AsianScientist

Issue 04
July 2018

SUPERCOMPUTING ASIA

RAISING A CITY'S IQ

WHY SMART CITIES NEED HPC

+

5 WAYS SUPERCOMPUTERS MAKE CITIES SMART

AI FOR EVERYONE

HELPING DATA GO AROUND THE WORLD

MPAS AWARDS 2017
Technology Trade Media of the Year
Science Trade Media of the Year

THE 10TH IDA INTERNATIONAL DESIGN AWARDS
Silver Award

MCI (P) 127/04/2018

CONTENTS

Issue 04
July 2018

COVER STORY

p. 16

RAISING A CITY'S IQ
Why smart cities need HPC

FEATURES

EDITOR'S NOTE

As far as megatrends go, few are as consequential as the rapid growth of urbanization. By 2050, cities are expected to accommodate an additional 2.5 billion people—90 percent of them in Asia and Africa.

Technologies like Internet of Things sensors, satellite images and even open source data are all being marshalled to make our cities safer, sustainable and more livable. But working behind the scenes to manage and make sense of all the data generated are supercomputers and high performance computing (*5 Ways Supercomputers Make Cities Smart*, p. 10).

Ultimately, however, it is not technology but people that make cities smart. Check out our cover story (*Raising a City's IQ*, p. 16) to see how supercomputers are helping city planners make data-driven decisions for the good of their citizens. And while you're here, don't miss our exclusive interview with Professor Satoshi Sekiguchi, where we get the lowdown on Japan's top AI supercomputer (*Artificial Intelligence for Everyone*, p. 22).

Rebecca Tan, Ph.D.
Editor-in-chief
Supercomputing Asia

/asianscientist @asianscientist
Asian Scientist Magazine

www.asianscientist.com

SUPERCOMPUTING ASIA

PUBLISHED BY
Wildtype Media Group Pte Ltd

DISTRIBUTED BY
Pansing Distribution Pte Ltd

HEAD OFFICE
Wildtype Media Group Pte Ltd
71 Ayer Rajah Crescent #07-03
Singapore 139951
Tel: 65-6909-2279
Website: www.asianscientist.com/subscribe

WILDTYPE MEDIA

Combining savvy communication with technical rigor, Wildtype Media Group is Asia's leading STEM-focused media company, spanning digital, print, custom publishing and events.

Brands under Wildtype Media Group include the flagship *Asian Scientist Magazine* and *Supercomputing Asia*, award-winning titles available in print and online. Through its Asian Scientist Intelligence consulting division, Wildtype Media Group provides a full suite of marketing and communication services to industry and academic clients hoping to engage Asia.

www.wildtype.media

Digital Dispatch

US$636 MILLION EARMARKED FOR EXASCALE IN US BUDGET REQUEST

The US Department of Energy has requested US$636 million to fund exascale supercomputing projects for fiscal year 2019, US$376 million more than the funding approved for fiscal year 2017. The amount earmarked for exascale supercomputing is part of the overall Department of Energy budget request for US$30.6 billion.

The increased funding puts the US on track to meet its goal of launching an exascale system by 2021. China, which has overtaken the US in terms of number of systems and percent of computing power on the TOP500 list, is thought to be leading the race to exascale, with Japan and the European Union close behind.

The funds will be split between the Office of Science and the National Nuclear Security Administration, which will receive US$473 million and US$163 million respectively. In addition to funding the development of an exascale software ecosystem, the money will be used to design and deploy the Aurora system at the Argonne National Laboratory and the Frontier system at the Oak Ridge National Laboratory.

EU ANNOUNCES BILLION-EURO SUPERCOMPUTING PLAN

In a bid to catch up with China and the US, the European Union has set aside €1 billion (~US$1.17 billion) of public funds for supercomputing under a new legal and funding structure called the EuroHPC Joint Undertaking.

"Supercomputers are the engine to power the digital economy. It is a tough race and today the EU is lagging behind; we do not have any supercomputers in the world's top ten," said Mr. Andrus Ansip, European Commission vice president for the Digital Single Market, in a press release.

"With the EuroHPC initiative we want to give European researchers and companies world-leading supercomputer capacity by 2020—to develop technologies such as artificial intelligence and build the future's everyday applications in areas like health, security or engineering."

The joint undertaking will involve research into supercomputing technology such as low-power microprocessors, as well as the establishment of two "world-class" pre-exascale and at least two "mid-range" machines. This infrastructure will be jointly owned and operated by the 13 signatories of the EuroHPC declaration and private members from academia and industry.

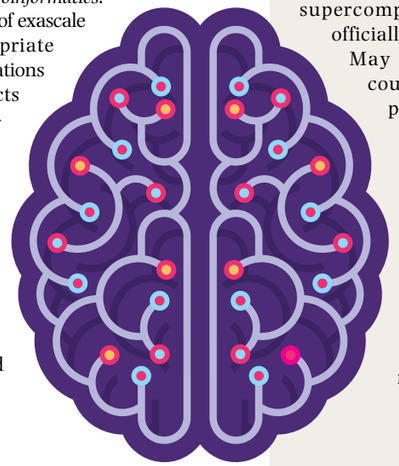
INTERNATIONAL COLLABORATION CLOSES IN ON EXASCALE BRAIN SIMULATION

An international team of researchers has developed software that takes them one step closer to simulating the human brain on future exascale supercomputers.

Their new algorithm allows larger parts of the brain to be represented using the same amount of computing memory, and also significantly speeds up brain simulations on existing supercomputers. The algorithm will be incorporated into NEST, an open source simulation software widely used in the neuroscience community, and a key tool for the European Human Brain Project.

The work, which involved researchers in Japan (RIKEN), Germany (Jülich Research Centre and Aachen University), Sweden (KTH Royal Institute of Technology) and Norway (Norwegian University of Life Sciences), has been published in the journal *Frontiers in Neuroinformatics*.

"The combination of exascale hardware and appropriate software brings investigations of fundamental aspects of brain function—like plasticity and learning —unfolding over minutes of biological time, within our reach," said Professor Markus Diesmann, director at the Jülich Institute of Neuroscience and Medicine and a lead author on the study.



TAIWAN ENTERS THE PETASCALE LEAGUE

Taiwan's 1.33 petaFLOPS supercomputer, the Taiwania, officially began operations in May 2018, marking the country's entry into the petascale league. The machine was built by the Ministry of Science and Technology's National Applied Research Laboratories (NARLabs) at a cost of NT$430 million (~US$14.4 million).

Named after a large coniferous tree native to east Asia, the Taiwania has seven times the processing capability and ten times the energy efficiency of its predecessor, the Advanced Large-scale Parallel Supercluster (ALPS), while taking up just one third of its footprint.

One of the Taiwania's many applications involves helping researchers and the government address poor air quality by more precisely monitoring the spread of air pollutants such as particulate matter 2.5, said Professor Shieh Ce-Kuen, director general of NARLabs' National Center for High-Performance Computing.

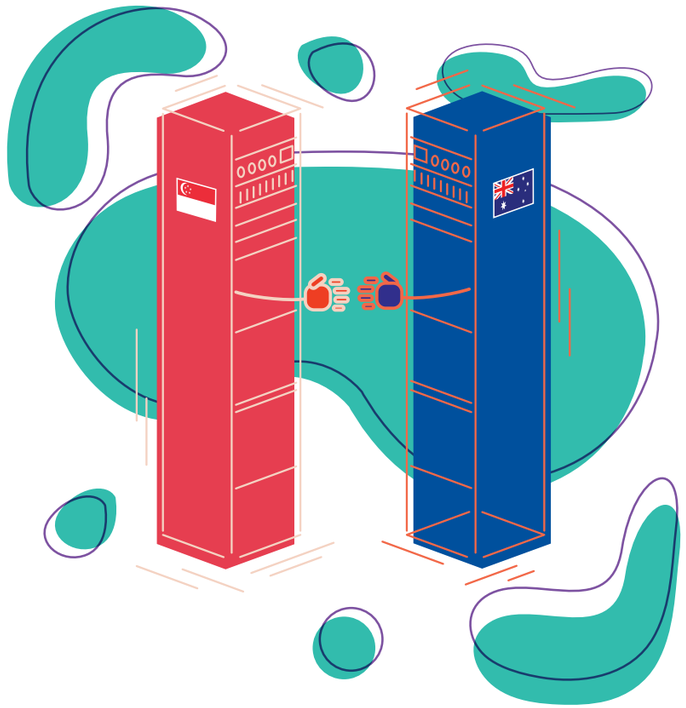
Digital Dispatch

PRATYUSH IS INDIA'S FASTEST SUPERCOMPUTER

India's supercomputing capabilities have received a boost with the January 2018 unveiling of the Pratyush, a 6.8 petaFLOPS supercomputer housed at the Indian Institute of Tropical Meteorology (IITM), Pune, and the National Centre for Medium Range Weather Forecast, Noida.

According to IITM, the Pratyush is the fourth fastest supercomputer in the world to be dedicated to weather and climate research, after machines in Japan, the US and the UK. The machine is a step up from India's current peak capacity of one petaFLOPS, and is expected to be ranked in the thirties on the TOP500 list.

The Pratyush will mainly be used for the forecasting of extreme weather events such as monsoons, tsunamis, cyclones, floods and earthquakes. The new system makes it possible to map regions in India at a resolution of three kilometers, and globally at a resolution of 12 kilometers, reported *The Hindu*.



SINGAPORE AND AUSTRALIA SIGN MOU

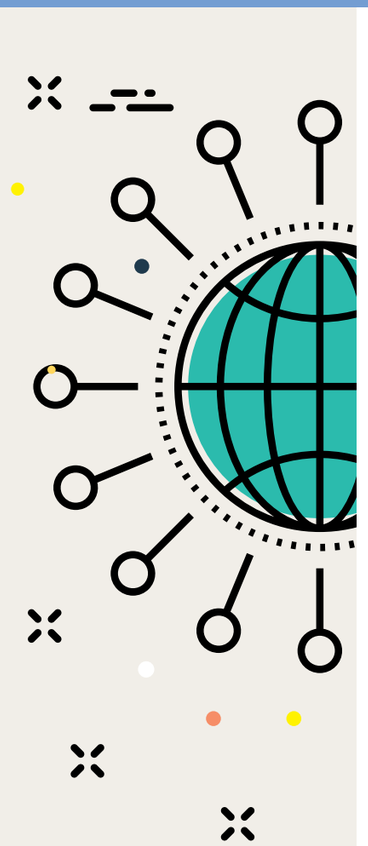
The National Supercomputing Centre (NSCC) Singapore and the Pawsey Supercomputing Centre in Australia have inked a memorandum of understanding (MOU) to collaborate in the fields of supercomputing, networking, data analytics, scientific software applications and visualization.

The two centers will work together on industry engagement, outreach activities, training and stakeholder management. Also on the agenda are governance matters, user support tools and methods, HPC software development and cybersecurity.

"With our national petascale HPC platform, NSCC's collaboration with international supercomputing centers such as the Pawsey will benefit and bring together people and researchers with different skillsets and expertise to solve problems of a scale previously not possible," said Professor Tan Tin Wee, NSCC's chief executive.

"We are convinced that this type of initiative will enhance the services each center provides, therefore researchers and their outcomes will be impacted. By working together, we aim to continue accelerating scientific discoveries in these two nations for the benefit of humanity," said Mr. Ugo Varetto, Pawsey's acting executive director.

WHAT'S UP!



SUPERCOMPUTINGASIA 2019: SAVE THE DATE

After a successful 2018 edition, Asia's premier HPC conference SupercomputingAsia (SCA) will be back next year from March 11–14, 2019 in Singapore.

Organized by the National Supercomputing Centre (NSCC) Singapore, SCA19 will once again feature keynote addresses from the world's HPC leaders in academia and industry.

In addition, expect an exciting calendar of research presentations, workshops, industry-specific showcases and other notable HPC events. Save the date and stay tuned for more details!

For more information, visit sc-asia.org

WHAT

SupercomputingAsia 2019

WHEN

March 11–14, 2019

WHERE

Singapore

LOOK OUT FOR SC18

This year's edition of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC18) will be held in Dallas, Texas, from November 11–16, 2018.

Since its inauguration in 1988, the SC conference series has brought participants up to speed with cutting-edge HPC developments through a highly competitive technical program, which includes talks, panels, research papers, poster presentations, tutorials and 'Birds of a Feather' sessions for audience-driven discussions.

Another highlight is the SC Awards, which recognize outstanding contributions and accomplishments in the field; these include the prestigious ACM Gordon Bell Prize and the IEEE-CS Seymour Cray Computer Engineering award.

Young researchers need not feel left out—through the Students@SC program, the conference also offers professional development programs, mentoring, job fairs and technical sessions targeted at students at both the undergraduate and graduate levels.

Notably, SC18 will also commemorate the conference's 30th anniversary with a series of special activities, including a history scavenger hunt, a special panel, invited speakers and a dedicated digital archive.

For more information, visit sc18.supercomputing.org

WHAT

SC18

WHEN

November 11–16, 2018

WHERE

Dallas, Texas, US

5 WAYS SUPERCOMPUTERS MAKE CITIES SMART

The brains behind bustling metropolises

Supercomputers are helping urban planners run simulations, make predictions and build models, bringing the dream of a smart city closer to reality.

By **Jeremy Chan**

EVERYTHING FROM MOBILE PHONES TO televisions are getting the prefix 'smart,' and cities are no exception. With the help of sensors monitoring parameters such as air quality and rubbish levels, city planners can now make data-driven decisions instead of relying on estimates or best guesses.

However, drawing insights from big data requires significant computing power. This is where supercomputers can make a difference, helping governments and private enterprises involved in city management cut through the noise and make decisions that improve how cities are run.

Illustrations by Allan Chan/*Supercomputing Asia*

1 SMARTER DESIGN AND PLANNING

The physical landscapes of cities are in constant flux as new structures emerge amid tight spaces. To make the best use of limited space, city planners must make good decisions during the design and planning stage.

In the past, two-dimensional blueprints would be the go-to source of information when deciding on the location of a new building or how it would look. Now, urban planners have a new tool with which to visualize buildings before construction begins—three-dimensional simulations.

China's Tianhe-1A supercomputer, which boasts 2.57 petaFLOPS of computing prowess, is helping city planners create simulations of huge construction projects before a single brick is laid. The simulations use data on building materials and energy consumption to optimize how resources can be mobilized. According to Mr. Meng Xiangfei, head of the applications department of the National Supercomputer Center in Tianjin, such simulations may reduce construction costs by up to 20 percent.

2

EMPOWERING ENERGY RESEARCH

Large cities are voracious energy guzzlers, and governments everywhere have made it a priority to find alternative sources of electrical power. Nuclear fission—the splitting of the nucleus of an atom—has been tabled as a compelling option for powering the smart cities of the future. The drawback? Radioactive waste produced by nuclear fission is damaging to humans and the environment.

As an alternative to nuclear fission, scientists have billed nuclear fusion as a cleaner and more abundant energy source. During nuclear fusion, the nuclei of two atoms collide to form one or more new, non-toxic products, releasing vast amounts of energy in the process. There are, however, existing knowledge barriers to nuclear fusion. For one, researchers need a better understanding of superhot gas called plasma, a critical enabler of nuclear fusion.

To simulate plasma behavior, the National Institutes for Quantum and Radiological Science and Technology in Japan has commissioned a supercomputer dedicated to nuclear fusion research. With a peak performance of around four petaFLOPS, it may not be the fastest supercomputer in the world, but it provides more than the necessary compute resources to shed light on the future of energy.

3

FORECASTING THE WEATHER

Looking into the future might sound like something straight out of fantasy fiction, but with supercomputers, smart cities are now able to forecast extreme weather and predict natural disasters ahead of time. This has wide-ranging implications, from day-to-day living to agriculture and emergency response.

In January 2018, India unveiled a Cray supercomputer consisting of two nodes—named Pratyush and Mihir—that are to be used for meteorological monitoring. Additionally, the pair of supercomputing nodes will allow researchers in India to crunch data from 146 river sub-basins, enabling more accurate flood forecasts.

Similarly, the Korea Meteorological Administration (KMA) of South Korea uses a Cray XC40 supercomputer to analyze weather patterns. The Cray XC40 generates 160,000 weather maps per day with a processing speed of 5.8 petaFLOPS. KMA has announced a US$56 million investment into building a 50 petaFLOPS supercomputer that will be able to produce one million weather maps per day when it is ready in 2020.

4

AUGMENTING TRANSPORTATION

Most modern cities are chock-full of people, and moving them around in urban centers can be challenging. The future of urban mobility therefore lies in optimizing the use of multiple modes of transportation.

Researchers at Singapore's Institute of High Performance Computing are tapping on data from the city's smart farecard system and combining it with land use data to predict transport ridership across the city. The insights obtained from these predictions may help improve the deployment of passenger-carrying vehicles in the city state or provide feedback for urban planning.

Meanwhile, Chinese technology company Baidu, in collaboration with American chipmaker NVIDIA and German smart mobility firm ZF, are developing another form of mobility altogether—autonomous vehicles (AVs).

While AVs are driven by sensors and artificial intelligence, they need to cope with enormous volumes of data as they assess and react to their surroundings. Hence, under the bonnet of Baidu's mass-produced AV model Apollo Pilot is NVIDIA's DRIVE Xavier™, a supercomputing chip capable of performing 30 trillion operations per second on a paltry 30 watts of power.

5

MONITORING AIR POLLUTION

Heavy vehicular traffic and industrial activity within cities often means that air quality takes a hit. Beijing experienced severe smog in January 2017, when the air quality index hit an all-time high of 482. Not only are readings above 300 considered extremely hazardous to health, visibility was also so poor that flights had to be cancelled.

Measures have since been taken to better monitor and control air pollution using supercomputers. The Tianhe-1A supercomputer is being used to run simulations and build computer models that include data on the various natural and man-made sources of air pollution.

Using data from 114 cities in China accumulated through 668 distributed sensors, scientists are hoping to be able to predict the onset of smog in advance, estimate its duration, as well as pinpoint the root causes of extreme cases of air pollution. This will help municipal authorities issue early warnings and craft policies to improve air quality.

RAISING A CITY'S IQ

Why smart cities need HPC

Smart cities represent a tremendous opportunity to improve the lives of millions, but many technical challenges stand in the way. Here's how high performance computing is being used to address several of the most difficult ones.

By **Rebecca Tan**

Since the dawn of human civilization, cities have been synonymous with opportunity. By bringing both people and resources together, cities enjoy economies of scale and accelerated economic growth. Drawn by these opportunities, an estimated three million people move into cities every week. By 2050, cities are expected to house an additional 2.5 billion people, roughly equivalent to the entire population of both India and China, the two most populous countries in the world.

The unprecedented speed and scale of urbanization poses an urgent and complex challenge to governments and city planners. In the absence of adequate action, sprawling slums and deteriorating infrastructure can transform once pleasant cities into unlivable ghettos, with knock-on effects on health, security and other measures of quality of life.

In an attempt to forestall or circumvent these consequences, leaders around the world have turned to technology. Equipped with an array of Internet of Things (IoT) sensors and increasingly powerful analytic capabilities, the smart cities movement promises to make city life safer, smoother and more sustainable. If executed well, smart cities have the potential to impact millions, if not billions, of lives.

"The timely analysis of sensor and IoT data would allow city managers to solve problems ranging from traffic prediction to flood monitoring and even food safety," said Professor Wang Lizhe of the Institute of Remote Sensing and Digital Earth at the Chinese Academy of Sciences. "With their ability to analyze data at lightning-fast speeds, high performance computing (HPC) makes smart cities possible."

THE POWER OF KNOWING

The first step toward building smarter cities is collecting information about what happens in it, thereby reducing the number of unknowns faced by decision makers. Currently, many decisions, even in well-developed cities, are made in the absence of accurate information—sometimes with tragic results.

In 2015, a hundred-meter-high construction landfill in Shenzhen, China, collapsed, causing a landslide that claimed the lives of 74 people and destroyed 33 buildings. "If officials had enough information about such a situation before it happened, for example if they had models that could warn them about the possibility of such an occurrence, they could have taken action and prevented such man-made disasters," Wang said.

"On the other hand, if you have no geological and geographical knowledge about the city, you will not be able to manage the city well."

To remedy this situation, Wang and his team have been working on developing a decision support system for the management of smart cities. The objective of the project, supported by the National Natural Science Foundation of China, is to help administrators control and maintain their cities, as well as plan for the future, he said.

Wang is now building up a comprehensive understanding of a city from four different sources of data: satellite data, ground-level data, underground data and crowd-sourced data. This data is then combined with simulations to calculate warning levels and projections to enable city planners to make informed decisions.

"The model can also help balance between all kinds of conflicts by calculating what would happen if we removed any of them," Wang explained. "Decision makers can use our simulations to test what can be done before committing resources to a particular course of action."

Last but not least, the decision support system can also help with resource allocation and planning, he said. "If limited resources such as space are wasted, there will not be enough space to build critical infrastructure such as subways or water pipes in the future."

DATA FROM THE GROUND UP

Building this smart city decision support system was no mean feat, with one of the main challenges being having to deal with the vast amount of data needed to understand something as complex as an entire city.

"A single satellite can generate several terabytes of high-resolution data within a day," Wang explained. He himself collects not only optical data based on visible wavelengths, but also synthetic aperture radar, light detection and ranging (LIDAR) and global positioning system data.

Furthermore, Wang and his team do not stop at a detailed understanding of the surface of a city's terrain but literally go deeper, incorporating underground data from subway systems and modern buildings and even crowd-sourced data such as pictures of subway incidents or building collapses uploaded by internet users.

"Well-developed cities—whether in Shenzhen, Shanghai or Singapore—all face the problem of limited space. Although we may not have paid as much attention to it in the past 20 years when there was still room to grow, underground data is increasingly important for large cities," Wang said.

To add to the challenge of dealing with large amounts of heterogeneous data, the data collected must be coupled with complex physical models like those used in climate prediction, he continued. Supercomputers are required at both steps: processing large amounts of data and combining the data with physical models.

"Sometimes the model is not correct as it is only an approximation of reality. Other times, the observations are not correct as they contain errors. If we simply put the model and data together we will definitely not get a good result," Wang said. "We first need to use a lot of computing resources to remove the errors, a process that we call data assimilation."

TRANSFORMING TRANSPORTATION

Another area where supercomputing is poised to make a positive impact on smart cities is the realm of transportation. For example, a study by market intelligence firm Juniper Research found that smart mobility technology like intelligent traffic systems and route optimization based on open data would save nearly 60 hours per person each year.

Yet the reality for many cities—particularly in Asia—is snarling gridlocks on a daily basis. Bangkok, which has consistently been ranked as one of the most congested cities in Asia, saw its residents spend more than 64 hours stuck in traffic in 2017, with Jakarta not far behind at 63 hours. In addition to the short-term loss of productivity, bad traffic also has long-term consequences on health and the environment.

Even in Singapore, where people spend just ten hours stuck in traffic a year, incidents can still happen. In 2015, the country's Mass Rapid Transit system experienced the worst breakdown in its 30-year history, when both the North-South and East-West lines broke down and left over 250,000 commuters stranded at rush hour. Although shuttle buses were deployed, movements of the unprecedented crowd were unpredictable, resulting in chaos and congestion.

"The Land Transport Authority (LTA) found that they did not have enough visibility into the crowd levels, especially when an incident happens. Because the crowd was not very well understood, the response could not happen as effectively as LTA wanted it to," said Dr. Laura Wynter, head of AI at the IBM Research Singapore Lab.

"To know how to deal with crowds when an incident happens, you will need to understand what the crowd levels and characteristics are, what the crowd is like usually, and how it is changing," she continued. "The problem was that there was no obvious data source."

Unlike the trains themselves, which are scrupulously monitored by an array of devices, people don't come equipped with sensors. Till now, there has not been a feasible way for LTA—or any metro system anywhere in the world—to know the exact number of people at each location in the transport network, Wynter added.

> "With their ability to analyze data at lightning-fast speeds, high performance computing makes smart cities possible."
>
> Professor Wang Lizhe
> Institute of Remote Sensing and Digital Earth
> Chinese Academy of Sciences

A FASTER FEEL OF THE GROUND

To get a better sense on the size of crowds at each station and the movement of people through the network, LTA teamed up with IBM Research on a project known as Fusion AnalyticS for public Transport Emergency Response, or FASTER. FASTER uses WiFi and other sources of data to provide rail operators with both situational awareness and decision support.

Before the FASTER project, the only sources of information on the crowd levels were farecard data and video feeds. Farecard data only provides information on entry and exit points, while video processing is very computationally intensive.

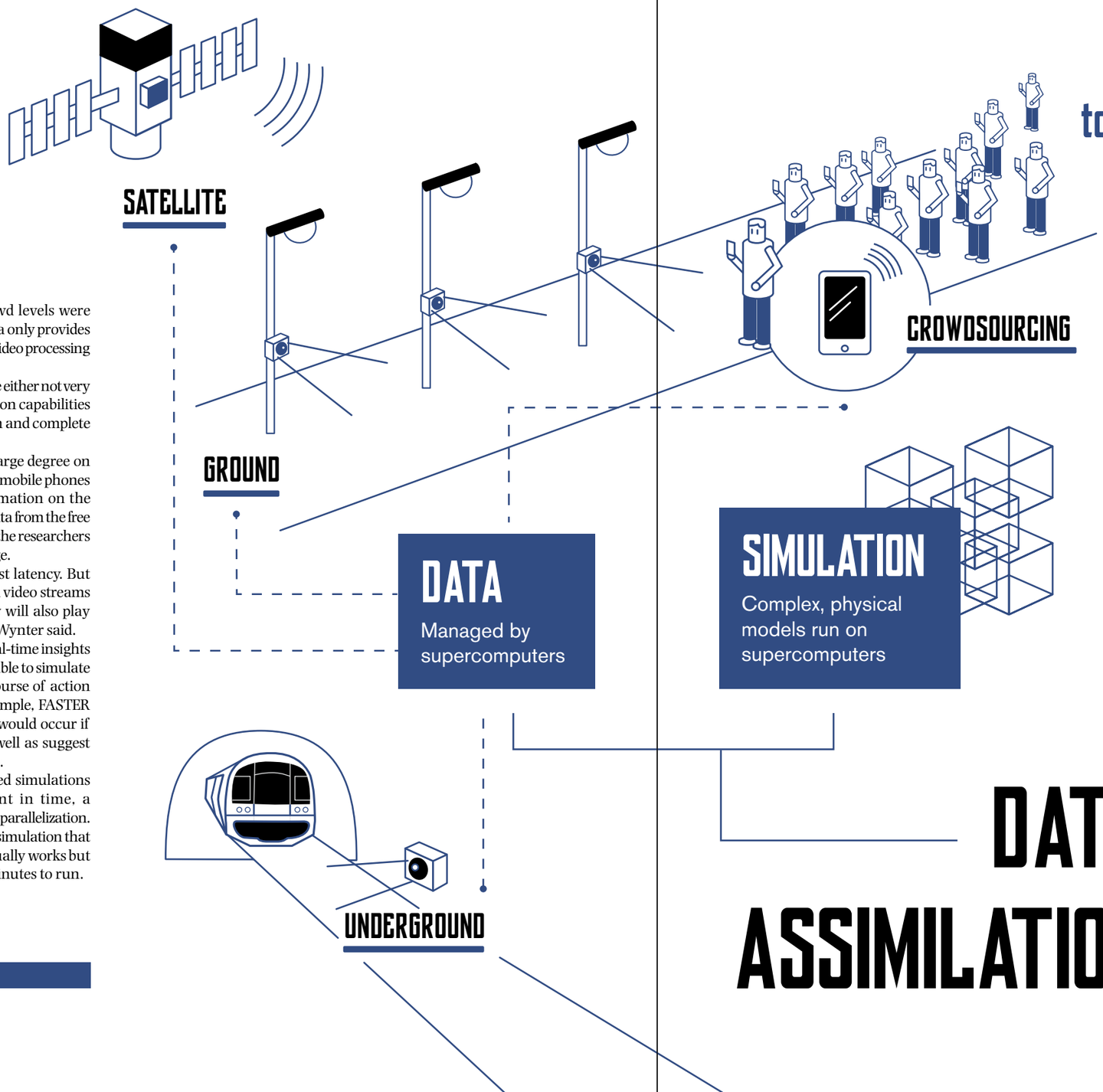
"Since these sources of information are either not very timely or incomplete, FASTER has simulation capabilities to ingest these partial pieces of information and complete the picture," Wynter explained.

At the moment, FASTER relies to a large degree on anonymized WiFi data, where commuters' mobile phones provide a rich source of geospatial information on the whereabouts of each anonymized device. Data from the free Wireless@SG broadband network supplies the researchers on the FASTER project with broad coverage.

"Right now, WiFi data has the lowest latency. But as other data sources such as farecard and video streams become available with low latencies, they will also play an important role in real-time analyses," Wynter said.

Beyond giving transport operators real-time insights into the levels of crowding, FASTER is also able to simulate 'what if' scenarios and suggest the best course of action in the face of service disruptions. For example, FASTER is able to predict the expected delays that would occur if a section of track becomes unusable, as well as suggest optimal shuttle service deployment routes.

These capabilities rest on agent-based simulations of about 400,000 agents at any point in time, a computationally intensive task that requires parallelization. The challenge, Wynter said, is developing a simulation that is faithful to how the transport system actually works but remains manageable and requires only minutes to run.



Illustrations by Allan Chan/*Supercomputing Asia*

> "We hope to help people leverage data and give them tools for effective data-driven decision support."
>
> Dr. Laura Wynter
> Head of AI
> IBM Research Singapore

"While situational awareness works at specific locations, simulations for planning purposes work at a network-wide level. Both these applications have their own computing challenges; situational awareness needs to be quick, while long-term simulations need to be more accurate," said Mr. Hassan Poonawala, a data scientist at IBM Singapore and one of the project leads on FASTER.

MAKING DATA-DRIVEN DECISIONS

Whether the decisions are about when and where to deploy shuttle buses in the event of a train breakdown or how to plan cities in general, HPC can help city planners make sense of a large amount of information and make better decisions as a result.

"In smart city settings, a decision maker isn't just a chief executive or operations manager; everybody is a decision maker at their own level. We hope to help people leverage data and give them tools for effective data-driven decision support," said Wynter.

To get there, however, will require a commitment to developing not just big data, cloud computing and IoT capabilities, but also supercomputing core facilities capable of turning data into insight.

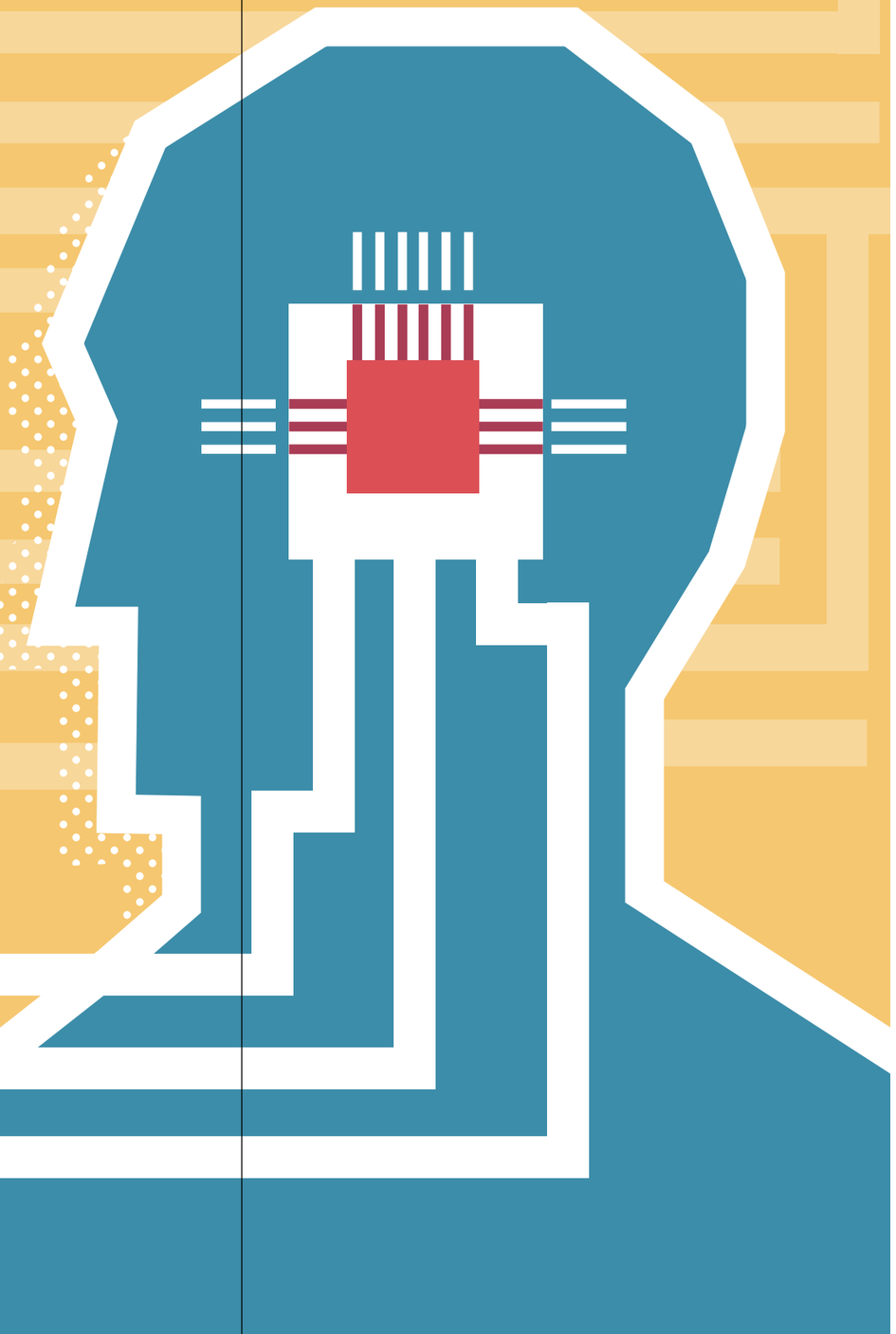
"Ever-larger HPC engines are necessary to meet the tremendous computational needs introduced by the huge scale of data involved in complex, long-term modeling." said Wang. "Supercomputers also accelerate the integration of multidisciplinary sources of information and nurture emerging disciplines, directly addressing the technical challenges faced in implementing smart cities."

AI FOR EVERYONE

As easy as ABCI

Rather than dominate the TOP500 rankings, the ABCI supercomputer was designed to bring computing capacity for AI applications to the masses, says AIST's Satoshi Sekiguchi.

By **Sim Shuzhen**



If you're a small tech company running some sort of deep learning application—whether it's medical diagnostics, price forecasting or image or voice recognition—chances are you're also looking to scale that application up. To do so, you'll need to beef up your company's computational muscle.

But as things stand now, your choices are limited. You could build your own cluster, but that would cost an arm and a leg to procure and maintain. Or you could tap into computing services offered by the likes of Amazon and Google—but that would mean ceding some measure of control over your data and computational resources to the whims of big tech.

Soon, however, you might be able to avail yourself of a third option: a powerful new artificial intelligence (AI) supercomputer known as the Artificial Intelligence Bridging Cloud Infrastructure (ABCI), built by Fujitsu for Japan's National Institute of Advanced Industrial Science and Technology (AIST).

Scheduled to launch by end 2018, the ABCI will offer industry and academic researchers cloud access to compute and storage capacity that is particularly suited for AI applications, said AIST vice president and ABCI project lead Dr. Satoshi Sekiguchi.

"Big data is getting bigger. Even if you have your own small cluster, it takes a month just to complete one round of training a network. That is not realistic; for research to be competitive, we need infrastructure that reduces the turnaround time of each job. We need a platform where advanced researchers can come together," Sekiguchi told *Supercomputing Asia*.

BREAKING RANK

When news of the ABCI broke in 2016, media speculation was rife that the Japanese were building a supercomputer that would topple China's Sunway TaihuLight from its perch atop the TOP500 ranking.

But this was not the intent, said Sekiguchi, adding that AI supercomputers like the ABCI play in a different arena. The TOP500 ranking is largely determined by how fast supercomputers can crunch double-precision (64-bit) floating point operations. But unlike traditional high performance computing, deep learning applications don't require such high precision—for training a network, half precision (16-bit) will suit you just fine.

Designed to maximize these less precise operations, the ABCI now clocks in at 550 half-precision petaFLOPS or 'AI-FLOPS'—an order of magnitude faster than the 47-peta-AI-FLOPS TSUBAME 3.0, its AI supercomputer brethren at the Tokyo Institute of Technology (TiTech).

"We haven't decided if we will submit our numbers to the TOP500, because our primary focus is on half precision rather than double precision at this moment," said Sekiguchi. That said, the ABCI is not too shabby even by TOP500 standards—its 37 petaFLOPS of double-precision speed would put it in the top spot in Japan and third globally, assuming the official LINPACK benchmark is run on it in the future.

AI FOR ALL

A widely respected high performance computing pioneer with a background in parallel computing, Sekiguchi has spent his career building grid- and cluster-computing infrastructure for Japan's research community, as well as developing advanced high-speed networking that has connected Japanese researchers with their international counterparts.

In 1997, for example, Sekiguchi was involved in setting up the Ninf network infrastructure, which allowed researchers to access data, hardware and software across a distributed computing environment—a project he remembers as particularly challenging as the team had to build everything, including security management and application programming interfaces (APIs), from scratch.

"But this kind of distributed software development was very exciting for us. We found good partners for taking care of the bottom of the stack, such as security and discovery. By working very closely with these other teams, we were able to take off the lower layer of our system," he recalled.

In many ways, the ABCI is a natural extension of Sekiguchi's career-long efforts at making connections and fostering collaborations. By providing access to AI computing on the cloud, the machine fills the gaping hole in computational resources that currently hampers closer collaboration between academics, companies and vendors of IT solutions, as well as narrows the widening chasm between small players and large companies with deep pockets, said Sekiguchi.

"Some companies have their own data, but not the skill [to analyze it] and understand what is happening. They might want to work with IT vendors, but IT vendors may not have the resources to support this kind of activity," he explained. "This is an opportunity for us to accelerate open innovation." While not yet finalized, the ABCI's prices will be cheaper than those offered by commercial cloud service providers, as the project does not intend to turn a profit, added Sekiguchi.

Companies from the manufacturing, autonomous vehicle and heavy equipment industries have already registered interest, as have academic researchers in areas as diverse as genomics, speech recognition and particle physics, said Sekiguchi. All users, both within the country and internationally, will be able to tap into the ABCI via high-speed networks such as Japan's 100-gigabit SINET5, which has links to Asia, Europe and the US.

THE GREENING OF AI

Aside from the promise of delivering AI capabilities over the cloud, another unique aspect of the ABCI is its high energy efficiency, said Sekiguchi. The data center, located on the Kashiwa campus of the University of Tokyo, will be chilled by a hybrid water-air cooling mechanism, designed with the help of green supercomputing pioneer Professor Satoshi Matsuoka of TiTech.

Since liquid conducts heat away more efficiently than air, this combination uses much less power compared to a fully air-cooled system. Further, the starting temperature of the water and air coolants can be as high as 32 and 35 degrees Celsius respectively—higher than the temperature of the ambient air except in the hottest months of the year—thus eliminating the need to use extra energy to cool them, said Sekiguchi.

▲ Dr. Satoshi Sekiguchi, vice president and director general of the Department of Information Technology and Human Factors at Japan's National Institute of Advanced Industrial Science and Technology.

Photo credit: Cyril Ng/ *Supercomputing Asia*

Like most buildings in Japan, the ABCI data center was designed with earthquakes in mind. But rather than shelling out a fortune to install anti-shaking dampeners on a multi-storey building, AIST decided to organize the data center across a single storey, atop a slab of hard concrete. This simple design took a mere month to build, said Sekiguchi.

CUTTING THE FAB

In addition to acting as a community resource, the ABCI also serves as a test of the concept of 'AI fabrication,' or 'AI fab' for short, said Sekiguchi. Just as chipmakers like Qualcomm farm out the expensive and finicky process of semiconductor fabrication to specialized facilities, researchers can now leave the AI fab (the building and maintenance of AI computational resources) to the ABCI—that is, they can go 'fab-less,' said Sekiguchi.

"At the end, the user receives the trained network, so they don't need to go through the whole process of building it and maintaining the computational resources needed to do so," said Sekiguchi. "This is deep learning for the future."

The idea of an AI-dedicated computational platform has caught the attention of other countries in the region, including South Korea, Singapore, Taiwan and Thailand, which are now starting to explore similar infrastructure, said Sekiguchi. He thinks that by making AI computing more affordable and accessible to smaller players, business models like the ABCI's can offer Asia a competitive edge over US tech giants, which now mostly cater to consumers or to other large businesses.

"[Small businesses] are kind of a niche right now, but the market is growing because small and medium enterprises really need to deploy AI to improve their business; yet they don't want to give their data to Amazon or Google. But if they have access to computing resources that they have control over, then there is a good opportunity for them to make things happen."

Sekiguchi's top priority over the next two years, he said, is to organize an ecosystem of industry and academic users around the ABCI, a task he views as more important than the addition of more computational resources. "The AI business we're focusing on isn't a single-capability business; it's more like a capacity offering, a system that people can use." §

THE ABCI SUPERCOMPUTER

NO. 1 IN JAPAN

NO. 3 WORLDWIDE pending official LINPACK benchmark

BUILT BY FUJITSU

1088 NODES

PER NODE:

4 NVIDIA TESLA V100 GPUs total 4,352

2 INTEL XEON GOLD PROCESSORS total 2,176

550 PETAFLOPS HALF PRECISION

37 PETAFLOPS DOUBLE PRECISION

HELPING DATA GO AROUND THE WORLD

A six-lane highway for scientific data

A high-performance, high-capacity network specially built for research purposes, the Global Research Platform will help scientists all over the world move and manage giant datasets.

By **Sim Shuzhen**

When something goes wildly viral—Beyoncé's triumphant comeback performance at Coachella, for example, or videos of oversized cats desperate to fit into tiny boxes—we often say that it 'broke the internet.' Yet, that dubious honor might well belong to the data-hungry scientific community, which on a daily basis generates and moves orders of magnitude more information across networks than Queen Bey and YouTube cat Maru's traffic combined.

For information to flow at that scale, biologists sequencing genomes, high-energy physicists tracking subatomic particles and astronomers training giant telescopes at the heavens need something more than the World Wide Web, said Dr. Joe Mambretti, director of the International Center for Advanced Internet Research (iCAIR) at Northwestern University in the US.

"The general internet cannot handle [such huge amounts of data]—it's not designed for it. In fact, the internet actually has blockages in it to prevent big flows of data, because that could mean a denial of service attack," he told *Supercomputing Asia*.

If the internet is a congested city street, Mambretti and his collaborators are assembling a six-lane highway for information: a sprawling computing architecture called the Global Research Platform (GRP), which will help researchers around the world access and exchange large datasets. "They need special channels and special capabilities for moving this data," explained Mambretti.

CONNECTING THE DOTS

The GRP has its roots in the Pacific Research Platform (PRP), a regional data-sharing system connecting research universities and supercomputing centers on the US West Coast, as well as several partners elsewhere in the US and in Asia.

The PRP consists of multiple science 'demilitarized zones' (DMZs)—a model developed by the US Department of Energy in which a secure, high-performance network for data-intensive science is kept separate from the general internet.

But while science DMZs already existed at multiple individual research campuses across the country, it wasn't until the PRP—funded by the US National Science Foundation in 2015—that these islands were linked up into a unified network, said Dr. Thomas DeFanti, one of the PRP's co-principal investigators and a research scientist at the University of California, San Diego's Qualcomm Institute.

To handle the huge flows of data coming through the network, DeFanti and his colleagues designed low-cost networked PCs called FIONAs (an acronym for Flash Input/Output Network Appliances), which today hum at many of the PRP's nodes. Equipped with flash memory and powerful GPUs, FIONAs send, receive, monitor and cache data, acting as a data 'capacitor' or interface between the high-speed network and the node's scientific instruments and computers.

In some cases, the PRP has sped network connections up by a thousand-fold over the general internet, said DeFanti, who is also distinguished professor emeritus in computer science at the University of Illinois. "We're very proud of that... but the question is how do we do this in more than just California and a few other places?"

GOING GLOBAL

Mambretti, DeFanti and their collaborators now want to work with partners around the world to replicate the PRP on a global scale. The GRP team holds regular demonstrations and workshops to get the word out, including at major high performance computing and networking events such as the Supercomputing 2017 conference and the Global Lambda Integrated Facility's Global LambdaGrid Workshop 2017.

The field is starting to move away from a one-size-fits-all approach to networking for data-intensive science, acknowledging that each scientific community has its own specialized needs and hence requires specialized capabilities, said Mambretti. The GRP's architecture will address this by allowing networks to be segmented—that is, sliced up to provide such differentiated services, he added.

Indeed, various scientific communities have already expressed interest in the new platform. The GRP team is planning to work with the astrophysics community on data from the Square Kilometer Array, the Large Synoptic Survey Telescope and the Nobel Prize-winning Laser Interferometer Gravitational-Wave Observatory; with the high energy physics community at facilities such as the Large Hadron Collider and Fermilab; and with precision medicine initiatives in countries such as the US, the UK and Taiwan.

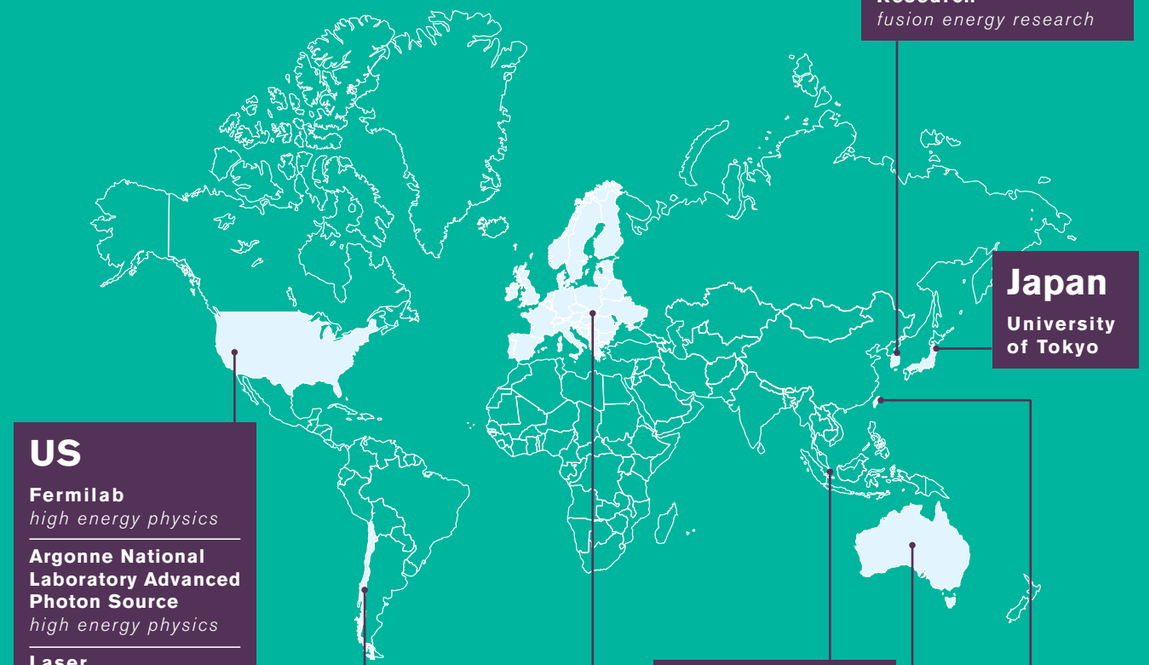
Research institutions and supercomputing centers in Asia—in China, Japan, Singapore, South Korea and Taiwan, for example—have also registered interest in being part of the network, said Mambretti.

Getting partners on board, however, is not without its challenges. The GRP's novel architecture has proved a stumbling block for more conservative organizations, said Mambretti. "They have their traditional tools, traditional ways of doing things and traditional policies, and they're suspicious of the new," he explained. "It's easier for [networking] to be one-size-fits-all... when people want different things, it's harder for you as a provider."

Time, perhaps, will convince doubters of the worth of the new platform. While the GRP is not yet fully up and running, Mambretti and DeFanti stressed that all the necessary tools, as well as many collaborations, are already in place for the network to take off. "We already have all the LEGO blocks—we just need to put them together," said Mambretti.

> **"WE ALREADY HAVE ALL THE LEGO BLOCKS—WE JUST NEED TO PUT THEM TOGETHER."**
>
> Dr. Joe Mambretti
>
> Director of the International Center for Advanced Internet Research
> Northwestern University

GLOBAL RESEARCH PLATFORM (GRP) PARTNERS INCLUDE:



South Korea
Korea Institute of Science and Technology

Korea Superconducting Tokamak Advanced Research
fusion energy research

Japan
University of Tokyo

US
Fermilab
high energy physics

Argonne National Laboratory Advanced Photon Source
high energy physics

Laser Interferometer Gravitational-Wave Observatory
astrophysics

Europe
Large Hadron Collider
high energy physics

Copernicus Program, European Space Agency
earth observation

Singapore
National Supercomputing Centre Singapore

Chile
Large Synoptic Survey Telescope
astrophysics

Australia
Square Kilometer Array
astrophysics

Taiwan
Academia Sinica
precision medicine

ASIA'S RISING AI TALENTS

Bringing HPC to the world

To help more countries embrace the potential of high performance computing and artificial intelligence, the HPC-AI Advisory Council is starting with students.

By **Rebecca Tan**

From predicting hurricanes to making cars safer, high performance computing (HPC) has quietly been a part of our everyday lives for decades. More recently, artificial intelligence (AI) has emerged as a technology that has a similar impact on the way we live, promising to transform virtually every sector of the economy.

To promote the global adoption of HPC and AI, the HPC-AI Advisory Council and the National Supercomputing Centre Singapore have come together to organize a competition targeted at students at universities and technical institutes. The inaugural APAC HPC-AI Competition aims to capture the imagination of the next generation and pit the region's best teams against each other in a two-part challenge.

Supercomputing Asia caught up with Mr. Gilad Shainer, chairman of the HPC-AI Advisory Council and vice president of marketing at Mellanox, to find out what participants can expect from the competition.

How are HPC and AI related?

Gilad Shainer: Both HPC and AI deal with large amounts of data and therefore rely on largely the same technology. In recent years, we've found that the work we've been doing at the HPC-AI Advisory Council also extended to AI, so it made sense for us to cover both topics and help people understand both sides of the technology.

You can think of it as a car with two different destinations; whether you are going towards HPC or AI, you still need to learn how to drive a car. The car is analogous to the common hardware: a large number of graphics processing units (GPUs) and central processing units (CPUs), as well as high-bandwidth, low-latency networks and fast storage. The software used is admittedly slightly different but they all use the same principles.

What makes competitions such an effective way of achieving the mission of the HPC-AI Advisory Council?

GS: The main mission of the council is to drive HPC and AI outreach and education. Firstly, competitions are a great way to spur people to learn new things and gain hands-on experience that they otherwise wouldn't have.

Competitions are also a lot of fun for the students. When I see the faces of the students and their excitement, it reminds me of when I was young and excited to learn. I'm very happy and appreciative to be able to inspire the next generation this way.

Last but not least, competitions help to create connections between the universities and industry, as teams get the chance to work with vendors to get hardware and so on.

Why do the competitions organized by the council focus on students?

GS: I firmly believe that if you want to develop people with the capability to come up with great ideas, you need to start at the university level or even earlier, when they are high school students. If people start at that age and are given the tools to develop themselves further, it will serve the next generation of scientists, researchers and innovators who are going to build great companies and so on.

We've seen students who have participated in our competitions go on to do good work in companies such as Baidu, Alibaba and Tencent, bringing new technologies and capabilities into those companies and helping them make the leap into HPC and AI.

Why have you launched this competition that focuses on the Asia-Pacific region?

GS: We have been running HPC competitions since 2010, when we ran a competition at the International Supercomputing Conference in Germany. Since then, we have been involved in starting regional conferences such as one in South Africa and one in China. We wanted to make such opportunities available to more people living in the Asia-Pacific region, home to 60 percent of the world's young people.

The results so far have been very encouraging. We have received submissions from places like China and Japan, which we expected because these countries have traditionally been strong at HPC and AI. But we have also seen teams coming from countries like Thailand and Bangladesh, which shows that they recognize the importance of these technologies and how competitions can help them progress.

What skills will you be looking out for in this competition?

GS: The competition has two main parts: there is an AI challenge and a HPC portion. Teams are required to build complete software solutions and run several pre-determined benchmarks. This requires that they understand all the software layers, from the application level all the way down to the hardware level.

The team that wins this competition will secure themselves a spot at the international competition in Germany. There they will face the additional challenge of building the hardware as well as the software. Understanding how the hardware components fit together and how to innovate on the hardware is the next step, and will give them a better understanding of how to develop comprehensive solutions. §

◀ **Mr. Gilad Shainer**
Chairman
HPC-AI Advisory Council

Photo credit: Cyril Ng/*Supercomputing Asia*

For more information on the APAC HPC-AI Competition, please visit:

http://hpcadvisorycouncil.com/events/2018/APAC-AI-HPC/

FINALISTS

COUNTRY	INSTITUTION
China	▪ Nanjing University ▪ Guizhou University ▪ Tianjin University ▪ Tsinghua University ▪ University of Science and Technology of China
Japan	▪ Tsukuba University
South Korea	▪ Gwangju Institute of Science and Technology ▪ Seoul National University
Singapore	▪ Nanyang Technological University ▪ National University of Singapore ▪ Singapore Polytechnic
Taiwan	▪ National Taiwan University ▪ National Cheng Kung University
Thailand	▪ Thammasat University ▪ Kasetsart University
Bangladesh	▪ BRAC University

AI AND HPC CONVERGE ON ASIA

Highlights from SupercomputingAsia 2018

Over 800 delegates from 24 countries convened in sunny Singapore for the inaugural SupercomputingAsia conference, a jam-packed conference bringing together workshops, an exhibition and five different co-located events. SupercomputingAsia 2018 (SCA18) was held at the Resorts World Convention Center on March 26–29, 2018, and builds on three successful years of the international Supercomputing Frontiers conference first run in 2015.

Organized by the National Supercomputing Centre Singapore (NSCC), the theme for SCA18 was the convergence of AI and HPC. It featured keynote speakers such as Dr. Goh Eng Lim, vice president and CTO of Hewlett Packard Enterprise; Dr. Rangan Sukumar, senior data analytics architect at the office of the CTO at Cray Inc; and Dr. Loy Chen Change, senior research consultant at SenseTime and adjunct assistant professor at the Chinese University of Hong Kong.

In addition to the technical program offered by the Supercomputing Frontiers Asia sessions, SCA18 also included the Asia Pacific Advanced Network Meeting (APAN45), a meeting of the largest and longest-standing community of advanced network experts; the inaugural Conference on Next Generation Arithmetic (CoNGA), focused on the latest developments in arithmetic that could replace the IEEE 754 floating point standard; Singapore-Japan joint sessions aimed at fostering bilateral collaborations; and a meeting of HPC and networking experts to discuss the groundwork for the Asia Pacific Research Platform (APRP), a planned network to bridge researchers across the region.

With industry talks by representatives from Cray Inc, Mellanox Technologies and Microsoft, sponsor-run exhibition booths and tracks dedicated to precision medicine and startups, SCA18 offered something for everyone. The conference also served as the launch platform for the inaugural APAC HPC-AI Competition, an opportunity for tertiary students to demonstrate their HPC and AI programming abilities (check out our interview with Mr. Gilad Shainer, chairman of the HPC-AI Advisory Council, on p. 30 to find out more!).

Last but not least, four prizes were given out during the opening ceremony under the auspices of the SCA18 Awards. Dr. Satoshi Sekiguchi, vice president and director general of the Department of Information Technology and Human Factors at the National Institute of Advanced Industrial Science and Technology, Japan, received the Asia HPC Leadership Award from Dr. Janil Puthucheary, Singapore's senior minister of state, Ministry of Communications and Information and Ministry of Transport.

Also receiving their awards from Puthucheary were Dr. Goh Eng Lim, a veteran of the HPC industry, who was awarded the Singapore Visionary Award; and Professor Lam Khin Yong, vice president (research) at Nanyang Technological University, Singapore, who received the Singapore Distinguished Service Award.

Lastly, the National Supercomputing Center in Wuxi, home to the Sunway Taihulight supercomputer, was recognized with the Asia HPC Outstanding Innovation Award.

1. Mr. Gerrit Bahlman, chairman of the Asia Pacific Advanced Network, giving the welcome remarks at the Resorts World Convention Center ballroom.
2. Dr. Janil Puthucheary, senior minister of state, Ministry of Communications and Information and Ministry of Transport, delivering the opening speech as the guest-of-honor at SupercomputingAsia 2018.
3. From left to right: Dr. Satoshi Sekiguchi, recipient of the Asia HPC Leadership Award; two representatives of the National Supercomputing Center in Wuxi, recipient of the Asia HPC Outstanding Innovation Award; Mr. Peter Ho (center), chairman of the NSCC steering committee; Dr. Janil Puthucheary, senior minister of state and guest-of-honor at SCA18; Dr. Goh Eng Lim, recipient of the Singapore Visionary Award; and Professor Lam Khin Yong, recipient of the Singapore Distinguished Service Award.
4. Dr. Goh Eng Lim, vice president and CTO of Hewlett Packard Enterprise, delivering his keynote lecture on the relationship between HPC and AI.
5. Professor Thomas DeFanti, distinguished professor emeritus at the University of Illinois and research scientist at Qualcomm Institute, delivering his keynote address.
6. Dr. Peter Lindstrom, project leader at the Center for Applied Scientific Computing of the Lawrence Livermore National Laboratory, US, delivering a keynote lecture at the Conference on Next Generation Arithmetic (CoNGA).

7. Dr. Satoshi Sekiguchi, vice president and director general of the Department of Information Technology and Human Factors at the National Institute of Advanced Industrial Science and Technology, shared Japan's plans for the world's first open public infrastructure for AI research, the ABCI system. Find out more in our interview with him on p. 22!

8. Mr. Song Qingchun, senior director of market development (Asia Pacific), Mellanox Technologies.

9. Professor Joe Mambretti, director of the International Center for Advanced Internet Research at Northwestern University. Check out our feature on p. 26 to find out more about the Global Research Platform, an initiative both DeFanti and Mambretti are leading.

10. Dr. Rangan Sukumar, senior data analytics architect at the office of the CTO at Cray Inc.

7

8

9

10


11

12

13


14

11. From left to right: Mr. Robert Hodges, senior director (medtech) of investments for SGInnovate; Dr. Mohammad Asif Khan, president of the Association for Medical and Bio-Informatics, Singapore (AMBIS); and Dr. Benedict Yan, director of the Molecular Diagnostics Centre at the National University Hospital, Singapore.

12. Ms. Vivien Chow, director of applied innovation and partnership at the Government Technology Agency of Singapore (GovTech), sharing about her organization's translational R&D grants program.

13. Dr. Sidney Yee, CEO of the Diagnostics Development Hub (DxD Hub) and executive vice president of Exploit Technologies (ETPL).

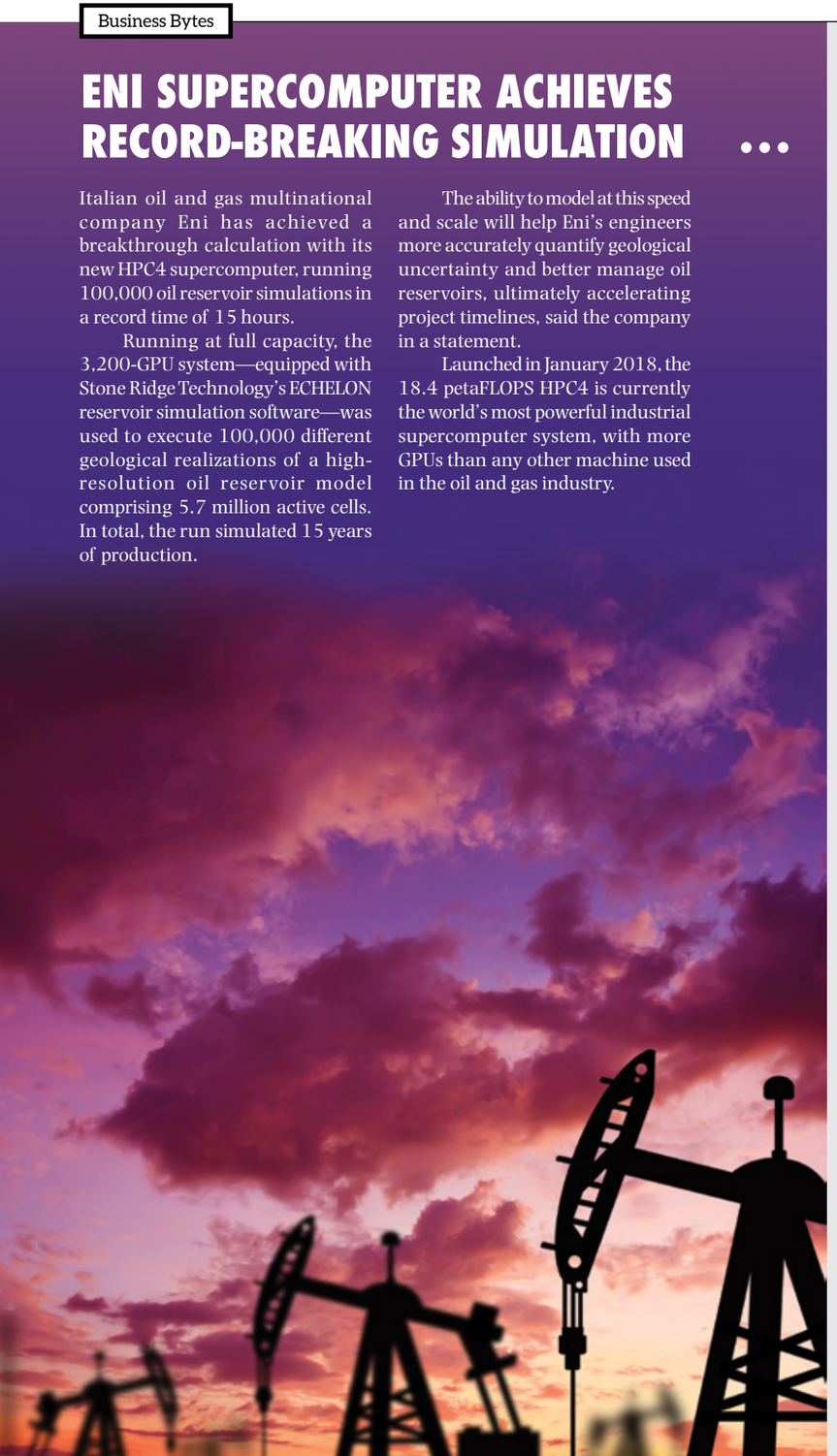
14. Professor Satoshi Matsuoka, director of the RIKEN Center for Computational Science (R-CCS), speaking on a panel.

Photo credits:

National Supercomputing Centre Singapore

Business Bytes

ENI SUPERCOMPUTER ACHIEVES RECORD-BREAKING SIMULATION



Italian oil and gas multinational company Eni has achieved a breakthrough calculation with its new HPC4 supercomputer, running 100,000 oil reservoir simulations in a record time of 15 hours.

Running at full capacity, the 3,200-GPU system—equipped with Stone Ridge Technology's ECHELON reservoir simulation software—was used to execute 100,000 different geological realizations of a high-resolution oil reservoir model comprising 5.7 million active cells. In total, the run simulated 15 years of production.

The ability to model at this speed and scale will help Eni's engineers more accurately quantify geological uncertainty and better manage oil reservoirs, ultimately accelerating project timelines, said the company in a statement.

Launched in January 2018, the 18.4 petaFLOPS HPC4 is currently the world's most powerful industrial supercomputer system, with more GPUs than any other machine used in the oil and gas industry.

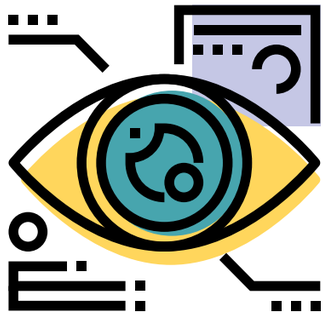
NVIDIA SETS ITS SIGHTS ON HEALTHCARE

Computer hardware company NVIDIA has revealed plans for Project Clara, a new medical imaging supercomputer. Announced in March 2018 at the GPU Technology Conference by NVIDIA CEO Mr. Jensen Huang, the supercomputer signals an entry into the healthcare industry for the company, which is better known for its GPU and autonomous vehicle technology.

Billed as a platform for medical imaging, Project Clara will allow medical professionals to access artificial intelligence-driven medical image analysis remotely via the cloud. While it would take decades to upgrade the three million medical instruments installed around the world, Project Clara will let NVIDIA "virtually update every system that's out there," said Huang in his keynote.

Further advantages of the platform include the ability to run many computational instruments simultaneously, as well as perform computations for any type of imaging, including computed tomography, magnetic resonance imaging, ultrasound, X-ray and mammography, wrote Ms. Kimberly Powell, vice president of healthcare at NVIDIA, in a company blog post.

CHINA'S SENSETIME VALUED AT US$4.5 BILLION



Industry insiders have put Chinese artificial intelligence (AI) company SenseTime Group Ltd's value at US$4.5 billion, after the company announced in April 2018 that it had raised US$600 million in a Series C funding round led by Alibaba Group Holding Ltd.

Known for its facial recognition technology, SenseTime also develops AI for other applications, including video analysis and autonomous driving. Other investors in the record-breaking funding round included Chinese e-commerce company Suning.com Co Ltd and Singapore state fund Temasek Holdings (Private) Ltd.

Now the world's most valuable AI startup, SenseTime has tripled its worth in less than a year. In July 2017, the company was valued at US$1.5 billion after a Series B funding round.

"SenseTime has established an AI ecosystem anchored with robust research, deep industry collaboration, and diverse partnerships. Our round C funding will maximize these advantages by accelerating the development of a global footprint with a larger ecosystem incorporating both domestic and overseas partners," said Dr. Li Xu, SenseTime's co-founder and CEO.

CAMBRICON LAUNCHES NEW AI CHIPS

Chinese AI chip maker Cambricon Technologies Corp Ltd has launched two new products: the MLU100, a cloud-based AI smart chip; and the 1M, an upgraded, third-generation version of Cambricon's AI processor.

The announcements, made at a launch event in May 2018, come amid China's intensified efforts to reduce its dependence on technology imports.

The MLU100 supports cloud-based machine learning applications, including visual, speech and natural language processing. The chip is capable of processing under complex scenarios, including big datasets, multi-tasking, multi-modality and low latency, with a peak performance of 166.4 teraFLOPS. Meanwhile, the 1M was designed to speed up a range of AI applications, and is available in three versions to meet different levels of processing needs.

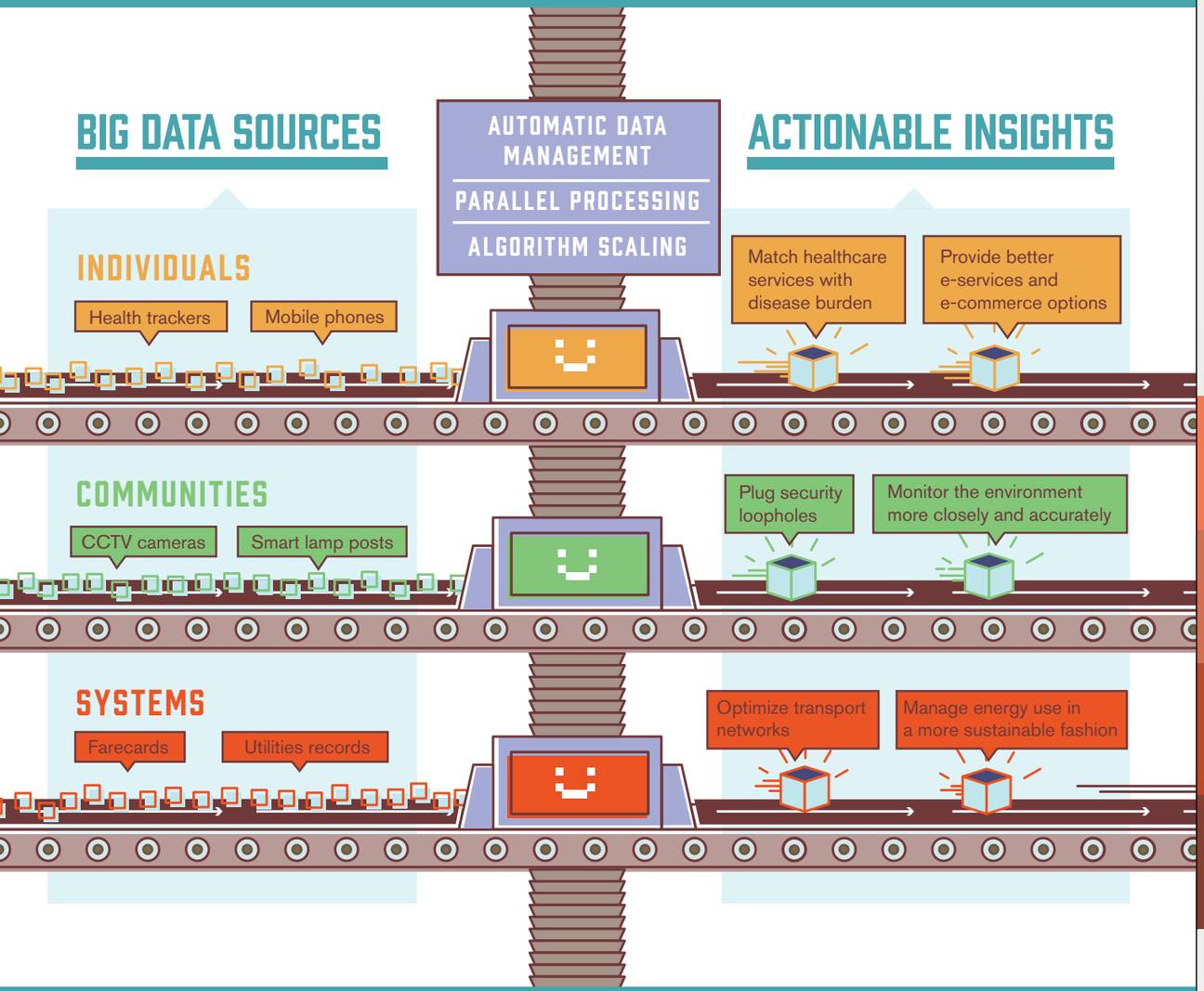
"The MLU100 chip is a milestone and a new starting point for Cambricon, and a benchmark for smart chips," wrote Cambricon CEO Mr. Chen Tianshi in an internal letter, reported *China Money Network*.

Super Snapshot

BUILDING BLOCKS FOR BETTER LIVING

From data deluge to deft decisions

Collecting data is just the first step—for the data to make sense and be useful, it needs to be converted from its raw form into actionable insights. The challenge is immense, but supercomputers are up to the task.



To Lead in HPC & AI You Need to Join the Leading HPC & AI Network of Experts

HPC-AI ADVISORY COUNCIL

NETWORK OF EXPERTISE

With over ten years of ongoing success and collaborations, more than 400 member companies, five world-wide annual industry conferences & three technology competitions, the HPC-AI Advisory Council is a network of expertise.

Join now and benefit from

- An all access platform to deep expertise and the innovative technologies being used for HPC and AI
- A catalyst in continuing education, developing critical hands-on skills
- An open forum for sharing new ideas, novel research and revolutionary breakthroughs

For more information or to join write us at: info@hpcadvisorycouncil.com
Or go to: www.hpcadvisorycouncil.com