Spring 2012

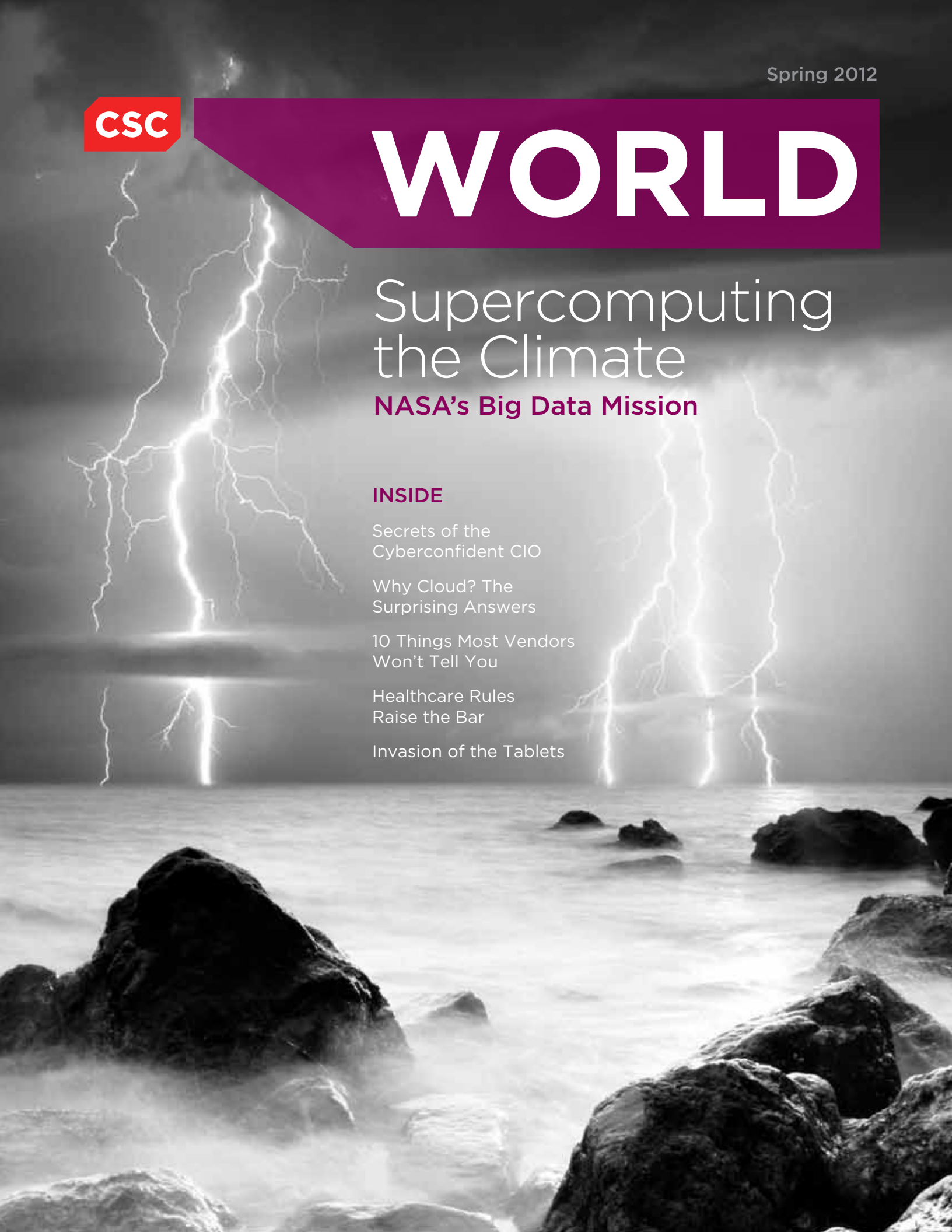
CSC

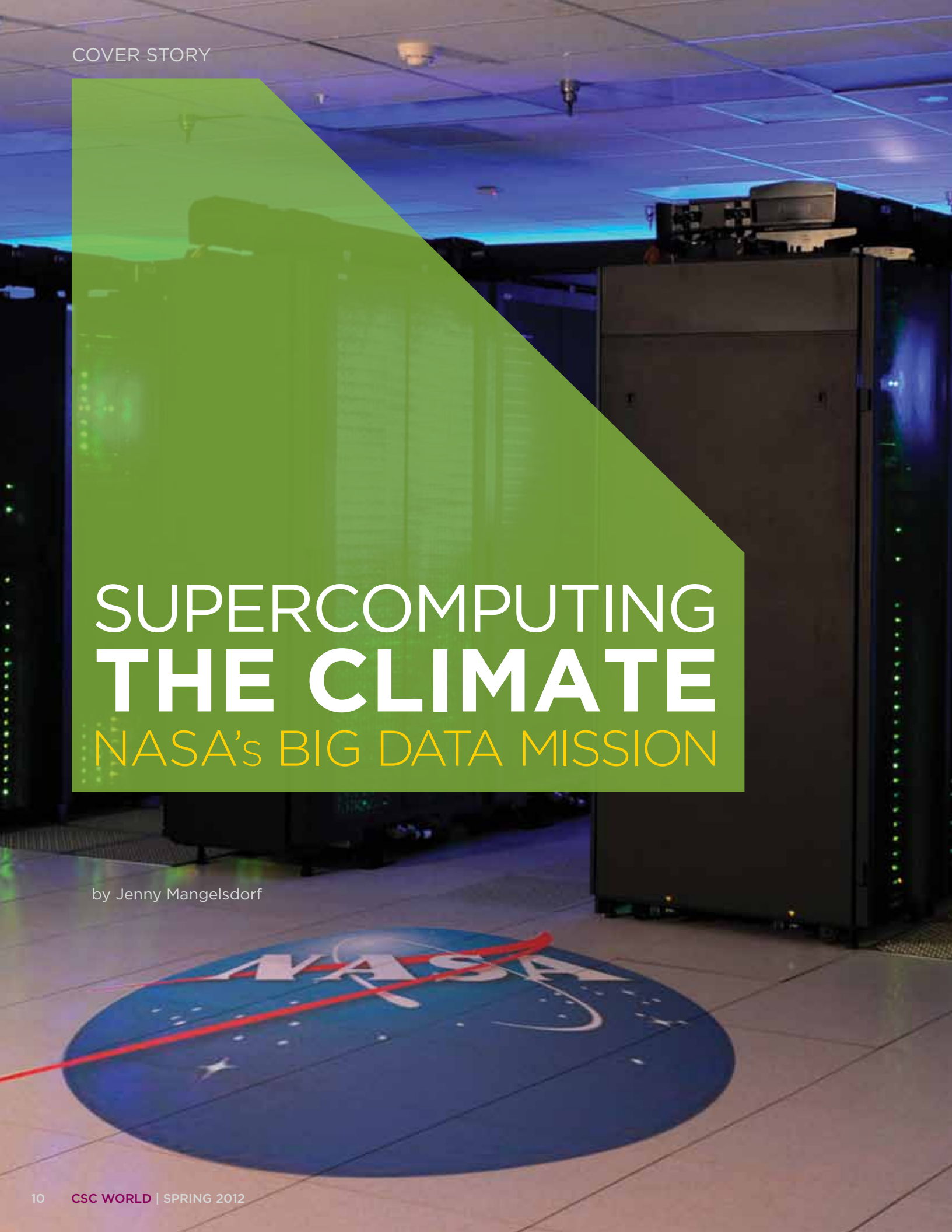
WORLD

Supercomputing the Climate

NASA's Big Data Mission

INSIDE





SUPERCOMPUTING THE CLIMATE

NASA's BIG DATA MISSION

by Jenny Mangelsdorf

Science needs data — and today's technologies are giving researchers more data than ever before. But making sense of all that data requires computing power on an extreme scale.

The NASA Center for Climate Simulation (NCCS) crunches massive amounts of climate and weather information, giving researchers eye-opening visibility into their data.

The center is based at NASA's Goddard Space Flight Center in Greenbelt, Md. Home to one of the largest contingents of earth scientists in the world, Goddard uses the NCCS' thousands of compute nodes for batch and interactive analysis. The facility has several groups of computers, each of which is tasked with a particular aspect of data-intensive supercomputing.

The NCCS Discover supercomputing cluster, which ranks among the top 100 supercomputers in the world, plays a central role in NASA's earth science mission and is the main system used for processing jobs that require significant computing resources.

Powering climate studies

The NCCS integrates supercomputing, visualization and data-interaction technologies to support research for more than 500 scientists at NASA centers, as well as researchers at laboratories and universities around the world.

"The computer is the climate scientist's tool — the better the tool, the better the scientific results, and the greater the understanding of what's happening in the complete earth system," says Phil Webster, head of Goddard's Computational and Information Sciences and Technology Office. "A key challenge for us is to build better machines because what we need doesn't exist."

"We can't just pick up a commercial off-the-shelf system, plug it in and we're good to go," adds Fred Reitz, CSC NCCS Support operations and deputy program manager. "We have to create the system image and all that goes along with it." For example, one of the center's latest scalable computer units that CSC helped build has 1,200 nodes and 14,400 CPUs.

CSC helps NCCS operate, maintain and improve its supercomputing systems. In the past five years, CSC has helped increase Discover's performance 130-fold. Today, it uses more than 35,000 processing cores to crunch more than 400 trillion floating-point operations per second. By comparison, it would take every person on Earth adding pairs of seven-digit numbers at the rate of one per second more than 17 hours to do what Discover can do in one second.

Working with Big Data

Besides having to build extremely powerful systems, another challenge the center faces is data management, or more accurately, Big Data management. Scientists using the center integrate millions of observations collected daily, reanalyze past observations and perform climate-model simulations, each of which can produce massive amounts of data. CSC helps administer Discover's archive system, which stores about 32 petabytes of data, with a total capacity of 37 petabytes. One petabyte equals one quadrillion bytes, or 1,000 terabytes.

"The Big Data problem is like finding a needle in a needle stack," says Scott Wallace, CSC NCCS Support program manager. "Finding your needle in a pile of 32 trillion needles is not significantly harder than finding it in a pile of one trillion needles because they're both effectively impossible, unless you build in a way to keep track of where each needle is located."

As the center generates and manages increasing quantities of data, it has turned to visualization technologies to help scientists see their research. A recent addition to the center is its Visualization Wall, driven by 16 Linux-based servers. These servers split images across the 17-by-6-foot wall, creating one huge, high-resolution medium on which scientists can display still images, video and animated content from data generated on Discover.

Phil Webster, head of Goddard's Computational and Information Sciences and Technology Office

"The wall gives scientists an important new tool because it lets them see their research in incredible detail," says CSC's Reitz. "It's pretty breathtaking to look at the climate models with the degree of resolution the high-performance cluster provides."

CSC, which has supported the center since 2000, helps the NCCS enhance the supercomputer cluster's compute and communications capabilities, and continuously add more processing cores to increase Discover's resolution.

As the center improves its capabilities, researchers continue to ask for more. For example, several groups of scientists have more than doubled their workload requests because of upcoming deadlines on key projects such as the Intergovernmental Panel on Climate Change's (IPCC) Fifth Assessment Report, which will provide an update of knowledge on the scientific, technical and socioeconomic aspects of climate change.

The two primary groups that use the NCCS are NASA's Global Modeling and Assimilation Office (GMAO), which aims to maximize the impact of satellite observations in climate, weather and atmospheric composition prediction; and the Goddard Institute for Space Studies (GISS), which researches global change, addressing natural and manmade changes in the environment that occur at various times and affect the planet's habitability. Both groups are major data providers to the IPCC's Fifth Assessment Report, which will be released next June.

"Both of these organizations can only sample the kinds of things they want to do," says NASA's Webster. "It's clear that the more computing they can get, the better the scientific results will be, because if they can work faster, they can evaluate more parameters and include physical processes that they may not have been able to include before to get a greater understanding of what's happening in the complete earth system."

To meet the IPCC's timeline, Discover is providing resources for thousands of simulation-years. Together, GISS and GMAO typically run more than 100 concurrent jobs, using more than 10,000 cores on Discover to simulate the breadth of assessment scenarios for greenhouse gas, aerosol and land-use changes.

Stretching the computing envelope

Today, Discover can compute in one day three simulated days in the life of the Earth at one of the highest resolutions ever attained — about 3.5-kilometer global resolution, or about 3.6 billion grid cells. The center's current "stretch" goal is to generate in one day a computation that covers 365 days at 1-km global resolution.

"Just in terms of electricity, that one computation would require 16 megawatts of power the way things are done today," says CSC's Wallace. "This isn't within reach now, but that's our distant grail. We're forever looking for better resolution and faster times."

Recently, CSC helped the center reach a new benchmark when Discover ran the highest resolution atmospheric simulation of its kind, modeling two years of the Earth's climate at 10 km globally. Being able to model the Earth's climate at this resolution is like giving scientists a high-resolution image to study versus a blurry one.

"To be able to run at resolutions like this is an example of taking a large step forward both technically and for climate science," says NASA's Webster.

To achieve advances like these, the center also taps CSC's High Performance Computing Center of Excellence for assistance. Established in 1999, the CSC center has more than 170 specialists operating systems that collectively provide capacity for more than 110 petabytes of data and have a capability of almost two petaflops, or two thousand trillion floating-point operations of computation per second.

The CSC center continually looks for new advances it can leverage to gain new efficiencies and overcome current challenges, such as how to take a month's climate data, pull out the temperature for a specific part of the world, calculate the average and analyze that data against multiple factors, such as soil and atmospheric temperature, to determine climate impact on that region — all in a reasonable amount of time.

"Climate research continues to stretch computing capabilities," says Donna Klecka, director of CSC's High Performance Computing Center of Excellence. "Through our center, we can further support NASA's center, bringing our deep computing expertise to innovate and create Big Data solutions that its climate scientists need." ■

Learn more about CSC's work with NASA at www.csc.com/NASA.

JENNY MANGELSDORF is a writer for CSC's digital marketing team.

CSC, DOE TEAM ON CLIMATE INNOVATION

CSC and the U.S. Department of Energy's Oak Ridge National Laboratory (ORNL) have agreed to collaborate on solutions that will help customers address the potentially adverse effects of climate change by using climate knowledge, data and tools.

Through the arrangement, CSC and ORNL, which is the DOE's largest science and energy laboratory, will explore projects of mutual interest, applying climate science and technologies to model the Earth system, analyze climate and environmental data, and develop decision support tools and actionable information.

CSC will work with ORNL's Climate Change Science Institute (CCSI), based in Oak Ridge, Tenn. The institute houses ORNL's significant expertise in climate change science, modeling and data systems, while CSC brings expertise in application and decision-support tool development, Big Data management and network integration.

"Partnering with companies such as CSC, which has a broad list of clients and understands their decision and information needs, enhances our ability to translate our world-class climate science findings into analyses and products that the public and private sectors need to be competitive in a changing business climate," says Jim Hack, CCSI director.

Changes in climate, which can cause rising sea water levels and drought, for example, will ultimately affect decisions made in a variety of industries, ranging from agriculture to civil construction. Both the public and private sectors are looking to determine the short- and long-term risks brought about by natural climate variability and long-term climate change.

In addition to serving as a major climate-modeling center, ORNL has considerable expertise in applied research, climate decision support and risk management, and supports DOE missions in scientific discovery and innovation, clean energy and nuclear security.

CSC's climate services span research and operations, including data collection, management, analysis and application solutions. The company also has a broad range of expertise in key fields related to climate impacts, such as air quality modeling, toxicology, chemistry, microbiology, geology, oceanography, and chemical, environmental and infrastructure engineering.

This new cooperative agreement enhances CSC's efforts to help its clients, such as those in the financial services and process manufacturing industries, make greater use of the deep expertise and information developed by federal and academic scientists in their study of extreme climate and weather conditions.