

The background of the cover is a grayscale photograph of two men in a professional setting. One man is seated at a table with his hands clasped, looking towards the camera. The other man is standing behind him, leaning over and looking down at something on the table. The image is partially obscured by a large, dark, diagonal shape that cuts across the upper left and middle of the page. The title text is overlaid on the white area of the page.

TOP 10 EVALUATION CRITERIA FOR COPY DATA MANAGEMENT & DATA VIRTUALIZATION

WHITE PAPER



Data cloning technology has advanced significantly with data virtualization solutions such as Copy Data Management (CDM) and Database Virtualization Appliances (DVA). These technologies have seen a rapid adoption in the past 4 years during which time more than 100 Fortune 500 companies have implemented data virtualization. Early adopters have achieved significant acceleration in their IT projects and massive reduction in their data storage requirements by leveraging the ability to produce clones of data in minutes with almost no storage overhead. This is achieved not by copying data, but instead by sharing one initial data copy across multiple environments, while at the same time allowing read/write access by storing changes privately for each environment.

Why are CDM and data virtualization so important to businesses? CDM and data virtualization solutions drastically reduce the data sprawl created by copies of data. The average enterprise creates 8-10 copies of every production data source, for app development, testing, user acceptance, production support, reporting, business intelligence, etc. Thus, a 5 TB production database leads to 40-50 TB of downstream copies, and a F500 firm might have more than 1,000 databases. The amount of storage required to effectively manage copy data is stunning. Data virtualization eliminates all the redundant copies of data while at the same time—and more importantly—reducing the time required to make copies. The result is accelerated IT projects at reduced costs.

Today there are only a few vendors offering products focused on copy data management, but it's already becoming difficult to navigate through vendor marketing and uncover the true capabilities and benefits of each vendor's products—and this challenge will only grow as more vendors enter the space. This becomes even more challenging still when you attempt to map a vendor's unique features to your specific use cases (such as application development, disaster recovery, or test data management). Savvy IT execs also realize that "virtualizing" data or simply managing copies of data will soon be the baseline functionality required in IT departments.

The time has already come where you should be asking questions like:

- What types of unique features does the vendor provide?
- How does the solution scale beyond my initial use cases?
- Can I use the solution both in my datacenter and across hybrid and public clouds?
- How much automation, self-service, and application integration is pre-built into the solution?
- How much custom work, such as scripting, manual management or using other products, will be required to use the solution?
- Does the vendor have customers of my size and in my industry using their solution today?
- How fast can the solution be tested and run in a Proof Of Concept (POC)?
- How rapidly can the solution be installed and rolled out?

So let's now review the top 10 POC considerations from companies already using these solutions. We'll break them down into two groups:

- The top 5 qualification questions to ask before the POC
- The top 5 things to test during the POC

TOP 5 QUALIFICATION QUESTIONS

1. DOES THE SOLUTION SUPPORT MY ENVIRONMENT?

The first and most obvious goal is to find a solution that will easily integrate with your company's infrastructure, data sources, and application stacks. This includes both on-premise environments and scaling into remote or cloud environments. You also want to make sure that you're not being locked into a solution that only supports a single source environment. Do you have a requirement for application support such as Oracle EBS, PeopleSoft or SAP? Do you have a requirement for more than one type of database such as SQL Server, Sybase, Oracle, MS SQL, etc.? Do you have requirements for multiple host operating systems such as Linux, AIX, HP/UX, Solaris and Windows? Finally, does the solution require specialized hardware or can it run on your existing system resources?

2. DOES THE SOLUTION HAVE SPECIFIC BUILT-IN FEATURES FOR MY INTENDED BUSINESS GOALS AND REQUIREMENTS?

Is this the solution that best aligns with my company's key business outcomes? Some example outcomes are:

- Accelerating application release cycles
 - Are there specific GUIs and features to support application developers?
- Streamlining business intelligence
 - Are there data alignment and fast data refresh capabilities that can eliminate batch windows, shorten ETL times and offload production?
- Ensuring data privacy and security
 - Does the solution include support for data masking, auditing and chain of custody?
- Accelerating integration testing
 - Is there support for fast QA environments and QA-specific features such as rollback for destructive testing?
- Migrating data to cloud environments
 - Does the solution support cloud infrastructures and enable replicating data from in-house sources to public/hybrid clouds and vice versa?
- Enhancing backup and DR strategies
 - Does the solution support long range and fine-grained RPO and fast RTO?

3. DOES THE SOLUTION ALLOW ME TO EASILY SCALE TO NEW USE CASES?

How well does the solution drive down TCO and provide a greater ROI by addressing other business needs throughout my company? (For example, I now have virtualized data that is synchronized with production for an integration testing use case. Can I now mask that data, migrate it to a public cloud and enable my analytics team to perform business intelligence on the same datasets?)

4. DOES THE VENDOR HAVE CUSTOMERS SIMILAR TO MY SIZE AND BUSINESS REQUIREMENTS?

Will I be forced to go through the growing pains of helping the vendor break new ground or do they have leaders in finance, retail, manufacturing, government, high-tech, and other industry verticals using their solution today?

5. CAN I EASILY AND QUICKLY POC THE SOLUTION?

Is the vendor willing to show me all of these features during an actual POC? Are they able to back up all of their sales and marketing claims during an onsite POC with clearly defined success criteria?

TOP 5 BEST PRACTICES FOR THE POC

1. POINT-IN-TIME PROVISIONING

Provision an environment to an exact point in time (in between “snapshots”). What is the process for finding an exact point in time? How easily does the solution allow me to provision data environments down to the minute, second, or transaction? Can provisioning be accomplished by the push of a button by an end user such as a developer or business analyst or does it require custom scripts and multiple people such as storage admins, database administrators or system administrators?

2. BRANCH, ROLLBACK AND RESET OF ENVIRONMENTS

Now that I’ve provisioned a parent environment (a replica of a production environment), I’d like to make some changes and provision a branch (or child) of that environment. After making some changes to the child environment, now I would like to rewind the child environment back an hour, or 6 hours. How is this accomplished? Finally, I would like to reset both environments to their original state. (These functionalities are critical for application development, integration, or test data management use cases).

3. REFRESH PARENT AND CHILDREN ENVIRONMENTS WITH THE LATEST DATA

I’ve created a number of parent environments and I’ve spawned a number of child environments off of those parents. What is the process for getting the latest data from production environments into my parent and child environments? Could this process be performed by a developer or analyst? What is the impact to production (if any)?

4. PROVISION MULTIPLE SOURCE ENVIRONMENTS TO THE SAME POINT IN TIME

I have a number of use cases (business intelligence, integration, DR, etc.) where I will need to align and provision multiple different data sources to a particular point in time. (For example: align all of my source datasets to Friday at 4pm, their local time). How is that achieved?

5. AUTOMATION / SELF-SERVICE / AUDITING CAPABILITIES

Can I perform all of the above tasks via a self-service GUI console? Is the GUI intuitive enough to provide self-service for my developers, analysts, and data owners? How robust is the CLI, and is there a full set of RESTful APIs for integration with DevOps tools? Finally, are all the tasks captured in an event log to provide me a source of record of access to my data?

SUMMARY

The number one take-away should be to make sure the vendors prove their solution and feature sets to you in the POC. Copy Data Management and Data Virtualization solutions have the potential to deliver massive advancements in data agility and data center utilization—advancements at a level not seen since VMware popularized server virtualization over ten years ago. For this reason, there will soon be many vendors attaching to this wave and claiming that their solution provides all of the capabilities outlined in this document. Your task of separating the contenders from the pretenders is potentially a simple one—just make them prove it!



Top 10 Evaluation Criteria For Copy Data Management & Data Virtualization January 2015

You can find the most up-to-date technical documentation at:

<http://www.delphix.com/support>

The Delphix Website also provides the latest product updates.

If you have comments about this documentation, submit your feedback to:

help@delphix.com

Delphix Corp.
275 Middlefield Road, Suite 50
Menlo Park, CA 94025
www.delphix.com

© 2015 Delphix Corp. All rights reserved.

The Delphix logo and design are registered trademarks of Delphix Corp.
in the United States and/or other jurisdictions.

All other marks and names mentioned herein may be trademarks of
their respective companies.