



Atlas 900
fastest AI training cluster
全球最快的AI训练集群
256~1024 PFLOPS@FP16

Huawei unveils computing strategy and world's fastest AI training cluster

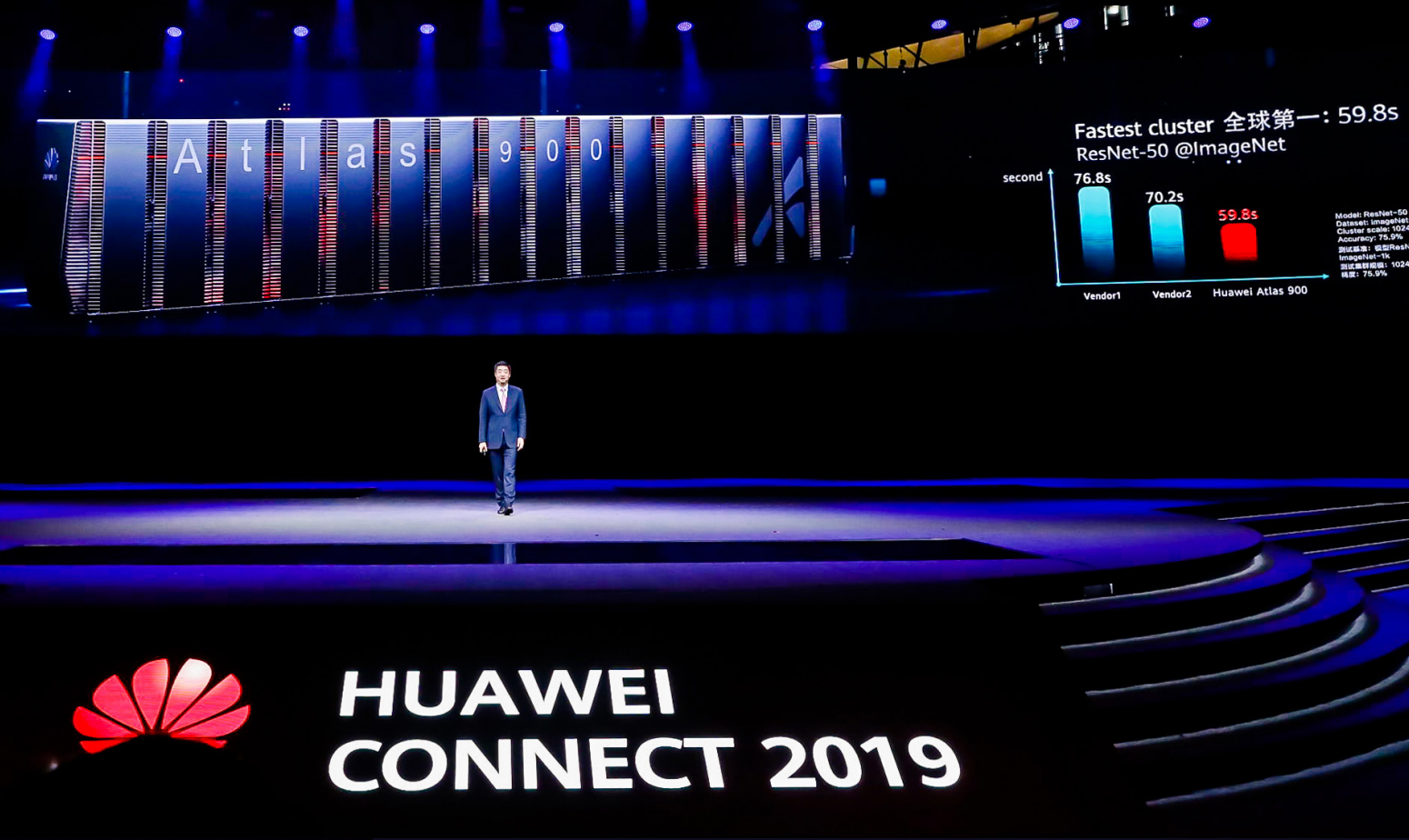
At HUAWEI CONNECT 2019, Huawei announced its strategy for the computing market and released Atlas 900, the world's fastest AI training cluster. Below is the keynote speech given by Huawei Deputy Chairman Ken Hu on September 18, 2019.

By Ken Hu, Deputy Chairman, Huawei

Two years ago we announced our company-wide mission: to bring digital to every person, home, and organization for a fully connected, intelligent world. In this world, we believe that connectivity and computing will be woven into the fabric of everything.

Two key technologies: connections and computing

When most people think Huawei, they think connections. It's true we've been investing nonstop in connectivity for the past 30 years. From fixed



networks to wireless, from 2G, 3G and 4G, all the way up to 5G, we've made quite a bit of progress in the industry.

But our work doesn't stop at connectivity. If our goal is to build an intelligent world, both connections and computing are key – they're inseparable. The two are interdependent, one pushing the other forward, both developing in step.

So wherever there's a connection, you'll have computing. And where there's computing, you'll have connections too. In terms of Huawei's investment, they're equally important. In the past, we mostly talked about connections. Now, I'd like to focus on computing.

People and computers are closer than ever

Allow me to share a few thoughts on the computing

industry.

After the birth of the first computer in 1946, we've seen incredible changes in form factor. From old mainframe computers to PCs, from desktops and laptops to tablets, to technology that travels with us – like mobile phones and wearables – computers are getting smaller, more powerful, and closer to us than ever before.

In effect, computers have become an extension of ourselves. Our capabilities. And based on this trend, it's clear that computing as an industry has boundless potential.

Our approach to computing is still evolving

For the past 70 years, our approach to computing has been evolving nonstop. I first came into contact with computers back in university. In programming class,

“

Computing and intelligence will be ubiquitous – not limited to the cloud, but present in everything from your headphones and smartphones to specialized edge computing for things like gene sequencing.

”

my teacher taught us that all you have to do is give a computer an equation and it will give you the result.

In the early days of computing, all the industry had was rule-based computing. You could compute anything as long as you could distill it down to a clear set of rules and parameters. And this was great for things like analyzing census data or calculating the trajectory of a moving object.

But for other types of problems for which you couldn't define clear rules and parameters, like voice recognition, image recognition, or real-time translation, rule-based computing didn't quite have what it takes.

To solve these types of problems, scientists developed statistical computing models. They are great for dealing with undefinable problems, and this laid the foundation for artificial intelligence.

Statistical computing will soon become the mainstream. We estimate that five years from now, AI computing will account for more than 80 percent of all the computing power we use around the world.

Computing in the age of

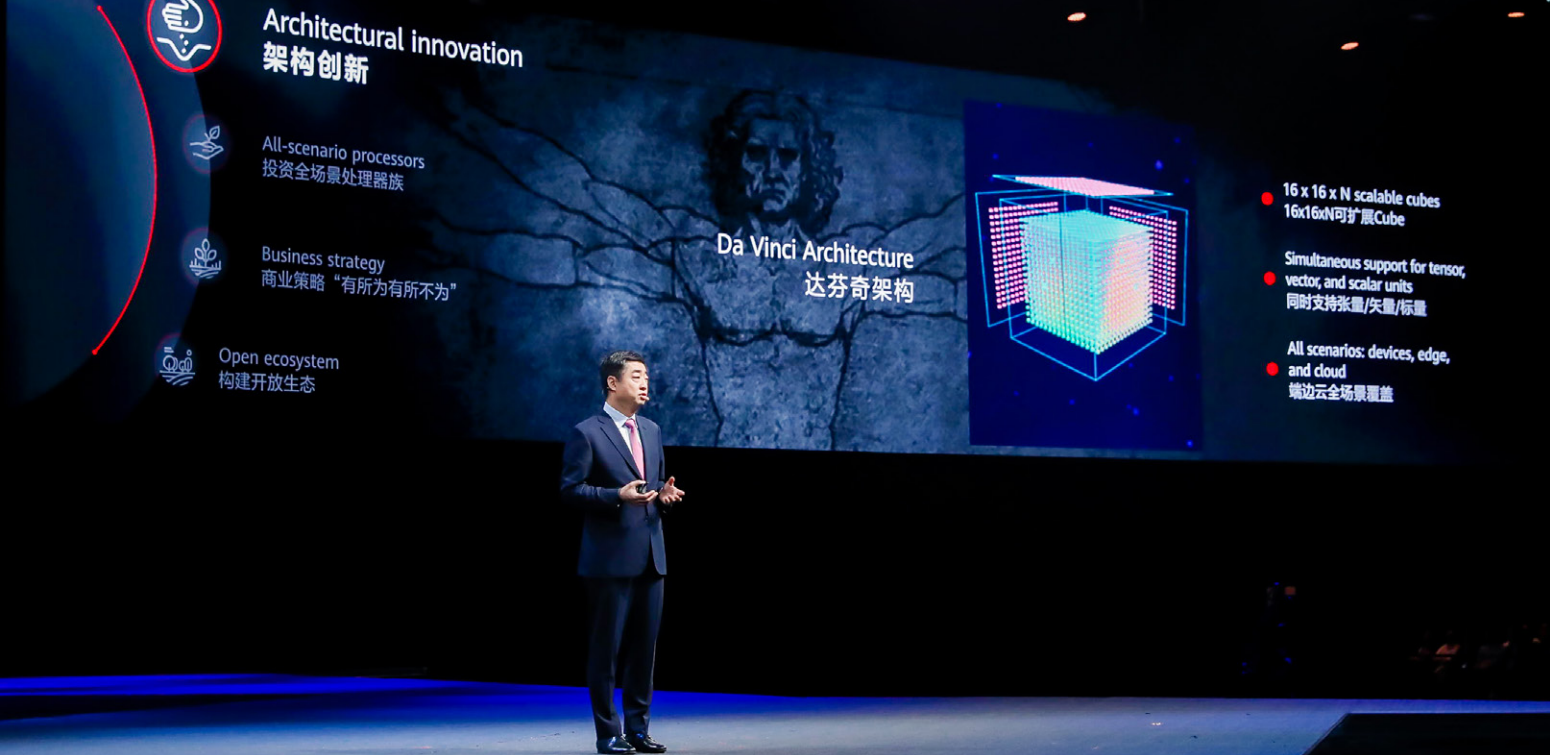
intelligence

In the age of intelligence, we'll see three major computing trends.

The first is demand for incredible computing power. Statistical computing is essentially a form of brute force computing; it eats up computing resources.

If you want to train an algorithm to recognize a cat, you need to feed it millions of images and let the system come to its own conclusions about what exactly defines “cat-ness”. This takes a metric ton of computing power. More complicated applications like autonomous driving, astronomy, and weather forecasting will take even more compute.

Second, computing and intelligence will be ubiquitous – not limited to the cloud, but present in everything from your headphones and smartphones, to specialized edge computing for things like gene sequencing. These three types of computing – on-device computing, specialized edge computing, and brute force computing on the cloud – form the computing landscape of the intelligent world.



Third, to better serve people in life and work, computing needs to be managed cooperatively across the computing landscape.

The cloud should only handle general-purpose model training, providing background support for personalized on-device AI and specialized edge computing. This not only improves experience, it's also better for protecting privacy.

We have a lot of challenges ahead of us. We need to beef up our computing power, explore new architectures, and develop processors that meet people's needs across all scenarios.

The US\$2 trillion market

But the bigger the challenge, the bigger the opportunity. According to Gartner, by 2023 the computing market will be worth more than US\$2 trillion. A massive blue ocean market.

Huawei's computing strategy

At Huawei, we've decided to invest more. Our strategy focuses on four areas:

Architecture innovation

In a future where computing and intelligence are everywhere, computing power will be the bedrock of everything.

The industry doesn't have nearly enough computing power to meet demand. Compute supply relies on processor performance. And since we're reaching the limits of Moore's law, if the industry wants to provide a steady and abundant supply of affordable computing power, we need to make breakthroughs in processor architecture.

At the same time, Huawei's portfolio covers networks, devices, and public cloud services. The ability to provide seamless intelligence across device, edge, and cloud is a core part of our business.

That's why we developed our Da Vinci architecture. We want to make computing and intelligence as readily available as oxygen, so developing this architecture was the natural next step. Da Vinci is the only processor architecture in the world that can support all-scenario intelligence across device, edge, and cloud, and it will pave the way for future growth

“

Our software will be open source. That includes software like server operating systems, databases, and AI development frameworks. This will help our partners develop better commercial software more easily.

”

in the computing industry.

We are investing in all-scenario processors

Processors are the basic building blocks of the computing industry. After years of hard work and investment, we've released several families of processors for different scenarios.

We have a full lineup: Kunpeng processors for general purpose computing, Ascend processors for AI, Kirin processors for smart devices, and Honghu processors for smart screens. Moving forward we will release more processors for different scenarios.

Our business strategy

Simply put, there are things we'll do and things we won't. To start with, we won't sell our processors directly. Broadly speaking, we will provide them to our customers in the form of cloud services and to our partners in the form of components, prioritizing support for integrated solutions.

We will open up hardware like AI servers, accelerator cards, and modules for our partners, giving them the

components they need to integrate AI computing into their own products and solutions.

Our software will be open source. That includes software like server operating systems, databases, and AI development frameworks. This will help our partners develop better commercial software more easily.

We also enable application development and portability. We won't develop applications ourselves, but we will provide tools and teams to help our partners develop and port their applications more efficiently.

By drawing clear lines between what we do and what we don't, we hope to better support the business development of our partners.

We will build an open ecosystem

Computing has always been an open industry. No single company can prop up the entire industry on its own; healthy growth requires an open ecosystem and global collaboration.

Back in 2015 we announced our Huawei Developer Program. Since then we have made great progress,

“

We will invest US\$1.5 billion in our dev program. We want to expand the program to 5 million developers and enable our partners around the world to develop the next generation of intelligent applications and solutions.

”

empowering more than 1.3 million developers and 14,000 ISVs around the world.

I'd like to thank Huawei's partners and developers for their ongoing support.

Moving forward we will invest another US\$1.5 billion in our dev program. We want to expand the program to five million developers and better enable our partners around the world to develop the next generation of intelligent applications and solutions.

General purpose computing strategy

Next, let's take a look at how we'll move forward with this strategy, starting with general purpose computing.

General purpose computing is all about Kunpeng. We will develop the Kunpeng series of processors into the industry's most competitive processors for general purpose computing.

Building on Kunpeng, we will invest in key technologies and products like accelerator cards, servers, operating systems, databases, compilers, and other tools. We will increase our investment to connect the whole value

chain, build out the Kunpeng ecosystem, and give our partners confidence in its growth potential.

Developing the Kunpeng ecosystem

Right now we're working with our partners to lay the foundation for the Kunpeng ecosystem.

Working together with local governments and partners, we're helping different communities make the most of their local industrial strengths by building Kunpeng innovation hubs and incubators. These hubs will bring together partners across the ecosystem, where we can carry out application pilots, cultivate talent, and develop standards as a team.

So far we have set up Kunpeng innovation hubs in cities like Beijing, Shanghai, and Shenzhen. We look forward to having more partners join us across the ecosystem.

AI computing: Full-stack, all-scenario

Last year, Eric Xu announced our full-stack, all-scenario AI portfolio. At that time we had only

“

Atlas 900 packs a ton of computing power, and we want to get it in your hands as soon as possible. So we've deployed it on Huawei Cloud as a cloud service.

”

launched the Ascend 310 processor for inference and our ModelArts application development platform.

We've made a ton of progress in just one year. This year we released an AI processor for model training, the Ascend 910. And just last month we announced MindSpore, our AI computing framework. After a year of hard work, we've managed to roll out our entire portfolio. So rest assured: we're ready to do this, and we'd love to have you join us.

I'm also excited to announce the release of a brand new, heavyweight product that brings together decades of technological expertise at Huawei. The Atlas 900.

Atlas 900: The world's fastest AI training cluster

The Atlas 900 is the world's fastest AI training cluster, combining the power of thousands of Ascend processors.

So how fast is it? Using the ResNet-50 model, which is industry standard for measuring AI training performance, we put our Atlas 900 to the test. It finished the entire training in less than a minute – 59.8 seconds, to be precise. This is about 10 seconds

faster than the previous world record.

You might think, 10 seconds, so what? But this is some pretty serious stuff. Imagine it like this: A sprinter crosses the finish line, and has enough time to drink a bottle of water before the second person arrives.

Atlas 900 is a powerhouse of AI computing, and it will bring new possibilities to different fields of scientific research and business innovation – anything from astronomy to oil exploration. For models that used to take several months to train, Atlas 900 can handle them in seconds.

In astronomy, we're working with the Shanghai Astronomical Observatory and Square Kilometre Array (SKA) to process epic amounts of data. In this field, you need as much computing power as you can get.

Space exploration is incredibly important for the entire world, and it's clear that there are a lot of challenges ahead.

Let's take a look at how Atlas 900 can help.

If we look at a map of the Southern Hemisphere, we can see more than 200,000 stars in a single image –



far more than you can see with the naked eye.

Before Atlas, if an astronomer wanted to find a celestial body with specific features in an area of sky that big, it would take 169 days of full-time work.

Atlas 900 can speed up this process significantly. It was able to scan through mountains of data to locate and identify a specific type of star in only 10 seconds.

From 169 days to 10 seconds. It's revolutionary. And this will free up scientists' time for more important work.

Announcing Ascend-based Huawei Cloud EI cluster services

Atlas 900 packs a ton of computing power, and we want to get it in your hands as soon as possible. So we've deployed it on Huawei Cloud as a cloud service.

We're making it available at a great discount for universities and research institutes around the world. If you're interested, go ahead and apply now – we'd love to have you try it out.

We're confident that Atlas 900 will help bring the power of AI to all industries.

An ocean of boundless potential is waiting

Today, all industries are rushing to go digital and artificial intelligence is all the rage. The computing industry is booming, and will only continue to grow.

We'll continue to invest. Starting with the most difficult challenge ahead of us – making breakthroughs in architecture – to developing processors, we're going to help expand the industry and build out the ecosystem.

We strongly believe that facing challenges head-on is the only way to build our competitive strengths – the only way to build out the market for our partners. We're going to use the best technology to solve the world's hardest problems and make the impossible possible.

This is a new age of exploration. An ocean of boundless potential is waiting, but just one ship won't cut it. So, we've launched a thousand ships.

Let's work together, seize this historic opportunity, and advance intelligence to new heights. 