Performance Audit of Snohomish County’s
Department of Information Services’
Data Storage Methods and Costs

June 2, 2015
# Table of Contents

Report Highlights ...........................................................................................................ii

A. Introduction and Background .................................................................................. 1

B. Scope and Methodology .......................................................................................... 8

C. Audit Observations and Conclusions....................................................................... 10
   - C.1 Current Data Storage Strategies........................................................................10
   - C.2 Tier One—Application-Centric Data .................................................................12
   - C.3 Tier Two—Unstructured Data............................................................................16
   - C.4 Tier Three—Continuous Business Operations and Disaster Recovery ........20
   - C.5 Conclusions and Recommendations................................................................23

Appendix A—Recommendations and Corrective Action Plan ............................. 25

Appendix B—Department of Information Services’ Response......................... 29
June 2, 2015

**OVERALL CONCLUSION**
Alternatives to current data storage methods exist through the cloud, representing the potential for cost savings. Consideration of each should be based on the type and use of data on the County’s network.

**AUDIT PURPOSE**
To assess existing data storage methods and costs and to determine whether alternative protocols could provide a more cost-effective approach in the future.

**BACKGROUND**
- In 2015, DIS will operate with a $2 million capital budget. In the past, 17 percent of capital costs have been used to support its disaster recovery and data storage infrastructure.
- The audit assessed alternatives for three distinct tiers of data:
  - **Tier One**—Application-centric data that is structured through databases.
  - **Tier Two**—Unstructured data, such as network drives, archives, legal hold documents, and public records; 88 percent of DIS data is unstructured.
  - **Tier Three**—Backup data, including continuous replication at a Mirror 1 data center and disaster recovery at a remote Mirror 2 data center.
- DIS experienced a 300 percent growth in data volume from 2012 to 2015, primarily due to the increase in Tier Two and Tier Three data.

**KEY FACTS & OBSERVATIONS**
Given that the Department of Information Services (DIS) recently completed significant capital improvements to its application-centric (Tier One) data storage, most near-term opportunities relate to unstructured (Tier Two) data and disaster recovery (Tier Three) data. Specifically:
- DIS began a project to migrate approximately 10TB of 150TB Tier Two data to Microsoft Office 365 cloud storage. This data includes user-specific files, such as Word, Excel, and PDFs.
- While user files will migrate to Office 365, DIS does not have specific plans to migrate the remaining 140TB of Tier Two data at this point, though alternatives have been found to be cost effective.
- DIS disaster recovery costs were generally consistent with comparable cloud-based disaster recovery solutions—roughly $160,000 per year—but there may be benefits to a cloud-based approach. While on-site disaster recovery may guarantee rapid recovery times, cloud solutions guarantee greater geo-diversity.
- Tier Two and Tier Three storage systems will reach the end of their lifecycles within the next two to three years requiring DIS to incorporate analyses of various alternatives in upcoming strategic plans.
- This audit did not identify specific cost savings associated with Tier One data storage—primarily due to the high cost variability in procuring applications and insufficient information available from current DIS system metrics. Nonetheless, potential options for cost savings exist, and are associated with migrating existing systems to a cloud-based platform or procuring systems that are hosted in the cloud.
- Any migration of Tier One or Tier Two data would result in a corresponding reduction in Tier Three storage.

**KEY RECOMMENDATIONS**
To ensure a cost-effective data storage strategy in the future, DIS should:
- Continue migration of selected files to Microsoft Office 365 cloud storage and evaluate the cost-benefit of extending the migration of remaining unstructured data to cloud-based solutions, such as SharePoint, cold storage, or an alternative service.
- Evaluate cloud-based alternatives to existing Tier Three data storage, particularly those that offer cost-savings and increase geo-diversity.
- Incorporate into each strategic plan, or other formal documentation, analysis that addresses the cost-benefit of current data storage methods as compared to recognized alternatives.
- Gather sufficient system and network data necessary to fully evaluate the cost-benefit of moving Tier One applications and data to the cloud.
## Summary of Potential Cost Savings

<table>
<thead>
<tr>
<th>Tier</th>
<th>Estimated Average Cost per TB</th>
<th>Estimated Average Cost per Year</th>
<th>Assumed Volume</th>
<th>Decision Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier One</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing</td>
<td>$3,508.68</td>
<td>$140,347</td>
<td>40TB</td>
<td>• Lifecycle ends in 2020, thus cost saving unlikely to occur prior to 2020.</td>
</tr>
<tr>
<td>SaaS</td>
<td>Undetermined</td>
<td></td>
<td></td>
<td>• Decision points occur when each application or physical server needs an upgrade or replacement, or when new business needs are discovered.</td>
</tr>
<tr>
<td>PaaS or IaaS</td>
<td>Undetermined</td>
<td></td>
<td></td>
<td>• Decisions will likely lead to an incremental migration to the cloud as determined by analysis of cost-benefit and business needs.</td>
</tr>
<tr>
<td>Potential Savings for Cloud</td>
<td>Undetermined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tier Two</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing</td>
<td>$330.27</td>
<td>$49,541</td>
<td>150TB</td>
<td>• Cost savings unlikely to occur prior to lifecycle end in 2018 unless migrating storage to the cloud before 2018 mitigates upgrades to on-premises storage before the lifecycle end.</td>
</tr>
<tr>
<td>Amazon WorkDocs</td>
<td>$300</td>
<td>$8,400</td>
<td>30TB</td>
<td>• Type of data (user-specific working files, department files, public records requests, legal holds, HIPAA, and CJIS) will determine which solution is most suitable.</td>
</tr>
<tr>
<td>Google for Work</td>
<td>$2,000</td>
<td>$56,000</td>
<td>30TB</td>
<td>• Decisions will likely lead to an incremental migration to the cloud as determined by analysis of cost-benefit and business needs.</td>
</tr>
<tr>
<td>Microsoft Office 365</td>
<td>$60</td>
<td>$1,680</td>
<td>30TB</td>
<td>• Cold Storage cost per TB starts as low as $120, but will also include costs depending on the frequency of data access.</td>
</tr>
<tr>
<td>Cold Storage</td>
<td>$120</td>
<td>$13,440</td>
<td>120TB</td>
<td>• 30TB assumed volume for Existing Microsoft License is a rough estimate based on 10TB of user-specific files plus 20TB of other data that would require frequent access.</td>
</tr>
<tr>
<td>Potential Savings for Cloud based on “Cold Storage” and “Existing Microsoft License”</td>
<td>$229</td>
<td>$34,421</td>
<td>150TB</td>
<td></td>
</tr>
<tr>
<td>Tier Three, Mirror 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing</td>
<td>$999.41</td>
<td>$189,888</td>
<td>190TB</td>
<td>• Lifecycle of the data center ends in 2035.</td>
</tr>
<tr>
<td>Tier One</td>
<td>Undetermined</td>
<td></td>
<td>40TB</td>
<td>• Mirror 1 cloud solution is dependent on Tier One and Tier Two solutions.</td>
</tr>
<tr>
<td>Tier Two</td>
<td>Dependent</td>
<td></td>
<td>150TB</td>
<td>• Tier One and Tier Two data migration will have a corresponding reduction to Mirror 1.</td>
</tr>
<tr>
<td>Potential Savings for Cloud</td>
<td>$229</td>
<td>$34,421</td>
<td>190TB</td>
<td></td>
</tr>
<tr>
<td>Tier Three, Mirror 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing</td>
<td>$1,130.59</td>
<td>$169,588</td>
<td>150TB</td>
<td>• Lifecycle ends in 2018, thus cost saving unlikely to occur prior to 2018; however, Tier One and Tier Two data migration will have a corresponding reduction to Mirror 2.</td>
</tr>
<tr>
<td>Barracuda (high)</td>
<td>$2,484</td>
<td>$372,598</td>
<td>150TB</td>
<td>• Cloud alternatives may offer cost savings and greater geo-diversity, but possibly at the expense of data security and time taken to rebuild systems.</td>
</tr>
<tr>
<td>Carbonite</td>
<td>$1,003</td>
<td>$150,484</td>
<td>150TB</td>
<td></td>
</tr>
<tr>
<td>Barracuda (low)</td>
<td>$792</td>
<td>$118,798</td>
<td>150TB</td>
<td></td>
</tr>
<tr>
<td>Potential Savings for Cloud based on “Barracuda (low)”</td>
<td>$339</td>
<td>$50,790</td>
<td>150TB</td>
<td></td>
</tr>
</tbody>
</table>
A. Introduction and Background

The mission of Snohomish County’s Department of Information Services (DIS) is to provide technical expertise and strategic insight to improve delivery of public services by all County departments, offices, and courts (herein collectively referred to as “departments”). DIS serves the County’s 24 other departments, over 2,700 County employees, and a population of more than 700,000 residents through centralized administration and support of information technology (IT) resources. As an internal services provider, DIS is headed by the Director of Information Services and is structured into three divisions—Administration, Applications, and Systems. Together these divisions carry out DIS’s core responsibilities, including:

- Delivering business functionality to County departments and offices through projects and upgrades, including the design and implementation of IT solutions; project management; and support for database and geographic information systems.
- Designing, installing, managing, and optimizing county data network components, servers, and databases with over 700 application services delivered and maintained.
- Addressing the need to keep current services and infrastructure functioning smoothly and continuously, including: maintenance of operations; and system security and stability.
- Staffing the service desk, including response to service requests for a wide range of changes, interruptions, or failures.
- Providing and maintaining information technology security, including protection from cyber threats and compliance with mandatory security requirements (such as HIPAA\(^1\) or CJIS\(^2\)).
- Planning for future IT improvements.
- Coordinating the delivery of copy, mail, imaging, records management, and Public Records Requests.

A.1. Department of Information Services’ Budget

To carry out these responsibilities, DIS employed between 83 and 91 full-time equivalent (FTE) positions from 2012 through 2015, which included five limited term FTEs. DIS also managed a budget ranging from $16.6 million to $19.2 million during the same period. DIS’s budget involves capital (Fund 315) and operating (Fund 505) components. The primary funding source for DIS’s budgets is Information Technology Rates, which are charged to the County’s departments according to a rate schedule. In some cases, individual departments contribute funds to DIS’s operating or capital budget beyond IT rates for projects or personnel that are typically not enterprise-wide. In other cases, individual departments fund IT capital projects and personnel that are not in the DIS budget; the scope of these projects and personnel are typically limited to the individual departments. The capital and operating budget for DIS, which does not

\(^{1}\) The federal Health Insurance Portability and Accountability Act of 1996 (HIPAA) requires special security requirements for personal health information that is stored electronically.

include IT projects and positions funded solely in other departments’ budgets, over the past four years is presented in Table 1.

Table 1: DIS Budget Expenditures, 2012-2015

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital Budget (Fund 315)</td>
<td>$1,542,383</td>
<td>$2,617,482</td>
<td>$1,638,269</td>
<td>$2,058,000</td>
</tr>
<tr>
<td>Operating Budget (Fund 505)</td>
<td>$15,008,071</td>
<td>$14,885,452</td>
<td>$17,591,299</td>
<td>$15,750,030</td>
</tr>
<tr>
<td>Total</td>
<td>$16,550,454</td>
<td>$17,502,934</td>
<td>$19,229,568</td>
<td>$17,388,299</td>
</tr>
</tbody>
</table>

Source: DIS draft Information Services Plan (ISP), May 2015

At more than $9.6 million, personnel costs (i.e., salaries and benefits) represents half of DIS’s annual operating budget expenditures in 2015, as illustrated in Figure 1.

Figure 1: Department of Information Services Expenditures (Fund 315 and 505)—2015

Source: DIS draft 36-month business plan, May 2015

From 2012 to 2014, data storage related expenditures consumed varying amounts of the DIS annual budget. Costs tend to spike upward during lifecycle upgrade and replacement years, such as replacements caused by differed maintenance, and drop during periods of baseline maintenance and intermittent capital improvements. Over the three-year period, the cost for data storage (including Tier Three disaster recovery) comprised an average of approximately 17 percent of DIS’s capital costs, as illustrated in Table 2.
Table 2: Actual Capital Infrastructure Expenditures, 2012-2014

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>Sub-Totals</th>
<th>Percent of Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Center Infrastructure</td>
<td>$8,394</td>
<td>$14,301</td>
<td>$0</td>
<td>$22,695</td>
<td>1%</td>
</tr>
<tr>
<td>Disaster Recovery (Tier Three – Mirror 2)</td>
<td>$11,972</td>
<td>$192,530</td>
<td>$0</td>
<td>$204,502</td>
<td>6%</td>
</tr>
<tr>
<td>Network Infrastructure</td>
<td>$10,098</td>
<td>$81,306</td>
<td>$397,779</td>
<td>$489,183</td>
<td>14%</td>
</tr>
<tr>
<td>Servers and Systems</td>
<td>$238,965</td>
<td>$170,439</td>
<td>$181,033</td>
<td>$590,437</td>
<td>17%</td>
</tr>
<tr>
<td>Storage Infrastructure (Tier One and Tier Two, includes Mirror 1)</td>
<td>$65,078</td>
<td>$78,276</td>
<td>$238,167</td>
<td>$381,521</td>
<td>11%</td>
</tr>
<tr>
<td>Workstation Assets Replacement</td>
<td>$580,968</td>
<td>$371,187</td>
<td>$428,170</td>
<td>$1,380,325</td>
<td>39%</td>
</tr>
<tr>
<td>Audio Visual</td>
<td>$111,225</td>
<td>$86,777</td>
<td>$9,205</td>
<td>$207,207</td>
<td>6%</td>
</tr>
<tr>
<td>Other</td>
<td>$76,558</td>
<td>$173,384</td>
<td>$48,403</td>
<td>$298,345</td>
<td>8%</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$3,574,215</td>
<td></td>
<td></td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: DIS Infrastructure Replacement YTD Expenditure Reports, 2012-2014

A.2 Program Overview

Information technology (IT) systems that support a large user base, such as the employees and residents of Snohomish County, involve many types of data intended for a wide variety of purposes. Data can be stored in several ways and the method of data storage depends largely on its type and purpose. In simple terms, data can be grouped into the following three categories, herein called “Tiers”:

- **Tier One: Application-Centric (or Structured) Data Storage.** Structured data refers to a set of data for which each datum is tagged with one or more identifiers. This “structuring” allows for efficient storing and then later analyzing or recalling of data in a query or report. Traditional forms of structured data include relational databases and spreadsheets, although newer technologies allow for more types of data to be structured. To have the data readily accessible for frequent reads and writes with little delay or error, structured data requires high performance—Tier One—storage. Examples of the County’s Tier One data include the data associated with the Ascend and Cayenta financial systems and with payroll and timekeeping data in the human resource management system, Highline.

- **Tier Two: Unstructured Data Storage:** Unstructured data refers to essentially all other data that is not structured. Examples of unstructured data in the County include collaborative working files (e.g., Microsoft Word or Excel, PDFs, and other enterprise documents stored on personal and shared drives), email and other archives. Tier Two
data can also include data with special security requirements such as HIPAA Data and CJIS Data.

- **Tier Three: Backup Data Storage:** Backups include data stored for the purpose of continuous business operations (herein “Mirror 1”) and disaster recovery (herein “Mirror 2”).
  
  - **Mirror 1.** The County’s continuous business operation backup is a secondary data center located at Snohomish County’s Everett Campus that allows workflow functionality to continue despite hardware failure—such as, if an email server goes down the secondary email server can be brought online immediately to avoid service disruption.
  
  - **Mirror 2.** The County’s disaster recovery backup is located in a more remote data center that is not designed to fully operate County systems, but is instead designed to store complete copies of all servers, systems, and data to mitigate the risk of long-term service disruption and data loss in the event of a catastrophic event at the operational data centers.

Historically, organizations with significant IT systems, such as the County, maintain large data centers. These facilities require many components—such as controllers, switches, firewalls, hubs, routers, servers, and data storage devices—that are interconnected on expansive networks. As technology needs grow and evolve, so do these networks. As equipment and applications become obsolete or are no longer maintained by the vendors, migration to updated components and approaches is needed.

**Figure 2: Where Data Resides in the County**

*Source: Auditor Generated based on observation and DIS representations.*
Figure 2 presents a simplified visual representation of where County data is stored, illustrating the general location and movement of data across the County IT infrastructure (as of April 2015).

The County’s Tier One data were nearly exclusively stored within its data center—with relatively few applications (about 14%) hosted by third parties in the cloud. Also, nearly all of Tier Two data were exclusively stored in this data center. DIS has a service agreement with the cloud storage provider, Box, which augments the County’s ability to make collaborative work documents accessible from any computer by adding about 0.033 terabytes\(^3\) (TB) in cloud-based storage space. The County maintains its Tier Three in separate DIS-operated data centers located at different County-owned facilities.

In addition to the primary and mirror data centers, there are three additional networks operated on County-owned hardware but not supported by DIS. For security purposes they are not connected to the general County network or the internet. According to DIS, these off-network systems include the water treatment and distribution system, a secured network mandated by the FBI and the State, and an internet solely for devices and environment management. Business needs of these systems require constraints on access, so they are not suitable for a cloud solution which by definition requires access by internet.


The advent of cloud-based application hosting and data storage presents an alternative delivery model of IT services (networks, servers, storage, applications, etc.). Conceptually, cloud-based services and applications can be rapidly acquired and released from a third-party vendor, and affords elasticity of scale. It also provides a model whereby a customer can expand or reduce “the customer’s systems” depending upon current and expected needs and can pay for only what is used, thereby significantly reducing costs associated with excess capacity. Cloud-based models offer alternatives that include the ability to purchase applications that are entirely administered on an external network (a hosted application), the ability to migrate a currently owned-application to a new platform entirely on an external network, and the ability to store data on external networks.

The costs of cloud solutions depend on many factors including the volume of data, the frequency and speed requirements of access, and the durability of the storage (i.e., resistance to data loss). Research reveals that cloud costs per TB have decreased significantly between 2009 and 2014, as illustrated in Figure 3.

---

\(^3\) One terabyte is equal to 1000 gigabytes (GB). One terabyte (1TB) is a measure of data storage volume equivalent to approximately 1,400 CDs, 250 DVDs, or 250 4GB USB flash drives.
While cloud-based storage costs have declined from almost $200 per TB per month in 2009 (or $2,400 per year) to about $50 per TB per month in 2014 (or about $600 per year), according to cloud service provider pricing guides, the “per TB” cost of cloud storage includes several related cost variables, including:

- The cost of operations performed on the data by viewing and editing it.
- The cost to download data from the cloud that some cloud storage providers charge.

Further, research suggests that additional costs for cloud solutions could require County upgrades to internet service. According to DIS, the decision was already made in 2014 to upgrade the County’s internet access driven by factors including the ability to host more applications in the Cloud, migrate to a network based phone system, the ability to provide uninterrupted access to email and personal files on the cloud, and the preparation for moving disaster recovery functions offsite. The additional costs of these internet upgrades were not analyzed as part of this audit because the required metrics to calculate the increase in internet traffic were not available (as further explained in the Tier One discussion, section C.2.a). These costs include:

- The cost of redundant internet service providers to ensure maximum available up-time to access cloud data through the internet.
- The cost of raised data caps for internet service plans to account for the increased traffic of data uploads and downloads.

Additionally, a cloud solution’s ability to meet the County’s business needs will vary—possibly providing greater accessibility or posing a security weakness. These business needs should also be taken into account when determining the true cost of a potential cloud solution.

A.4. Cloud Computing within the DIS “ISP”

Snohomish County Code, Section 2.350.070 Information Services Plan, requires that the DIS “director shall prepare and submit a 36-month information services plan” (ISP) that shall be
reviewed and considered for comment by the County’s interdepartmental Information Technology Advisory Committee (ITAC). The Executive and Council consider the ISP as part of preparing and adopting the County’s annual budget. Among other required elements, the ISP is required to identify opportunities to improve IT service delivery and reduce IT costs.

In May 2014, DIS issued its 2015–2017 (ISP), in which DIS outlined specific cloud-based solutions included in its implementation plans or slated for testing between 2015 and 2017, such as migrating Tier Two data—email, SharePoint services, and two of the County’s network drives used for working documents (the ‘P’ and ‘S’ drives)—to Microsoft Office 365, allowing for the collaboration and sharing of information with broader accessibility. The ISP also indicated that DIS will consider the use of cloud services to augment the County’s current model for Tier Three data—system recovery, backup, and data durability (i.e., prevent the loss of data).
B. Scope and Methodology

On February 11th, 2015 Snohomish County Council passed Motion No. 15-064 approving the 2015 Audit Plan and directing Council staff to issue a Notice to Proceed to the Interim Performance Auditor—George Skiles of Sjoberg Evashenk Consulting (SEC)—to complete a performance audit of the County’s IT data storage methods and costs (Audit Topic 6 in the 2015 Audit Plan). The scope of this audit included an evaluation of data storage practices and costs from 2012 through the end of audit fieldwork with the objective of determining whether alternative data storage protocols could be more cost-effective than existing protocols. To meet this objective, the SEC audit team performed the following procedures:

B.1 Interviewed key DIS personnel to identify existing methods of storing, backing-up, and securing data, and identified key factors that contributed to the County’s decision to implement the existing data storage methods.

B.2 Reviewed policies and procedures related to the County’s data storage, DIS 36-month ISP (2012–2015), and technical standards.

B.3 Identified DIS’s inventory of existing systems and existing backup protocols. This included the extent to which the systems are integrated on the County’s network, including systems requiring specialized security measures such as those involving justice partners or those containing personal or medical information, and other factors related to existing data storage methods.

B.4 Identified operations, maintenance, and labor costs to support existing data storage methods. This included a review of DIS’s budget, fiscal records (Funds 315 and 505), labor management reports, expenditure summaries, and technical replacement plans to identify costs.

B.5 Identified potential costs of alternative data storage cloud solutions by using vendor pricing guides and online calculators.

B.6 Identified alternative data storage methods through discussion with a limited selection of counties and through research of best and leading practices. In addition to contacting selected counties, this research included a review of counties’ strategic business plans, budget presentations, and other informational materials including industry research on cloud storage and related topics.

B.7 Compared the implementation and use of alternatives to existing data storage over time. Compared the use and establishment of alternatives (including using “cloud storage technologies”) to existing data storage over time.

Audit fieldwork was performed between March 2nd and May 15th. On May 18th, DIS management was provided a draft of this report and, on May 26th, SEC discussed the report findings and recommendations in an exit conference with DIS management. Responses and input provided by DIS were considered and incorporated where applicable in the final report. DIS management was in general agreement with the conclusions and recommendations of this report. DIS’s formal response is included in Appendix B of this report.
We conducted this performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our conclusions based on our audit objectives.
C. Audit Observations and Conclusions

Through its this strategic planning process, DIS noted in 2012 a likely significant increase in data storage needs estimating that by the end of 2015 it would need the capacity for 1500 terabytes (TB) of storage space (with 30TB dedicated solely to Public Records Request storage). According to DIS this projection was based on the business expectation that the County would deploy body-cams to the sheriff deputies, which was not done. Actual storage needs increased from 120TB in 2012 to 495TB in 2015 with 1.9TB dedicated to public records requests and 41.9TB dedicated to legal holds. While actual growth was lower than expectations, DIS’s approach to accommodate this growth in an affordable way was to adopt two significant changes at its data centers:

- First, in 2008, DIS decided to consolidate much of its server environment by purchasing what is known as Virtual Machines (VM)—which can be vastly expanded—and migrate applications from aging stand-alone servers to this equipment.

- Second, and in conjunction with the transition started in 2008, in 2012 DIS implemented a data hierarchy. That is, instead of using a high-performance storage system for all data, the solution was to use cheaper data storage for unstructured data that would not be affected by lower-performance hardware.

The related initiatives to migrate to VM and to implement a new hierarchy of data storage devices are nearly complete and each is expected to have about a six-year life-span. This audit noted that in the draft 2016–2018 ISP, DIS outlined specific cloud-based solutions either expected for initiation or slated for testing over the next three years. Also, the ISP indicated that DIS will consider the use of cloud service to augment the County’s current model for system recovery, backup, and data durability.

Although these efforts will likely meet the County’s near-term data storage needs, this audit revealed that alternatives exist. Among the alternatives to be considered, the audit noted potential efficiencies with each tier, including shifting to hosted applications instead of DIS-administered applications, migrating unstructured data to a variety of cloud services suited for the differing uses and sensitivities of Tier Two data, and considering cloud-based disaster recovery to better enhance geo-diversity. As discussed in the remainder of this report, each option offers benefits and potential cost savings, but must be evaluated by application or by type and anticipated use of data.

C.1. Current Data Storage Strategies

Key to a discussion related to data storage is the understanding of what those terms entail. Data storage is basically all places where data or bits of information reside. The frequency of use or retrieval, speed of accessing the data, type of data, and sensitivity (such as personal health information) all have a bearing on how this data should be stored. To address these differences in types of data, DIS employed a hierarchical storage system that separates data storage based on “performance needs”, such as how quickly the data are needed and how often the data are used. As of April, 2015, DIS’s actual storage needs reached about 495TB of data, about one third of the volume estimated back in 2012.
Although the volume was lower than the projected levels, there was approximately a 300 percent growth from 2012 to 2015. This growth can be primarily attributed to the rapid increase in “unstructured data” that is a result of the increasing need to store public records requests and information related to legal matters, the rapid increase and use of video and audio files, the general transition from paper documents to electronic documents, and the decision to transition from tape backups to disk backups. Due to this growth, the vast majority of the County’s data consists of unstructured data, as illustrated in Figure 5.

**Figure 5: Volume in Terabytes of Tier One versus Tier Two Data**

![Figure 5: Volume in Terabytes of Tier One versus Tier Two Data](source)

*Source: Department of Information Services, April 2015*
Fundamental to the discussion of data storage are the key “tiers” of data managed by the County. The following discussions of the County’s data storage approach and options for the future will be assessed by each of the three tiers of data.

C.2. Tier One—Application-Centric Data

Tier One data is frequently accessed structured data with high-performance needs. Each application which accesses this data is housed by the County on one of two platforms. A significant majority of County applications (302 of 343 applications, or 88 percent) was maintained and administered on county-owned servers in the primary data center supported by DIS. A small number of applications (41, or 12 percent) were purchased from third-party service providers that host these applications on servers outside of the County’s network and accessible through the internet—i.e., in the cloud. Both are discussed below.

- On-Site Servers and Storage Area Networks (SAN): In the past, DIS servers existed on individual physical machines, as many as 120 or more of them, each of which required maintenance and support. Since 2008, DIS has methodically migrated servers on most of these machines to newer “virtual machine” (VM) technology. Through this VM technology, DIS can house 60 individual servers on a single VM that supports numerous and varied applications. As of May 2015, DIS maintained 210 servers on four VMs. Unlike the stand alone servers that housed applications along with the related data on which DIS previously relied, this consolidated virtualized environment is intended to provide the platform on which multiple servers could operate but was not intended to store application-specific data. Instead, data used by these applications are stored on “fast access” storage hardware, specifically, Hewlett Packard 3PAR. Figure 6 illustrates DIS’s new Tier One data storage infrastructure, including its shift from more than 120 servers to a virtualized operating environment and centralized data storage.

![Figure 6: DIS Application Servers and Tier One Data Storage](source: Auditor-generated based on observations and DIS representations.)
The consolidation of stand-alone or individual servers to the virtual machines yields several benefits, including reducing the number of physical machines, reducing wasted processing capacity and storage that would typically remain underutilized on each of multiple machines, and requiring less physical floor space and less cooling—each of these elements generally contribute to lower IT costs. DIS remains in the process of upgrading and migrating to the VM and anticipates completing these initiatives by June 2015. After this time, the County will have as many as 70 percent of servers and applications hosted on VMs. According to DIS virtual servers are easier and much less expensive to migrate to an offsite disaster recovery location.

- **Cloud-Hosted Applications:** The County’s cloud-hosted applications are accessible through the web and include solutions for both public-facing services as well as internally used applications. Public facing applications include employee recruitment (NeoGov), building permits (MyBuildingPermit.com), crime mapping (Citizen Crime Mapping), campsite reservations (Reserve America), and all services accessed through the County’s hosted website (such as the bid listing). Internal applications are those used by County employees such as for time keeping (KRONOS Time Entry), infrastructure management (StreetSaver), and police training record management (PoliceOne). The County does not actively track the volume of data associated with these applications because this “marginal volume of data” has no impact on the cost of service.

In terms of County capacity for data storage, the audit determined the capacity of application-specific data stored at the data center on 3PAR. As illustrated in Figure 7 approximately 37.4TB of the 39.2TB, or about 95 percent of the 3PAR structured data storage infrastructure was in use as of April, 2015—very near to maximum capacity. However, 17.4TB of the volume on 3PAR was utilized by the VM for virtualized servers. DIS plans to free the 17.4TB of space with the 2015 acquisition of DataOn hardware, which will provide the platform on which its virtual machines will operate.

![Figure 7: Volume of Tier One Data on 3PAR](image)

*Source: Department of Information Services, April 2015*
Since 2012, DIS has spent nearly $800,000 to deploy and maintain this approach to Tier One data storage, as illustrated in Table 3. These costs include over $330,000 in capital expenditures, more than $291,000 in operations and maintenance costs, and an additional $177,500 in labor costs. A total of 80TB purchased—40TB for Tier One and 40TB for Mirror 1. Using a five-year projection based on historical averages for labor and intermittent capital improvements plus the five-year amortization of the 3PAR upgrade, the average annual cost of Tier One data storage was estimated at $3,509 per TB per year, or a cost of close to $140,350 per year for the primary data center’s 40TB-capacity of Tier One data storage. Actual Tier One infrastructure costs are summarized in Table 3:

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>$62,908</td>
<td>$27,363</td>
<td>$239,749</td>
<td>$330,020</td>
</tr>
<tr>
<td>Maintenance</td>
<td>$83,255</td>
<td>$110,735</td>
<td>$97,382</td>
<td>$291,372</td>
</tr>
<tr>
<td>Labor</td>
<td>$74,080</td>
<td>$47,860</td>
<td>$55,593</td>
<td>$177,533</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$220,243</strong></td>
<td><strong>$185,958</strong></td>
<td><strong>$392,725</strong></td>
<td><strong>$798,925</strong></td>
</tr>
</tbody>
</table>

*Source: Auditor Generated based on expense reports, fiscal records, and Labor Distribution Reports provided by DIS.*

C.2.a. Cloud-Based Alternatives to the Tier One Model Exist

As a result of its migration from individual servers to virtual servers, and the related consolidation of all application-centric data to the 3PAR storage, DIS established an effective method of maintaining sufficient storage capacity through the system’s lifecycle, approximately six years. While this approach has resulted in a system that meets the DIS established needs, alternatives exist for future consideration. Since application-centric data are dependent upon the applications that access the data, consideration of Tier-One data storage alternatives requires an evaluation of the applications themselves. As a general rule, the data should be kept on a similar platform to that of its application or program. That is, it is not an industry best practice to operate a program or application on-premises and maintain the related data in a cloud environment. Because of this, there are two primary cloud-based options available to the County. These include:

- **Software as a Service (or SaaS).** A third party service provider hosts an application—such as Oracle’s Salesforce—and sells “licenses” for the application to its customers (e.g., the County). Licensing an application is much like leasing as it eliminates the need for the County to purchase and house the application in its data center, maintain storage space for the application, administer the application, and purchase maintenance. Data storage associated with the application typically constitutes a marginal cost compared to the overall license fee. Therefore, DIS does not usually evaluate the cost of Tier One storage when making comparisons for Tier One cloud solutions. As of April 2015, the County utilized a SaaS approach for a small number of its applications, only 41, or 12 percent.
 Costs associated with SaaS solutions are primarily associated with acquiring new or replacing existing information systems through new license agreements, and thus can only be determined through issuing requests for proposals (RFPs) and/or negotiations—a process that can only occur on an application-by-application basis and as needs arise. In such a case, the cost of data storage is only one of many factors that require consideration, but a County-wide strategic plan that addresses cloud storage could guide such considerations.

- **Platform or Infrastructure as a Service (PaaS or IaaS).** This alternative allows a business or government agency that already has commercial off the shelf (COTS) or other home-grown systems to lease the required server space in the cloud to run the applications and store the related data. For example, DIS pays CivicPlus to host various County websites built and administered by its employees. This option requires fewer County staff to maintain the storage infrastructure but also contributes to less immediate control over the data and infrastructure. Using this type of external service requires DIS to routinely monitor cloud usage and the level of service needed in order to ensure that excess costs are minimized and delivery of data access, security, and durability is satisfactory.

Costs associated with a PaaS or IaaS solution are dependent on a wide range of highly variable factors, including the desired operating system (Microsoft Sever 2012 versus Linux), number of dedicated CPU cores to be used, the required gigabytes of RAM, the number of hours per day the platform will operate, level of activity (i.e., number of times data was opened and edited), and/or the potential need to upgrade the application with internet-compatible interfaces.

Based on audit research and inquiries, it appears common for IT agencies to lack this type of information unless cloud-based solutions for Tier One storage are under active consideration. In part, this is due to the fact that gathering some of this data requires network performance monitoring and auditing, a process that could potentially have negative impact on system performance as a whole. Nevertheless, this key information was not gathered or evaluated, nor was the cost-benefit of such an approach analyzed, prior to DIS’s acquisition of a new Tier One data storage infrastructure in 2014 and 2015.

Given DIS’s recent acquisition of an entirely updated Tier One storage infrastructure, DIS has sufficient Tier One capacity for the foreseeable future, suggesting there is no immediate cost-benefit to shifting Tier One storage to the cloud. Despite this, to facilitate future considerations, this audit attempted to gather sufficient information that would allow for a reasonable cost estimate for a PaaS platform solution. However, because of the complexities described above and the amount of time needed to generate the data, DIS was not able to gather the sufficient data for the auditors to conduct the cost analysis.
C.2.b. Existing Hardware and Software Life Cycles Suggest Software and Platform Alternatives Should Remain a Consideration in Future ISPs

DIS has established a standard six-year life cycle for most of its data storage infrastructure components. With the purchase of 3PAR in 2014 and the recent acquisition of DataOn in 2015, DIS projects that the County will have sufficient Tier One data storage capacity through 2020. Before this horizon, DIS will face the decision to begin the process of reducing, on an incremental basis, the volume of its on-site storage in favor of cloud alternatives, which has the potential to reduce future Tier One storage costs.

Similar to hardware, applications also have finite life cycles after which their functionality and usefulness may no longer meet the changing needs of the County and, in the case of purchased systems, may no longer be supported by the vendor. The life cycle for each of the 302 county-administered applications varies, with some applications already at the end of their effective life cycle. These looming situations create continuous decision points for DIS; critical decisions should be made to determine the best course of action on an application-by-application basis. For instance, in the 2015–2017 ISP, DIS indicated that a new human resources system for payroll and timekeeping was needed. As each aging application reaches the end of its usefulness, the County must consider whether to upgrade the existing application or procure a new solution; if an updated or new application will operate in the DIS data center or be hosted in the cloud; and whether the application should be developed in-house, purchased as a COTS system, or procured as a third-party service. Each of these considerations impacts the method and cost of Tier One data storage, and DIS’s overall approach to IT service delivery.

According to DIS, a cost benefit analysis is performed for each project. However, DIS was not able to provide sufficient documentation of this analysis. Therefore, DIS should incorporate into its ISPs (or other formal documentation) fully developed evaluations and analyses of applications as they reach the end of their useful life. These evaluations should include detailed analysis of options, application and supporting hardware life cycles, and cost factors (programming, implementation, migration, administration, and ongoing maintenance, storage volume). These deliberative analyses should be conducted in coordination with the application “owners” or key users and should become a standard element in each ISP moving forward. Further, DIS should begin to strategically evaluate how it will ensure cost-effective Tier One data storage. This cost benefit evaluation should commence with its 2016 ISP to ensure that by 2019, when the County’s data storage infrastructure will reach the end of its useful life, the County is well postured. These efforts should include an evaluation of SaaS and PaaS as an alternative to procuring a new on-site Tier One data storage infrastructure.

C.3. Tier Two—Unstructured Data

The entirety of the County’s permanent storage of Tier Two data was maintained on county-owned media located in County-owned facilities. Recognizing that some of this data did not require the expensive storage hardware that would allow the rapid access required for databases and other structured data, DIS procured lower-cost hardware—by Coraid—for its Tier Two unstructured storage.

As illustrated in Figure 8, approximately 70TB of the 214TB capacity of Coraid, or nearly 33
percent, remains available for future storage of unstructured data. Of the nearly 144TB stored on Coraid, 99.9TB, or 69.5 percent, consists of shared working documents, including collaborative working files, HIPAA-protected data, CJIS-restricted data, archives (aged files and email archives), PDFs, and photos. This data represents files that continue to be used on a daily basis (e.g., working files on County shared drives) as well as those less frequently accessed such as archived emails.

Figure 8: Volume of Tier Two Data on Coraid

Since 2012, DIS invested more than $476,000 building and expanding its Tier Two data storage infrastructure, with $257,500 (54 percent) relating to capital costs and nearly $176,000 for related labor costs. Tier Two storage capacity includes 214TB at the primary data center as well as about 150TB of Tier Two data at Mirror 1. Using a five-year projection based on historical averages for labor and assuming there will be no additional capital upgrades plus the five-year amortization of actual capital and maintenance, the average annual cost for the 214TB-capacity Tier Two data storage in the primary data center is approximately $70,600, or $330 per TB per year. Actual Tier Two data storage costs are summarized in Table 4.

Table 4: Tier Two Unstructured Data Storage Costs, 2012-2014

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>$99,256</td>
<td>$123,612</td>
<td>$34,659</td>
<td>$257,527</td>
</tr>
<tr>
<td>Maintenance</td>
<td>$9,979</td>
<td>$21,204</td>
<td>$11,749</td>
<td>$42,932</td>
</tr>
<tr>
<td>Labor</td>
<td>$74,080</td>
<td>$56,948</td>
<td>$44,899</td>
<td>$175,927</td>
</tr>
<tr>
<td>Total</td>
<td>$183,315</td>
<td>$201,764</td>
<td>$91,307</td>
<td>$476,386</td>
</tr>
</tbody>
</table>

Source: Auditor Generated based on expense reports, fiscal records, and Labor Distribution Reports provided by DIS.
C.3.a. Cloud-Based Alternatives to this Approach Exist

The current approach to data storage for Tier Two data is meeting the DIS established needs for these files. The current Coraid system was purchased in 2012 and 2013; on a six-year lifecycle the next major upgrade will take place around 2018. While the original vendor, Coraid, is no longer in operation, DIS was able to procure maintenance service through a new vendor for the duration of the system’s lifecycle. Generally, the cost for Tier Two cloud storage is advertised in the “per gigabyte (GB) per month” format. However, depending on the service provider, there may be additional costs for accessing, modifying, and downloading the data. Similar to the conclusions relating to Tier One storage, a number of alternatives for Tier Two data exist for future consideration. The cloud solution differs for the types of data specified below.

- **Working documents.** Unlike some cloud services, the audit found that storage solutions for working documents were not “pay-as-you-go”. That is, typically, service providers charged “per user” and not “per volume”. For example, Amazon WorkDocs pricing is five dollars per user per month with each user able to store 200GB (0.2TB). As it related to Snohomish County, moving employee files to a service like Amazon WorkDocs would cost an estimated $13,500 per month for 2,700 employees (or $162,000 per year) based on its published costing model, and it would allow a total of 540TB—much more than the County currently uses. The cost breaks down to $300 per TB per year, slightly less than the $330 per TB per year for the existing Tier Two storage infrastructure.

Google and Microsoft offer similar cloud storage plans for working documents which also start at $5 per month per user. However, each service provider bundles the plan with slightly different services, such as email through Gmail or Outlook. Table 5 shows a basic comparison of cloud services designed for storage of personal working documents.

<table>
<thead>
<tr>
<th>Provider</th>
<th>Google web apps including Gmail, Sheets, Docs, Calendar, Slides, and video calls.</th>
<th>Microsoft Office Online including Outlook, Excel, Word, and video conferencing.</th>
<th>Supports most common file formats, including .doc, .docx, .xls, .xlsx, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provider</strong></td>
<td>Google</td>
<td>Microsoft</td>
<td>Amazon</td>
</tr>
<tr>
<td><strong>Cost per User per Month</strong></td>
<td>$5</td>
<td>$5</td>
<td>$5</td>
</tr>
<tr>
<td><strong>Storage Space per User</strong></td>
<td>0.03TB</td>
<td>1TB</td>
<td>0.2TB</td>
</tr>
<tr>
<td><strong>Cost per TB per Year</strong></td>
<td>$2,000</td>
<td>$60</td>
<td>$300</td>
</tr>
</tbody>
</table>

*Table 5. Tier Two Cloud Service Breakdown for Working Documents*
At the time of audit fieldwork, DIS had already secured licenses for Microsoft Office 365 and planned to roll-out the cloud storage solution to all County employees by 2016. The phased implementation intends to allow time to establish standards of use and to provide training and support to employees while they learn the new system. The roll-out for cloud storage includes employee personal drives and department shared drives (‘P’ and ‘S’ drives), which DIS estimated to be 10TB of its unstructured data—or, approximately 7 percent of the 144TB currently stored on Coraid—at no additional cost. However, this does not include other working documents such as other network drives, including ‘G’, and various departmental drives. If DIS were to migrate these other working documents to Office 365 or other available products in the Office 365 suite such as SharePoint, it could continue to reduce on-site Tier Two storage by 80TB in addition to the 10TB already planned.

- **Cold Storage.** The County may be able to store infrequently accessed data, such as archived data, in online “cold” storage. As indicated by DIS, only a small portion of unstructured data is accessed on a daily basis. Although an exact statistic was not available, DIS estimated that about 80 percent (about 120TB) of unstructured data were not regularly used. Depending on the exact business needs for this data, and the frequency with which it needs to be accessed, cold storage may provide an inexpensive alternative. Cloud providers typically designate “cold” storage options for data that are infrequently accessed. As a lower-cost option, there may be increased latency or short delays in accessibility; however, this does not impact the durability—or resistance to loss—of the data. Cold storage was offered by Google “Nearline” and Amazon “Glacier”—both of which started at $120 per TB per year, but with additional charges for viewing and editing the data (with Google) or for deleting the data sooner than 90 days (with Amazon). Metrics needed to determine the portion of data, if any, that would be suitable for cold storage were not tracked by DIS; similar to the explanation in C.2.a, it appears common for IT agencies to lack this type of information unless cloud-based solutions are under active consideration.

- **Public Access Documents.** Data stored on the County’s network is not generally available to the public. To comply with public records and legal requests for data, DIS makes available certain sections of County data in various formats. In one particular case, DIS conducted a pilot study with Microsoft Azure. Specifically, when a large amount of public records files related to the Oso mudslide were being requested multiple times, DIS made a business decision to reduce the staff costs of repeating the requests multiple times by procuring the cloud-based storage service Microsoft Azure to host the data. According to DIS the use of Microsoft Azure in this instance was not cost-effective based on the number of views the data received. However, DIS was not able to provide documented analysis to support this claim. Despite this, Cloud service providers make various alternatives available that should continue to be considered as potential methods to make records available to the public in the future.

Even with the aforementioned options, there may still be data remaining on County storage—at least in the near-term—which may include legal holds and public records holds. Further, a cloud storage solution would need to meet any special security requirements—such as health records
protected by HIPAA and criminal justice records protected by CJIs—which according to DIS was about 12TB of the data in the County’s primary data center.

C.3.b. Coraid’s Remaining Two- to Three-Year Life Cycle Expectancy Requires that DIS Address the Replacement of Coraid and the Consideration of these Alternatives in its ISP

As described in the previous section, DIS is in the process of migrating a portion of its unstructured data to the cloud without incurring additional costs. DIS had not yet established a plan for the possible migration of the remaining 130TB in unstructured data housed on Coraid including 42TB in legal holds, 2TB in public records, and 90TB of other unstructured data (including department network drives, and HIPAA and CJIS data). With only two to three years of effective life cycle remaining, DIS must evaluate whether accelerated migration of all working documents to Office 365 is warranted. Based on this analysis, Office 365 appears to offer the County a cost-effective alternative to on-site storage of working documents, and presents an opportunity for the County to reduce Tier Two storage costs in the long term.

However, achieving a significant reduction in Tier Two storage costs will require more than a migration of working documents to Office 365. If the other categories of data (e.g., legal hold files, public records documents, archives, etc.) remain on Coraid, DIS may experience a situation where Coraid will be significantly underutilized, with potentially as little as 44TB used on the 214TB-capacity system. By the time Coraid reaches the end of its useful life, DIS will still be in a position to either replace Coraid with another low-cost solution appropriately sized to the volume of legal hold and public records data maintained at the time or with cloud-based alternatives, such as cold storage, that ensure capacity flexibility. The cost-benefit of the partial or full migration of working documents, as well as the alternatives for the remaining unstructured data, should be addressed in upcoming DIS ISPs and/or other formal documentation.

C.4. Tier Three—Continuous Business Operations and Disaster Recovery Backups

DIS established two methods of backing up its Tier One and Tier Two data to recover from any system failure or disaster. The two methods include a process of continuous data replication utilizing a second full-service county-owned data center (Mirror 1) and a nightly, point-in-time incremental backup to a third and more remote data center designed only for disaster-recovery purposes (Mirror 2). As backup systems, the two mirrors are intended for separate functions, as described below.

- Mirror 1—for Continuous Business Operation—is a real-time replica or a redundant system of the primary data center in terms of hierarchical storage. Mirror 1 is a “pseudo-warm site” located in a different County building than the primary data center. Generally in IT nomenclature, a “warm site” refers to a location which operates a real time backup of business data and where staff could immediately move should the original site become unusable or fail in some fashion. While the Mirror 1 facility lacks the physical space to house County staff, it does replicate all the Tier One and Tier Two
data as well as communication servers in the primary data center. Should the primary
data center “go down”, email and the voice over IP phone system would continue to
operate from Mirror 1. Other business critical systems, such as financial, could be
rebuilt in as few as four hours and DIS estimated the complete rebuild to take about
four days.

As previously mentioned, the cost per TB in Mirror 1 "mirrors" the costs per TB for
both Tier One and Tier Two—about $3,509 per TB for approximately 40TB of Tier
One data in Mirror 1 and about $330 per TB for approximately 150TB of Tier Two data
in Mirror 1 for a total of approximately $190,000 for Mirror 1 data storage.

Because Mirror 1 backup is essentially a replica on a real-time basis of Tier One and
Tier Two data, limited alternatives to fulfill this role exist independently from those for
Tier One and Tier Two respectively. Nonetheless, for continuous business operations
backup, options include: replication on servers and data storage in another county-
owned facility, as is current practice; replication on servers and data storage in a
colocation site (“rented space” in a third-party data center); or, replication on cloud-
based servers and data storage through an Platform as a Service (PaaS) or
Infrastructure as a Service (IaaS) model (which again relies on decisions made for Tier
One data). Because DIS has already invested in its primary and Mirror 1 data centers,
which have life cycles up to 20 or more years, colocation does not appear to be a cost-
effective option at this point or in the foreseeable future. However, any data migration
efforts for Tier One or Tier Two will have a corresponding reduction on the amount of
data stored at Mirror 1.

- Mirror 2—for Disaster Recovery—represents a remote backup that is incrementally
updated each night. This remote facility is also on County premises, but is located at a
distance from the County’s primary data center to maintain geo-diversity—or distance
between primary and backup data centers—a key feature of a disaster recovery site.
This second backup is intended to house data needed for restarting County business
should a significant disaster occur and render the primary and redundant data centers
non-functional. Mirror 2 data was stored on a low-cost Coraid system, similar to that
used to store Tier Two unstructured data, which lacks the capability to support Tier One
system functions. Rather, this Coraid system is used to store all Tier One and Tier Two
data located on each of DIS’s servers and data storage devices in a compressed format,
allowing DIS to rebuild a functioning data center if there was a catastrophic event. As
with Mirror 1, there will also be a reduction on the amount of data stored at Mirror 2
corresponding to any data migration from Tier One or Tier Two to the cloud.

In 2013, DIS upgraded its disaster recovery system to switch from tape backup to disk backup,
which included purchasing close to 144TB in Coraid storage for approximately $92,000 and a
total cost of $636,000 in overall upgrades to the data center. Using a five-year projection based
on limited labor costs in the future and assuming no additional capital improvements, this
equates to about $1,131 per TB per year. Actual Mirror 2 data storage costs between 2012 and
2014 are illustrated in Table 6.
Table 6. Tier Three, Mirror 2—Disaster Recovery Costs, 2012–2014

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital (Infrastructure)</td>
<td>$11,972</td>
<td>$100,986</td>
<td>$0</td>
<td>$112,958</td>
</tr>
<tr>
<td>Capital (Storage)</td>
<td>$0</td>
<td>$91,545</td>
<td>$0</td>
<td>$91,545</td>
</tr>
<tr>
<td>Labor</td>
<td>$74,080</td>
<td>$207,958</td>
<td>$149,693</td>
<td>$431,730</td>
</tr>
<tr>
<td>Total</td>
<td>$86,052</td>
<td>$400,489</td>
<td>$149,693</td>
<td>$636,233</td>
</tr>
</tbody>
</table>

Source: Auditor Generated based on expense reports, fiscal records, and Labor Distribution Reports provided by DIS.

C.4.a. Cloud Based Alternatives for Tier Three Data Exist

Alternatives for Mirror 1 are mainly dependent on Tier One and Tier Two migrations. For example, the proposed migration of user-specific Tier Two data will move about 10TB of Tier Two data to the cloud, which means there will also be 10TB less data to be stored on the Mirror 1 back-up. This applies to Mirror 2 as well but for a slightly lower volume of data as Mirror 2 data is stored in a compressed format to take up less space.

However there are also direct alternatives for Mirror 2. Alternatives to the County’s approach include utilizing a third-party service provider—such as Barracuda or Carbonite, among others—that specialize in disaster recovery storage. For instance, Carbonite offers a disaster recovery backup service for workstations as well as physical servers and virtual machines. Carbonite software installed on County hardware and would automatically back up data to the cloud. The storage is offered with unlimited sharing between the different devices. The cost would be $1,499.94 per TB per year for the first TB and $999.90 for each additional TB. To store a backup of the County’s data (approximately 150TB as of April, 2015) it would cost $150,484 per year. This is about equal to current County costs of $169,588 based on the $1,131 per TB per year for Tier Three data storage considering a cloud option still requires some DIS labor for testing and monitoring.

Barracuda offers a different backup service model. Barracuda sells hardware that would reside on County premises, and would automatically coordinate the continuous backup of data to the cloud. The on-premises hardware has a limited amount of backup space so one to three Barracuda machines could be required based on the volume of the County’s Tier Three data as of April 2015. The number of machines also depends on the County’s decision between having all data or just 50TB of critical data backed up in an immediately accessible format. The cost of one machine with no pre-paid service maintenance for 50TB of on-site backup and unlimited cloud backup for a year would cost $226,798. For three machines, five years of prepaid service maintenance, five years of updates, the total cost would be $1,862,991. The comparisons for backup costs are summarized in Table 7.
Table 7. Tier Three Disaster Recovery Backup Solutions Comparison

<table>
<thead>
<tr>
<th>Backup Solution</th>
<th>TB on Premises</th>
<th>TB in the Cloud</th>
<th>Cost per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current County Model—Coraid (Remote County premises)</td>
<td>150</td>
<td>0</td>
<td>$169,588</td>
</tr>
<tr>
<td>Carbonite (Cloud)</td>
<td>0</td>
<td>150</td>
<td>$150,484</td>
</tr>
<tr>
<td>Barracuda—zero years pre-paid service maintenance (Hybrid)</td>
<td>50</td>
<td>Unlimited (one year)</td>
<td>$118,798*</td>
</tr>
<tr>
<td>Barracuda—five years pre-paid service maintenance (Hybrid)</td>
<td>150</td>
<td>Unlimited (five years)</td>
<td>$372,598*</td>
</tr>
</tbody>
</table>

Source: Auditor Generated based on expense reports, fiscal records, and Labor Distribution Reports provided by DIS; vendor online pricing guides.
*Note: cost of hardware amortized over five-year period.

The costs of disaster recovery models differ depending on the amount of data stored on County premises versus in the cloud. On-premises storage has the advantage of a faster recovery time, but also has the risk that a catastrophic event could damage both the primary and remote data centers. This is why geo-diversity—or distance between primary and backup data centers—is important. Cloud storage solutions provide the greatest geo-diversity, but with the drawback of a likely longer time until recovery as the result of downloading the backup data through the internet. Further, a cloud storage solution would need to meet any specific business needs for the data, such as special security requirements which DIS estimated. DIS should evaluate the cost-benefit of each of these scenarios to provide strategic insight to the County in upcoming ISPs or other formal documentation.

C.5. Conclusions and Recommendations

Overall, this audit revealed DIS has established a data storage infrastructure with the capacity to meet the County’s needs. At the same time, evidence suggests that cost-effective alternatives for the future exist—particularly in relation to Tier Two, unstructured data—and should be evaluated in future deliberations and strategic planning efforts. Specifically, DIS should:

C.5.a Continue to migrate selected Tier Two working documents to Microsoft Office 365 cloud storage and incorporate in the ISP or other formal documentation the cost-benefit of extending this migration to include all other working documents.

C.5.b Conduct pilot testing of alternatives for the remainder of the Tier Two data held on Coraid, including Legal Hold, Public Records documents, archives, and other types of data not suitable for Microsoft Office 365, with the intention of identifying the most cost-beneficial solutions for each type of unstructured data by the time Coraid reaches the end of its useful life.

C.5.c Evaluate cloud-based alternatives to existing Tier One data storage methods, including:
C.5.c.i. Incorporating into the ISP, or other formal documentation, cost-benefit analyses of potential cloud-based solutions when considering replacing an existing information system or when acquiring a new system. This evaluation should account for the impact of each alternative on on-site data storage and the cost-benefit of hosting the application and all related data to the cloud.

C.5.c.ii. Gathering sufficient system and network data necessary to fully evaluate the cost-benefit of moving Tier One applications and data to a cloud-based platform. This should include (a) determining and evaluating key cost factors, and (b) including the analysis and results in the ISP or other formal documentation.

C.5.c.iii. Including the analyses and results of all past prototype and pilot testing conducted by DIS, and their impact on future planning decisions. The results should explain the cost-benefit of the new service, including whether the pilot should or should not be considered to have a positive return on investment, and should address both financial and non-monetary benefits to each alternative reviewed.

C.5.d Evaluate cloud-based alternatives to existing Tier Three data storage methods, such as Carbonite and Barracuda, among others.

C.5.e Incorporate into each future ISP, or other formal documentation, an analysis that explicitly addresses the cost benefit of contemporary data storage methods as compared to recognized alternatives at the time. In doing so, DIS and County officials will be able to monitor the cost-benefit of its data storage investments as compared to the rapidly changing alternatives available through the cloud.
# Appendix A – Summary of Recommendations and Corrective Action Plan

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Page</th>
<th>Responsible Entity</th>
<th>Priority</th>
<th>Entity’s Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C.5.a</strong></td>
<td>21</td>
<td>DIS</td>
<td>A</td>
<td>This has been and will continue to be part of formal decision practice. There are other factors which will also be a part of that analysis.</td>
</tr>
<tr>
<td><strong>C.5.b</strong></td>
<td>22</td>
<td>DIS</td>
<td>A</td>
<td>This evaluation has already begun, and specific characteristics of the data – such as frequency of access, frequency of reference, ability to search, support for legal retention, and other document management criteria – will be part of this decision.</td>
</tr>
<tr>
<td><strong>C.5.c</strong></td>
<td>22</td>
<td></td>
<td></td>
<td>This is done with every new system acquisition or major upgrade.</td>
</tr>
</tbody>
</table>

Continue to migrate selected Tier Two working documents to Microsoft Office 365 cloud storage and incorporate in the ISP or other formal documentation the cost-benefit of extending this migration to include all other working documents.

Conduct pilot testing of alternatives for the remainder of the Tier Two data held on Coraid, including Legal Hold, Public Records documents, archives, and other types of data not suitable for Microsoft Office 365, with the intention of identifying the most cost-beneficial solutions for each type of unstructured data by the time Coraid reaches the end of its useful life.

Evaluate cloud-based alternatives to existing Tier One data storage methods, including:
<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Page</th>
<th>Responsible Entity</th>
<th>Priority</th>
<th>Entity’s Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.5.c.i.</td>
<td>22</td>
<td>DIS</td>
<td>A</td>
<td>The cost of storage are included in every decision, but are only a part (usually small) of the complete consideration factors.</td>
</tr>
<tr>
<td>C.5.c.ii.</td>
<td>22</td>
<td>DIS</td>
<td>B</td>
<td>The potential cost savings, including the costs of storage, are part of the overall IS strategy. IS also considers the impact in time, money and business disruption in assessing whether to move a current application.</td>
</tr>
<tr>
<td>Recommendation</td>
<td>Page</td>
<td>Responsible Entity</td>
<td>Priority</td>
<td>Entity’s Response</td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>---------------------</td>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>C.5.c.iii. Including the analyses and results of all past prototype and pilot testing conducted by DIS, and their impact on future planning decisions. The results should explain the cost-benefit of the new service, including whether the pilot should or should not be considered to have a positive return on investment, and should address both financial and non-monetary benefits to each alternative reviewed.</td>
<td>22</td>
<td>DIS</td>
<td>A</td>
<td>IS will include the results of pilot testing until there is a significant change in either business needs or technology. At that point, the results of past pilots become less relevant.</td>
</tr>
<tr>
<td>C.5.d Evaluate cloud-based alternatives to existing Tier Three data storage methods, such as Carbonite and Barracuda, among others.</td>
<td>22</td>
<td>DIS</td>
<td>A</td>
<td>[DIS did not provide response to this recommendation]</td>
</tr>
<tr>
<td>C.5.e Incorporate into each future ISP, or other formal documentation, an analysis that explicitly addresses the cost benefit of contemporary data storage methods as compared to recognized alternatives at the time. In doing so, DIS and County officials will be able to monitor the cost-benefit of its data storage investments as compared to the rapidly changing alternatives available through the cloud.</td>
<td>22</td>
<td>DIS</td>
<td>A</td>
<td>The IS Plan forecasts future business needs and analyzes the trajectory of potential technological change and impact. Information Services focuses on making appropriate investments, with costs as one factor. Most IS investments are for a multi-year period, unless business needs or technology needs change.</td>
</tr>
</tbody>
</table>
A – **High Priority**: The recommendation pertains to a high priority conclusion or observation. Due to the seriousness or significance of the matter, immediate management attention and appropriate corrective action is warranted.

B – **Medium Priority**: The recommendation pertains to a moderately significant conclusion or observation. Reasonably prompt corrective action should be taken by management to address the matter. Recommendation should be implemented no later than six months.

C – **Low Priority**: The recommendation pertains to a conclusion or observation of relatively minor significance or concern. The timing of any corrective action is left to management's discretion.

N/A: Not Applicable
Appendix B – Department of Information Services’ Response

Information Services agrees with the majority of recommendations of this report. The Storage landscape is changing on a day to day basis almost. In 2014 online storage pricing dropped to a level that became competitive in the marketplace, evaluating both the cost and the added advantage of GEO-Diversity. That became a benchmark for making future storage decisions. Other business factors also drive value – the ability to continue business operations is just one of those considerations. Moving 10 terabytes of use for P:\ and S:\ storage space has more business values a business resumption tool than any cost savings will.

IS acknowledges the great difficulty in isolating and analyzing part of a very complex environment. To some degree, the focus on cost – while relevant and important – overshadowed the business context in which value is created. IS always focuses on costs, and has a good track record of making strategic investments which saved the county millions of dollars. However, there are many other factors (security, scalability, continuity, stability, business functionality, legal mandates, etc.) which IS considers when evaluating technology purchases. Making a cost evaluation without that context is like deciding to buy a house because you like its yard. Typically, many others factors are considered.

IS believes that every new project and every hardware of software purchase should ask; is it cheaper to buy, build or contract? This also applies to cloud. Is it cheaper to be in the cloud? Even if it is not cheaper, is the business solution’s functionality or availability in the cloud worth the added expense? This has been a standard part of the Information Services purchase evaluation process which requires vendors to say whether their solution is on-premises or hosted or both.

Business needs will drive any movement of applications to a hosted or cloud based environment. Lower costs may be a part of the business needs, but so will the ability to support current business process, adding or replacing functionality, access to data, integration with the other county applications, business continuity, interoperability, vendor capability and services, and risk management. The business decisions and fitness for the intended purpose always have constraints like costs, resources, and time. In the end all IS solutions need to focus on the business and business value. Save where possible, move at the right time and leverage each purchase to its fullest potential.

IS wants to thank the performance auditor staff for their perseverance, dedication to purpose, professionalism, and focus on their objective.