IMPLEMENTING CLOUD STORAGE METRICS TO IMPROVE IT EFFICIENCY AND CAPACITY MANAGEMENT

Contributor

Sudip Chahal
Intel IT
Intel Corporation

Krishnamurthy Anandarao
Intel IT
Intel Corporation

Chris Peters
Intel IT
Intel Corporation

Ajay Chandramouly
Intel IT
Intel Corporation

Shane Healy
Intel IT
Intel Corporation

Nigel Wayman
Intel IT
Intel Corporation

Steve Owen
Intel IT
Intel Corporation

Like many of our enterprise IT peers, we are being challenged with rapid growth in storage demand. In 2011 alone, we faced the need for a 53-percent increase in storage capacity to 38.2 PB from 2010 and have been averaging 35-percent year-over-year growth since 2007. The continued build-out of our office and enterprise private cloud could further increase demand and complexity of our storage environment. Clearly we could not increase costs linearly with demand. Through a variety of techniques described here, including implementation of capacity metrics, improving utilization, thin provisioning, and tiering, we have been able to support significant capacity and performance improvements while saving 9.2 Million US Dollars.

These techniques allow us to meet steep storage demand growth in a cost-efficient manner while not compromising on quality of service in our virtualized and multitenant computing environment.

Introduction

As we migrate to the private cloud, we seek to optimize the efficiency of the storage area network (SAN) environment by increasing storage capacity utilization. We have been further challenged by the simultaneous demand for storage capacity increasing unabated by 35 percent year over year since 2007. As a result, capacity purchases represent a significant and growing business cost. By increasing storage utilization, we can meet storage demand growth while not increasing costs linearly.

We are evaluating a wide array of technologies and approaches to improve SAN efficiency and utilization over time, including thin provisioning. Since many customers use only a small portion of their storage allocation, thin provisioning technology allows IT to over-allocate capacity and thereby increase utilization of storage resources. However, over-allocation also requires improved capacity and risk management so that we have adequate capacity to meet customer requirements.

We therefore recognized a need for enterprise-wide storage capacity metrics that accurately measure storage utilization and allocation across our SAN environment. Over the past year, we have developed and begun refining a storage metrics methodology that could span both our new private cloud and our traditional enterprise environment.

These metrics focus on three related areas: efficiency, capacity management, and risk management. We are building business intelligence and reporting

capabilities based on these metrics. Our storage capacity metrics are designed to:

- Establish a clear link between cost (the purchase of raw capacity) and value (the use of that capacity to store internal customer data).
- Reflect efficiency gains due to new technology adoption and storage efficiency approaches such as thin provisioning and data de-duplication.
- Establish operational thresholds, based on allocation and utilization levels, that alert us when we need to add or reallocate capacity before we jeopardize customer service levels.
- Remain uniform over time, providing us with a consistent view of efficiency across supplier product lines and technology generations.

We are using these metrics to analyze and compare the efficiency of our private cloud at multiple levels, from individual storage pools to a global view.

Custom reports provide an efficiency overview for senior IT managers, help data center managers plan capacity purchases, and enable operations engineers to respond to day-to-day capacity requests.

"Custom reports provide an efficiency overview for senior IT managers, help data center managers plan capacity purchases."

Potential future enhancements include customized risk management thresholds based on a predictive algorithm that can consider factors such as historical growth in customer storage requirements. We also anticipate being able to apply our metrics across our traditional office and enterprise network attached storage (NAS) environment.

As we scale our storage environment and increase utilization, performance health metrics will be essential to understand application and system requirements, identify potential bottlenecks, and assist in successful deployment. Accordingly, we are in the process of incorporating storage performance health into our capacity management and efficiency views.

Business Challenge

Intel IT is undertaking a major transition to an enterprise private cloud that will support our office and enterprise computing applications. This phased, multiyear initiative is designed to enable greater agility and efficiency.

We are building this multitenant environment on virtualized infrastructure as a service (IaaS). This infrastructure is based on clusters of Intel® Xeon® processor-based servers accessing shared pools of storage through storage area networks.

"Infrastructure is based on clusters of Intel® Xeon® processor-based servers accessing shared pools of storage."

As we transition to the private cloud, SAN storage is becoming an increasingly important element of Intel's IT storage infrastructure. Currently, Intel has a total about 50 PB of global storage; Intel's office and enterprise environment accounts for about 13 PB of this. Within this office-enterprise environment, SAN storage is used to support our enterprise private cloud as well as mission-critical and other applications.

"The cost of SAN storage is becoming a significant factor in overall IT costs."

The cost of SAN storage is becoming a significant factor in overall IT costs because SAN storage is a more expensive technology than locally attached storage while a larger proportion of our storage is becoming SAN based. Key factors in SAN capacity growth include server virtualization and optimizing SAN capacity utilization.

Server Virtualization

To build the foundation for our private cloud, we began accelerating the pace of server virtualization in late 2009. This drives incremental demand for SAN storage, because data that was previously stored on physical servers' local drives is now migrated to the shared SAN environment. We are currently 68 percent virtualized as measured by OS instance and are on-track to be 75 percent virtualized by the end of this year.

We are experiencing a rapid growth in storage demand with a 35 percent annual growth rate in raw data requirements due to trends such as increasing use of video and other graphical information, legal retention requirements, larger databases for operational and business intelligence use, and backup needs.

The Value of Optimizing SAN Capacity Utilization

With our increasing use of SAN storage, we saw the need for strategies and tools that would enable us to measure and optimize efficiency by maximizing SAN capacity utilization. Increasing utilization can reduce cost by helping to curb growth in the amount of capacity we need to purchase.

"Due to the large size of the Intel IT environment, even small efficiency improvements can result in significant financial savings."

Due to the large size of the Intel IT environment, even small efficiency improvements can result in significant financial savings because capacity purchases are the largest factor in Intel IT storage costs. Other costs such as operating costs, power and cooling, SAN switches, and cabling represent a much smaller percentage of overall total cost of ownership (TCO).

The Need for Enterprise-wide Metrics

To manage capacity and improve efficiency, we need to be able to accurately measure capacity utilization across our SAN environment. To date, this has been challenging for several reasons:

- Storage suppliers often use different terminology and methodologies for measuring capacity utilization. It is difficult to combine information from multiple suppliers into a single view that reflects Intel IT business requirements.
- Existing metrics do not accurately reflect efficiency improvements achieved using new technologies such as thin provisioning.
- Due to the complexity of the environment and the lack of a standard method for measuring storage utilization, we have found inconsistencies in the way we measure efficiency across different groups and teams within Intel IT.

We decided to implement a single set of metrics that provides standard measures of utilization across our diverse and constantly changing technology base.

Strategy Focus Areas

These enterprise-wide metrics must account for three interrelated focus areas within our storage management strategy: efficiency, capacity management, and risk management. The relationship between these three areas and their importance to different groups within Intel IT are described below and shown in Figure 1.

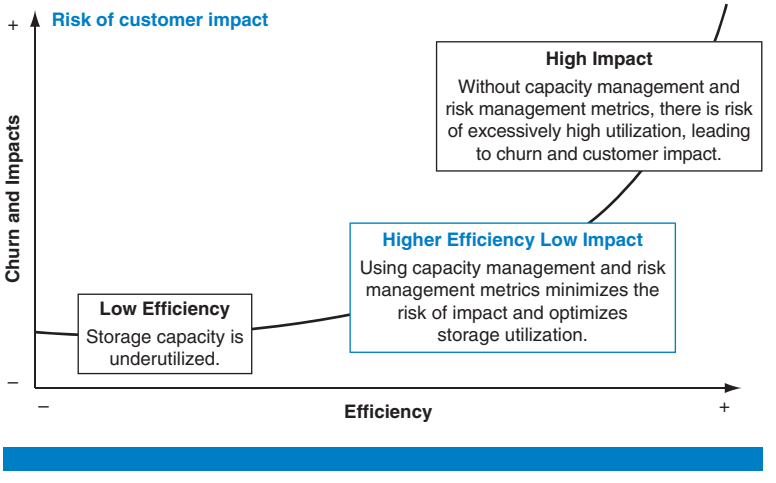

Figure 1: Metrics—efficiency, capacity, and risk management (Source: Intel Corporation, 2012)

Efficiency

Clearly, our goal is to achieve high storage capacity utilization without impacting our internal customers or increasing IT operational costs. Storage efficiency metrics provide visibility into how efficiently our storage capacity is being used to store customer data. Efficiency metrics are particularly interesting to Intel IT executives and senior managers, who view storage efficiency as a key IT indicator.

"Storage efficiency metrics provide visibility into how efficiently our storage capacity is being used to store customer data."

Capacity Management

As we increase efficiency and utilization, capacity management becomes increasingly important. We need to be able to analyze utilization growth rates and accurately anticipate and determine when we need to purchase more capacity. This enables us to make purchases in the most timely and cost-effective way.

Capacity management metrics are used by a variety of people within Intel IT. These include the data center managers, capacity planners, and operations engineers who are responsible for making sure there is enough SAN capacity to meet demand.

"Running out of capacity can result in an unacceptable disruption to business applications."

Risk Management

As we drive storage efficiency and utilization to higher levels, there is an increasing risk of customer impact. In an extreme case, running out of capacity can result in an unacceptable disruption to business applications. Very high storage utilization can also impact application performance.

There is also the potential for increased IT operational costs. To avoid customer impact, Intel IT operations engineers may devote a considerable amount of time to manually migrating data from heavily utilized to less-utilized pools.

These potential issues mean that as we improve efficiency, we also need to develop metrics that enable us to manage risk and have capacity available when needed. These include capacity utilization and allocation thresholds that can be used to trigger alerts before utilization reaches unacceptably high levels.

Risk management metrics are of primary interest to Intel IT storage operations engineers, who are responsible for making capacity available to support day-to-day business needs.

Improving Storage Efficiency with Thin Provisioning

We are applying a number of approaches to reduce cost and curb storage growth by utilizing capacity more efficiently. These include thin provisioning, reclaiming unused storage, and sharing storage infrastructure across multiple projects.

Thin provisioning is a key technology, especially within our private cloud. The concept is analogous to the way that airlines overbook flights. As shown in Figure 2, we can oversubscribe, or over-allocate, capacity to provide customers with adequate headroom, based on the assumption that they typically ask for more capacity than they need. The effect of oversubscription is to increase capacity utilization and enable us to avoid purchasing more capacity than customers actually need.

Because we are oversubscribing available capacity, our storage capacity metrics must be able to measure the efficiency benefits achieved with thin provisioning and help us manage capacity and risk.

We are widely deploying thin provisioning to increase efficiency across our private cloud storage environment.

We are also exploring the use of data de-duplication, which recognizes when multiple instances of the same data are stored by different customers. The technology consolidates the multiple data sets into a single copy, freeing space that can then be allocated for other use.

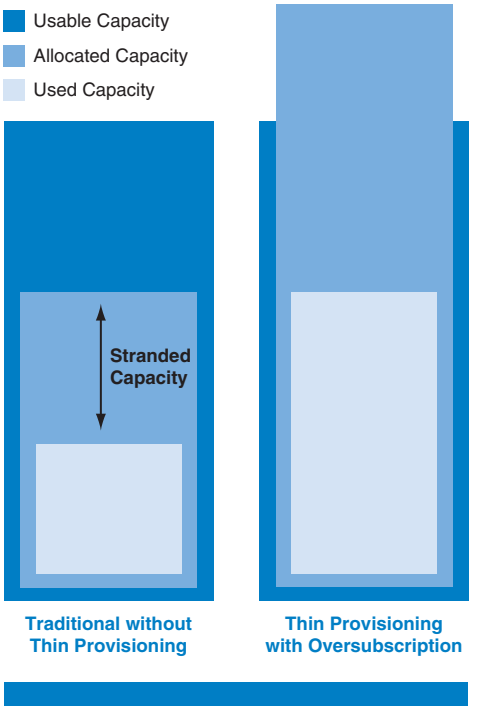

Figure 2: Thin provisioning enables oversubscription
(Source: Intel Corporation, 2012)

Storage Metrics

At the core of this framework is a single set of capacity utilization metrics that we can apply across our entire SAN environment to support efficiency, capacity management, and risk management. In the future, we also anticipate being able to apply this approach across our traditional office and enterprise network attached storage (NAS) environment.

We designed our metrics so that we can aggregate data to create different views that enable us to analyze storage use and efficiency across the entire storage environment. We can analyze the metrics data by data center, storage tier, and storage frame. In the future, our goal is to show by individual customers.

Efficiency Metrics

We defined three core metrics that measure the storage efficiency of different aspects of the environment. We believe that these metrics represent a minimum set necessary to measure the efficiency of storage capacity use and thereby help contain cost. The metrics are orthogonal, or independent; to maximize efficiency throughout the environment, we need to optimize all three:

"The metrics are orthogonal, or independent; to maximize efficiency throughout the environment."

- *Slot Utilization.* The ratio of storage frame slots that are populated with drives compared to the total available storage frame slots.
- *Overall Storage Efficiency.* The ratio of customer stored data compared to the raw storage capacity.
- *Low-Cost Storage Percentage.* The ratio of customer data stored on our low-cost storage tier compared to the total customer data stored.

Slot Utilization Percentage

Frames are expensive and occupy costly data center capacity. Therefore, efficient utilization of frame capacity is an important factor in minimizing overall storage cost.

"Efficient utilization of frame capacity is an important factor in minimizing overall storage cost."

Our measure of frame capacity utilization is the Slot Utilization Percentage metric. Each frame includes a manufacturer-defined maximum number of slots. Intel IT storage operations groups may restrict the number of these slots that are available for use to provide good performance for Intel workloads. As shown in Figure 3, we define Slot Utilization Percentage as the percentage of these available slots that are populated with drives.

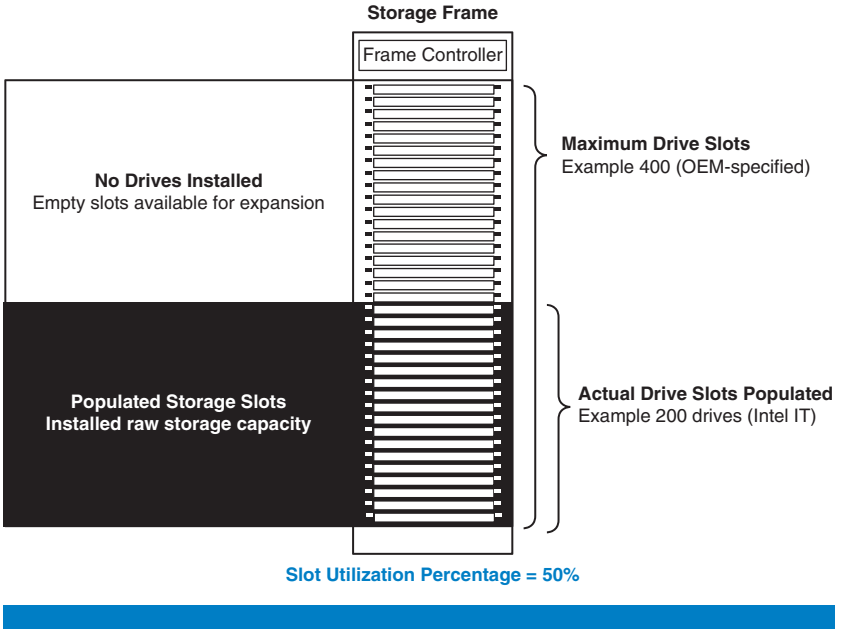

Figure 3: Slot utilization
(Source: Intel Corporation, 2012)

Overall Storage Efficiency

The goal of our storage strategy is to store data required by our internal customers in a cost-effective manner. Therefore, we defined a metric that measures how efficiently storage capacity is utilized for storing customer data. This metric, Overall Storage Efficiency, is defined as the ratio of stored customer data (Used Capacity) to raw storage capacity (Raw Capacity), as shown in Figure 4. Improving Overall Storage Efficiency is a primary focus of our storage strategy.

"Measuring Used Capacity presents some complexity for IT organizations because of data duplication."

Measuring Used Capacity presents some complexity for IT organizations because of data duplication (multiple instances of business data stored by one or more customers) and inefficient management of orphaned data (stored data that is no longer accessed by customers). The challenge is that the customer's view of the amount of data stored—the Used Capacity—may differ from

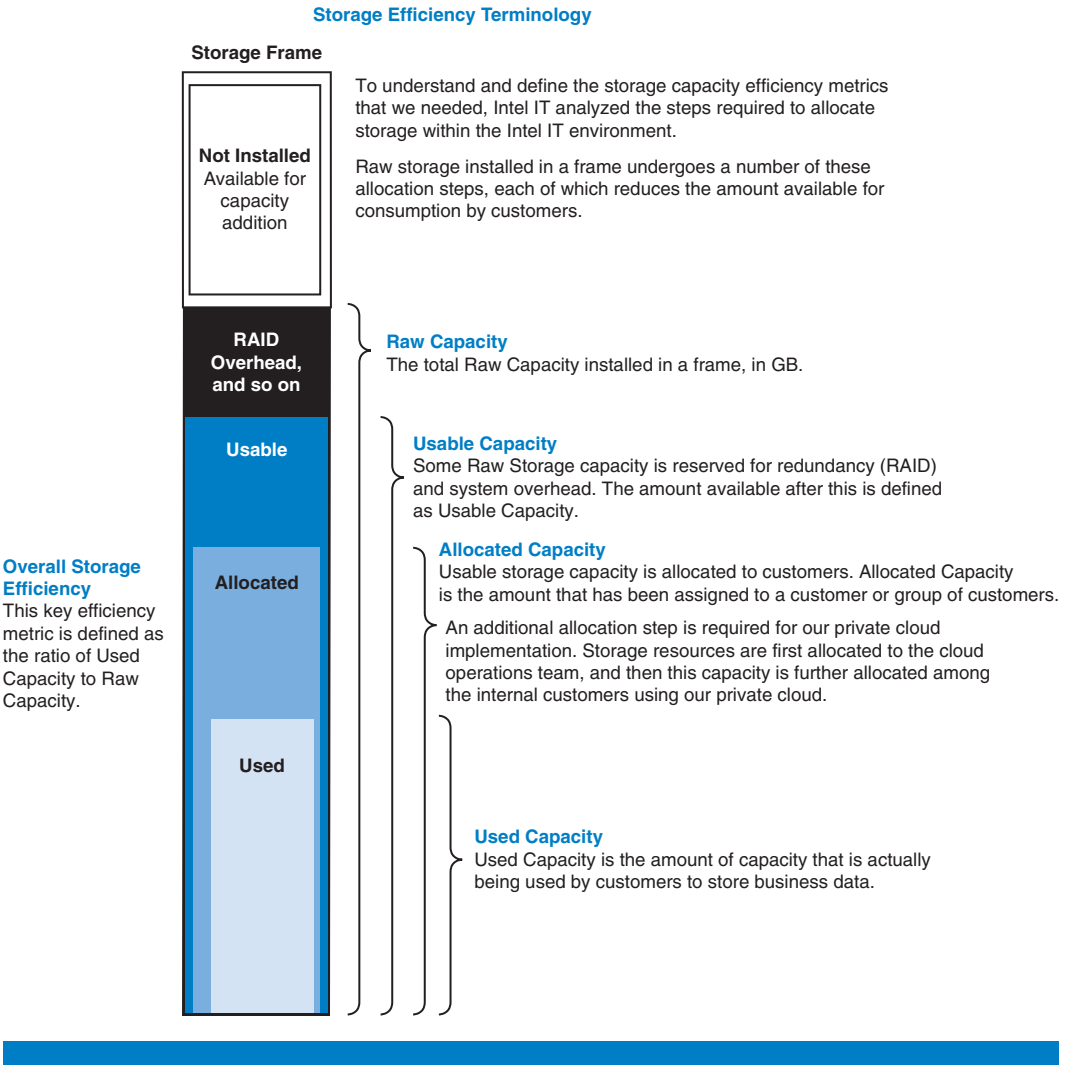

Figure 4: Storage efficiency metrics
(Source: Intel Corporation, 2012)

the storage frame view. As we evaluate data de-duplication technology, we are initially focused only on the storage frame view of Used Capacity as a reasonable approximation for customer stored data.

Low-Cost Storage Percentage

This represents the percentage of customer data that is stored on the lowest-cost tier, which is currently our M2 tier. Intel IT is planning storage tier initiatives with the goal of moving less critical business data to lower cost tiers while continuing to support customer needs.

Capacity Management and Risk Management Metrics

Some storage efficiency techniques, such as thin provisioning, increase utilization but also increase the risk of having insufficient capacity to meet growing and changing business needs. If we oversubscribe capacity (allocate more than is actually available), it becomes even more important to carefully monitor how much of this capacity is being used, how quickly usage is growing, how much more capacity we can safely allocate, and when we need to add or rebalance capacity.

"The metrics measure allocation and utilization relative to Usable Capacity, not Raw Capacity."

Our key capacity management and risk management metrics are designed to support these requirements. The metrics measure allocation and utilization within each storage pool, but can also be aggregated to provide summary views at the frame and data center level. The metrics measure allocation and utilization relative to Usable Capacity, not Raw Capacity, since this is what matters to the IT operations team and our customers. The metrics are shown in Figure 5 and described below.

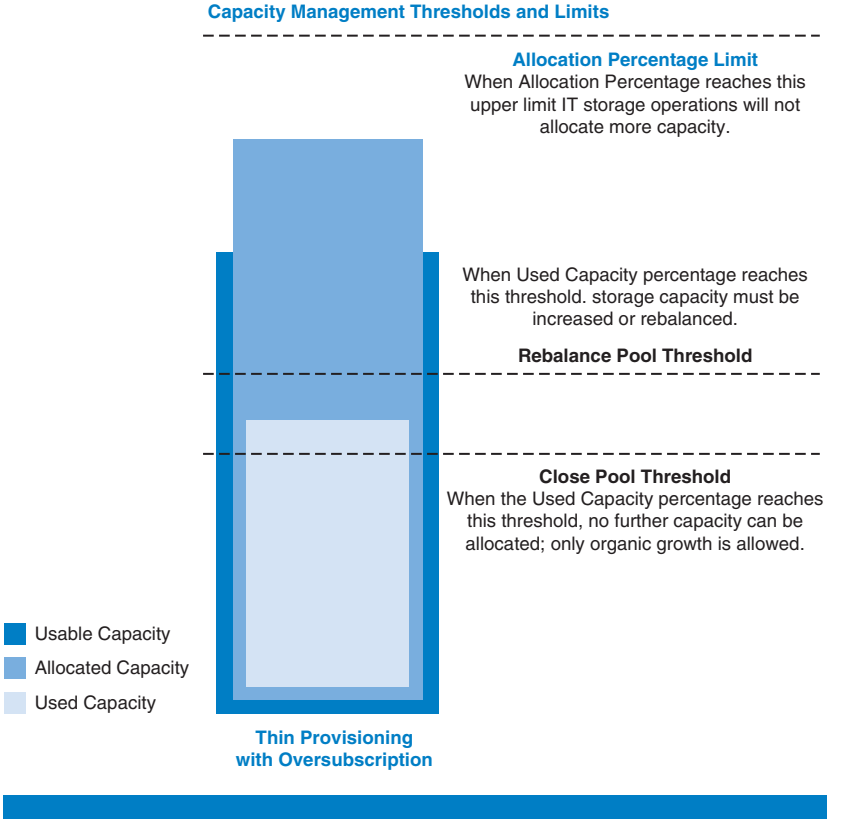

Figure 5: Capacity and risk management metrics (Source: Intel Corporation, 2012)

To minimize risk, we need to be alerted when allocation or utilization levels reach predefined thresholds. We have established several initial policy-based thresholds, as described below. We expect to refine these policies as we gain more experience in production. For example, we anticipate being able to incorporate better forecasting of customer utilization behavior and trends in storage demand growth, enabling us to include "time to full" projections. We also anticipate including performance health monitoring.

"To minimize risk, we need to be alerted when allocation or utilization levels reach predefined thresholds."

Used Percentage

This is the percentage of Usable Capacity that is used to store customer data. It is defined as the ratio of Used Capacity to Usable Capacity. By definition, this metric can never exceed 100 percent. We currently define two operational thresholds to manage risk as storage pools reach high utilization levels:

- *Close pool threshold.* When a pool's Used Percentage reaches 50 percent, the pool is considered closed. No further capacity can be allocated from this pool; only organic growth of existing customer storage workloads is allowed.
- *Rebalance pool threshold.* When the Used Percentage for a pool reaches 70 percent, storage capacity must be reassigned to other pools to avoid the risk of customer impact.

Allocation Percentage

This is the percentage of Usable Capacity that has been allocated to customers. With thin provisioning, this metric can exceed 100 percent. We have established an allocation limit of 150 percent. Once a pool's Allocation Percentage reaches this level, IT operations will not allocate more capacity, regardless of how much of the capacity is actually being used by customers.

Customer Utilization of Allocated Capacity

This represents how much specific customers have used the storage capacity that has been allocated to them. Ideally, all customers would consume 100 percent of their allocated capacity. However, most customers use only a small percentage of their allocation. Higher levels of storage pool oversubscription are possible in pools with lower customer utilization. We often refer to this metric simply as Customer Utilization.

"Most customers use only a small percentage of their allocation."

Allocation Headroom

Ultimately, the operations team needs to know whether they can allocate more storage from a pool in response to customer demand and, if so, how much storage can be allocated. The Allocation Headroom metric is designed to enable IT storage operations engineers to quickly and easily determine this.

Allocation Headroom is defined as the additional storage capacity, in gigabytes, that can be allocated from a pool. It is calculated based on several factors, including Allocated Percentage, Used Percentage relative to the thresholds described above, and Customer Utilization of Allocated Capacity. We define policies that alert operations engineers when a pool's Allocation Headroom has decreased to a predetermined low level.

"Our goal is to proactively analyze and manage SAN storage across our office and enterprise private cloud."

Analyzing Efficiency: Business Intelligence and Reporting

We have begun building business intelligence and reporting capabilities based on our metrics. Our goal is to proactively analyze and manage SAN storage across our office and enterprise private cloud.

We created a storage resource management (SRM) tool that automatically gathers storage capacity-related data across our SAN environment. Using the SRM tool, we can pull data from individual storage pools or aggregate data by frame or by data center. This allows managers to compare efficiency across different data centers, frames, and pools.

As of May 2012, we have established regular on-demand reporting across 130 pools in 16 data centers, representing over 3 PB of raw storage within our private cloud.

We use the information gathered by the SRM tool to create customized reports for different purposes and users within Intel IT.

For example, executives may receive a high-level overview of average efficiency and other summary data. Engineers responsible for day-to-day operations need reports that are at a more granular level and that alert them to potential problems such as over-allocated storage pools.

Some example reports are shown below. These show actual Intel IT operational data; we have changed some identifying details to protect Intel proprietary information. We are applying thin provisioning to increase efficiency within these data centers.

Executive Storage Efficiency Report

This report, shown in Table 1, includes the Overall Storage Efficiency metric and also provides a capacity overview for Intel IT executives and senior managers at the data center level.

Data Center	Raw Capacity (GB)	Used Capacity (GB)	Overall Efficiency %
DC1	114,174	13,999	12.26%
DC2	40,260	13,328	33.10%

Table 1: Executive Report
(Source: Intel Corporation, 2012)

Data Center Manager Report

These managers need to know how much usable capacity exists, how that capacity is being allocated and used, and how much headroom is available in each storage tier.

This information is critical for planning purchases of more storage capacity based on established IT operations policies and customer service-level agreements (SLAs).

While this report provides a useful overview at data center and storage tier levels, data center capacity managers typically also need information that helps them manage capacity at the level of individual frames.

"Data center capacity managers typically also need information that helps them manage capacity at the level of individual frames."

This frame-level report tells data center managers that they need to take action by working with the operations team to rebalance storage utilization among the existing frames or by acquiring more storage capacity. If they identify that more storage capacity is required, they can examine slot availability to determine whether there are available slots in existing frames or whether new frames need to be purchased.

Pool Status Report

Operations engineers responsible for day-to-day storage management need more granular information at the storage pool level. This enables them to respond to requests for more storage by determining which pools have enough capacity.

Conclusion and Next Steps

Storage capacity utilization metrics enable us to improve the efficiency of our private cloud by providing a framework we can use to analyze and manage SAN storage use across our diverse IT environment. Metrics also provide information required by Intel IT for capacity management and risk management.

"Storage capacity utilization metrics enable us to improve the efficiency of our private cloud by providing a framework we can use to analyze and manage SAN storage."

We find that the reports and metrics help our storage operations engineers and storage capacity managers work together to maximize storage use efficiency. The reports help them identify the best course of action. For example, when a pool's utilization or allocation reach high levels, they can determine whether they need to purchase new capacity or can simply rebalance storage across existing pools, frames, or data centers.

We have established on-demand reporting across 130 pools in 16 data centers, representing more than 3 PB of raw storage within our private cloud. We continue to increase the reporting coverage across our private cloud, and, in the future, we also anticipate being able to apply this approach across our traditional office and enterprise NAS environment.

We continue to improve our metrics. For example, we have developed an adaptive algorithm that can create customized allocation headroom estimates for each pool based on factors such as the rate at which a customer's capacity utilization is increasing.

As we scale our storage environment and increase utilization, performance health metrics will be essential to understand application and system requirements, identify potential bottlenecks, and assist in successful deployment. Accordingly, we are in the process of incorporating storage performance health into our capacity management and efficiency views. These

performance metrics include concepts such as response time, queue length, and storage processor utilization.

While we roll-out thin provisioning, we are also exploring the use of other efficiency technologies such as data de-duplication. Our metrics will enable us to measure the efficiency improvements resulting from the implementation of these technologies.

Author Biographies

Sudip Chahal is a principal engineer and compute and storage architect in Intel IT Enterprise Architecture group where he and his team work on IT architectures, strategies, and roadmaps for the enterprise compute and storage services. Sudip attended the Indian Institute of Technology, Kanpur and the University of Wisconsin, Madison earning a B. Tech. EE and MS CS respectively.

Chris Peters has more than 21 years' experience in business and technical fields ranging from information technology, to product marketing, manufacturing, supply chain, nuclear power, and consumer products. Chris is currently Director of Industry Engagement with Intel's Information Technology Team, where he and his team work closely with senior IT peers worldwide to share IT best practices with the goal of delivering better business value from IT innovation and investment. Chris attended the University of Rochester and University of Connecticut earning a BS Physics and MBA respectively.

Ajay Chandramouli has over 13 years of experience in the technology industry with over 10 years of experience at Intel Corporation. Ajay has held a variety of IT, software and hardware engineering positions while at Intel and the Lawrence Livermore National Laboratory. Ajay's current role is to share Intel IT's Cloud and Data Center best practices with his senior IT peers across the industry. Ajay holds both an MBA and MSE from UC Davis.

Steve Owen has held a wide variety of positions in IT Operations since joining Intel Corporation in 2000. Steve's current role is in the Storage Operations team designing and developing automation tools.

Krish Anandarao is currently Storage and Data Restoration Service Manager within IT Operations. He held a wide variety of positions at Intel IT since 2001, including Storage Capacity Management and Database Administration previous to his current role.

Shane Healy is Storage and Data Restoration Service Configuration Manager and Architect within IT Operations. He is based out of the Intel Ireland site.

Nigel Wayman has 22 years' experience in designing and administering IT infrastructure systems. Nigel is currently an IT technology specialist providing customer support for the SAN systems based on service-level agreements. He supports internal employees, and/or external customers/suppliers globally.

Nigel documents and troubleshoots SAN related problems for resolution and specializes in SAN performance issues. Nigel collaborates with other groups and organizations to provide a service for infrastructure changes, foster partnerships with other levels of support, elevate data quality concerns, share best known methods and information, communicate known system problems, and keep abreast of upcoming developments.