NASA'S DAY IN THE SUN: SPACE AGENCY SPEEDS ANALYSIS OF SOLAR IMAGES BY 150X USING DATA SCIENCE WORKSTATIONS

Scientists accelerate data analytics and computations that would've taken years on CPUs to less than a week with RTX-powered Z by HP data science workstations.

Featured image, courtesy of NASA, is a visual of the sun from SDO on May 16, 2012

NASA is using Quadro RTX GPUs to shine some sunlight on data analytics.

The U.S. space agency's Solar Dynamics Observatory collects images of the sun to help scientists and researchers gain insight into the different types of solar variations and how they affect life on Earth.

This data is a valuable asset for the research community, but with more than 18 petabytes of images collected, analyzing this information is a massive challenge.

With Quadro RTX-powered Z by HP data science workstations, however, the NASA team can easily sort through the data and analyze images up to 150x faster than on CPUs.

NASA's Big Data Challenge

The observatory collects data by taking images of the sun every 1.3 seconds. Researchers have developed an algorithm that removes errors from the images, such as bad pixels, then places them into an archive that's growing every day.

The algorithm is highly accurate, but with nearly 20 petabytes of images, there are billions of pixels that have been misclassified as errors. So, the NASA team needed to comb through 150 million error-files, in all containing about 100 billion individual detections, and find a way to sort and label the good pixels versus the bad ones.

With conventional computing, it was nearly impossible — using a CPU would take up to a few years to see any results. Even with the best multi-threaded CPU algorithm they could create, it would

still take about a year to compute and analyze all the data.

"For scientists, a year still wouldn't be enough time because we like to explore and iterate the results we find," said Raphael Attie, solar astronomer at NASA's Goddard Space Flight Center. "Even with one year of computation, it would still take us up to 10 years to find concrete results."

Needing results in a much shorter time frame, the NASA team started looking at the parallel processing capabilities that were available using NVIDIA GPUs.

Big Data Gets a Bigger Solution

Supercomputing resources at NASA are heavily restricted — researchers need to provide details as to how much computing resources they require

and how long they'll need to use it. However, this becomes challenging when the team isn't sure how much computing resources, they need in order to experiment with massive amounts of data.

But with the Z8 G4 desktop workstation powered by two Quadro RTX 8000 GPUs, the NASA researchers were able to get supercomputing resources right at their desks. They started to explore the project using big data analytics techniques and using NVIDIA's accelerated computing libraries to fully unlock the power of NVIDIA GPUs.

The data science workstations allowed the team to analyze the images and achieve results in less than a week.

"The data science workstations completely changed the field of possibility for us," said Michael Kirk, research astrophysicist at NASA. "These computations that previously weren't imaginable, we can now do 10-150x faster than we thought possible."

The NASA team conducts a broad range of work, leveraging AI, machine learning and data analytics to learn the sun's secrets. Most of their data science workflows are based in 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.

With the data science workstations, the team utilizes the power of GPUs to enhance their analytics workflows, allowing the researchers to explore and iterate calculations to get quicker results.

Once the NASA team completes their project of filtering and analyzing the current data, their next step is to use this information to analyze other pixels that were initially marked as good to make sure that they really are good in order to validate the entire data set.

A Change of Space for GPUs

In AI and big data analytics, projects can be severely impacted from non-responsive workflows in cloud environments. These interruptions break momentum, productivity and motivation in the long run. This is why Attie recommends having a local GPU-powered workstation or laptop that has enough memory to accommodate a subset of your data processing for comfortable prototyping.

"I find that a necessary condition for a responsive workflow is to have the input data rapidly accessible by your GPU devices," said Attie. "If it's not possible to have the data locally in the same machine as the GPU device, the network needs to be very fast and resilient, as AI applications often need fast access to the data."

The results of Attie and Kirk's projects get shared through publications and specialized journals. During seminars and conferences, they'll have discussions with colleagues and deliver presentations on how they obtain data with specific frameworks or customized codes. And as more people are working from home, the NASA team is getting more familiar with remote tools to connect with others and share findings from their latest projects.

THE HP Z8 G4 SELECTED FOR THE PROJECT FEATURED THIS CONFIGURED SOLUTION:
- HP Z8 G4 1125W PSU
- Dual Intel® Xeon® 6240 2.60GHz 18C CPU
- Dual NVIDIA® RTX 8000 graphics with 128 processing cores
- 384GB (12 x 32GB) DDR4 2933 ECC registered memory
- 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)

LET US HELP YOU CREATE SOME AMAZING BUSINESS SOLUTIONS TODAY

CONTACT US