



# MONITORING THE SURFACE OF THE SUN WITH NASA'S SOLAR DYNAMICS OBSERVATORY, Z BY HP, AND NVIDIA

Author: Mike Leone, ESG Senior Analyst; and Josh Clark, ESG Research Associate



## ABSTRACT

At NASA, data is everything. From object detection to mission enablement, data collection and rapid insight are paramount to mission success. And the challenge with analyzing the data is not just due to its size, but its type. The data is quite literally out of this world, including images of galaxies, stars, and planet surfaces from across the solar system. As NASA and its contractors look for faster and more reliable ways to collect

and analyze data, they are increasingly turning to artificial intelligence (AI) as the answer. ESG recently spoke with NASA scientists to understand their AI efforts and how Z by HP has enabled them to transform the way they analyze data.

## NASA AND AI

Based on the amount of data collected for any given NASA mission or project, it's no surprise that AI is an enabling technology. AI is primarily used today to help with the detection of "things" and model enhancement. At NASA, AI is leveraged to monitor regions of space and automatically detect if something interesting just happened or is going to happen. Based on collected data and scientific models that describe a physical process, AI can be exploited to make scientific models more robust and reliable by enabling wider parameter exploration.

NASA is just scratching the surface in its use of AI for mission enhancement and enablement. The amount of data collected on a spacecraft to finetune humanity's understanding of solar physics and derived models is massive, amounting to tens, if not hundreds of TBs per day. In fact, it's impossible to analyze all of this data simply due to its size and the limited computing capabilities on a spacecraft hundreds of miles away, all but forcing NASA to prioritize sending certain types of telemetry data down to earth, which vary depending on the mission. This presents opportunities for AI to be utilized on board satellites, rovers, planes, or balloons.

### Monitoring the Sun's Atmosphere

ESG spoke with Michael Kirk (ASTRA llc.), research astrophysicist, and Raphaël Attié (George Mason University), solar astronomer at NASA's Goddard Space Flight Center to understand their mission of monitoring the sun's atmosphere and the type of technology they rely on to complete this ongoing mission using AI. The first aspect of their mission is classification. They leverage AI to map localized plasma motions on the sun to classify regions based on magnetic activity, enabling them to focus on areas prone to solar flares. And while the unimaginable temperatures and heat that a solar flare lets off are not necessarily impactful to the earth or its atmosphere, electromagnetic radiation

and energetic particles can enter and alter the Earth's upper atmosphere, disrupting signal transmission from GPS satellites that orbit the earth. In addition to monitoring active magnetic regions of the sun, the scientists also focus on mapping the flow of plasma on the sun. For this, the team collaborates with Benoit Tremblay, scientist at the National Solar Observatory (NSO), to use a deep neural network trained on simulation data to understand plasma flows and their relation to magnetic fields.

## WORKFLOWS AND CHALLENGES

The Solar Dynamics Observatory collects data by taking images of the sun every 1.3 seconds. Kirk and Attié have repurposed an existing algorithm that remove errors from the images, such as bad pixels, and study the repository that is growing every day. To understand the magnitude of this task, it may help to know that tens of petabytes of images, totaling over 150 million error-files and 100 billion individual detections, must be accurately sorted, labeled, and qualified as containing good versus bad pixels. One of the first problems that the team has had to overcome is the lack of computing power colocated with the data. That means data would constantly have to move between archives, local laptops/workstations, and/or remote computing clusters. Put it all together and the delays were becoming untenable, especially as more and more data was being collected and analyzed.

### Archives in the Hundreds of TBs

Due to cloud computing and network limitations, pulling out complete data archives in the 100s of TBs simply was not feasible. Between the amount of data movement that was required and the limited computing capabilities, it would take up to a few years for

the team to see any results. Additionally, for the mission monitoring and analyzing the surface of the sun, 1.5 TBs of new data was collected daily. Delays and wait time were becoming unacceptable. To counter the delays in data movement and long processing times, the team attempted to leverage a desktop workstation designed for general GPU computing with just under 1 TB of SATA-based storage, 32GB of RAM, and one NVIDIA 1080 TI GPU. But because of the largeness of the data sets, Kirk and Attié were forced to leverage external storage via USB, creating significant I/O limitations. Leveraging a cloud computing cluster would seem to be a natural solution with plenty of storage and compute. However, the cloud presented a new challenge: navigating security protocols.

### Security

Security within NASA is becoming more stringent. While data sets are public, layers of security associated with access, movement, and computational consumption can sometimes serve as a deterrent for adopting new technologies in the data science space. And while some of the new technologies and software stacks are appealing, especially in cloud environments, security processes all but forced Kirk and Attié to reevaluate on-premises AI workstations.

### Reliability and Support

Pre-trained neural net used to detect objects in normal images simply wouldn't work on the team's images of the sun. The type of images that they needed to analyze represented 1.5 TBs of daily data. NASA images are saved as a

We quickly realized [with desktop workstations], that you can have a really high-powered engine. But if you don't have the transmission (i.e. optimized software stack) to put the power to the road, you know you might as well just give up because you're never going to get anywhere.

Michael Kirk, Research Astrophysicist at NASA Goddard, ASTRA llc.

scientific filetype called FITS, as opposed to more standard image files, such as JPG and TIFF image files. While it seems only natural to turn to IT to help, NASA's IT team is regularly busy. They don't have time to constantly troubleshoot, test, and implement. The research team's software stack and data science workflows consisted of Python, using TensorFlow, Dask, CuPy, and other apps for heavy data processing; Pandas, RAPIDS, and CuDF for statistical exploration; and a variety of 2D and 3D visualization tools. After fighting the existing system for over a year and working with IT to install the right drivers to support their requirements, they scrapped the idea of leveraging the workstation and were forced back to the cloud operating model.

### Requirements for a New AI Workstation

With the previous workstation unable to provide the proper amount of storage or speed for a single data set, and with the delays in data movement between archives, devices, and compute clusters that would take more than a year to see results, it was only logical for NASA to seek an alternative that would meet their individual needs and better support their mission objectives. They needed a powerful system on-premises that could support their software stack and custom workflows, ensure security compliance was maintained, minimize data movement, and deliver the right level of performance. On top of basic storage and compute performance, parallelism was critical, ensuring not only better use of available GPU

capacity, but putting extra CPU cores to work too. And based on their last experience, both the researchers and IT wanted better support for the exiting NASA computing requirements from the vendor they selected.

## ADDRESSING NASA'S REQUIREMENTS WITH THE HP Z8 WORKSTATION

HP delivered a complete Z8 workstation that enabled the use of powerful NVIDIA GPUs and available CPUs to satisfy their software requirements and data science workflows. In a desktop-size tower, NASA gained a system with high-density storage, mixing fast NVMe disks with enterprise-class spinning disks to store dozens of TB of data, the availability of thousands of NVIDIA GPU cores through two NVIDIA Quadro RTX 8000 GPUs, and two Xeon CPUs used for data preparation and data interactivity. And for the most part, it all just worked.

### Accessibility

For NASA researchers exploring the surface of the sun, the Z by HP has delivered a way to more easily access data. And it's not only faster, it's more consistent.

*"When you are in a large organization, you often cannot predict how fast your network is going to be. Sometimes I was accessing at 10 MB/s, sometimes at 1 MB/s, sometimes even less. I'm now able to store the dozens of TB data I need locally, and have it available all at once for statistical analysis that requires looking at all of it, and I can do so by leveraging either CPU-based or GPU-based parallelism, without worrying about a monthly cost of a cloud-based solution. This is a much less stressful*

*workflow, as I'm not worried about getting rid of data that took ages to download and that are sometimes not permanently available in the remote data archives that we use."* - Raphael Attié, solar astronomer at NASA Goddard

### Interactivity

Because data was now local on the Z8 workstation, it was more accessible. And because it was more accessible, interaction with the data became easier. What used to take days to move, process, move again, and interact with data is now taken for granted. Instead of leveraging different systems for processing, analyzing, and interacting, the team can leverage the Z8 system to fulfill all tasks. And the effect was a reduction in time to results. What had previously taken or was expected to take a year or more could now be done in less than a week.

*"Because it [the remote compute cluster] is not a visually interactive system, I had to download intermediate results locally to interact with a subset of what was processed. This was preventing any smooth workflow. Now I can include both computing and visualization within one pipeline. It gave me back more direct access to my science."* - Raphael Attié, solar astronomer at NASA Goddard

### Performance

Given the previous struggles to move 12TB chunks of data, the team knew that performance was key for any statistical analysis. It took weeks to download data, and because it was stored externally, the compute power could not leverage parallelism (either CPU or GPU). In the most extreme case, using the HP Z8 Workstation, they turned what would have taken several months of purely compute processing into a job that could finish in barely a week's time.

*"My typical benefit [using an HP Z8 workstation] is a speed up of about 1 order of magnitude with respect to my previous workflow using remote computer clusters. I'm able to train neural nets in seven minutes when it was taking an hour on a [remote] computer cluster, and inconceivable on a laptop."* - Raphael Attié, solar astronomer at NASA Goddard

The team witnessed one very specific example of a performance gain when comparing the results of computing on the CPU cluster with the NVIDIA Quadro RTX 8000. For about 130,000,000 files, each containing thousands of coordinates of energetic particles hitting the camera onboard the Solar Dynamics Observatory (SDO), it took approximately three seconds to analyze each file. Linearly extrapolating that measurement over 50 cores typically used in their remote compute cluster equates to 90 days to fully analyze the entire data set. With the NVIDIA Quadro RTX 8000, they could analyze each file in 20ms. That proved to be a 150x improvement and took the time to compute the entire data set from 90 days down to 1.5 hours.

### Flexibility

Understanding that some problems are better solved by a CPU or a GPU, the new Z by HP workstation enabled Kirk and Attié to rapidly prototype software and workflows in multiple compute environments all on the same box.

*"Trying out different workflows is immensely beneficial. In the cloud environment you don't have that flexibility. Now we can prototype simultaneously in a big sandbox, so we go from an idea to a prototype significantly faster."* - Raphael Attié, solar astronomer at NASA Goddard

### Support

Troubleshooting computer technology is not a scientist's job. As such, Kirk and Attié required assurance that if something went wrong, it could be fixed quickly. Whether upgrading a component or updating a driver, technology support from Z by HP was pretty much instantaneous.

*"We didn't have to rely on it [Z by HP support] a lot. The internal HDDs weren't big enough to start and support was right there. Putting in a new internal drive is not always easy. The new drive was delivered quickly, and we were able to simply plug and play the new drive. It just worked."* - Michael Kirk, Research Astrophysicist at NASA Goddard

### Security

By not having to leverage the cloud for compute, the team no longer had to deal with IT security delays based on what data was being moved into the cloud and what software was being used in the cloud. While they must maintain an effective security posture on premises, the system gives them more freedom to use the software they want to do the job they need to most effectively.

*"The layers of security that keeps piling up is a deterrent for the adoption of this industry, it defeats the original purpose of having something faster and more convenient to crunch through big data sets where each can be up to several TB. Now, when I compute TB-sized datasets locally, my momentum is not broken by the IT security bottlenecks of a large organization."* - Raphael Attié, solar astronomer at NASA Goddard



## THE BIGGER TRUTH

As data sets continue to grow and data movement further restricts the timeliness of results, the scientists at NASA's Goddard Space Flight Center studying the surface of the sun were at an impasse. They could accept the time it took to leverage a cloud computing cluster and deal with constant delays in moving data in and out of systems or look for an alternative. The Z by HP workstation proved to be the solution, providing

a powerful technical foundation to enable better, faster, and interactive data analysis more collaboratively on growing data sets. As NASA continues to turn to AI to help transform how they analyze otherworldly data, it's a good bet that HP's Z8 workstation with NVIDIA GPUs will be there delivering the performance, flexibility, and reliability required to continue exploring the next frontier.

### THE Z BY HP SYSTEM USED BY THE NASA RESEARCHERS INTERVIEWED FOR THE CASE STUDY IS CONFIGURED AS FOLLOWS:

#### Hardware:

- HP Z8 G4
- Dual NVIDIA® RTX 8000 graphics
- 384GB (12 x 32GB) DDR4 2933 ECC registered memory
- Dual Intel® Xeon® 6240 2.60GHz 18C CPU
- HP Z Turbo Drive M.2 1TB TLC SSD
- 2 x 4TB 7200 SATA Enterprise 3.5in
- Premium Front I/O (2 x USB3.1 Type C, 2 x USB3 Type A)

#### Software:

- Ubuntu Linux® 18.04
- Python 3.6 (TensorFlow, NumPy, among other scientific computing packages)



**Enterprise Strategy Group** is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

LET US HELP YOU CREATE SOME AMAZING BUSINESS SOLUTIONS TODAY

CONTACT US



© Copyright 2021 HP Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Intel and Xeon are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries. NVIDIA is a trademark and/or registered trademark of NVIDIA Corporation in the U.S. and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. All other trademarks are the property of their respective owners.

4AA7-9682ENW, January 2021

