U.S. Environmental Protection Agency

DATA CENTER CONSOLIDATION STRATEGIC PLAN

April 2019

TABLE OF CONTENTS

Table of Contents

1 Executive Summary

In 2016, the Office of Management and Budget (OMB) launched the Data Center Optimization Initiative (DCOI) as a successor to the Federal Data Center Consolidation Initiative (FDCCI). On August 1st, 2016, the Federal Chief Information Officer (CIO) issued memorandum M-16-19 which provided the framework for agencies to achieve data center consolidation and optimization requirements. Under memorandum M-16-19 agencies were required to develop, monitor and report on data center strategies to:

- Consolidate inefficient infrastructure,
- Optimize existing facilities,
- Improve security posture,
- Achieve cost savings, and
- Transition to more efficient data center infrastructures, such as cloud services and interagency shared services.

The EPA made considerable progress towards achieving the targets as established in M-16-19. As of March 2019, the EPA has closed 22 non-tiered data centers under DCOI and achieved energy metering requirement for non-advanced energy metering. As indicated in GAO-17-448: "Data Center Optimization" (August 2017); the EPA maintained advanced energy metering at one of the four tiered-data centers and achieved the M-16-19 targets for power usage effectiveness (PUE) and virtualization targets set forth in M-16-19. As of March 2019, none of the EPA's tiered data centers have experienced downtime in the two reporting periods and EPA expects to maintain a high level of availability due to redundancy inherent to a tiered data center.

Continuing the progress mentioned above, throughout 2019 and 2020, the EPA is on target to close 3 non-tiered data centers in 2019 and will work towards the continued expansion of its use of virtualization. To ensure virtualization remains a top priority for EPA, the CIO will expand the Virtual-First strategy currently in use at the EPA's National Computer Center (NCC) to the rest of the agency. The EPA will also explore the potential expansion of advanced energy metering at its remaining tiered data centers. The EPA is also working to expand the utilization of its ScienceLogic (EM7) tool to identify opportunities to strengthen its automated monitoring of server utilization and virtualization.

In 2019, OMB released a draft update to DCOI in response to the extension of the Federal Information Technology Acquisition Reform Act (FITARA) through November 2020. This document outlines EPA's strategy to address the requirements set forth in the draft document as written in March 2019.

2 Background

In August 2016, the Office of Management and Budget (OMB) released memorandum M-16-19 that established the Data Center Optimization Initiative (DCOI) which superseded the previous Federal Data Center Consolidation Initiative (FDCCI). As of March 2019, OMB released a draft memorandum that may supersede M-16-19. The draft memorandum establishes new optimization requirements and rescinds several M-16-19 requirements, such as, reporting for non-tiered data centers, PUE and DCEP. DCOI's consolidation and optimization requirements have a targeted completion date of September 2020 and are summarized as follows:

- Continue to close non-tiered data centers where viable
- Increase the use of virtualization
- Continue to replace manual collections with automated tools
- Evaluate and report data center costs, savings and avoidances.

3 Introduction

Consolidation and optimization of federal data centers continues to be a priority for the EPA. At the inception of DCOI, EPA had 58 non-tiered data centers and computer rooms that supported business operations for EPA facilities located across the country. Per OMB criteria, four of EPA's data centers are classified as tiered data centers. These include the following:

- National Computer Center (NCC) in Research Triangle Park, North Carolina.
- Potomac Yards Data Center in Washington, DC.
- Region 5 Data Center in Chicago, Illinois.
- National Enforcement Investigations Center (NEIC) in Lakewood, Colorado.

To improve efficiency and consistency of data center operations, EPA implemented a Physical-to-Virtual (P2V) initiative at the primary tiered data center (NCC) requiring offices to convert existing physical servers to virtual servers wherever possible. EPA also defined server and software standards for virtualized platforms and established an enterprise platform for infrastructure monitoring. EPA implemented strategic sourcing initiatives to pool resources and negotiate optimal pricing for IT products and services.

EPA is pursuing opportunities to establish shared data center services and encourage consolidation of data center functions. EPA established centralized resources for continuity of operations (COOP) and disaster recovery (DR) in the tiered data centers. Region 6 and Region 9 are currently leveraging these services and the Region 4 implementation is being operationalized. The General Services Administration (GSA) also established a five-year agreement with EPA to leverage NCC facilities and hosting services.

For non-tiered data centers, EPA plans to leverage its existing EM7 tool to support system monitoring requirements of server utilization and virtualization. EPA's efforts to consolidate and optimize its data centers have been significant. However, it is important to acknowledge that these efforts involved considerable resource commitments to upgrade data center infrastructure, consolidate data center assets and perform tool assessments. These efforts will enable EPA to

better monitor and manage its energy consumption and report on OMB's optimization metrics, but they may not produce cost savings in the near-term. Additionally, EPA must also balance consolidation efforts with network costs and application performance requirements. For example, some EPA regional offices, research centers, labs and other facilities host local infrastructure such as telecommunications infrastructure, to support specialized lab and research systems, and COOP and DR. The distributed nature of this infrastructure and the need to support these critical requirements can make further consolidation challenging. As EPA moves forward to achieve DCOI requirements, it will need to continue to balance the benefits of consolidation with operational requirements and costs. The remainder of this document provides more detail on EPA's progress and plans for meeting DCOI requirements.

4 Data Center Consolidation and Closure Targets

Achieved

The EPA has worked diligently to consolidate and close data centers. At the inception of DCOI, the EPA maintained 58 data centers as defined in M-16-19. As of March 2019, the EPA has successfully closed 22 non-tiered data centers and reclassified one data center as invalid due to modifications to the definition of a data center. Per OMB's criteria, EPA classifies four of its data centers as tiered. The tiered data centers support EPA's enterprise information technology (IT) operations. EPA's primary tiered data center is the NCC, located in RTP, North Carolina. In 2018, the EPA downgraded the Region 8 tiered data and promoted the federally owned National Enforcement Investigations Center (NEIC) in Lakewood Colorado to tiered data center status. This data center will serve as EPA's western presence data center. Under FDCCI, EPA closed 21 of its data centers. For DCOI, EPA has closed 22 data centers, 1 is in progress with an anticipated completion in FY2019.

Table 1 depicts a summary of EPA progress from FDCCI through FY 2019.

Table 1 - Completed and Planned Data Center Closures by Fiscal Year

Completed and Planned Data Center Closures by Fiscal Year											
FY 2010	FY 2011	FY 2012	FY 2013	FY 2014	Q1/2 2015	Q3/4 FY 2015	FY 2016	FY 2017	FY 2018	FY 2019 (Closed)	FY 2019 (Planned)
0	1	14	3	1	2	3	3	13	4	2	1
21 Total FDCCI Closures						22 DCOI Closures Completed					
43 Total Closures, 1 Planned											

Activities Underway

For those data centers targeted for closure, DCOI points of contact (POCs) are reviewing their data center IT assets and determining which to decommission, consolidate and/or virtualize. They will then execute consolidation plans to move applicable data center IT assets to the Cloud, an EPA tiered data center or the data center retained for that geographic area. The final phase of the effort will require DCOI POCs to surplus or excess data center IT assets and scale remaining data center assets to align with reduced capacity requirements. Efforts are also underway to align additional closures with lease timelines and funding availability.

Although the EPA does not plan to close any tiered data centers in FY 2019, we are in the planning phase to close the Potomac Yards tiered data center in FY 2020. The EPA anticipates

some cost savings or cost avoidance due to the reduction of a lease payment as the data center assets are expected to migrate to other federally-owned facilities. At project completion, the EPA will have three (3) tiered data centers that will be operated in a standardized fashion and will implement configurations that maximize power and cooling efficiency.

Challenges and Risks

In some cases, EPA's regional offices, research centers, labs and field offices host local infrastructure data center assets such as specialized lab and research support, emergency response (ER) and COOP that may have non-severable configurations and potentially, must remain co-located to operate and fulfill mission functions. The distributed nature of these offices and the continuation of essential localized mission functions require EPA to balance consolidation efforts with continuity of business functions, application performance requirements, cost and security. Testing is required to validate operational functionality before migrating these systems away from the local site. EPA has identified several laboratories with servers running Laboratory Information Management Systems (LIMS) that interface with LAN-based scientific equipment. To migrate the LIMS servers, the EPA is required to test the WAN capabilities of these systems as well as procure network and security hardware to ensure the scientific equipment is isolated due to security concerns. Budgetary uncertainties and suspension of federal services have extended execution of efforts required to perform validation and procurement activities into FY 2020.

5 Cost Savings

Initial Costs for Consolidation and Optimization

Given the upfront investments for infrastructure upgrades, tool assessment and validation, the EPA does not expect significant near-term cost savings under DCOI. The EPA structured its data center consolidation work to promote cost savings in areas of increased facility utilization, reduced energy consumption, maximized server and storage use, and reduction in the long-term growth of IT infrastructure costs. EPA is continuing to refine and enhance DCOI cost savings calculations to ensure accuracy. Though OMB agreed there would be negligible savings in the near-term, the EPA has begun gathering additional information to track any realized costs or savings within our quarterly progress data calls.

Life Cycle Costs Savings and Other Improvements

Projected costs needed to fulfill identified DCOI goals include planning and design activities, updates to data center facilities (e.g., uninterruptable power supply, power distribution unit, and computer room cooling equipment), validation testing of local applications across the Wide Area Network (WAN), DCIM tool and security device procurements, training and/or hiring. Additionally, the EPA required investment in WAN bandwidth upgrades to accommodate additional traffic that is no longer isolated to the Local Area Network (LAN) infrastructure. These investments, while critical to support long-term DCOI objectives, offset any near-term cost savings. In the longer-term, however, the Agency expects these improvements will reduce energy consumption leading to greater efficiency and demonstrable cost reductions.

Historical Costs, Cost Savings and Cost Avoidances

Agency	Total achieved savings for 2016 through 2018	OMB Savings Target	Additional Planned through 2019
Enviornmental Protection Agency	$0	$0	$0

6 Data Center Optimization Metrics

OMB established several optimization metrics under DCOI to include: virtualization, advanced energy metering, server utilization and availability. Energy efficiency and facility utilization metrics from M-16-19 were removed in OMB's draft DCOI document and will no longer be reported by the EPA under DCOI. EPA's implementation strategy for the DCOI metrics are outlined below.

6.1 Updated Metric: Virtualization

OMB updated the virtualization metric under the new draft guidance. Moving forward OMB requires the EPA to report the number of servers and mainframes that are currently serving as hosts for virtualized or containerized systems at EPA-managed data centers.

- **Planned**

  The EPA will expand the NCC Virtual-First strategy agency-wide. EPA regions and program offices will prioritize a virtualized server deployment, the EPA expects to naturally gravitate toward a greater agency virtualization portfolio.

- **Achieved**

  The EPA maintained a 4-to-1 or greater ratio of virtual operating systems to physical servers as calculated by M-16-19 standards. As data is gathered under the new DCOI virtualization requirements, the EPA will adjust their achieved virtualization metrics to align with the new guidance.

6.2 Updated Metric: Advanced Energy Metering

Under the new draft DCOI guidance, OMB requires agencies to provide energy metering numbers only for data centers that have advanced energy metering. Additionally, previous requirements for advanced energy metering at all tiered data centers under Executive Order 13693: Planning for Federal Sustainability in the Next Decade, was revoked by the Executive Order 13834: Efficient Federal Operations. The EPA continues to evaluate available tools and their planned return on investment to ensure the agency is maximizing value of tax payer money.

- **Planned**

  The EPA has advanced energy metering installed in one of the four (4) tiered data centers. The EPA explored adding additional tools such as a Data Center Infrastructure

Management (DCIM) at the NCC. Due to the high cost for implementation, available staff, and low return on investment, the EPA does not expect to deploy advanced metering or a DCIM at the other three (3) tiered data centers in FY 2019.

- **Achieved**

Under M-16-19 the EPA maintained an average 1.6 Power Usage Effectiveness (PUE). As this metric was recently modified, the EPA does not have any achieved value for advanced energy metering. Future cost savings and avoidance through efficient energy usage and improvements is expected for FY 2020.

6.3 Updated Metric: Server Utilization

M-16-19 set specific guidelines for server utilization calculations that created challenges in providing accurate measurements. The EPA identified its Science Logic tool (EM7) for providing automated reporting for server utilization across EPA data centers. Before the EPA could leverage EM7, it first had to upgrade licensing and add additional collectors to handle the expected increases in reporting. The EPA expects to have EM7 collection for all tiered data centers in place by the end of FY 2019.

- **Planned**

In FY 2018 the EPA made progress in upgrading EM7. The EPA plans to leverage this tool to provide dashboards of each data center and equipment they contain. By using EM7 and defined efficiency measures, the EPA expects to have a tool in place by FY 2020 that will automatically identify underutilized servers at EPA data centers. The EPA plans to use this reporting capability to identify, justify or remediate through consolidation or cloud migration any server that is not utilized properly.

- **Achieved**

The EPA made progress in upgrading the licensing and capabilities of EM7 in FY 2018. Due to difficulty in calculating M-16-19's utilization metric, the EPA did not meet or have ability to provide the required server utilization metrics under M-16-19. The EPA will continue to provide numbers for underutilized servers manually until EM7 is configured to automatically generate agency reports.

6.4 New Metric: Availability

OMB added data center availability as a new metric for DCOI. The EPA will report planned hours of data center availability, and actual downtime from the previous quarter (in hours).

- **Planned**

In FY 2019, the EPA expects to achieve 100% uptime at the four (4) tiered data centers. Due to the redundant configuration in place at the EPA's tiered data centers to include generators, uninterruptable power supplies (UPS), and network redundancy the EPA does not anticipate its data centers to experience any significant downtime.

- **Achieved**

As of March 2019, all tiered data centers report 0 hours of downtime and expect 99.99% availability for the upcoming calendar year.

7 Conclusion

Significant progress has been made to optimize and consolidate EPA data centers. The remaining planned improvements will enable the EPA to establish a baseline to achieve the vast majority of OMB's DCOI objectives, such as increased virtualization through a virtualization-first agency initiative and increased visibility towards server utilization using EM7. The EPA will work to balance the achievement of the DCOI objectives with the business needs, and work to overcome technical issues and resource constraints associated with achieving advanced energy metering at all tiered data centers as recommended by OMB. Investing in further consolidation will enable the agency to better monitor and reduce energy consumption but may not produce cost savings in the near-term.

EPA must balance consolidation efforts with network costs and application performance requirements. Because some EPA regional offices, research centers, labs and other facilities host local infrastructure such as telecommunications infrastructure, specialized lab and research systems, and COOP and DR in their data centers, further consolidation can be challenging. As EPA moves forward to achieve DCOI requirements, it will need to continue to balance the benefits of consolidation with operational requirements and implementation costs.