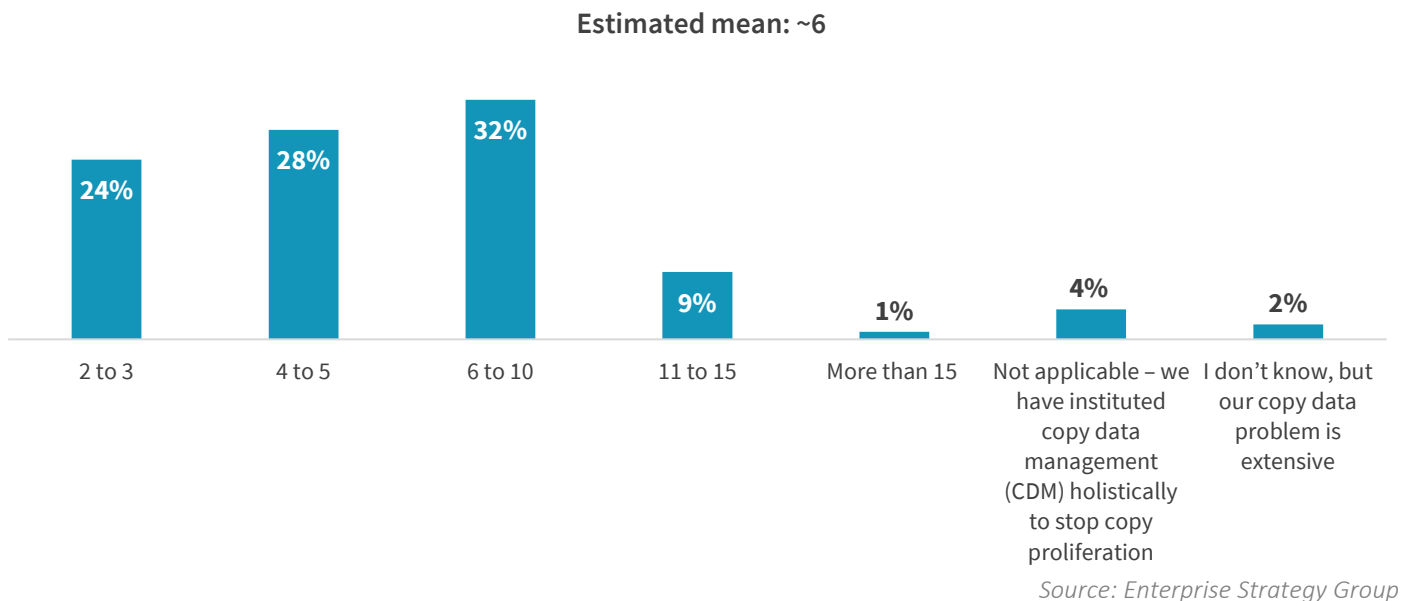
Estimated mean: 36%

Category	Percent
0%	1%
1% to 20%	31%
21% to 40%	34%
41% to 60%	18%
61% to 80%	10%
More than 80%	7%

Source: Enterprise Strategy Group

The Prevalence of 'Copy Sprawl'

When ESG asked respondents the number of times a typical data set is stored and copied, two key trends emerged. First, as already discussed, in the aggregate, organizations are experiencing significant "copy sprawl." Second, CIOs are alarmed at the amount of copies they have under management: on average, respondents holding a senior IT title estimate that 6.76 copies is typical versus middle management and staff who estimate 5.18, on average. Executives are more acutely aware of the problem than IT rank and file, which is logical as they are paid to understand the bigger picture.

Figure 2. On Average, Organizations Store Six Copies of the Same Data

For the typical data set, please estimate how many copies of the same data are created in your organization (i.e., on average, how many times is a data set copied and stored separately across all locations)? (Percent of respondents, N=200)

Estimated mean: ~6

Category	Percent
2 to 3	24%
4 to 5	28%
6 to 10	32%
11 to 15	9%
More than 15	1%
Not applicable – we have instituted copy data management (CDM) holistically to stop copy proliferation	4%
I don't know, but our copy data problem is extensive	2%

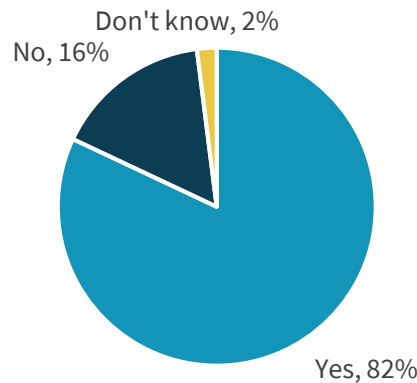*Source: Enterprise Strategy Group*

Fragmentation Across Locations

As noted, 73% of respondents reported their organization stores data in two or more public clouds today. These multi-cloud organizations report that "copy sprawl" is certainly not an on-premises-only phenomenon. When ESG asked respondents if their organization stored redundant copies of the data that is cloud-resident, 82% answered "yes" (see Figure 3). Sixty-one percent (61%) of respondents reported that redundant copies were stored on another public cloud and 38% reported redundant copies were stored in the same cloud as the copied data. In either case, the high degree of data redundancy within and across clouds results in budget wastage.

Figure 3. Over Four-fifths of Cloud Users Store Redundant Copies of Cloud-hosted Data

Does your organization store alternate or redundant copies of the data it stores on public cloud infrastructure? (Percent of respondents, N=188)


Don't know, 2%

No, 16%

Yes, 82%

Source: Enterprise Strategy Group

Heterogeneous IT Environments Contribute to Fragmentation

Data growth, redundant storage practices, and multi-cloud realities all contribute to MDF, but so do organizations' data center solutions. Different proprietary platforms are not engineered to share information or present data in a unified manner to end-users. The result is yet another silo in which insight can be trapped. When ESG asked respondents how many separate vendor solutions their organization uses across all of its secondary data operations, the average number was five (see Figure 4). Once again, enterprises appear to have the greatest degree of fragmentation with a significantly higher number of vendors reported compared to their midmarket counterparts (5.5 versus 3.8, on average).

Figure 4. On Average, Organizations Employ Five Separate Vendors Across All Secondary Data Operations

Approximately how many separate vendor solutions does your organization use for all of its secondary data operations (non-primary data e.g., backup applications, storage targets, archiving, disaster recovery, test and development provisioning, file serving, analytics)? (Percent of respondents, N=200)
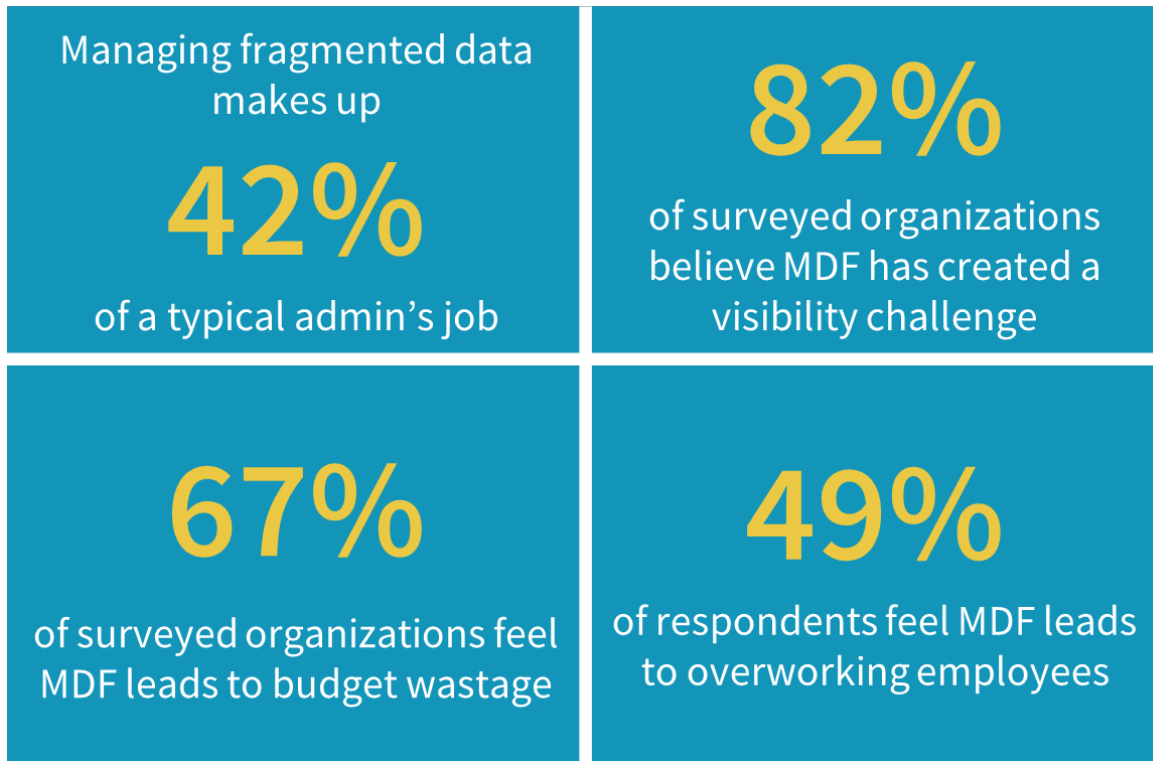

Estimated mean: ~5

1 to 3	4 to 5	6 to 10	More than 10	Don't know
38%	31%	22%	8%	2%
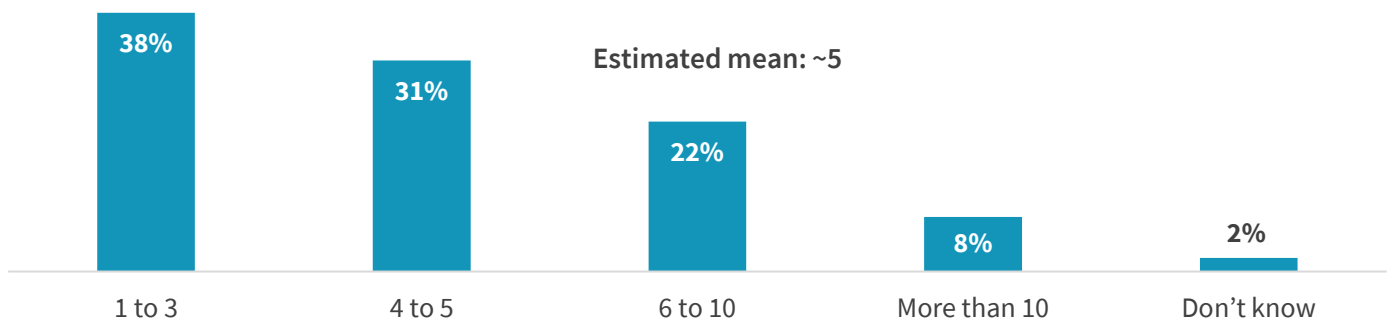
Source: Enterprise Strategy Group

Does Mass Data Fragmentation Matter? Yes!

The research clearly shows that organizations today are dealing with extremely fragmented data assets, but why does that matter? ESG delved into the ramifications of MDF in a series of questions and the summarized findings in Figure 5 are noteworthy.

Figure 5. Summarized Outcomes of Mass Data Fragmentation

Managing fragmented data makes up

42%

of a typical admin's job

82%

of surveyed organizations believe MDF has created a visibility challenge

67%

of surveyed organizations feel MDF leads to budget wastage

49%

of respondents feel MDF leads to overworking employees

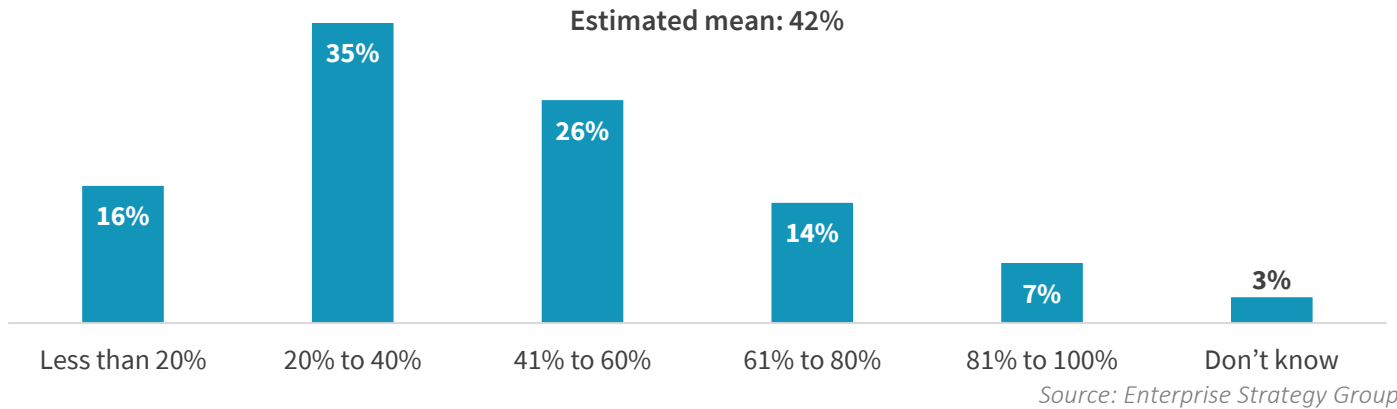*Source: Enterprise Strategy Group*

MDF Saps Productivity

ESG asked respondents what percentage of administrators' day-to-day tasks are dedicated to managing their organization's secondary data, applications, and copies across on-premises and cloud environments. The answer: 42%, on average (see Figure 6). If 42% of a typical IT admin's job is managing fragmented data, then 42% of their cost is ultimately wasted in non-productive, non-profitable endeavors.

Worse, 49% of respondents surveyed believe that MDF leads directly to overworked employees. If you consider that we know employees seek intelligent, meaningful tasks as a condition of employment and job satisfaction, we are essentially killing ourselves by continuously demanding skilled workers to perform menial, mundane jobs.

Figure 6. Percentage of Tasks Dedicated to Managing Secondary Data, Applications, and Copies

For the typical data storage/data protection administrator, what percentage of typical day-to-day tasks are dedicated to managing your organization's secondary data, applications, and copies across on-premises and cloud environments? (Percent of respondents, N=200)

Estimated mean: 42%

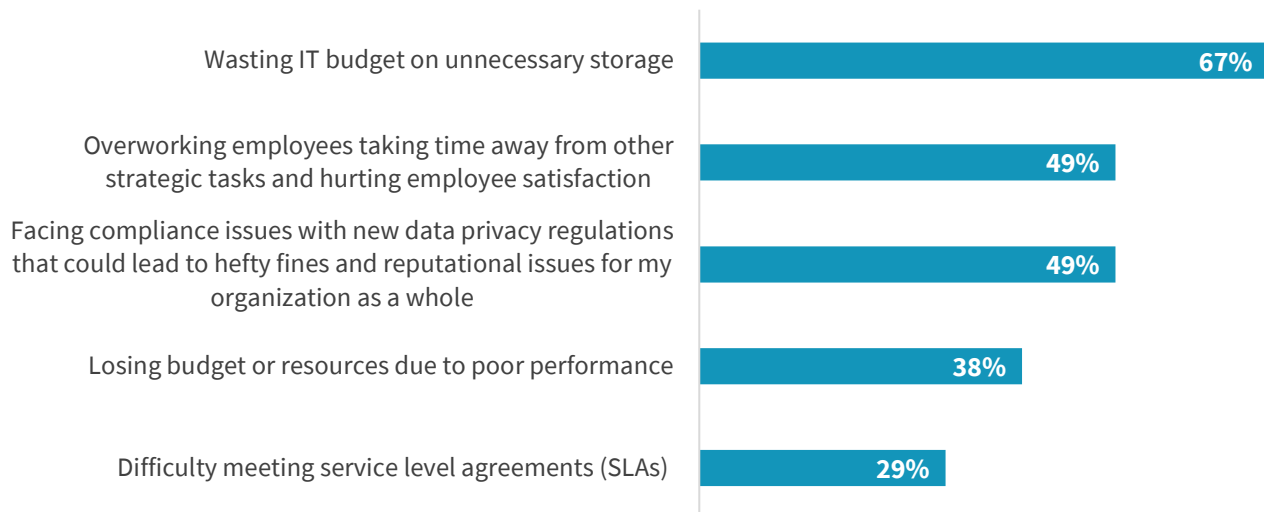|	
Less than 20%	16%
20% to 40%	35%
41% to 60%	26%
61% to 80%	14%
81% to 100%	7%
Don't know	3%

Source: Enterprise Strategy Group

MDF Creates Expensive Blind Spots

ESG asked respondents if they felt their organization had problems getting a holistic view of their data because of the numerous fragmented silos in the environment. The answer? A resounding yes: 82% reported the affirmative. Having 82% of organizations believing that MDF creates a data visibility challenge means that we are not even able to see—let alone leverage—all of the data assets in our organization. How can we be confident in our decision support systems if we are not able to see or access all of the data? The answer is we can't. At best we are flawed. At worst, we are wrong.

Beyond impacts to decisioning efficacy, respondents reported additional challenges associated with MDF, most often that it leads to wasted IT budget (see Figure 7).

Figure 7. Risks IT Organizations Face Due to Data Fragmentation Challenges

What are the risks your IT organization faces due to its data fragmentation challenges (i.e., the inability to get a holistic view of all its data because it has too many data silos)? (Percent of respondents, N=162, multiple responses accepted)

Wasting IT budget on unnecessary storage	67%
Overworking employees taking time away from other strategic tasks and hurting employee satisfaction	49%
Facing compliance issues with new data privacy regulations that could lead to hefty fines and reputational issues for my organization as a whole	49%
Losing budget or resources due to poor performance	38%
Difficulty meeting service level agreements (SLAs)	29%

Source: Enterprise Strategy Group

The Bigger Truth

MDF is a problem. It wastes productivity, makes staff miserable, and diminishes organizational intelligence, all while wasting money. And the problem will continue to get worse organically due to data growth. The proliferation of copied data across every aspect of our organization (and outside of it, via the cloud) is not only a significant costly management nightmare, it keeps us from harnessing the value of our data.

But there is hope. People are beginning to pay attention to MDF and do something about it. In the last few years, companies have formed, built from the ground up, to tackle this problem. The combination of new technologies with new processes, awareness, and more intelligent internal policies for dealing with MDF and the associated data copies gives us hope.

There is no panacea or magic bullet to eliminate the MDF problem but understanding that you have MDF and starting to question how to deal with it is the most important first step. For new application deployments or new use cases, organizations should be designing and architecting a plan for the real lifecycle of the data that is generated. Data lives and breathes in our organizations and requires more intelligent planning for how we are going to treat it at its various life stages.

We need to consider how we are going to store, protect, access, govern, and retire our data based on each data operational area—and stop just assuming "someone else will do it." Inconsistencies in policies will lead to "data chaos" within your organization. A more intelligent plan can save incalculable time and money, and lead to much better knowledge worker productivity—all of which mean a better bottom line.