



HENG CHERKENG: MAKING DATA SCIENCE HIS LIFE'S PASSION

Kaggle Discussion Grandmaster uses Z by HP to transform data science performance, reducing processing time by more than 50% across multiple projects





“Wow, a computer program can actually do that,” was the light-bulb moment when Heng CherKeng realized data science would be his career and life’s passion. He was working as a senior staff engineer in Panasonic’s R&D center in Singapore on digital camera face-detection algorithms. It was “love at first sight” when he saw how a camera could recognize an image with accuracy.

“Ever since I was young, I have loved problem solving, especially math problems. To me, developing image recognition algorithms is like solving a math puzzle. It excites me to search and find a solution. After my face detection solution was adopted as a company product, I knew I was working in the right field,” says CherKeng.



AT A GLANCE

HENG CHERKENG, SINGAPORE

- Freelance data scientist
- Former senior staff engineer, Panasonic's R&D center, Singapore
- Specializes in deep learning, computer vision and robotics
- Bachelor of Science Degree in Electrical and Electronics Engineering, National University of Singapore
- Kaggle Discussion Grandmaster and Competition Master
- Four Kaggle gold medals and eight silvers



As a Z by HP Global Data Science Ambassador, Heng Cherkeng's content is sponsored and he was provided with HP products.

That was 20 years ago after he graduated from the National University of Singapore with a bachelor's degree in Electrical and Electronics Engineering. Now CherKeng is a freelance algorithm developer in Singapore working on a variety of deep learning, computer vision and robotics solutions for his clients. Typically, these engagements run between six months and a year. One project that he is currently working on is collaborating with a local hospital's doctor to analyze prostate cancer MRI scans to improve diagnosis and treatments.

It's no wonder that CherKeng's technology journey found him squarely

with experts in many different professions." Since joining Kaggle, CherKeng has become a Kaggle Discussion Grandmaster, ranking three out of 161,364, as well as a Competition Master with four gold medals and eight silver competition medals, and a top rank of 34 out of 147,089.

Kaggle in the business environment

While most people see Kaggle helping to solve commerce and real-world problems through its competitions, CherKeng sees a subtle difference between competition and business. "Kaggle challenges are crafted carefully to enable data scientists to

"THE HARD TRUTH IS THAT I CANNOT COMPETE IN KAGGLE WITHOUT POWERFUL MACHINES LIKE Z BY HP."

Heng Cherkeng, Z by HP & NVIDIA® Data Science Global Ambassador

pointed in the direction of the online data science community Kaggle. One of the main reasons he joined Kaggle is its variety and scope. Although he sees it as a hobby, he really likes that you can be one thing one day and something completely different the next. He says, "You can be a computational biologist when you are solving a DNA Kaggle challenge. Then you can become a computational chemist the next day if you are trying to figure out molecule properties. This is an interesting part of being a data scientist and the reason why I joined Kaggle. I can try different problems and collaborate

compete. For example, data is cleaned and evaluation metrics are well defined. In real world applications, you spend a lot of time deciding if and how to apply data science in the first place since we are interested in the worst performance of unseen data— basically the generalization and robustness of an algorithm. We want to know when the algorithm will break down so we don't use it incorrectly."

However, CherKeng still believes Kaggle has an important role to play in the business community. "I think companies should encourage their data scientists to interact



SOLUTION HIGHLIGHTS:

HP Z8 G4 1125W PSU

- Dual Intel® Xeon® 6240 2.60GHz 18C CPU
- Dual NVIDIA® RTX 8000 graphics
- 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)
- Z by HP Data Science Software Stack

Z38C DISPLAY

- 37.5-inch diagonal curve
- 3840 X 1600 at 60 Hz resolution

HP ZBOOK STUDIO G7

- Approx: 4 lbs/2 Kilos
- i9 – 10885H processor
- 32GB RAM
- NVIDIA® RTX 5000 (16 GB VRAM)
- 15.6-inch display
- Ubuntu 20.04
- Z by HP Data Science Software Stack

PRE-LOADED, POPULAR DATA SCIENCE TOOLS, TURN-KEY & READY OUT OF THE BOX:

Z by HP Data Science Software Stack

Learn more:

Z by HP Data Science Brochure

and take part in Kaggle,” he says. “It enables them to see different applications and algorithms. For example, e-commerce data engineers may spend a great deal of time working on retail systems. But they may never have a chance to work on medical imaging projects. The core of an algorithm is universal and the basic principle can be applied broadly. Maybe the breakthroughs in medical imaging can also be used in other areas?”

Z by HP is “essential to data science and AI”

Prior to becoming a Z by HP & NVIDIA® Data Science Global Ambassador, CherKeng had little experience with HP, knowing only that it provided high-performance computing solutions for enterprises.

Since he started using Z by HP, CherKeng says, “My experience with HP has been great and I hope I can continue the ambassadorship in the future. Being a data science influencer is new to me and I love to share the experience of data science, particularly how it applies to the commercial world. It has brought me many opportunities to meet different people. This is very important because data science is about providing solutions. Having the best hardware and software to do that is critical.”

Every single day, CherKeng is conducting data science experiments. He needs to analyze results, write software code and develop algorithms against a list of criteria like business objective, achievability, time to market and resources needed to succeed.

“The power and performance that Z by HP delivers is absolutely important



“ONCE YOU HAVE EXPERIENCED A POWERFUL MACHINE LIKE THE Z8 WORKSTATION, YOU WON'T WANT TO GO BACK. THE Z8 IS THE MOST POWERFUL GPU I HAVE EVER USED AND IT HAS CHANGED THE WAY I WORK. IT CONSUMES LARGE AMOUNTS OF DATA, LETS ME RUN AI MODELS IN A VERY SHORT TIME AND DELIVERS GREAT PERFORMANCE.”

Heng CherKeng, Z by HP & NVIDIA® Data Science Global Ambassador



and essential to data science and AI. It shortens my development cycles. It determines how fast experiments can be done and the more experiments you do, the higher chance of success you have. It governs whether a certain method can or cannot be applied.”

CherKeng describes Z by HP hardware as “addictive.” He says, “Once you have

experienced a powerful machine like the Z8 workstation, you won’t want to go back. The Z8 is the most powerful GPU-accelerated workstation I have ever used and it has changed the way I work. It consumes large amounts of data, lets me run AI models in a very short time and delivers great performance.” He uses the ZBook Studio to work on Kaggle activity when away from home.

For CherKeng, the software stack on the hardware is very convenient and easy to work with, and has replaced some of his previous workflows. For instance, PyCharm is his favorite IDE python editor. He uses PyTorch for deep learning and NVIDIA® RAPIDS has replaced several time-consuming operations in python. He is starting to use specialized NVIDIA® CUDA

(Compute Unified Device Architecture) to accelerate libraries such as DALI for GPU image augmentation and TensorRT for fast, deep network inferences.

Having the software stack was important for CherKeng as he says, “I don’t have to go through the hassle of installing software myself. What is most significant is that if I mess up, I can simply and easily reinstall it.”



KAGGLE: LEARN, COMPETE, CHANGE THE WORLD

- Over 6 million registered Kagglers solve data science problems, gain access to powerful tools and resources, and compete for prizes.
- There are five tiers: Novice, Contributor, Expert, Master, and Grandmaster. Currently, there are 211 Grandmasters.
- Kaggle has run hundreds of competitions, from improving gesture recognition to improving the search for the Higgs boson at CERN. Competitions have resulted in successful projects including furthering HIV research and traffic forecasting. The learnings that result from the competitions and shared on Kaggle are being transferred into enterprise workflows to transform how business works.

Visit www.kaggle.com to learn more.

There is another, less obvious benefit, especially in a business environment. “The HP Z workstation doesn’t have noisy fans so it’s really quiet. This is important because you can imagine what it would be like trying to concentrate and having something like a vacuum cleaner turned on for 24 hours in the background.”

CherKeng is keen to extend his relationship with the Z by HP & NVIDIA® Data Science Global Ambassador program by getting involved in things like testing and evaluating HP’s related technology for other data science environments.

Delivering huge reductions in data processing time

CherKeng uses Z by HP primarily for Kaggle. He says, “The hard truth is that I cannot compete in Kaggle without powerful machines like Z by HP.” Recently, Kaggle competitions have started using large amounts of data like medical images or videos. In one example involving millions of images, the Z8 workstation reduced process time from a week down to one or two days.

One Kaggle competition that CherKeng cites is Rainforest Connection Species Audio Detection which aimed to automate the detection of bird and frog species in a tropical

soundscape. The Z8 workstation enabled him to complete the competition and achieve a reasonable ranking in 13 days instead of a month. In another instance, Z by HP meant he could process two million images—performing an image-to-text translation—in one hour instead of four.

CherKeng loves data science but he does see it has challenges. “Data science needs more successful stories and fortunately, we are beginning to see some emerging. The recent success of mRNA COVID-19 vaccines shows how AI can help in drug discovery. That has resulted in a tremendous increase in health and medical sector investments. In November 2020, Google’s DeepMind developed AlphaFold, an artificial intelligence program which performs predictions of protein structure. It was a breakthrough that solved a 50-year-old question in biology.”

He adds, “As long as there is positive news, there will be more investment. That leads to more innovations and I hope this healthy spiral of progress will continue. Data science doesn’t live by itself. In the next few years, I think we are going to see data science in post-pandemic innovations, especially health and medical, but also in next generation areas like 5G and green energy.”

THE POWER OF Z:

Essential to effective data science and AI

Millions of images processed in one day instead of one week

One hour versus four to perform an image-to-text translation

13 days instead of a month to complete a Kaggle competition

Most powerful GPU ever used

Unleashes data science expertise



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.

As a Z by HP Global Data Science Ambassador, Heng Cherkeng's content is sponsored and he was provided with HP products.

4AAB-0210ENW, October 2021

