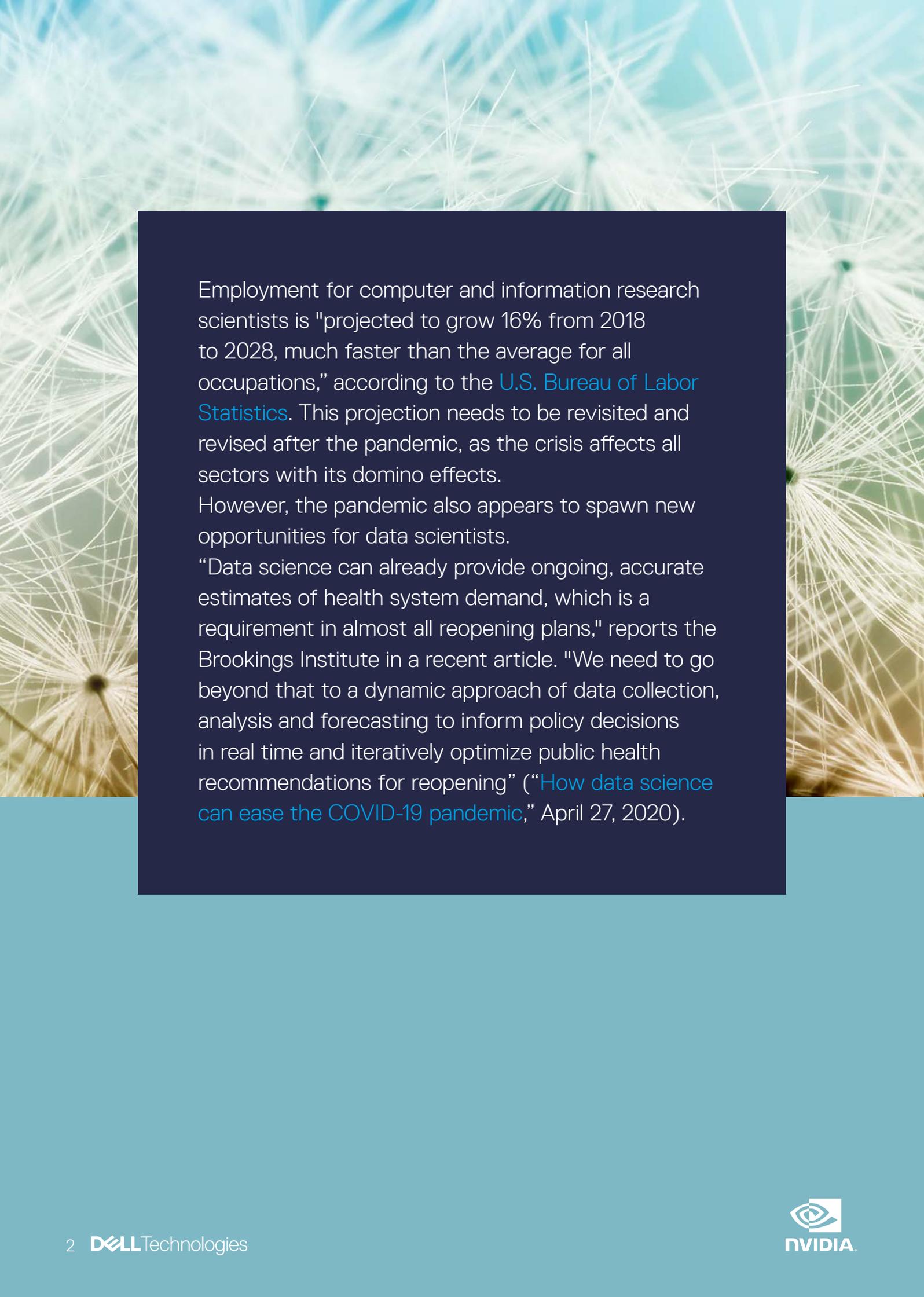




# DATA SCIENCE FINDS NEW ROLE IN THE FIGHT AGAINST COVID-19

GPU-accelerated data visualization and analysis could lead to a better understanding of the virus.



Employment for computer and information research scientists is "projected to grow 16% from 2018 to 2028, much faster than the average for all occupations," according to the [U.S. Bureau of Labor Statistics](#). This projection needs to be revisited and revised after the pandemic, as the crisis affects all sectors with its domino effects.

However, the pandemic also appears to spawn new opportunities for data scientists.

"Data science can already provide ongoing, accurate estimates of health system demand, which is a requirement in almost all reopening plans," reports the Brookings Institute in a recent article. "We need to go beyond that to a dynamic approach of data collection, analysis and forecasting to inform policy decisions in real time and iteratively optimize public health recommendations for reopening" ("[How data science can ease the COVID-19 pandemic](#)," April 27, 2020).



# THE DATA SCIENCE WORKSTATION

Data science is a mixture of big data, hardware and software. Data sets collected over a long period of time tend to be large. If the raw data involves, graphics, audiovisuals, and 3D data (which is usually the case with autonomous vehicle developers), loading and processing it on a consumer-class PC is out of the question. It may even bring a standard CAD workstation to its knees.

This is the reason professionals in the field are turning to purpose-built hardware, such as the data scientist's workstation offered by GPU maker NVIDIA and its partners. The special workstation was introduced by NVIDIA in March 2019, at the annual [NVIDIA GPU Technology Conference \(GTC\)](#).

“Enterprises are eager to unlock the value of their business data using machine learning and are hiring at an unprecedented rate data scientists who require powerful workstations architected specifically for their needs,” says Jensen Huang, founder and CEO of [NVIDIA](#). “With our partners, we are introducing NVIDIA-powered data science workstations—made possible by our new Turing Tensor Core GPUs and CUDA-X AI acceleration libraries—that allow data scientists to develop predictive models that can revolutionize their business.”

“Enterprises are eager to unlock the value of their business data using machine learning and are hiring at an unprecedented rate data scientists who require powerful workstations architected specifically for their needs,



# DATA-HUNGRY WORKFLOWS

At [Big Compute 20](#), held in February in San Francisco just before the lockdowns began, [Microsoft](#) and its partner [Ansys](#) teamed up to discuss the use of simulation in autonomous vehicle development.

“You have machine-learning scientists, data scientists, who have created these perception models. You are constantly running these driving scenarios you have collected through these AI (artificial intelligence) models, constantly tweaking your AI models to have better accuracy. Once you have some sense of what your accuracy is, you want to verify that before you put it in your vehicle,” says Nidhi Chappell, head of Product, Microsoft Azure.

These models involve the car making autonomous decisions based on real-time sensor data, programmed policies, and physics-base simulations of vehicle movement and speed. “Generally, a Level 2 car will have eight or more sensors. To validate the sensor model, you need

2,200 CPUs, and 200 GPUs. That’s just for a Level-2 car. When you go to a Level 4 or 5, you are looking at 30 sensors,” says Chappell.

The compute-insatiable workflow points to the rise of distributed computing at a much larger scale, with the data center as the backbone.

“*Generally, a Level 2 car will have eight or more sensors. To validate the sensor model, you need 2,200 CPUs, and 200 GPUs. That’s just for a Level-2 car. When you go to a Level 4 or 5, you are looking at 30 sensors*

NVIDIA Clara for healthcare includes the use of AI-based models, developed jointly with the National Institutes of Health. Image courtesy of NVIDIA



# DATA SCIENCE AND DATA CENTERS

Due to the lockdowns, this year's NVIDIA GTC keynote was delivered in prerecorded videos, shot in NVIDIA CEO Huang's home. (It was his "first-ever kitchen keynote," he quips.) "The applications we're using now are so large they don't fit in any computer," Huang observes. "In fact, the server is no longer the computing unit. The data center is the computing unit."

The next NVIDIA GPU called the NVIDIA A100, based on the new Ampere architecture, "provides up to 1.5 Terabytes frame buffer bandwidth. This is the first processor in the history that comfortably provides over a Terabyte per second of bandwidth," says Huang.

In cluster environments, the A100 GPUs can be linked together using the third generation NVIDIA NVLink and NVSwitch technologies, making them function like a single entity with faster

inter-GPU communication. NVIDIA Drive, the GPU maker's hardware-software combination for autonomous vehicle training and development, will be upgraded with Ampere-class GPUs for better performance.

“*The applications we're using now are so large they don't fit in any computer*”

# DATA SCIENCE AND COVID-19

Joining the quest for a vaccine, NVIDIA recently made certain software tools available for free to data scientists analyzing and studying virus data. The company offers 90-day licenses of its genome-sequencing software Parabricks to qualified scientists.

The NVIDIA Rapids team, devoted to developing CUDA-based software libraries to execute data science and analytics pipelines, uploaded Plot.ly's Dash, a virus spread visualization tool, to GitHub.

"Oxford Nanopore, for example, was able to sequence the genome of the virus in just seven hours using our technology," said Huang in his keynote. "Working with Plot.ly, we can now do real-time infection rate analysis."

As states and countries begin to ease their lockdowns to restart the economies that have been put on pause, a new normal is starting to emerge. For the foreseeable future, apps for real-time tracking of infection, autonomous robots to deliver food and monitor patients, and the race to develop a vaccine or a cure appear to be part of life. In all of these, data science is the backbone and the brain.



## MORE INFO

[GPU-accelerated Covid-19 Tools](#)