

# 计算机运力世界大比拼

时代在变。十年前,英国《经济学家》曾用以工业用电量为主的一套指标来评估中国 GDP,并将其命名为“克强指数”。

今天,中国经济实力最强的北上广深等大城市都在全力比拼一个新的指标——计算力。因为在数字经济时代,这个指标将决定它们在未来城市大战中的生死。甚至,像安庆这样的三四线城市,也在“着急上火”,因为慢一步,就会被狠狠甩在身后。

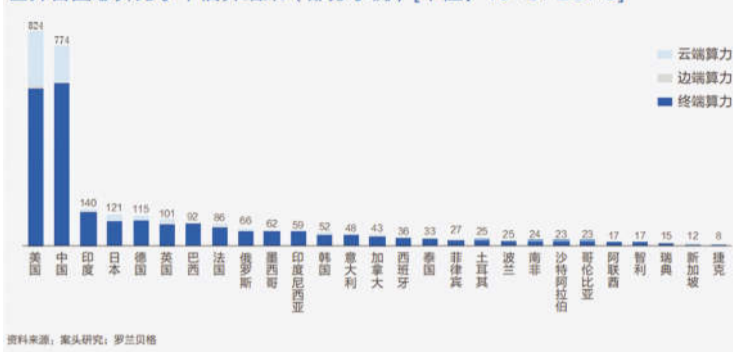
□ 张静波



多元化的格局。

## 突飞猛进

世界各国总算力水平估算结果(部分示例)[单位: 10<sup>9</sup> GFLOPS]



全球各国总算力排行榜。

今天,我们所面对的场景和需求,远比识别一张图片更复杂、多样化。从VR/AR到无人驾驶,从工业机器人到远程医疗……AI正改变我们的生活。而它背后的算力支撑,也从云到边缘,到设备端,变得无处不在。

19世纪70年代,发电机的问世催生了第二次工业革命。电力成为驱动全球经济发展的强大动力,人类由此迈进电力时代。一百多年后的今天,算力正变得跟电力一样触手可及。未来,在数字经济时代,谁掌握更强的算力,谁就将拥有世界。

来自罗兰贝格等国际权威机构的数据显示,在全球各国总算力的排行榜上,中美两国遥遥领先。紧随其后的,也都是英、法、德、印、日等经济强国。

算力不但是衡量国家实力的

指标,也正在中国各大城市间引发一场热战。从北上广深到杭州、南京……大家都憋着一股劲儿,建设数字经济城市,争当算力之都。就连安庆这样的三四线城市,也在拼命学习、追赶。

上世纪90年代,公路、铁路、电力等旧基建,奠定了中国制造业大国的地位。而今,围绕城市间的算力之争,一大批集约化的大型数据中心正拔地而起。这背后,是一场规模空前的新基建运动。

在这场运动中,算力正取代电力,成为新基建的核心。因为算力,是数字经济时代的基础设施,是新的生产力。以传统制造业为例,在算力上每投入1美元,可带来10美元的产出。未来,随着新基建的大规模落地,算力的提升将为各行各业带来质的飞跃,并筑起中国数字经济的新底座。

## 数据大爆炸

出的是近年来一个迅猛增长的计算趋势:传统CPU一家独大的格局正在终结,全世界正迎来一个新的异构计算时代。

在这个新的计算时代,各种非x86架构的算力百花齐放,以满足复杂场景下,对算力更加多样化的需求。比如,FPGA和ASIC。

此前,无论CPU还是GPU,都属于通用芯片,它们基于半个多世纪前的冯·诺依曼结构设计,可胜任大多数的通用计算场景。但这种基于指令集的计算架构,需要在内存和处理单元间传输数据,不但影响了效率,还增加了功耗。

FPGA和ASIC不同,它们是一种专用芯片,没有指令集,也无需共享内存,而直接以并行和流水线的方式处理数据。不但速度快,功耗也低得惊人。当然,代价也不菲。由于是专用芯片,它们只能用于特定目的。其中,ASIC完全针对客户定制;而FPGA属于半定制,现场可编程,可重复使用。

自从1985年,赛灵思(Xilinx)创始人罗斯·弗里曼发明FPGA以来,这种芯片已广泛用于许多要求低时延的场景。比如,5G通信基站

和自动驾驶,后者几毫秒的刹车时延就可能致命。还有金融市场高频交易、军用导弹制导系统等。

不仅如此,今天,FPGA正与GPU一起,成为全球高性能计算(HPC)和大型数据中心的加速器。前者以速度取胜,后者擅长大量数据处理,它们如同人的四肢。CPU则形同大脑,负责统筹、调度各种计算资源。

2010年以前,微软必应(Bing)搜索引擎基本上靠CPU驱动。但大数据和人工智能时代的到来,让传统设备不堪重负。最终,微软不得不用FPGA来加速计算。结果,无论数据吞吐量,还是延迟,都得到10倍级的提升。

如今,微软已在自家数据中心、一半以上的服务器部署了FPGA加速卡。亚马逊、IBM等全球科技大佬也纷纷入局。就连英特尔,也顾不上身段,斥资167亿美元,收购了全球第二大FPGA芯片制造商Altera,试图续写神话。谷歌则基于ASIC架构,开发了一款专门用于加速神经网络计算的TPU芯片。

众多非x86架构的芯片,共同托起了一个崭新的计算时代。

## 新计算时代

20年前,人类进行一次全基因组测序,需要数年之久,现在只需要1天。在这背后,是成千上万倍的算力提升。

新计算时代,给人类带来的远不止算力的提升。在算力需求还不大的年代,功耗并非大问题。但如今,随着全球HPC和大型数据中心的兴建,高耗能正成为一个棘手问题。

另一方面,手机等小型终端设备,也对续航和功耗,提出越来越苛刻的要求。传统CPU,无论硬件设计还是软件设计,功耗都不占优。硬件层面,同等算力下,采用流水线设计的ASIC和FPGA,功耗比CPU低很多。

软件层面,x86 CPU采用复杂指令集,与采用精简指令集的ARM等芯片比,功耗天生就吃亏。由此可以预计,未来ARM、FPGA、ASIC等低功耗的绿色计算,将大行其道。

多元化的格局,也为全人类提供了一种更安全的计算,同时避免对某种单一计算架构的过度依赖。从这个意义上讲,未来的计算将满足四个特征:异构、集约化、绿色和安全。

这样一个新计算时代,对今天的中国意义尤其深远。

1882年9月4日,爱迪生亲手合上世界上第一个商用电力系统的电闸,电流沿电线从珍珠街电站流向曼哈顿金融区,照亮大半个纽约。4年后,西屋电气创始人威斯汀豪斯又建造了美国第一个商用交流电电力系统。

一轮接一轮电力大基建,使得美国人率先开启了电气时代,并为美国经济注入强劲动力。最终,美国取代欧洲,成为新的世界中心。

一百多年后,随着数据大爆炸,算力正成为驱动经济的新动力。在中国,12.9亿手机用户、近1亿个体户和

2000多万家企业,每天都要靠算力来完成各种线上线下交易。这背后,每年高达数十万亿元的电商交易,要通过云端的算力来处理。

不仅如此,算力还支撑着从VR/AR到自动驾驶,从人工智能到工业互联网……所有我们生活中的新科技、新产业,正在改造着传统产业的面貌。

因为有了更强的算力,智慧工厂、智慧农场、智慧交通、智慧医疗、智慧政务……不断从理想变为现实,并由此带来更高的效率和更高的产出。

在美国,随着智能机器人等基于智能计算的新技术不断应用在传统制造业,未来预计将为美国制造业带来1700亿美元的额外产出。与此同时,智能电网的大规模建设,将为其带来1.3万亿—2万亿美元的收入。

因为算力对经济的这种强大驱动力,全世界正掀起一场大规模的新基建。在中国,新基建更是已上升至国家的发展大战略。

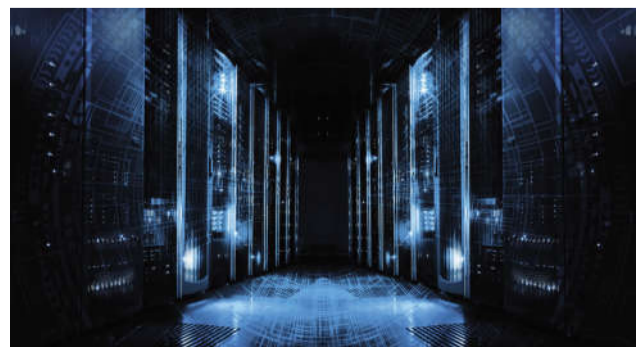
从5G、特高压、高铁、充电桩,到AI、大数据中心、工业互联网……七大领域,背后都需要算力作支撑。

也因此,短短数年,全中国各地掀起一场以兴建大数据中心为代表的算力大比拼。北京、杭州、深圳、上海,你方唱罢我登场。就连昔日偏远的贵州,也凭借华为、苹果、腾讯等公司的大数据中心落地,上演了一场GDP逆袭。

在中国国家七大超算中心,国产自研的申威26010处理器正在支撑着全世界仅次于美国的最强算力。一个动力充沛、多样性的算力时代正在来临。

未来,这个新的算力时代不但将催生一个万亿级产业,更将通过新基建,源源不断地为中国经济注入澎湃的动力。

(《华商周刊》、《新财经》)



多样性的算力时代正在来临。

自从1946年,冯·诺依曼发明第一台电脑ENIAC以来,算力一直在突飞猛进。

半个世纪前,第一次将人类送上月球的阿波罗11号飞船,其所搭载的电脑,CPU主频只有0.043MHz。今天,任何一部普通的手机,主频都在2500MHz以上,是登月电脑的5万倍!

即便如此,仍无法满足人们对算力的需求。在IDC/EMC发布的一份报告中,最近十年来,全世界算力的增长远远滞后于数据的增长。导致这一算力瓶颈的,是两场革命:一个是互联网革命,另一个是AI革命。

互联网让更多设备联接在一起,未来,物联网(IoT)还将带来千亿级的设备接入。海量的设备,叠加复杂的应用场景,引发数据大爆炸。

1992年,全人类每天只产生100GB数据。而今天,全球70亿人,每人每天产生的数据高达1.5GB。仅一辆自动驾驶汽车,一天就能产生64TB数据,足以填满32块硬盘。

爆炸式增长的数据,哺育了人工智能(AI),使得深度学习等过去难以实践的各种算法,得以喂养、训练,并大规模应用。这反过来,对算力提出巨大的需求。

2012年,谷歌大脑为了从1000万张图片中识别出一只猫,整整动用了1000台电脑、16000个CPU。

一方面,数据大爆炸和人工智能对算力提出前所未有的需求。另一方面,传统的计算架构却走进了死胡同。

过去数十年,全球算力基本上是CPU一家独大。尤其1978年,英特尔推出x86架构以来,CPU的算力更是以摩尔定律——每隔18个月翻一倍的速度,突飞猛进。

然而近年来,随着半导体工艺制程逼近极限,CPU算力的增长也渐趋平缓。曾经推动全球计算产业大发展的CPU,如今正陷入算力瓶颈。

21世纪头十年,研究人工智能的科学家们曾到处寻找一种高效且低成本的算力,以便帮他们完成大吞吐量的数据处理,训练算法。最终,他们大多数选择了英伟达的GPU,而不是英特尔的CPU。

原因是,这种非x86架构的图形处理芯片,拥有动辄几百个内核,能够并行地同时处理数千个线程。正是这批人,用GPU推动了一场人工智能革命。

作为这场革命的引领者,谷歌大脑之父吴恩达后来重新做了一次识别猫脸的实验,结果只用了16台电脑、64个GPU,比用CPU整整节省了近百倍。

GPU只是一个缩影,它反映