**IT@Intel White Paper**

**Intel IT**
Data Centers and IT Innovation
September 2013

# Improving Business Continuity with Data Center Capacity Planning

By getting visibility into data center capacity issues over a 3-year planning horizon, we can strategically plan ahead to accommodate evolving business needs and market changes.

## Executive Overview

**To meet Intel's business computing needs more effectively and ensure continuity of service, Intel IT has developed a predictive data center capacity planning process. Our approach helps us to accurately forecast demand on our data center facilities and proactively identify capacity shortfalls years in advance.**

We use the predictive planning approach during our long-range planning (LRP) process, which happens twice a year. We conduct the analysis over a 3-year planning horizon to understand business growth trends and the need for future data center capacity.

Our LRP process involves three key stages:

- **Gather data from our business segments.** Each business segment reviews their business plans and provides demand forecasts based on historical trends and growth estimates. The goal is to generate a 3-year capacity demand forecast on the compute capacity requirements for all business segments.
- **Analyze the data.** We map the forecasted data against baseline data—the currently available space, cooling, and power—in each data center. Our primary goal is to identify constrained data centers, those that will have more demand than available capacity at some point over the planning horizon.
- **Report the results.** We report our findings to senior management. Our LRP reports are designed to bring visibility into upcoming data center capacity issues and spur discussion of remedial plans to address these issues. We propose solutions to optimize the performance and utilization of each data center.

A shortfall in data center capacity can impact business goals and budgets as well as overall growth and innovation. By getting visibility into data center capacity issues over a 3-year planning horizon, we can strategically plan ahead to accommodate evolving business needs and market changes.

Monica To
Data Center Planner, Intel IT

Kevin Emery
Data Center Planner, Intel IT

## Contents



## IT@INTEL

The IT@Intel program connects IT professionals around the world with their peers inside our organization – sharing lessons learned, methods and strategies. Our goal is simple: Share Intel IT best practices that create business value and make IT a competitive advantage. Visit us today at www.intel.com/IT or contact your local Intel representative if you'd like to learn more.

# BUSINESS CHALLENGE

**Data centers are at the heart of Intel's massive worldwide computing environment. Intel IT operates 67 data centers[1] housing approximately 72,000 servers that underpin the computing needs of more than 100,000 employees. These facilities enable the innovation that fuels Intel's business growth.**

Responding quickly to the changing demand for computing power from Intel's business groups—our internal customers—is critical to Intel's success. Our data centers serve five business segments—Design, Office, Manufacturing, Enterprise, and Services. Providing the right amount of computing power at the right time saves costs, prevents overspending, and enables business agility.

Running out of data center capacity can impact business goals and budgets as well as overall growth and innovation. A shortfall in data center capacity or any loss of needed computing power can have a significant impact on Intel's revenue streams—at the rate of millions of U.S. dollars (USD) per day or more.

In order to meet Intel's ever-changing computing needs and keep operations running smoothly, Intel IT requires an accurate and long-range view of business computing requirements, as well as data center capacity and utilization across the enterprise. Data center capacity planning, however, is challenging because it requires identifying "constrained" data centers years in advance due to long lead times for adding new infrastructure. A constrained data center is not only one that is at full capacity today, but also one that may be fully utilized within a 3-year planning horizon.

Building a new data center is a huge investment that involves significant resources and funding, and typically takes about 18 months or longer to complete. Even retrofitting a data center with new hardware to increase capacity can take up to six months or more. Adding to the planning challenge, by the time a data center is complete, capacity demands from our business segments may have changed or grown as market opportunities evolve.

Given these challenges, Intel IT has developed and optimized our approach to data center capacity planning over several years (see Figure 1). In 2004, we didn't have a centralized data center capacity planning strategy. Our data centers grew as needed to address immediate business needs. Consequently, the number of data centers grew to one or two at every site. By 2007, our data center planning team created a centralized and formal process to consolidate data from different business segments and create a concise view of global demand.

[1] This number reflects the data center count at the time of this writing. In 2012, we achieved significant efficiency in data center operations by reducing the number of data centers from 87 to 68, closing some and reclassifying others. Further cost reductions came from refreshing older compute and storage servers with fewer higher-performing servers based on the latest Intel® Xeon® processors and Intel® Solid-State Drives (Intel® SSDs). See *Intel IT Performance Report, 2012-2013.*

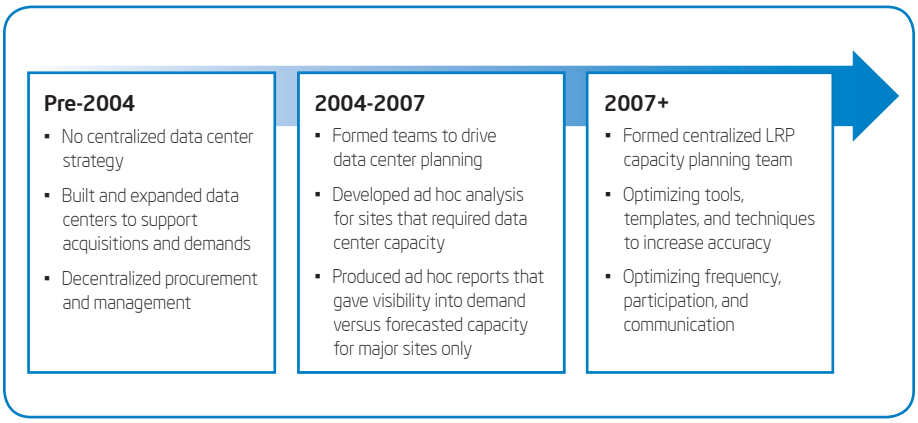


Figure 1. Intel's long-range data (LRP) center capacity planning has evolved to optimize the process.

Since then, we've continued to optimize and evolve our processes. Our current approach focuses on efficiently meeting capacity needs by identifying future capacity gaps with sufficient time to implement the best solutions to meet future demand.

## SOLUTION

**To effectively meet Intel's business computing needs, Intel IT developed a predictive data center capacity planning approach and implemented a long-range planning (LRP) process. Our approach proactively identifies capacity shortfalls, helping us to accurately forecast demand, and thus enables us to plan ahead to accommodate evolving business needs and market demands.**

Key elements for success in data center capacity planning:

- Interactive communication with business partners within a well-defined process
- Support from senior-level management to invest the time and resources necessary for a successful outcome
- A history of accuracy in projections that builds trust and cooperation between our business partners and IT

## Predictive Planning Approach

The goal for data center capacity planning is to proactively predict capacity shortfalls across the company. This predictive approach provides a breakdown analysis by business segment, identifies when data centers will become constrained, and enables us to implement remedial plans. By analyzing the data from Intel's business segments on planned utilization and comparing it to current and projected capacity, Intel IT's data center planning process provides reports and recommendations per data center and per business segment for the most effective approach going forward.

Figure 2 depicts our predictive approach conceptually. The graph shows the forecasted utilization of a data center, on a quarterly basis over three years, as compared to available data center infrastructure.

In our process, a constrained data center is not only one that is at full capacity today, but also one that may be fully utilized within our 3-year planning horizon. Because each server within a data center requires space, cooling, and power—including an uninterruptible power supply (UPS), generator, and utility power—these key constraints can limit capacity in a data center. The limits reached on any one or all of those infrastructure components cause a data center to become constrained.

The forecast is based on the number of devices, such as servers or blades, that the business segments anticipate are required to support their business. These forecasts, which are gathered by business segment, include required devices for new projects and reduction of devices through our planned server refresh process. Then we calculate and map the forecasted number of servers against the available data center infrastructure (space, cooling, and power) to determine the gaps.

Figure 2 shows infrastructure capacity for UPS power in this data center as the horizontal black line. "Positive headroom" occurs when the forecasted usage is lower than infrastructure capacity, which means that our forecasts predict that the data center has sufficient capacity until the first quarter of the second year. "Negative headroom" occurs when the forecasted usage is greater than infrastructure capacity—when we have excess demand from a business segment and not enough capacity to handle the forecasted demand.

"Zero headroom" is the inflection point where the data center is predicted to be fully utilized and no longer able to accommodate additional devices. At this inflection point, the data center is forecasted to be constrained and remedial plans must be put in place to address the excess demand for compute capacity. In our example, this occurs between the first and second quarters of year 2. Any new design projects, server refresh funding, or other approaches to capacity constraints can pull in or push out the inflection point.

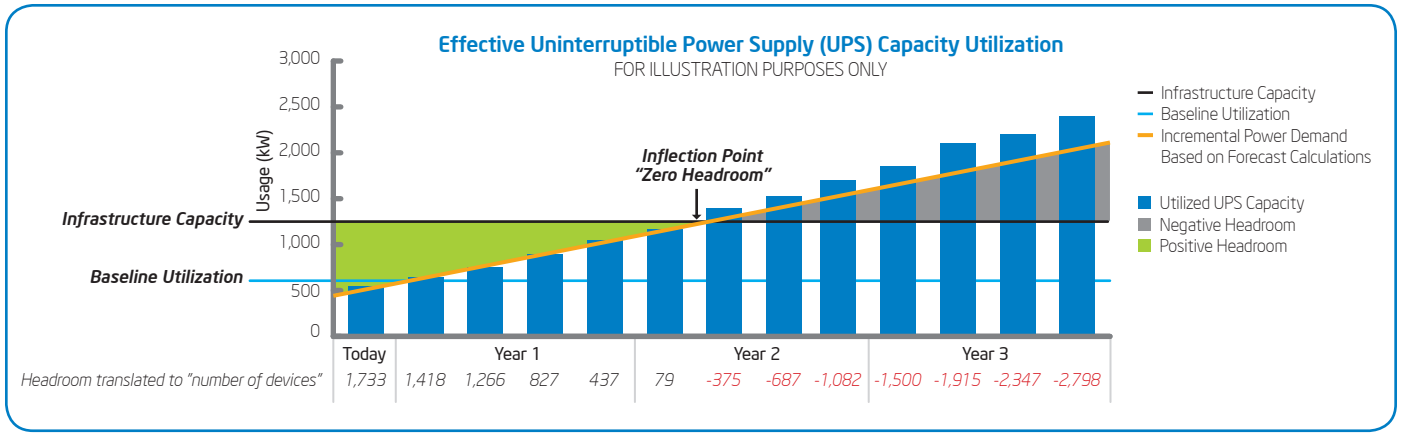


Figure 2. Intel IT's predictive planning approach identifies constrained data centers over a 3-year planning horizon, giving us time to remediate the anticipated shortfalls. The information in this sample data center shows the incremental power demand and the inflection point where our forecast predicts we will run out of capacity. The negative numbers indicate the magnitude of the shortfall between forecasted capacity demand and available power. FOR ILLUSTRATION PURPOSES ONLY

## Long-Range Planning Process

We use our predictive planning approach during our LRP process. Through this process we perform a breakdown analysis by business segment, identify when data centers will become constrained, and implement remedial plans as needed.

Shown in Figure 3, our LRP process takes place twice a year with three main phases—Initiation and Planning (Pre-LRP), Execution (LRP), and Closing (Post-LRP).

Throughout the process, we communicate to our stakeholders to ensure they are aware of the progress and decisions being made. This active engagement facilitates prompt responses and captures any input or changes that may be required. We involve both organizational managers and individual contributors to capture different levels of expectation and to ensure the roles, responsibilities, and deliverables are defined clearly.

In the Initiation and Planning (Pre-LRP) phase, the focus is on identifying stakeholders and planning detailed tasks to prepare the LRP tasks for the Execution phase. In the Closing (Post-LRP) phase, we conduct post-mortems and identify lessons learned from the entire LRP process. The most important phase for the entire LRP process is Execution, which includes three stages (see Figure 4, on the next page).

### DATA GATHERING

The stakeholders in our five business segments have varying workflows and capacity growth patterns. The goal for this stage is to generate a 3-year capacity demand forecast that predicts the compute capacity requirements for all business segments.

Each business segment reviews their business plans and provides demand forecasts based on historical data, approved funding, and their growth estimates. The review process may vary based on their individual business requirements. For example, the Design segment is dynamic and project-based. Their growth fluctuates in response to rapid market changes and demands. On the other hand, the Manufacturing segment has more predictable growth in computing demand. As a result, we customize our data gathering and forecasting techniques for each business segment.

We tailor our process for each business segment to streamline the data collection process. Every business segment has their own tools and templates for collecting the forecasted based on their business needs. Once the forecast is completed and delivered by the business segment, we analyze the demand data and identify any added or retrofitted pieces of equipment needed. Then the space, cooling, and power requirements for each data center are calculated.

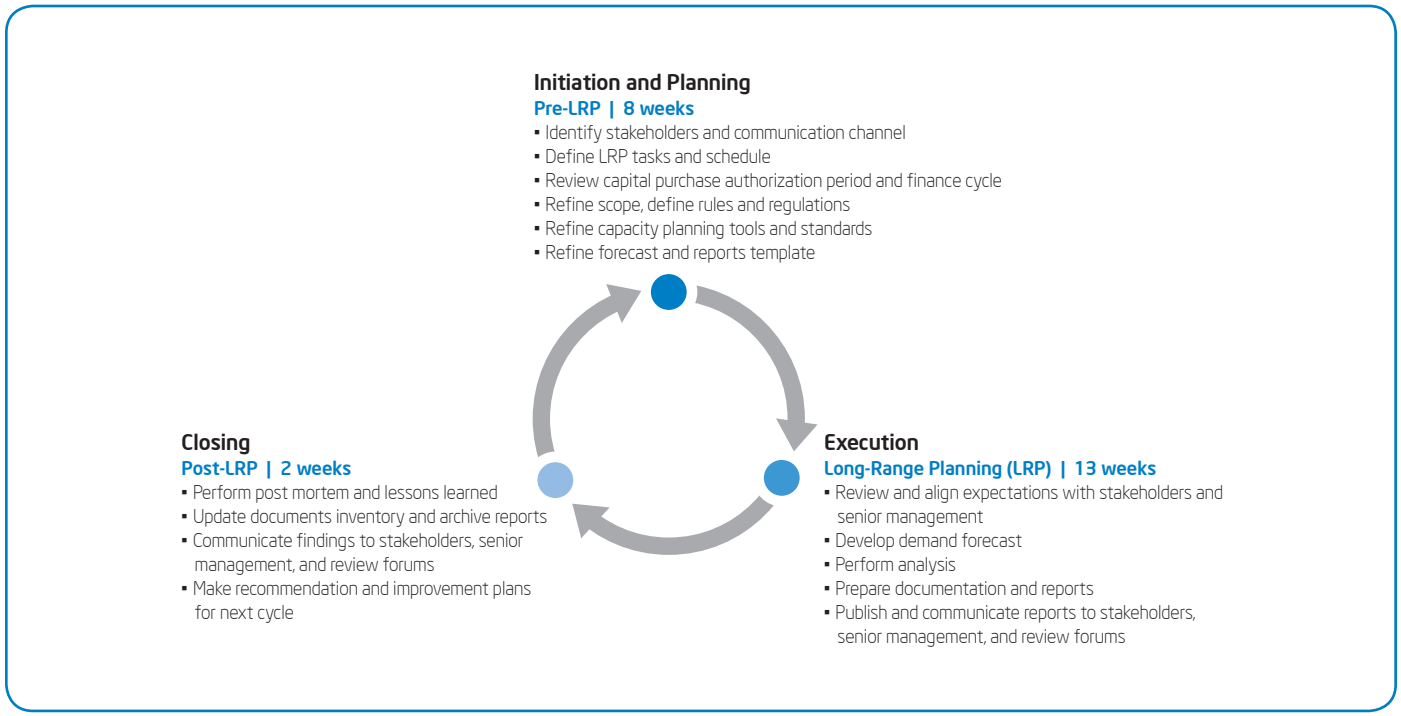


Figure 3. The three phases in a complete long-range planning (LRP) process together provide visibility into capacity by identifying constrained data centers.

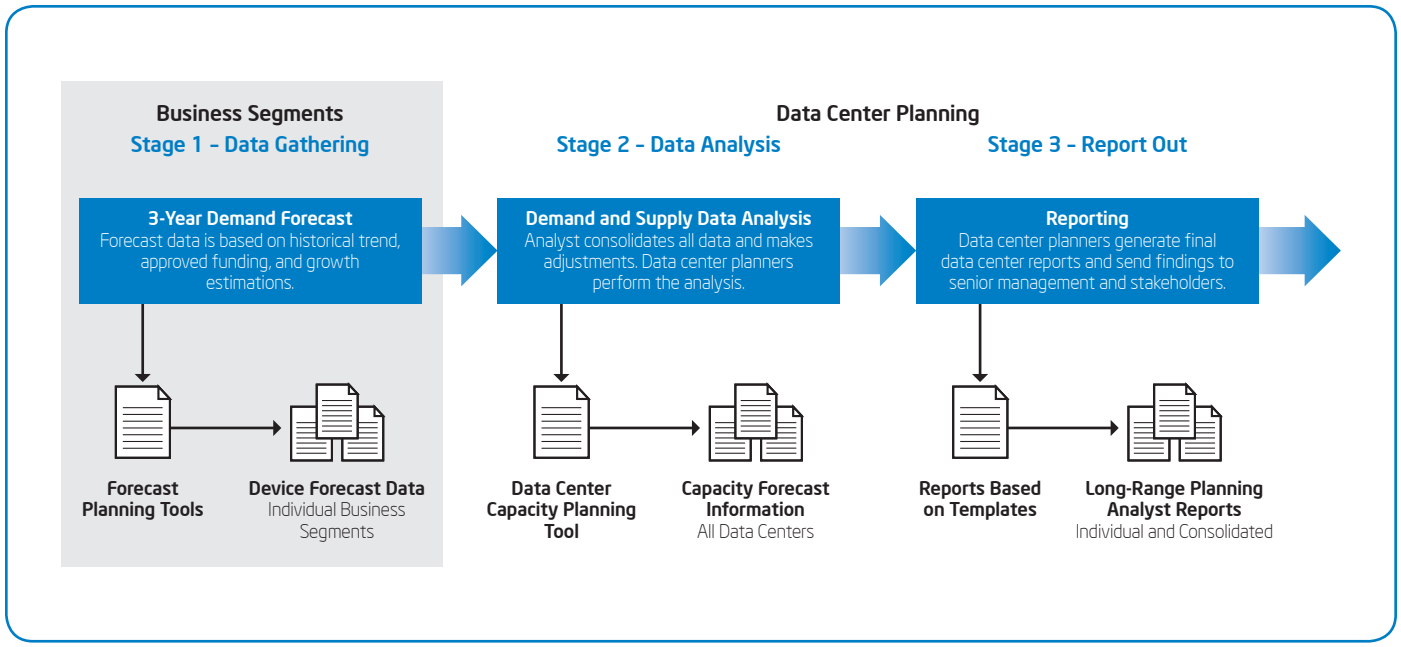


Figure 4. The Execution phase of the long-range planning (LRP) process has three stages: data gathering, data analysis, and report out.

## DATA ANALYSIS

During this stage, our primary goal is to identify constrained data centers, those that have more demand than available capacity. We analyze the data over the 3-year planning horizon to understand business growth trends and the future needs for additional data center capacity. Also, we create remedial plans to optimize data center capacity and resolve capacity shortfalls.

One of the analysis tools we use contains models for each of the IT devices. Each model contains the specifications for that particular device in terms of space, power, and cooling required in the data center. The forecast and models are consolidated to produce a capacity demand. That demand is analyzed in conjunction with existing data center infrastructure capacity to determine the future impact.

When data discrepancies and unrealistic forecasted data are identified, we manually adjust the data to ensure the highest degree of integrity. For example, forecasted demand may be ascribed to data centers planned for decommissioning or show unrealistic growth for the projected business. The forecasted demand may not be adjusted for data center construction in progress. Before the final reports are distributed, we review the results with the stakeholders and reconcile any discrepancies.

## REPORT OUT

During this stage, we report the findings from our analysis to senior management in the IT organization and our key stakeholders. The analysis includes the capacity performance by business segment and capacity availability over the next three years. We propose solutions and make recommendations for data centers that are constrained or underutilized, with a goal to optimize the performance and utilization of each data center. Our LRP reports are designed to bring visibility into upcoming issues to help implement remedial plans to address these issues. Each report highlights the projected capacity status along with risks, an action plan, deliverables, targets, and due dates.

Two main types of reports are generated:

- Individual data center LRP capacity reports
- A consolidated data center performance report (across all business segments)

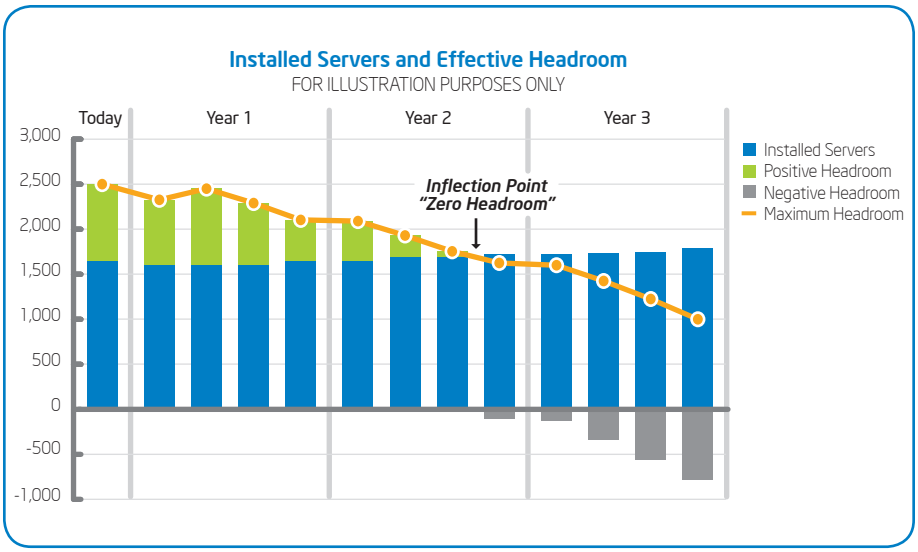


Figure 5. This sample data center capacity chart identifies when the data center becomes constrained, enabling us to create remedial plans. FOR ILLUSTRATION PURPOSES ONLY

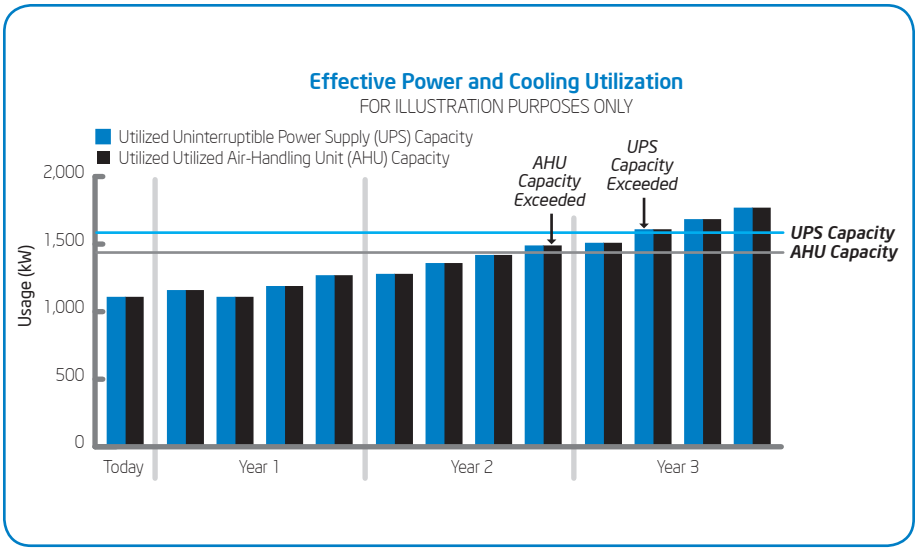


Figure 6. This sample capacity report shows a data center becoming constrained by two different infrastructure elements: Power (UPS) Capacity and Cooling (AHU) Capacity. We can more strategically plan our remediation effort, possibly a retrofit, by addressing both issues at the same time. FOR ILLUSTRATION PURPOSES ONLY

### Individual Data Center LRP Capacity Report

Figures 5 and 6 show examples of charts for the types of LRP capacity report we may generate for an individual data center.

Figure 5 shows the forecasted demand and the change in headroom over a 3-year period. For example, the first column shows an installed base of about 1,600 servers with spare capacity (positive headroom) for an additional 860 blades. This headroom declines over time and goes negative in the fourth quarter of the second year. Before that happens, we need to create a remedial plan and take action to mitigate this constrained data center.

Figure 6 shows the utilization level for cooling and power in a sample data center and the point in time where our forecasted demand will exceed capacity. As shown, the forecasted demand for air-handling units (AHUs) will exceed capacity for cooling in the fourth quarter of the second year and UPS power in the second quarter of the third year. If a data center infrastructure retrofit is the preferred solution, we would need to address both capacity issues before they become constrained. Two retrofits could be conducted separately, but the data center manager may wish to minimize the impact and risk by coordinating both retrofits simultaneously.

### Consolidated Data Center Performance Report

Table 1 shows an example of a consolidated data center performance report.

The report shows all of the data centers across all business segments that are forecasted to become constrained over the next three years, with the ones that become constrained earliest at the top of the chart. This report helps us prioritize our remedial efforts by showing the degree of urgency and the degree of the constraint (in the negative headroom numbers highlighted in grey). For example, the first two data centers at the top of the chart are a concern because they are already constrained today but the fourth data center down in the Year 1/Q1 is a bigger problem because the gap between forecasted demand and available capacity is over 1,000 servers as compared to 5 servers.

Having multiple constrained data centers in the future can reflect a positive outcome of planning. If no data centers are constrained within the planning horizon, data center infrastructure may have been over-engineered. At Intel IT, our planning is designed to efficiently address infrastructure capacity shortfalls in a "just-in-time" manner.

**Table 1. This sample consolidated data center performance report helps us prioritize our remedial efforts by showing all of the constrained data centers across all business segments—indicating the degree of urgency and the degree of constraint (using negative headroom numbers).** FOR ILLUSTRATION PURPOSES ONLY

| | Today | Year 1 | | | | Year 2 | | | | Year 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| Data Center 1 | -34 | -5 | 13 | -63 | -144 | -224 | -301 | -359 | -446 | -511 | -580 | -671 | -671 |
| Data Center 2 | -66 | -144 | -136 | -141 | -141 | -141 | -141 | -175 | -175 | -175 | -175 | -209 | -209 |
| Data Center 3 | 17 | -16 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | -3 |
| Data Center 4 | 17 | -1110 | -1320 | -1416 | -1581 | -1633 | -1693 | -1794 | -1849 | -1901 | -1962 | -2024 | -2024 |
| Data Center 5 | 309 | -74 | -285 | -349 | -405 | -475 | -531 | -604 | -665 | -727 | -786 | -848 | -848 |
| Data Center 6 | 106 | -28 | 12 | 11 | 14 | 8 | 13 | 18 | 28 | 26 | 39 | 52 | 52 |
| Data Center 7 | 80 | -321 | -462 | -494 | -517 | -533 | -561 | -587 | -613 | -637 | -656 | -672 | -672 |
| Data Center 8 | 129 | -15 | -63 | -68 | -78 | -86 | -93 | -101 | -110 | -119 | -130 | -137 | -137 |
| Data Center 9 | 730 | -399 | -521 | -891 | -1228 | -1556 | -1877 | -2200 | -2531 | -2845 | -3171 | -3490 | -3490 |
| Data Center 10 | 949 | 5 | -245 | -481 | -525 | -560 | -594 | -638 | -684 | -728 | -761 | -804 | -804 |
| Data Center 11 | 877 | 252 | -19 | -437 | -472 | -577 | -617 | -865 | -899 | -971 | -1024 | -1058 | -1058 |
| Data Center 12 | 971 | 289 | 73 | -119 | -253 | -409 | -557 | -733 | -863 | -1031 | -1172 | -1314 | -1314 |
| Data Center 13 | 91 | 67 | 13 | 6 | -1 | -8 | -15 | -23 | -31 | -40 | -48 | -56 | -56 |
| Data Center 14 | 71 | 7 | 6 | 3 | -1 | -3 | -5 | -7 | -9 | -11 | -13 | -17 | -17 |
| Data Center 15 | 43 | 23 | 20 | 13 | 8 | -4 | -8 | -14 | -19 | -29 | -32 | -34 | -34 |
| Data Center 16 | 1709 | 971 | 571 | 311 | 40 | -140 | -326 | -500 | -687 | -861 | -1071 | -1289 | -1289 |
| Data Center 17 | 54 | 36 | 17 | 11 | 5 | -1 | -7 | -13 | -18 | -24 | -30 | -36 | -36 |
| Data Center 18 | 89 | 98 | 114 | 73 | 58 | 12 | -25 | -61 | -96 | -138 | -169 | -199 | -199 |
| Data Center 19 | 474 | 483 | 453 | 379 | 301 | 221 | 141 | 71 | -3 | -76 | -149 | -222 | -222 |
| Data Center 20 | 206 | 186 | 167 | 139 | 114 | 90 | 65 | 41 | 16 | -9 | -46 | -82 | -82 |
| Data Center 21 | 463 | 326 | 321 | 276 | 249 | 227 | 205 | 147 | 84 | 24 | -46 | -116 | -156 |
| Data Center 22 | 120 | 117 | 117 | 106 | 93 | 79 | 65 | 55 | 38 | 22 | 5 | -11 | -11 |
| Data Center 23 | 74 | 66 | 62 | 49 | 45 | 41 | 37 | 22 | 17 | 8 | 3 | -3 | -9 |

## Choices for Meeting Data Center Capacity Needs

Long-range planning (LRP) and analysis provides the needed visibility into computing capacity issues with enough time to respond to projected gaps between supply and demand. Building new data centers is expensive and time-consuming. Intel IT applies a variety of other approaches for running data centers efficiently and effectively with a goal to mitigate under or over capacity.

One solution Intel IT employs is server refresh, swapping in new, more powerful servers that take up less space, use less power and require less cooling. This solution can have dramatic impacts on both capacity and total costs. Sometimes shortening the anticipated refresh cycle is sufficient to meet demand. At Intel, we've achieved significant cost reductions from refreshing older compute and storage servers with fewer higher-performing servers based on the latest Intel® Xeon® processors and Intel® Solid-State Drives (Intel® SSDs).

In some cases, we've had individual data centers near or over capacity while others are under capacity. We've solved this by load balancing between data centers globally and spreading workloads across multiple data centers. Re-routing traffic geographically can help ensure business continuity and provide a seamless user experience while meeting demand. In other situations, our business groups have chosen project locations based on where both the project team expertise and the data center capacity are located near each other. Once data center constraints are identified, the capacity mitigation plan may include a combination of these and other approaches.

## CONCLUSION

**Our predictive data center capacity planning and LRP process enables us to better support the computing needs for our key business segments and accommodates change in response to fluctuating market demands.**

This approach brings visibility to data center capacity issues, enabling IT and our business segments to strategically address the shortfalls through a systematic and repeatable process. Our business segments can adjust and control their demand for data center capacity, alter spending forecasts, plan new design projects, obtain server refresh funding, or create expansion or reduction plans. The analysis reports show a snapshot that helps business groups make decisions and take action within a 6-month cycle. Business segments and data center planners can refine both data center plans and business plans to avert issues and increase efficiency of data center resources.

Running out of data center capacity can impact business goals and budgets as well as overall growth and innovation. In a dynamic business environment, our predictive data center capacity planning approach is crucial to delivering uninterrupted computing capability and business continuity across Intel.

## FOR MORE INFORMATION

**Visit www.intel.com/it to find white papers on related topics:**

- "Facilities Design for High-density Data Centers"
- "Solving Intel IT's Data Storage Growth Challenges"

### ACRONYMS

| | |
|---|---|
| AHU | air-handling unit |
| SSDs | Solid-State Drives |
| LRP | long-range planning |
| UPS | uninterruptible power supply |

**For more information on Intel IT best practices, visit www.intel.com/it.**


Printed in USA ♻ Please Recycle 0913/WWES/KC/PDF 329130-001US