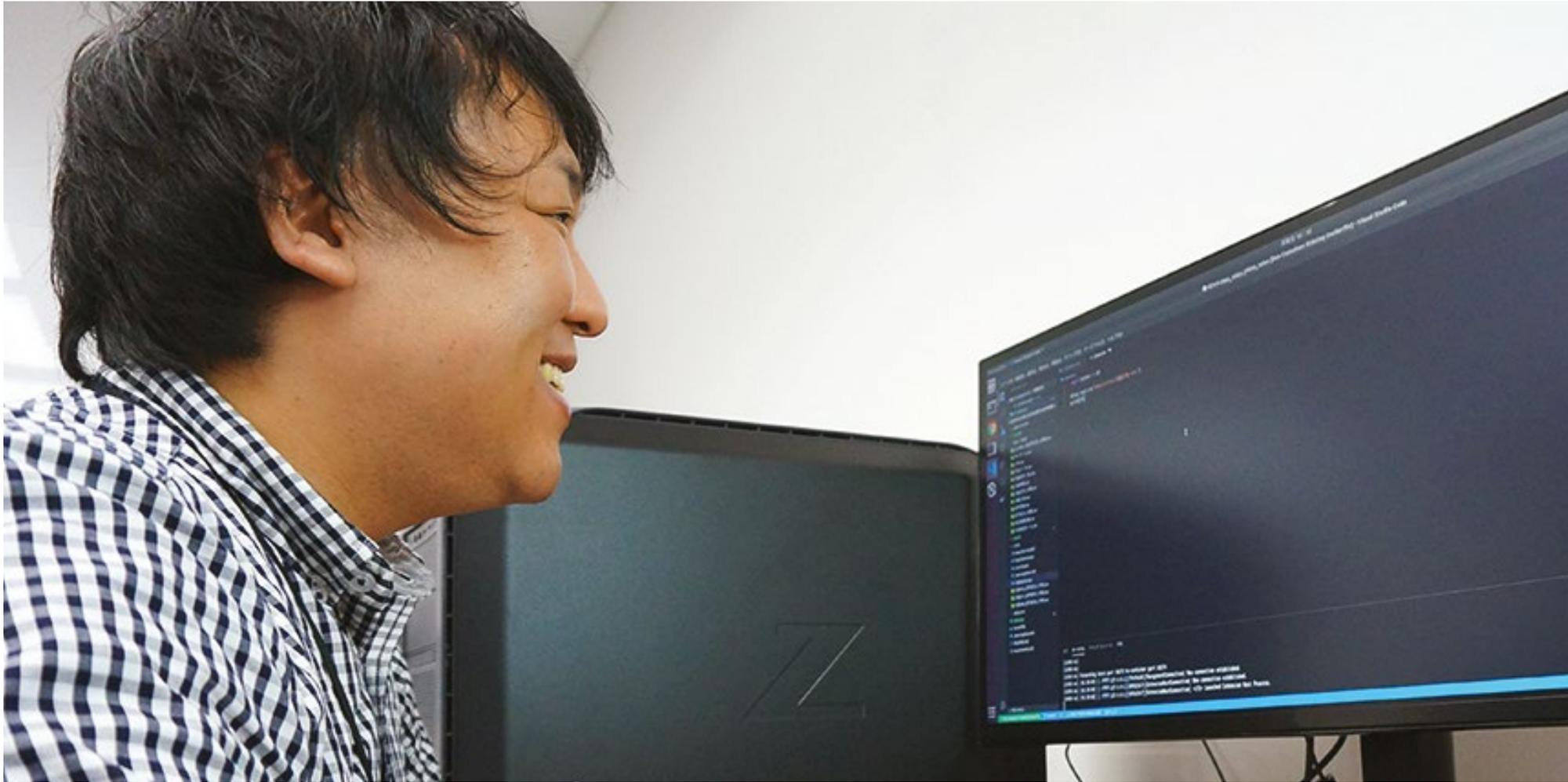RETAILERS RELY ON RELIABLE SALES
FORECASTS BACKED BY AI DEVELOPMENT

Information technology has become popular in the Japanese retail industry. Self-registration and semi-self-registration, which are being introduced in many supermarkets, are typical examples.

VINX Corporation, which provides many IT solutions for such retailers, creates intelligent solutions using HP workstations in its system development environment.

IT makes up for human resources shortage

VINX is a company that provides a wide range of IT services to support the business activities of distribution companies like retailers. The company continues to develop systems to offer timely, secure, and reasonable services to industries that are susceptible to change.

"There is one big issue in the retail industry in recent years. It is a shortage of human resources, and it is difficult for people to gather at any company or store. Therefore, it is necessary to promote operational efficiencies using IT," says Kazuki Noto, Product Planning Department, Sales Headquarters at VINX.

The three key elements of the retail business are the cash register for directly exchanging money with consumers; ordering for replenishing goods as needed so that sales are not interrupted; and display for arranging delivered products on the shelves.

"The cash registers are the most IT-enabled. I think that the cash register is already being automated even at your local supermarket," Noto says. "However, displaying products on shelves must be lined up directly based on visual appeal to humans, so it will be difficult to automate this process until high-performance robots become more widely used. The other task of ordering goods actually requires large skill differences based on individual abilities, and until now, it has been difficult for IT to intervene."

When ordering a product, the person in charge predicts and analyzes consumption behavior in a few days, considering factors including the weather, season and people's tastes, ultimately settling on the appropriate number of items. In many cases, these personnel rely on their expertise and intuition, and those who are said to be good at ordering rely on their innate senses.

"It was a so-called personal task, but with the advent of machine-learning technology, it has become possible to make highly accurate predictions. In other words, the era in which IT is finally useful in ordering operations has arrived, " Noto adds.

An ordering business system with artificial intelligence

Retail forecasting is unique in that each end user has very different feature value.

"In order to improve the accuracy of the forecast, it is necessary to generate various features and try combinations to build the optimum model," Noto says.

To achieve this, it is essential to properly combine libraries that enable high-speed processing using GPU computing with libraries that do not.

"As a development environment, we need a computer with a powerful CPU and a computer with a powerful GPU," Noto says. "You can proceed while rearranging computer resources in the cloud, but the higher the specifications, the higher the usage fee. It's a drawback. It also puts pressure on developers because they tend to work with time efficiencies in mind to keep costs down."

VINX tried the HP Z8 G4 Data Science Workstation (with NVIDIA GPU acceleration) which is equipped with the necessary tools, and the optimum specifications—operating system, software, and the like—so that data scientists with ever-increasing needs can greatly improve efficiency from pre-processing to learning.

The Z8's processor is equipped with dual Intel Xeon 6240. The GPU has two NVIDIA RTX 8000s connected by NVLink and a vast GPU memory of 96GB. At 384GB, the main memory allows for maximum response even when a heavy load and altitude calculation are continuously processed.

"IF WE HAD FOUR MORE HOURS IN THE EVENING TO FINISH THE PROCESS, WE WOULD HAVE TO WORK OVERTIME UNTIL THE CALCULATIONS WERE COMPLETE TO AVOID UNNECESSARY EXPENSES WITH HOURLY LOANS. BUT WITH THE Z8 IN HAND, IF THE SAME SITUATION ARISES, WE CAN SIMPLY LOOK AT THE RESULTS FIRST THING THE FOLLOWING MORNING, SO WE DON'T HAVE TO WORK OVERTIME."

Kazuki Noto, Product Planning Department, Sales Headquarters, VINX

> "IN ACTUAL DEVELOPMENT, THE PROCESSOR RUNS 100 PERCENT OF THE RESOURCES CONTINUOUSLY FOR A WHOLE DAY OR MORE, WITH NO ISSUES. THIS HIGH STABILITY IS FOR DEVELOPERS, AND IT'S VERY REASSURING AND RELIABLE."

Kazuki Noto, Product Planning Department, Sales Headquarters, VINX

"For forecasting for the retail industry, we mainly use GBDT (Gradient Boosting Decision Tree) and combine various libraries such as a neural network system library, but this machine does not have to switch one by one," Noto says. "This computer can process both CPU-only libraries and GPU-compatible libraries at high speeds."

In addition, the workstations can be operated on site, freeing them from time constraints.

"For example, if we had four more hours in the evening to finish the process, we would have to work overtime until the calculations were complete to avoid unnecessary expenses with hourly loans. But with the Z8 in hand, if the same situation arises, we can simply look at the results first thing the following morning, so we don't have to work overtime."

Improving efficiency in processing capacity to over one-half

Though it depends on the type and amount of calculation to be processed, in some cases, processing capacity can be improved by about a third to one-half compared to the conventional system, he observes. There is not a benchmark because it is not yet feasible to exactly duplicate the same processing parameters during development work. However, significant efficiencies can now be achieved.

"In actual development, the processor runs 100 percent of the resources continuously for a whole day or more, with no issues. This high stability is for developers, and it's very reassuring and reliable, " Noto explains.

HP workstations like the Z8 are built for tough use. They have been commercialized through repeated rigorous tests to optimize airflow and waste heat treatment.

"When I opened the case, I was impressed when I saw that all the parts were arranged in an orderly manner. The design is easy to maintain," Noto says.

He also realizes the value of operating a computer on the premises.

"We are dealing with important customers' data," Noto says. "We are less reluctant to upload data to the cloud, but we must always manage and utilize it with the utmost care. If you use the Z8 on-premises, you can manage your data locally without having to upload it to the network. It is also advantageous that you don't have to worry about confidentiality. You can try various things and work efficiently without worrying about the billing cost resulting from the transfer time."

Expanding services for an evolving retail industry

The Z8 was easy to install and run immediately.

"I build a development environment with Linux OS on a daily basis, but the Z8 has the necessary software on the Ubuntu environment from the start. It was almost pre-installed. I only put in Microsoft Visual Studio Code," Noto adds with a smile.

Another advantage of the Z by HP Data Science Workstation is that the development environment is prepared in advance and can be used immediately after installation.

"Currently, in order to enable demand forecasting by machine learning for retailers, we are conducting a demonstration experiment to link with MDware provided by our company," Noto says. "In the future, we will predict congestion based on the number of orders and other factors. I think that the need for services that predict the appropriate number and allocation of staff on an hourly basis will also increase. As the retail industry will continue to permeate technology using machine learning, we will deliver better services in the future."

HP will continue to provide assistance to back up the VINX development environment which in turn continues to support the retail industry.

LET US HELP YOU CREATE SOME AMAZING WORKSTATION SOLUTIONS TODAY

CONTACT US

VINX