



开放数据中心委员会
Open Data Center Committee

数据中心算力白皮书

[编号 ODCC-2020-01008]

开放数据中心委员会
Open Data Center Committee



目录

1. 研究背景.....	1
2. 研究现状.....	3
2.1. 超算算力评估.....	4
2.1.1. TOP 500	4
2.1.2. Green 500	5
2.2. 常规服务器算力评估.....	7
2.2.1. SPEC CPU	7
2.2.2. SPEC power.....	7
2.2.3. MLPerf	8
2.2.4. 服务器能效规范	9
2.3. 电能利用效率（PUE）评估	10
2.4. 小结.....	11
2.4.1. 算力构成复杂	11
2.4.2. 算力测试难以开展	11
2.4.3. 尚无合理完善的评估体系	11
3. 算力及算效指标.....	13
3.1. 算力的定义.....	13
3.1.1. 通用计算能力	14
3.1.2. 高性能计算能力	19
3.1.3. 存储能力	27
3.1.4. 网络能力	28
3.2. 数据中心算力模型（CP）	31
3.2.1. 方法	31
3.2.2. 模型	33
3.3. 数据中心算效模型（CE）	36
4. 数据中心情况分析	37
4.1. 我国机架总体情况.....	37

4.2.	我国上架率情况.....	40
4.3.	服务器出货量.....	41
4.3.1.	CPU 服务器.....	41
4.3.2.	GPU 服务器.....	43
4.4.	数据中心能效情况.....	44
5.	我国数据中心算力情况.....	46
5.1.	算力分析.....	46
5.2.	算效分析.....	48
6.	发展趋势.....	48
6.1.	异构加速计算需求旺盛.....	48
6.2.	大型、边缘两极化发展.....	49
6.3.	计算、存储、网络深度融合.....	49
6.4.	对更高计算效能的追逐.....	50
6.5.	量子计算大有可为.....	50
7.	相关后续工作.....	51
8.	总结.....	52

前言

2020年3月4日，中共中央政治局常务委员会召开会议，明确指出“加快5G网络、数据中心等新型基础设施建设进度”，将数据中心纳入“新基建”范畴。4月20日，国家发改委明确新型基础设施的范围，数据中心作为算力基础设施，成为信息基础设施的重要组成部分。

数据中心是一个汇集大量服务器、存储设备及网络设备的基础设施，是人工智能、物联网、区块链等应用服务的重要载体。数据中心算力水平的提升将会带动全社会总体算力的提升，满足各行业的算力需求。

本白皮书梳理算力的研究现状并通过分析，提出一种数据中心算力和算效的衡量方式，据此测算出我国当前的数据中心算力和算效的水平，并基于现阶段的研究成果开展数据中心算力和算效的相关评估。对数据中心算力及算效进行衡量与评估将为数据中心产业发展提供重要的指导，数据中心监管部门、第三方服务商及相关从业人员能够根据数据中心算力和算效情况判断行业发展趋势，同时为数据中心未来算力规划和部署提供思路。

顾 问

中国信息通信研究院： 何宝宏

中国电信北京研究院： 杨明川

编写组

中国信息通信研究院： 郭亮、吴美希、李宁东、邱奔

中国电信北京研究院： 王峰、赵继壮、刘倩

Intel： 龚敏、王栋

AMD： 谷建、王文栓、曲大健

美团： 朱洪涛、胡湘涛、王中平

此外，感谢为本白皮书做出贡献的华为、希捷、西部数据、三星等企业的相关专家。

审 校

中国信息通信研究院： 李洁

数据中心算力白皮书

1. 研究背景

2020年3月4日，中共中央政治局常务委员会召开会议，明确指出“加快5G网络、数据中心等新型基础设施建设进度”，将数据中心纳入“新基建”范畴。4月20日，国家发改委明确新型基础设施的范围，数据中心作为算力基础设施，成为信息基础设施的重要组成部分。今年两会上，政府工作报告明确指出要“加强新型基础设施建设，发展新一代信息网络，拓展5G应用，建设数据中心”。¹

以人工智能、物联网、AR/VR等技术为核心的智能化应用场景正在不断地改善着人们的生产方式和生活水平，同时这些新兴互联网应用的快速普及也对算力水平提出了更高的要求。算力已经成为推动计算产业发展的重要动力，对算力的投入不仅能够带动相关产业的科技创新，同时还能够极大地推动经济发展，提升国家综合竞争力，世界各国高度重视算力的发展，并且通过政策引导及行业联盟等形式加强对算力资源的管控，希望能够进一步提升国家算力总体供给，提升算力资源的利用率。

海外巨头在算力方面动作不断：Intel以167亿美元收购FPGA生产商Altera。Intel预计到2020年，30%以上的服务器CPU芯片将配备一个FPGA协处理器；AMD经过一系列业务分拆逐步从IDM芯片

¹ 中国信息通信研究院&开放数据中心委员会，《数据中心白皮书（2020）》

商转型为 Fabless 设计商，目前已经在 Rome 系列使用 7nm 工艺；2020 年 4 月，NVIDIA 以 69 亿美元成功收购 Mellanox（迈络思），并推出网卡新品 NVIDIA Mellanox。

新基建发布后，国内头部企业在算力方面在之前的积累上，也在不断加大投入。腾讯云发布首款拥有完全自主知识产权的服务器产品星星海 SA2，该云服务器实例已经被应用于腾讯内部包括腾讯会议、腾讯教育、春节 QQ 红包、微信、视频转码、广告检索等产品和业务中，表现出亮眼的性能和可靠的稳定性²。阿里巴巴在 ODCC 牵头发起的方升服务器开源项目，旨在打造云基础设施行业的 COCI(China Open Cloud Infrastructure)技术标准，引领行业发展，共建中国云数据中心硬件生态，目前已经部署超过 10 万节点。百度推出的 FPGA 版本百度大脑，运用到包括语音识别、广告点击率预估模型、DNA 序列检测以及无人车等业务中，性能提升 3 倍以上。

数据中心是一个汇集大量服务器、存储设备及网络设备的基础设施，是人工智能、物联网、区块链等应用服务的重要载体。数据中心算力水平的提升将会在很大程度上推动全社会总体算力供给，满足各行业数字化转型过程中的算力需求。对数据中心算力水平及算力能效进行评估将为数据中心产业发展提供重要的指导，数据中心监管部门、运营商及相关从业人员能够根据数据中心各地区算力分布及算力供

² <https://cloud.tencent.com/act/pro/starlake#%E5%BA%94%E7%94%A8%E6%A1%88%E4%BE%8B>

需状况判断行业发展趋势，同时为数据中心未来算力规划和部署提供更好的解决思路。

2. 研究现状

目前，关于算力的研究主要集中在以下几个方面，一是算力评估指标的构建，二是某区域、某行业或某个应用场景算力需求及供给规模的评估，三是对算力投入与经济回报关系的评估，四是对改善算力资源利用效率方法的研究。

在上述研究活动中，算力评估指标构建是评价算力水平的基础，算力评估指标能够为算力评估和分析提供统一口径，在以往的算力研究中，研究人员更加关注超算及常规服务器算力的测试及评估，对数据中心的算力测试及评估研究则相对较少。超算算力评估主要是针对单机开展的，国际知名的超算榜单 TOP 500 根据超算执行 Linpack 程序的浮点运算能力对超算算力进行评价，常规服务器的算力评估大多是针对单机或芯片开展的，SPEC CPU 套件能够对 CPU 整点运算及浮点运算水平进行评估。

本章首先对目前发展和应用相对成熟的超算算力评估及常规服务器算力评估方法进行分析，在此基础上，对数据中心算力评估现状及其算力评估过程中存在的难点进行分析。

2.1. 超算算力评估

2.1.1. TOP 500

超级计算机拥有海量数据处理能力和高密度计算能力，是衡量一个国家先进计算水平的代表性设施。人们普遍用浮点运算速度（FLOPS）对超算性能进行评价。

TOP 500 排行榜主要以超算系统运行 LINPACK(线性系统软件包) 基准测试所能达到的最高性能 Rmax（单位：TFLOPS）对 500 个超算系统进行排名，该排行榜还提供超算地点、核数等基础信息。TOP 500 每年 6 月和 11 月更新一次，表 1 展示了 2020 年 11 月 TOP 500 排行榜中排名前六位的超算系统，其中我国神威·太湖之光排名第四，天河二号排名第六。

表 1 2020 年 11 月 TOP 500 排行榜超算系统前十³

排名	地点	超算系统	核数	Rmax ⁴ (TFLOPS)
1	理化学研究所（日本）	Supercomputer Fugaku	7630848	442010.0
2	橡树岭国家实验室（美国）	Summit	2414592	148600.0
3	劳伦斯利弗莫尔国家实验室（LLNL）（美国）	Sierra	1572480	94640.0
4	无锡国家超算中心（中国）	Sunway TaihuLight	10649600	93014.6
5	英伟达公司（美国）	Selene	555520	63460.0
6	广州国家超算中心（中国）	Tianhe-2A	4981760	61444.5

³ <https://www.top500.org/>

⁴ 数值为双精度浮点运算次数（FP64）

TOP 500 测试过程简单易操作，但是也有不少研究人员对 TOP 500 选取的测试标准产生质疑，TOP 500 并没有对超算通信子系统运算性能及其处理数据密集型应用的运算性能进行评估。针对 TOP 500 测评指标在超算性能评估方面的缺陷，有研究人员提出了新的测试标准，如 Graph 500。Graph 500 对超算每秒遍历边缘的数量进行了评估，其单位为 Gsteps（每秒 10 亿遍历边缘数）。Graph 500 榜单的发布表明了超算算力测评标准的多样化发展趋势，在不同的应用场景下，超算运算性能可能会有所差别。

2.1.2. Green 500

在很长一段时间内，从业人员普遍将超算运算速度视为评价超算性能的决定性指标，随着世界能源的日益紧张，超算的能耗问题也受到了广泛的关注。2007 年，Green 500 榜单发布，该榜单以用电效率（GFLOPS/W）为评估指标对 500 个超算进行排名。

表 2 展示了 2020 年 11 月发布的排行榜超算系统前十榜单，从该榜单中可以看出 Green 500 和 TOP 500 在超算系统排名上存在很大的差异性，一些在 TOP 500 上排名靠后的超算在 Green 500 榜单上排名靠前，如位于美国的超算 NVIDIA DGX SuperPOD 和日本的超算 MN-3 在 TOP 500 上排名分别为 170 和 330，但是在 Green 500 上排名靠前，排名分别为第一和第二。我国在 TOP 500 上排名靠前的超算并没有出现在 Green 500 榜单前十位置，这说明我国超算在用电效率方面

还存在很大的进步空间。

从 TOP 500 到 Green 500，超算算力评价指标逐渐从以运算速度为主转变为运算速度和用电效率兼顾，这充分说明世界各国在先进算力竞争中从一味追求运算速度到追求算力能效的理性转变。TOP 500 和 Green 500 提出的浮点运算能力及浮点运算用电效率指标为其他行业算力评估工作提供了重要的参考。

表 2 2020 年 11 月 Green 500 排行榜超算系统前十⁵

排名	TOP 500 排名	地点	超算系统	核数	Rmax ⁶ (TFLO PS)	Power (kW)	Power Efficiency ⁷ (GFLO PS/W)
1	170	NVIDIA Corporation (美国)	NVIDIA DGX SuperPOD	19840	2356.0	90	26.195
2	330	Preferred Networks (日本)	MN-3	1664	1652.9	65	26.039
3	7	Forschungszentrum Juelich (FZJ) (德国)	JUWELS Booster Module	449280	44120.0	1764	25.008
4	146	Atos (法国)	Spartan2	23040	2566.0	106	24.262
5	5	NVIDIA Corporation (美国)	Selene	555520	63460.0	2646	23.983
6	239	Fujitsu Numazu Plant (日本)	A64FX prototype	36864	1999.5	118	16.876
7	29	Rensselaer Polytechnic Institute Center for	AIMOS	13000	8339.0	512	16.285

⁵ <https://www.top500.org/lists/green500/2020/11/>

⁶ 数值为双精度浮点运算次数 (FP64)

⁷ 数值为双精度浮点运算次数 (FP64)

		Computational Innovations (CCI) (美国)					
8	8	Eni S.p.A. (意大利)	HPC5	669760	35450.0	2252	15.740
9	458	MIT /MGHPCC Holyoke, MA (美国)	Satori	23040	1464.0	94	15.574
10	1	RIKEN Center for Computational Science (日本)	Supercomputer Fugaku	7630848	442010.0	29899	15.418

2.2. 常规服务器算力评估

2.2.1. SPEC CPU

SPEC CPU 是由全球权威性能评估机构 SPEC (Standard Performance Evaluation Corporation) 推出的一款行业标准的 CPU 测试基准套件, 该套件最新版本为 SPEC CPU 2017, 主要通过 SPECspeed 2017 Integer、SPECspeed 2017 Floating Point、SPECrate 2017 Integer、SPECrate 2017 Floating Point 这四个套件的 43 个测试项目, 对 CPU 整点运算能力、浮点运算能力、整型并发速率和浮点并发速率进行测试。SPEC CPU 套件将会根据测试结果为 CPU 整数运算及浮点运算能力进行打分, 用户能够通过打分结果直观地看出不同 CPU 的性能差异, 并且根据业务需要选择合适的 CPU。

2.2.2. SPEC power

SPEC 早在 2006 年成立了 SPEC power 工作组, 工作组将能效基

准测试工具的开发作为主要工作目标。2007年，在美国环保总署和能源使用效率协会的赞助及 AMD、Dell、Intel 等厂商的参与下，SPEC power 委员会推出了一款里程碑式的能效基准测试套件——SPECpower_ssj2008，该基准致力于构建一个符合 IT 实际工作环境的性能/功耗评价基准，这将在很大程度上转变当前业界普遍存在的只重视系统性能而忽视系统能效的服务器性能评价观念。

由于 SPECpower_ssj2008 测试本身被限制为 JAVA 应用负载，而 JAVA 在桌面以及工作站领域缺乏相关的应用。SPEC power 委员会在 2013 年正式发布了 SERT (Server Efficiency Rating Tool, 服务器效率评级工具)。它由数十个被称为 worklet 的负载组件组成，并在运行时分别对服务器的 CPU、内存、存储组件进行测试。SERT 对 SPECpower_ssj2008 中的 SPEC PTDaemon 进行了复用。

2.2.3. MLPerf

MLPerf 起源于 2018 年，是业内首套测量机器学习软硬件性能的基准套件。该基准套件囊括了一组关键的机器学习训练和推理的工作负载，代表了重要的生产级别用例。对于训练 (Training)，涵盖了图像和自然语言处理，以及推荐系统和强化学习总共 7 个测试项目；对于推理 (Inference)，涵盖了图像、自然语言处理 2 种计算任务在 4 个应用场景下的测试项目。截止于 2020 年 4 月，MLPerf 已经发布了两轮训练 (Training) 测试结果以及一轮推理 (Inference) 测试结果。2020

年7月,MLPerf组织发布第三个版本MLPerf Training v0.7 基准测试。

ID	Submitter	System	Processor	#	Accelerator	#	Software	Benchmark results (minutes, smaller is better)		
								CosmoFlow	DeepCAM	Details Code
Available on-premises										
0.7-404	CSCS	piz_daint_gpu_n128_t2.2.0	Intel® Xeon® Processor E5-2690 v3 @2.60GHz	128	NVIDIA P100-PCIe-16GB	128	TensorFlow 2.2.0	461.01		details code
0.7-405	CSCS	piz_daint_gpu_n256_t2.2.0	Intel® Xeon® Processor E5-2690 v3 @2.60GHz	256	NVIDIA P100-PCIe-16GB	256	TensorFlow 2.2.0	327.01		details code
0.7-406	Fujitsu	ABCI PRIMERGY CX2570 M4	Intel® Xeon® Gold 6148 Processor @2.40 GHz	512	NVIDIA V100	1024	PyTorch 1.6.0		11.71	details code
0.7-407	Fujitsu	ABCI PRIMERGY CX2570 M4	Intel® Xeon® Gold 6148 Processor @2.40 GHz	256	NVIDIA V100	512	TensorFlow 2.2.0	34.42		details code
0.7-408	Fujitsu/RIKEN	flugaku_512xA64FX_tensorflow_closed	FUJITSU Processor A64FX	512	N/A	0	TensorFlow 2.2.0 + Mesh TensorFlow	268.77		details code
0.7-409	Fujitsu/RIKEN	flugaku_8192xA64FX_tensorflow_closed	FUJITSU Processor A64FX	8192	N/A	0	TensorFlow 2.2.0 + Mesh TensorFlow	101.49		details code
0.7-410	LBNL	congpu_n64_pt1.6.0	Intel® Xeon® Gold 6148 Processor @2.40 GHz	16	NVIDIA V100	64	PyTorch 1.6.0		139.29	details code
0.7-411	LBNL	congpu_n64_ft1.15.0	Intel® Xeon® Gold 6148 Processor @2.40 GHz	16	NVIDIA V100	64	TensorFlow 1.15.0	364.73		details code
0.7-412	LBNL	corikni_n512_ft1.15.2	Intel® Xeon® Phi™ Processor 7250 @1.40GHz	512	N/A	0	TensorFlow 1.15.2	536.06		details code
0.7-413	NCSA	hal_v100_n16_ft1.15.0	IBM POWER 9 model 2.2	32	NVIDIA V100	64	TensorFlow 1.15.0	265.59		details code
0.7-414	TACC	Frontiera-RTX	Intel® Xeon® Processor E5-2620 v4 @2.10GHz	32	NVIDIA Quadro RTX 5000	64	TensorFlow 1.15.2	602.23		details code

图 1 MLPerf Training v0.7 基准测试

2.2.4. 服务器能效规范

ODCC（开放数据中心委员会）于2019年发布了《服务器能效评测规范》，该测试规范将服务器能效定义为服务器计算性能与功耗的比值，并且将服务器综合能效视为电源模块效率、服务器空闲能效及服务器工作能效的加权平均数。在服务器工作能效测试过程中，该测试规范将服务器性能测试划分为CPU、内存及存储三个部分，利用Benchmark软件对服务器各部分性能及功耗值进行记录，在不同负载条件下，得到服务器工作状态功耗。服务器工作状态的效率（ Eff_{server} ）定义为如下公式计算：

$$Eff_{server} = \exp [W_{cpu} \times \ln (Eff_{cpu}) + W_{Memory} \times \ln (Eff_{Memory}) + W_{Storage} \times \ln (Eff_{Storage})] \quad (1)$$

其中： W_{cpu} 、 W_{Memory} 和 $W_{Storage}$ 分别是CPU、内存和存储的权重。

该规范根据服务器能效指标测评结果进行打分，按照分数高低划

分不同能效等级。

2.3. 电能利用效率（PUE）评估

PUE（Power Usage Effectiveness，电能利用效率）是 TGG（The Green Grid，绿色网格）发布的一项用于评价数据中心能效的指标，该指标已经得到业界的广泛认可。

PUE 在数值上等于数据中心总耗电与 IT 设备耗电的比值。在整个数据中心中，IT 设备是对外提供服务的主体设备，是产生算力的源泉。PUE 值越小则表明数据中心 IT 设备能耗占比越大，有更多电能被用于产生算力资源。

尽管数据中心能效与算力具有上述关联，但是这并不意味着提升数据中心整体能效水平就一定能够提升数据中心算力能效，数据中心算力能效除了与电能供给有关，还与 IT 设备的硬件性能、虚拟化技术的应用等因素有关，例如，对于不同性能的 CPU 芯片，消耗相同电能产生的算力可能并不相同。虚拟化技术的应用能够提升服务器资源的利用率，在相同的时间周期内，采用虚拟化技术的服务器往往能够处理更多任务，其实际算力水平自然能够提升，算力能效也会有所提升。

2.4. 小结

2.4.1. 算力构成复杂

数据中心算力评估与超算、常规服务器算力评估有很大不同，数据中心算力水平不仅取决于服务器的算力，同时还会在很大程度上受到存储及网络设备算力水平的影响，存储及网络设备算力水平无法满足要求而引发的“木桶效应”可能会拉低整个数据中心的实际算力水平。数据中心算力是服务器、存储及网络设备合力作用的结果，计算、存储及网络传输能力相互协同能够促使数据中心算力水平的提升。单独讨论服务器算力水平并不能反映数据中心的实际算力水平。

2.4.2. 算力测试难以开展

在数据中心算力水平测试方面，对于正在运行的数据中心，让其单独运行某一测试程序从而判断其算力水平不仅会带来大量的测试工作量，同时也会使数据中心承担一定的运行风险。考虑到数据中心服务器、存储及网络设备之间具有较强的联动性，单独对某一台或几台典型服务器算力水平进行测试也难以反映数据中心整体算力水平。在数据中心算力能效评估方面，数据中心算力水平研究的空白使得算力能效测试及评估难以开展。

2.4.3. 尚无合理完善的评估体系

在算力评估体系的构建方面，针对超算及常规服务器的算力评估

体系构建相对完善，在超算领域，TOP 500 和 Green 500 分别从浮点运算速度和浮点运算用电效率两个角度为超算算力评价提供了统一的业界标准，在常规服务器方面，SPEC CUP 对 CPU 整数运算、浮点运算性能进行评价，SPEC power 则构建了服务器能效测试基准。尽管现有的超算及常规服务器算力测试方法及评估指标可能并不完善，也无法满足每个用户的算力评价需求，但是从某个角度上来看，这些算力测试及评估指标为超算及常规服务器算力性能评价提供了一套相对受认可的行业标准，这将为超算及常规服务器的算力发展提供参照。

与超算及常规服务器相比，数据中心算力评估体系的研究则相对匮乏，目前还没有权威组织和机构提出可用的能够被业界广泛认可的算力评估指标及体系。尽管人们已经认识到数据中心算力是服务器、存储及网络设备合力作用的结果，且这三者相互影响，但是数据中心算力评估体系构建的难度仍不容忽视，在数据中心算力评估体系构建过程中依然有很多问题亟需解决。

总体来看，数据中心算力及算力能效评估较为困难，构建一套数据中心算力及算力能效评估指标体系将成为当前数据中心算力研究的重点。

3. 算力及算效指标

3.1. 算力的定义

数字化经济时代，万物万联产生各种各样的大数据，在数据已成为劳动、资本、土地和技术之外的第 5 个生产要素的背景之下，一个企业数字化的过程，就是数据逐渐成为企业核心要素和核心竞争力的过程。可以说，数字化转型离不开数据，数据离不开数据中心。

大数据中心建设成为未来国家战略中的一个重要衡量指标。从本质上说，数据中心为应对 5G、人工智能、工业互联网的大数据需求而生，构成了新基建的“基础”，为数据存储和应用提供平台支撑，是推进数字化转型的关键资源，是海量数据交换最为集中的关键节点所在。而数据中心所提供的最核心的能力，即算力，也称计算力，是数字化时代下的核心生产力。

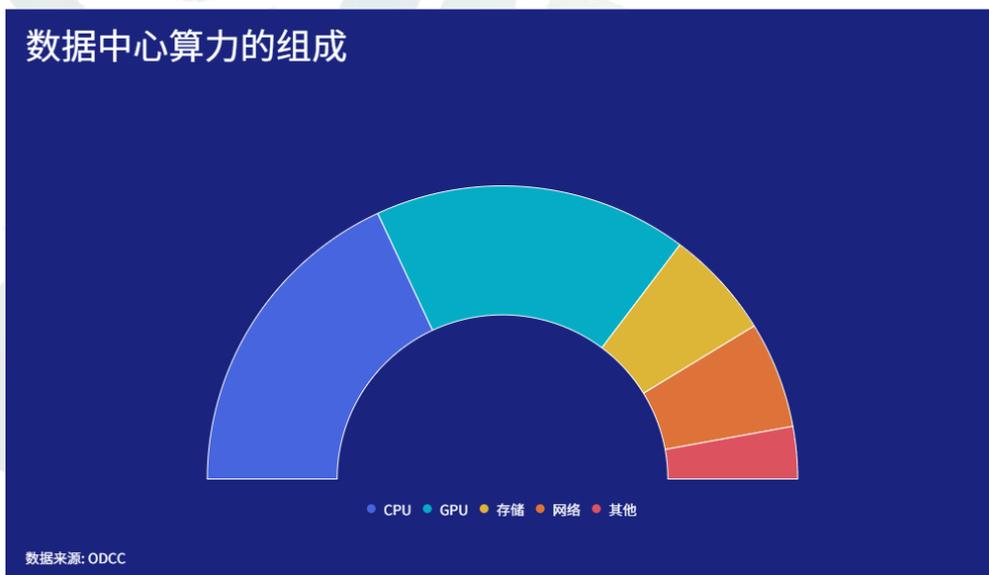


图 2 算力的组成

数据中心算力是数据中心的服务器通过对数据进行处理后实现结果输出的一种能力。在服务器主板上,数据传输的顺序依次为 CPU、内存、硬盘和网卡,若针对图形则需要 GPU。所以,从广义上讲,数据中心算力是一个包含计算、存储、传输(网络)等多个内涵的综合概念,是衡量数据中心计算能力的一个综合指标。

数据中心算力由数据处理能力、数据存储能力和数据流通能力三项指标决定。其中,数据处理能力,在应对以大数据、AI 人工智能为代表的新一代数字化技术产业趋势过程中,又可以区分为以 CPU 为代表的通用计算能力,和以 GPU 为代表的加速计算能力。前者主要用作执行一般任务,后者主要承担图形显示、大数据分析、信号处理、人工智能和物理模拟等计算密集型任务⁸。

综上,数据中心算力指标包含 4 大核心要素,即:通用计算能力、高性能计算能力、存储能力、网络能力。

3.1.1. 通用计算能力

CPU (Central Processing Unit, 中央处理器) 作为通用处理器,更偏重支持控制流数据。CPU 每个物理核中大部分的硬件资源被做成了控制电路和缓存,用来提高指令兼容性和效率,只有小部分是真正用来做计算的逻辑运算单元 (ALU)。在没有 AI 或其它高计算力要求时,CPU 可以应付得绰绰有余,在 AI 或高计算力要求时,CPU 在

⁸ John Nickolls and William J Dally. The gpu computing era. IEEE micro,30(2):56-69, 2010.

异构系统当中，扮演和发挥重要的指挥统筹，控制核心的功能。

CPU 的芯片又分为多种架构，主要包含 x86、ARM 等等。其中，x86 为主流架构，几乎占据全部市场份额，代表厂商为 Intel 和 AMD。

3.1.1.1.CPU（x86）

（1）简介

x86 架构 CPU 处理器是由 Intel 首先开发并主导的基于 x86 指令集的一系列处理器统称，属于传统通用型 CPU，适合逻辑控制、串行运算，并支持通用型负载。因其高性能，强扩展能力，良好的软件生态，占据着处理器市场很大一部分份额。

深度学习是从神经网络模型发展而来，随着 AI 技术的发展，往往通过深度学习作为切入点来分析各个芯片的性能。例如 Intel 的第二代至强可扩展处理器，集成了深度学习加速技术，不仅以优化的微架构、更多及更快的内核和内存通道带来了计算性能的提升，更面向 AI 应用提供了更为全面的硬件加速能力。

（2）案例

X86 架构 CPU 芯片计算能力与对应的核数、主频和对应的微架构计算能力息息相关，作为当今世界生产 CPU 的最大厂商 Intel 和 AMD，各自有生产的专门针对服务器的 CPU 芯片，Intel 主要是至强系列芯片，AMD 主要是霄龙系列芯片。Intel 和 AMD 的服务器所用的部分芯片的参数详见表 3。

表 3 X86 架构典型芯片处理器参数⁹

序号	处理器编号	厂家	发行年	内核数	处理器频率
1	E5-4648V3	Intel	2015	12	1.70GHz
2	E5-2660V4	Intel	2016	14	2.00 GHz
3	6140	Intel	2017	18	2.30 GHz
4	6240	Intel	2019	18	2.60GHz
5	AMD EPYC™ 7601	AMD	2017	32	2.20 GHz
6	AMD EPYC™ 7F72	AMD	2019	24	3.20 GHz

x86 架构 CPU 计算能力与对应的微架构和指令集息息相关，以 Intel 的 Haswell 微架构为例进行分析，Haswell 架构上计算单元有 2 个 FMA，每个 FMA 可以对 256bit 数据在一个时钟周期中做一次乘运算和一次加运算，所以对应单精度浮点计算能力为 32 FLOPS/cycle。

(3) 实测

SPEC CPU 2017 包括在测试整型速度、浮点运算速度的性能测试 SPEC speed 2017 Integer、SPEC speed 2017 Floating Point 以及测试整型并发速率和浮点并发速率 SPEC rate 2017 Integer 和 SPEC rate 2017 Floating Point。其中 SPEC speed 用于比较 CPU 完成单个任务的能力，而 SPEC rate 用于度量执行多个任务时 CPU 的运行速率。针对 SPEC CPU 2017 测试，选取了 4 大类测试的几个典型 CPU 测试结果，详见表 4。

⁹ 数据来源：intel、AMD 官网

表 4 SPEC 测试结果¹⁰

系统名称	测试类型	基本核	处理器			测试结果	
			启用核	启用芯片	线程/核心	基础	峰
XXXX 1288H V5 (Intel Xeon Gold 5218R, 2.10 GHz)	SPEC speed 2017 Floating Point	40	40	2	1	126	127
XXXX 1288H V5 (Intel Xeon Gold 5220R, 2.20 GHz)		48	48	2	1	138	140
XXXX 2288H V5 (Intel Xeon Gold 6252N)	SPEC rate 2017 Floating Point	96	48	2	2	273	—
XXXX 2488H V5 (Intel Xeon Gold 6244)		64	32	4	2	312	312
XXXX 1288H V5 (Intel Xeon Gold 5218R, 2.10 GHz)	SPEC speed 2017 Integer	40	40	2	1	11.3	11.6
XXXX 1288H V5 (Intel Xeon Gold 5220R, 2.20 GHz)		48	48	2	1	11.5	11.8
XXXX 2488H V5 (Intel Xeon Gold 6244)	SPEC rate 2017 Integer	64	32	4	2	287	296

¹⁰ 数据来源: <https://www.spec.org/cpu2017/results>

3.1.1.2.CPU（ARM）

（1）简介

ARM 架构 CPU，开始于低功耗、计算量小的场景，如智能手机、穿戴设备、IoT 等领域。随着 ARM 技术不断进步，多核性能大幅提高，尤其是开放的生态，ARM 也从端和边缘计算走向服务器和数据中心。当下，ARM 架构发挥在多核、低功耗等方面的优势，在面向大数据、分布式存储和 ARM 原生应用等场景，为企业构建高性能、低功耗的新计算平台，这也是计算发展的必然趋势。

由于 ARM 架构 CPU 采用 RISC 精简指令集，内核结构简单小巧，器件的功耗低，具有以下额外突出特点：同样功能性能下，芯片占用面积小、功耗低、集成度更高，具备更好的并发性能，这种多核高并发优势在大数据、存储等领域带来性能优势；支持 16 位、32 位、64 位多种指令集，能很好的兼容从 IOT、终端到云端的各类应用场景。

（2）案例

ARM 架构 CPU 计算能力与对应的微架构和指令集息息相关，以 ARM 的 Cortex-A57 微架构为例进行分析，Cortex-A57 架构上计算单元有 1 个 FMA，每个 FMA 可以对 128bit 数据在一个时钟周期中做一次乘运算和一次加运算，所以对应 32bit 单精度浮点计算能力为 8 FLOPS/cycle。

ARM 架构 CPU 芯片计算能力与对应的核数、主频和对应的微架构计算能力息息相关。因为 ARM 架构 CPU，开始于低功耗、计算量小的场景，如智能手机、穿戴设备、IoT、等领域，所以当前 ARM 架构的芯片主要还是用于移动设备，主要厂商有高通、三星、华为等。随着服务器的发展，华为公司聚焦于发展基于 ARM 架构的鲲鹏处理器的核心能力，通过战略性、长周期的研发投入，吸纳全球计算产业的优秀人才和先进技术，借助鲲鹏计算产业打造“新基建”底座。部分 ARM 架构的 CPU 芯片情况详见表 5。

表 5 ARM 架构典型芯片（高通、华为、飞腾）¹¹

序号	处理器编号	厂家	内核数	处理器频率
1	骁龙810	高通	8	2.00GHz
2	APQ8096SG	高通	4	2.34GHz
3	鲲鹏920	华为	64	2.60GHz
4	FT-2000/4	飞腾	4	3.00GHz

3.1.2. 高性能计算能力

近些年来，随着摩尔定律逐步趋于失效，硅芯片的物理性能已经逼近极限，经济成本也在逐步增加。通用处理器将无法满足人工智能等新兴技术对高性能计算能力的需求。在此背景下，GPU、FPGA、ASIC 和其它加速器（Accelerator）便得到了快速发展，所支持的高并

¹¹ 数据来源：<https://www.qualcomm.cn/>
<https://www.hisilicon.com/cn/>
<https://www.phytium.com.cn/>

行、高密度的异构加速计算将成为支撑未来 AI 等复杂应用的必然选择。

3.1.2.1.GPU

(1) 简介

GPU (Graphics Processing Unit, 图形处理器) 具有强大而高效的并行计算能力, 对于海量训练数据, 用 GPU 来训练深度神经网络, 所使用的训练集可以更大, 所耗费的时间能够大幅缩短, 占用的数据中心基础设施也更少。此外 GPU 还被广泛用于云端进行分类、预测和推理, 从而在耗费功率更低、占用基础设施更少的情况下能够支持远比从前更大的数据量和并发吞吐量。与单纯使用 CPU 的做法相比, GPU 具有数以千计的计算核心, 可实现 10-100 倍应用吞吐量。

相比于传统的 CPU 运算, GPU 并行运算有自身独有的特点。GPU 运算核心的数量要远远多于 CPU, 虽然 GPU 每个运算核心的工作频率不及 CPU, 但 GPU 计算单元的算力要远远强于 CPU, 且 GPU 的芯片面积比、功耗比等总体性能也比 CPU 好。在传统的 CPU 运算中, 线程之间的切换开销很大, 开启大量线程的算法效率比较低。但是 GPU 线程之间的切换相对廉价, 能够通过大量并行线程之间的交织运行隐藏全局的延迟, 同时还拥有大量的通用寄存器、局部存储器 and 高速缓冲存储器, 可以用来提升外部存储的访问性能。

(2) 案例

1) NVIDIA GPU

NVIDIA 是最早将 GPU 用于人工智能领域的公司，创造性的推出了 CUDA 通用并行计算架构。作为一项同时支持硬件和软件的技术，CUDA 利用图形处理器中的多颗计算核心进行通用计算处理工作，计算性能可获得显著提升。从软件生态上，CUDA 不仅支持自身的 API，也支持其他通用计算 API 比如 OpenCL/DX Compute 等。为适应近些年来 AI 的发展，CUDA 也增加了一系列的标准库函数来支持人工神经网络。但 CUDA 架构及生态只支持 NVIDIA 公司的 GPU 产品，不开源，所以在有国产化要求的 AI 应用中有一定的安全隐患。

2) AMD GPU

AMD 于今年推出了全新设计的 AMD CDNA GPU 架构，专为数据中心工作负载而优化，其采用面向高性能计算和人工智能的全新 Matrix Core 技术、搭载百亿亿级计算核心，可提供卓越的性能和能效。为 HPC 和 AI 打造全新 Matrix Core 技术，为单精度和混合精度矩阵运算（如 FP32、FP16、bFloat15、Int8 和 Int4）提供超强性能，促进 HPC 和 AI 的融合。CDNA 架构还包含 AMD Infinity 架构，为紧密连接的 GPU 系统提供了一个先进的运行平台，让工作负载可以快速高效地共享数据。

AMD 也正式发布了首款基于全新 CDNA 架构的 Instinct MI100 加速显卡，率先超越 10TF FP64 大关的数据中心 GPU，为行业提供领先的 11.5 TFLOPS（FP64）峰值性能、23.1 TFLOPS（FP32）理论

峰值性能。

3.1.2.2.FPGA

(1) 简介

FPGA (Field Programmable Gate Array, 现场可编程逻辑门阵列) 是一种高性能、低功耗的可编程芯片, 可以根据客户定制来做针对性的算法设计。FPGA 不采用指令和软件, 是软硬件合一的器件。对 FPGA 进行编程要使用硬件描述语言, 硬件描述语言描述的逻辑可以直接被编译为晶体管电路的组合。所以, FPGA 直接用晶体管电路实现用户的算法, 而没有通过指令系统的翻译。

FPGA 灵活性介于 CPU、GPU 等处理器和专用集成电路 ASIC 之间, 在硬件固定的前提下, 允许使用者灵活使用软件进行编程。它的开发周期比 ASIC 短, 不过相对于 ASIC 的批量出货, 单个 FPGA 的成本更高一些。

相对于 CPU、GPU 等处理器, FPGA 具有以下优势和特点:

1) 高度可定制。开发者可以使用 Verilog 或 VHDL 硬件描述语言对 FPGA 的电路逻辑和功能进行灵活开发。由于其高度可定制, 以前 FPGA 常被用作 ASIC 芯片流片前的硬件验证方法。

2) 低功耗与高性价比。低功耗是 FPGA 最引人注目的优势。FPGA 计算的绝对性能并不如 GPU, 但是企业级 GPU 的功率通常高达数百瓦, 而 FPGA 的功率可以低至 20 瓦以下。低功耗带来的直接收益是

省电，而在数据中心的场景中，低功耗的意义会更大。

3) 计算并行性与流水线控制。FPGA 设计可以做到全定制，设计者可以精确控制 FPGA 内部所有的逻辑单元、寄存器和数字运算单元在任何一个时钟周期的行为。因此，FPGA 上所有的运算单元都可以同时工作，在计算时具有天然的并行性。除此之外，开发者还可以精确地控制流水线，最大程度地对计算单元加以利用。FPGA 在计算方面具有的并行性，使其适用于以深度学习为代表的各种并行计算的业务场景。

ASIC 芯片与 FPGA 相比较，单纯的 ASIC 芯片无法完全替代 FPGA，也不具备 FPGA 的灵活性。在传统领域，ASIC 无法完全替代 FPGA，在 AI 领域，FPGA 与 ASIC 会分而治之。而在智慧工业机器人、智慧通信、智慧医疗、智慧装备等领域，除 FPGA 以外其他 AI 芯片都难以介入。

(2) 举例

全球最大的 FPGA 芯片生产厂商赛灵思和 Intel，都确立了相似的战略布局，将重心放在了数据中心市场，并将更多可编程应用逻辑元件集成在芯片上，满足 AI 等应用对强大计算能力的需求。

随着大数据发展和算力的提升，越来越多的企业正在最大程度挖掘 FPGA 使用范围，充分发挥 FPGA 的优势。微软将 FPGA 作为核心 AI 计算平台，从使用装满 FPGA 的专用服务器集群，到用专用网络连接 FPGA 加速卡集群，再到将共享数据中心网络的大规模

FPGA 云服务，都在寻找将 FPGA 优势发挥的最优路径。

Intel 针对 FPGA 开发了一种经济高效的可重复编程平台，具有定制性能、定制功耗、高吞吐量和低批量延迟的功能，满足各种业务规格要求。并且可提供高度精细的片上带宽，提升内存密集型工作负载的性能，加快从网络边缘到数据中心的应用。

3.1.2.3.ASIC

(1) 简介

ASIC (Application Specific Integrated Circuit, 特殊应用专用集成电路) 是为了某种特定的需求而专门定制的芯片。ASIC 芯片的计算能力和计算效率都可以根据算法需要进行定制，所以 ASIC 与通用芯片相比，具有以下几个方面的优越性：体积小、能耗低、计算性能高、计算效率高、芯片出货量越大成本越低。但是缺点也很明显：算法是固定的，一旦算法变化就可能无法使用。按照定制程度来看，ASIC 芯片可分为全定制 ASIC 芯片、半定制 ASIC 芯片及可编程 ASIC 芯片三类。

从目前众多行业巨头的布局来看，都在积极的布局定制化芯片，特别是在 AI 芯片布局方面，已经成为了众多行业巨头的重心，因为包括深度学习、机器学习、大数据分析及判断、自动决策等各种 AI 应用如雨后春笋般出现，而针对不同应用打造的特殊应用芯片 ASIC 的需求也自然呈现爆发态势。

(2) 案例

TPU (Tensor Processing Unit, 张量处理器) 是 Google 为机器学习定制的专用芯片 (ASIC), 专为 Google 的深度学习框架 TensorFlow 而设计。TPU 在算法架构上介于 CPU 和全定制化 ASIC 之间, 兼具桌面计算设备与嵌入式计算设备功能。TPU 算法具备较宽容错性, 在硬件组成上相对 CPU 类通用芯片更加简洁。相同数量晶体管条件下, TPU 算法架构的 ASIC 芯片可完成更高运算量。相对同级别 CPU、GPU, 该类 ASIC 芯片可提高运算性能 15 倍至 30 倍, 并提高能耗效率 30 倍至 80 倍。Google 针对 AI 很早就设计了自己的 TPU 芯片, 之前的 Alpha Go 就有采用。此外, Google 自己的手机 Pixel 2 也首次采用了自研芯片 Pixel Visual Core 以提升拍照性能。

半导体芯片设计公司 eASIC 解决了 ASIC 设计复杂、费用高昂等障碍, 降低了 ASIC 的进入门槛, 同时解决了 FPGA 静态泄漏和高功耗。凭借其高性能、低功耗、低费用和快速上市时间等特点, 未来有望在快速增长的定制化芯片市场上占据一席之地。

3.1.2.4. NPU 等 (AI 芯片)

随着人工智能新兴产业的高速发展, 芯片技术是人工智能产业发展的核心因素。人工智能芯片的研制, 会将传统芯片技术和人工智能技术加持的优势结合, 将科技发展推向一个更高的阶层。面向 AI 应用的芯片, 除了上面介绍的 GPU、FPGA、ASIC 之外, 还有其他架构

的类脑芯片。当前技术条件下的 AI 芯片主要是针对神经网络的训练与推理所研发的加速器，其底层架构是具备乘加功能的计算单元。人工智能芯片目前还处于起步阶段，它的发展还有很长的一段路要走，各国都在积极摸索研究 AI 芯片的发展方向，在我国也涌现出了一批 AI 芯片企业。

NPU（Neural-network Processing Unit，网络处理器）采用数据驱动并行计算的架构，用于加速神经网络的运算，解决传统芯片在神经网络运算时效率低下的问题。在实际应用中，特别擅长处理视频、图像类的海量多媒体数据，是一款专门为人工智能而设计的处理器。寒武纪作为国内专注 AI 芯片研发的年轻公司，发布了多款 NPU。2016 年发布的寒武纪 1A 处理器具有完全自主知识产权，可用于计算机视觉、语音识别、自然语言处理等关键人工智能任务，具备出色的通用性和效能化。

华为一直以来致力于昇腾系列 AI 芯片的研发。2018 年华为正式发布了昇腾 910 和昇腾 310 两款 AI 芯片，其中昇腾 910 是目前单芯片计算密度最大的 AI 芯片，适用于 AI 训练。而昇腾 310 是目前面向边缘计算场景最强算力的 AI SoC，适用于 AI 推理，最大功耗仅 8W，是一款高效计算低功耗 AI 芯片，非常适合边缘计算的低功耗要求场景。

燧原科技在 2019 年发布了首款面向云端数据中心的 AI 训练芯片“邃思”和计算编程平台“驭算”。邃思芯片主要面向数据中心，驭算

平台针对邃思芯片进行深度优化，支持主流深度学习框架，包括常用的算子数据库和高效灵活的调度机制，可提供视觉、自动驾驶、大数据、高性能计算等多领域的解决方案。

科大讯飞在 2019 年发布了家电行业专用语音 AI 芯片 CSK400X 系列，芯片集成了讯飞的语音 AI 算法，可以有效通过深度神经网络算法解决家居中的噪音问题。

3.1.3. 存储能力

目前数据中心的特点是数据量爆炸性增长，数据总量呈指数上升，确保数据存得下，传得快、无篡改，是存储关心的主要课题。数据存储能力由存储容量、存储性能、存储安全三方面共同决定。

存储容量：数据中心存储系统不仅要有大量的现实容量，还应该具有很好的可扩展性，能根据数据量的增长提供无缝的、不停机的容量扩充。

存储性能：数据是具有时效性的，及时获得所需数据非常关键；对于 ICP 而言，较高的访问速度是服务质量的重要指标。对于宽带应用，存储系统的带宽要与网络带宽相适应。因此，存储系统的响应速度和吞吐率是 IDC 数据中心存储系统应该密切关注的问题。从历史上看，计算机速度的瓶颈已逐渐从 20 世纪 80 年代的 CPU 和 90 年代的网络带宽转移到 IO 子系统。因此，要提高 IDC 数据中心存储系统的整体性能，存储系统的性能提高是一个关键问题。

存储安全：数据中心存储系统存储了企业大量的关键数据，必须保证这些数据始终是安全可用的。在任何情况下，例如系统产生错误或遇到意外灾难，数据都不能丢失。系统应具有快速故障恢复能力，保证应用系统永不停机（7×24 小时不间断工作），数据始终保持完整性和一致性。

3.1.4. 网络能力

3.1.4.1. 网络作用

AI时代已经来临，随着人工智能应用数据的急速增加，AI算法规模不断增加，AI模型的训练任务提出海量规模的计算量需求。根据OpenAI¹²的分析数据，自2012年起，最新的模型训练所需的计算量每3、4个月翻一倍。模型的训练时间也随之延长，严重影响AI模型部署的迭代周期。为了满足AI业务市场需求，缩短模型训练时长，对AI算力提出越来越高的要求。

为了满足 AI 时代模型训练的算力需求，AI 加速硬件 GPU 的算力持续着近乎摩尔定律的极限增长速度。然而，GPU 算力的增长速度仍然远低于 AI 应用对计算量的需求的增速。因此，分布式 AI 集群系统成为了满足 AI 应用算力需求的首选。

12 Is Network the Bottleneck of Distributed Training, Johns Hopkins University & AWS

InfiniBand（简称 IB）和 RoCE 是目前 AI 领域网络技术的重要网络支撑技术。IB 在实际的大规模集群中使用了动态路由的网络，在美国最快的前两大超级计算机上都用到了动态路由技术，根据美国橡树岭国家实验室、Lawrence Livermore 国家实验室和 IBM 公司公布的在 Summit 上的测试结果¹³，在静态路由的时候，9,200 个通信节点的平均通信效率只有 50%到 60%，利用动态路由，他们达到了平均 96%的 MPIGraph¹⁴ All2All 网络通信效率，是目前全球最高的网络通信效率。

通过对超算网络的流量模型进行分析，可以得到超算网络的典型特性：80%以上均为小于 256 字节¹⁵Payload 的报文，极端考验设备的静态时延。典型的以太网芯片静态时延约 500ns，IB 芯片时延约 90ns。从静态时延角度看，以太相比 IB 存在劣势，这种劣势通过网络与计算的协同可以改善甚至消除。

传统集合通信过程中的计算在服务器侧完成，网络只负责转发；网算一体技术采用交换设备对集合通信加速，卸载一部分集合通信的计算过程，可以有效地提高集合通信的效率，从而降低总的任务完成

13 《The Design, Deployment, and Evaluation of the CORAL Pre-Exascale Systems》，Sudharshan S. Vazhkudai†, Bronis R. de Supinski‡, Arthur S. Bland†, Al Geist†, James Sexton*, Jim Kahle*, Christopher J. Zimmer†, Scott Atchley†, Sarp Oral†, Don E. Maxwell†, Veronica G. Vergara Larrea†, Adam Bertsch‡, Robin Goldstone‡, Wayne Joubert†, Chris Chambreau‡, David Appelhans*, Robert Blackmor etc. SC18, November 11-16, 2018, Dallas, Texas, USA

14 MPIGraph All2All 是用来测试网络中的任意两个 MPI 进程直接的并发通信效率的工具

15 《Characterization of MPI Usage on a Production Supercomputer》

时间。ODCC 针对基于以太的网算一体交换机开展了性能测试。其采用 8 台 TAISHAN 服务器（鲲鹏 920 芯片）以及 NVIDIA 的 100GE 网卡，测试软件采用了业界主流的 OSU。测试结果如下图所示，相对于传统 RoCE 网络，网算一体技术大幅度降低了 HPC 计算的任务完成时间，平均降幅超 20%。

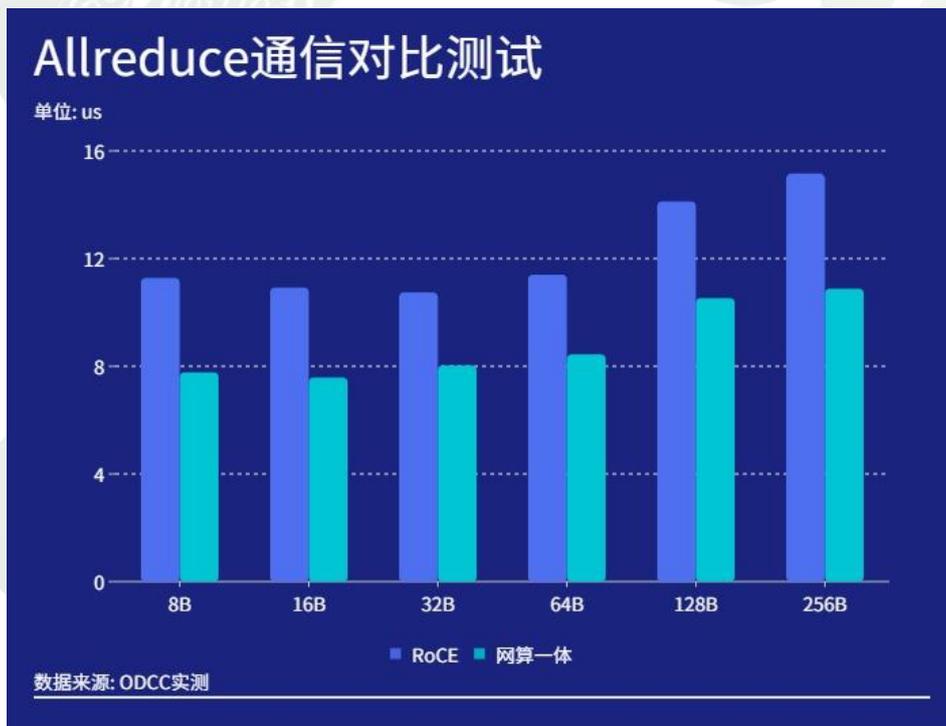


图 3 八节点 768 核 allreduce 通信 ODCC 对比测试

3.1.4.2. 网络性能参数

数据中心中，网络起着承上启下的作用，将计算和存储资源连接在一起，并以服务的形式对内部及外部的提供数据访问能力。

带宽：带宽越高意味着数据中心可以具有更强的处理能力，可以

完成更多的业务应用。它取决于数据中心网络整体的性能，除高带宽网络设备外，还需根据数据中心网络的实际流量需求，划分内部各应用服务器之间交互及外部交互分别需要的带宽，构建适宜的网络结构，达到最优的网络带宽。

延迟：网络延迟也是体现数据中心网络性能的重要参数，指报文在传输介质中传输所用的时间，即从报文开始进入网络到它开始离开网络之间的时间，单位为毫秒，对于一些对延迟非常敏感的关键业务延迟的高低具有决定性的作用。由于网络的复杂性、网络流量的动态变化和网络路由的动态选择，网络延迟随时都在不停的变化。网络延迟和网络延迟的抖动越小，那么网络的性能就越好。

丢包率：数据在网络中是以数据包为单位传输的，数据包的传输存在一定的损失，网络将根据协议来补包和重传该包。但是，如果网络线路较差，数据的损失量就会非常大，补包工作也不能百分之百完，此时数据的传输就会出现空洞，造成丢包。丢包率是数据包丢失部分与所传数据包总数的比值。丢包率越低，网络性能越好。

3.2. 数据中心算力模型（CP）

3.2.1. 方法

目前数据中心内部的服务器芯片类型以 CPU 和 GPU 这两个类型为主。前者主要用作执行一般任务，后者主要承担图形显示、大数

据分析、信号处理、人工智能和物理模拟等计算密集型任务¹⁶。

在计算机系统的发展过程中，曾经提出过多种方法表示计算能力，目前为止使用最广泛的是“浮点运算次数表示法”。FLOPS（floating-point operations per second）为每秒执行的浮点运算次数，是对计算机性能的一种衡量方式。FLOPS 的概念最早由 Frank H. McMahon¹⁷在其报告中提出。国内外不少文献以及服务器产品参数都采用浮点运算次数对算力进行描述，例如，Yifan Sun¹⁸等使用 FLOPS 作为度量标准，以评估 CPU 和 GPU 的单精度和双精度计算能力。

“浮点运算次数表示法”利用科学计数法来表达，包含三种常见类型：

- 1) 双精度浮点数（FP64）：采用 64 位二进制来表达一个数字，常用于处理的数字范围大而且需要精确计算的科学计算；
- 2) 单精度浮点数（FP32）：采用 32 位二进制来表达一个数字，常用于多媒体和图形处理计算；
- 3) 半精度浮点数（FP16）：采用 16 位二进制来表达一个数字，NVIDIA 从 2002 年 CUDA7.5 开始，支持 16 位浮点数的存储和计算，有效降低数据传输和存储成本，适合在深度学习中应用。

¹⁶ John Nickolls and William J Dally. The gpu computing era. IEEE micro,30(2):56–69, 2010.

¹⁷ McMahon, F H. The Livermore Fortran Kernels: A computer test of the numerical performance range. United States: N. p., 1986. Web.

¹⁸ Yifan Sun et, al. Summarizing CPU and GPU Design Trends with Product Data.2019

本白皮书使用“每秒浮点运算次数”(Floating-point operations per second, FLOPS)来评估数据中心的通用算力和高性能算力。同时,与 Linpack 仅关心双精度的浮点计算(FP64)能力不同,我们将给出双精度(FP64)和单精度(FP32)浮点计算能力算法,以便更加清晰地辅助判断数据中心适合的计算场景:

- 1) 用双精度浮点计算能力(FP64)评估数据中心的高性能计算能力;
- 2) 用单精度浮点计算能力(FP32)评估数据中心的通用计算能力。

除了 FP64 和 FP32 之外,其他的计算精度也越来越广泛被用于计算领域。对于人工智能来说,FP16 大有后来居上的趋势。主流的 AI 芯片和 AI 软件都已经支持 FP16 精度用于深度学习训练。同时 INT8 也越来越多用于深度学习推理领域。在本白皮书中,我们目前仅采用 FP64 和 FP32 两种精度衡量数据中心算力和算效,未来考虑加入更多的精度已更加全面地衡量数据中心的算力。

3.2.2. 模型

数据中心算力(CP, Computational Power)的模型如下:

$$CP = f(\text{通用算力, 高性能算力, 存储能力, 网络能力}) \quad (2)$$

其中,

- 1) 通用算力 $=\sum(\text{某型号 CPU 服务器存数} \times \text{该型号服务器 CPU 算力})$ (3)

以主流 CPU 型号为例，算力如表 6 所示：

表 6 主流 CPU 服务器算力¹⁹

序号	型号	FP64
1	Intel® Xeon® Processor E7 Family	921.6 GFLOPS
2	Intel® Xeon® Processor E5 Family	774.4 GFLOPS
3	Intel® Xeon® Scalable Processors	1.6 TFLOPS
4	AMD EPYC™ 7002 Series Processors	2.6 TFLOPS

2) 高性能算力= \sum (某型号 GPU 服务器存数×该型号服务器 GPU 算力) (4)

以 NVIDIA 主流 GPU 型号为例，算力表 7 所示：

表 7 NVIDIA 主流 GPU 型号算力²⁰

序号	型号	FP32
1	NVIDIA Tesla K40	5 TFLOPS
2	NVIDIA Tesla M40	6.8 TFLOPS
3	NVIDIA Tesla P100(PCIe)	9.3 TFLOPS
4	NVIDIA Tesla V100(NVLink)	15.7 TFLOPS
5	NVIDIA Tesla V100(PCIe)	14 TFLOPS
6	NVIDIA Tesla A100(NVLink)	19.5 TFLOPS

3) 存储能力

SSD 硬盘由于使用了高速的闪存颗粒做为物理存储资源，并且使用 PCIe 等高速传输协议/接口做为主流数据交换的物理通道，其在 IOPS 和带宽方面远优于传统的 HDD 硬盘。

¹⁹ 根据官网参数计算，取每系列最高值

²⁰ 数据来源：NVIDIA 官网

以企业级 PCIe SSD 卡和企业级 SAS HDD 硬盘来比较，PCIe SSD 卡的 4K 随机读的 IOPS 为 1M 以上，而 SAS HDD 硬盘的 IOPS 为 700 左右；带宽方面，PCIe SSD 可达到 7000MB/s 以上，而 SAS HDD 仅为 200MB/s 左右。从测试数据可以看到，不管是 IOPS 还是吞吐量，PCIe SSD 的测试结果均为 SAS HDD 的 1000 倍以上。

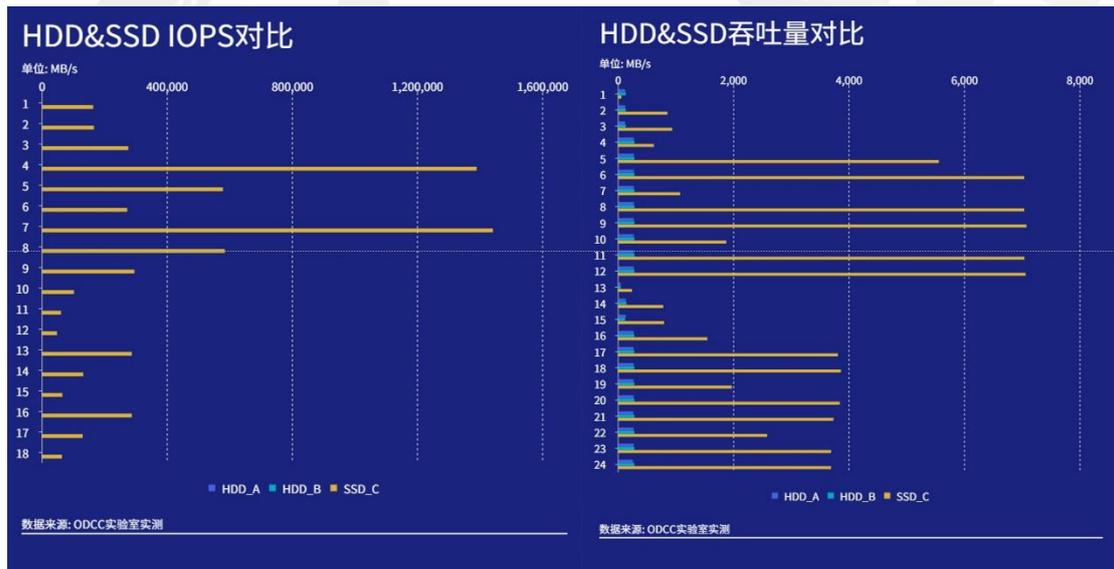


图 4 HDD&SSD 性能 ODCC 实测数据对比

存储对算力的贡献，一方面体现在高速存储对高性能计算的支撑，另一方面，体现在对海量数据的存储。该部分模型将在本白皮书的后续更新中体现。

4) 网络能力

随着 AI 训练集群规模的增大，以及单节点算力的增长，分布式 AI 集群系统已经逐渐从计算约束转换为网络通信约束。一方面，在过去 5 年，GPU 算力增长了近 90 倍，而网络带宽仅增长了 10 倍。另一方

面，当前的 AI 集群系统中，当 GPU 集群达到一定规模以后，随着计算节点数的增加，由于分布式 AI 集群节点之间的通信代价的增加，可能导致集群每秒训练的图片数量不增反减。

当前存储介质 SSD 的访问性能相比传统分布式存储 HDD 已提升了 100 倍，对于采用 NVMe 接口协议的 SSD（简称 NVM 介质）时，访问性能相比 HDD 甚至可以提升 10000 倍。在存储介质的时延已大幅降低的情况下，网络的时延占比已从原来的小于 5% 变化到 65% 左右，也就是说，宝贵的存储介质有一半以上的时间是空闲通信等待。如何降低网络时延成为提升 IOPS（Input/Output Operations Per Second，每秒读写次数）的核心。

网络将成为数据中心计算、存储能力能否充分发挥的重要支撑。该部分模型将在本白皮书的后续更新中体现。

3.3. 数据中心算效模型（CE）

受摩尔定律的影响，CPU 的算力提升方法通常有两种，一是增加“数量”，即增加核心的数量；二是提高“质量”，即提高单核心的运算效率，即提高主频。但主频的提高并不是无限制的，会受到功耗的制约。所以，数据中心算力功耗也是一个非常重要的方面。

从各类数据来看，数据中心无疑是“能耗大户”。从微观上看，一个超大型数据中心每年的耗电量超过亿度；从宏观上看，全世界数据中心的耗电规模为 205TWh，占全世界耗电规模的 1%。但脱离了算

力而单纯的指责数据中心耗电多而制约其发展是片面的，我们要将算力与功耗结合来看，单位功耗的算力是评价数据中心计算效果更为准确的一个指标。

我们定义**数据中心算效**（**CE, Computational Efficiency**）为**数据中心算力与所有 IT 设备功耗的比值**，即“**数据中心 IT 设备每瓦功耗所产生的算力**”（单位：**FLOPS/W**），这是同时考虑数据中心计算性能与功耗的一种效率。

公式如下：

$$CE = \frac{CP}{\sum IT \text{设备功率}} \quad (5)$$

提升算效，不仅能够降低服务器本身的运行成本，同时还可以降低数据中心制冷设备的运行成本，从而降低数据中心整体供电负担，降低整体的能耗。

4. 数据中心情况分析

4.1. 我国机架总体情况

数据中心总体算力水平与数据中心机架规模密切相关，数据中心机架上承载着各类服务器、存储设备及网络设备，这些设备共同构成了数据中心的算力基础。在保证上架率的情况下，数据中心机架规模越大表明数据中心能够提供的算力越高，算力资源供给更为充足。在数据中心算力能效测算方面，数据中心 IT 设备能耗之和可以用机架功耗之和近似替代，在得到 IT 设备能耗后可以进一步利用总算力与

能耗的比值求出数据中心总体算力能效。



图 5 近五年我国数据中心机架规模及预测²¹

图 5 反映了近几年全国数据中心在用机架规模和大型规模以上机架的变化情况，从图中可以看出，从 2016 年到 2019 年，我国数据中心在用机架规模以 30%左右增速逐年增长。机架规模的增长充分表明我国企业及用户对数据中心算力的潜在需求较为旺盛，与此同时，这种高速增长的算力需求进一步推动了我国机架规模的增长，逐年增加的算力资源将为云计算、人工智能、物联网等应用服务的开展提供重要保障。

²¹ 数据来源：中国信息通信研究院&开放数据中心委员会，《数据中心白皮书（2020）》

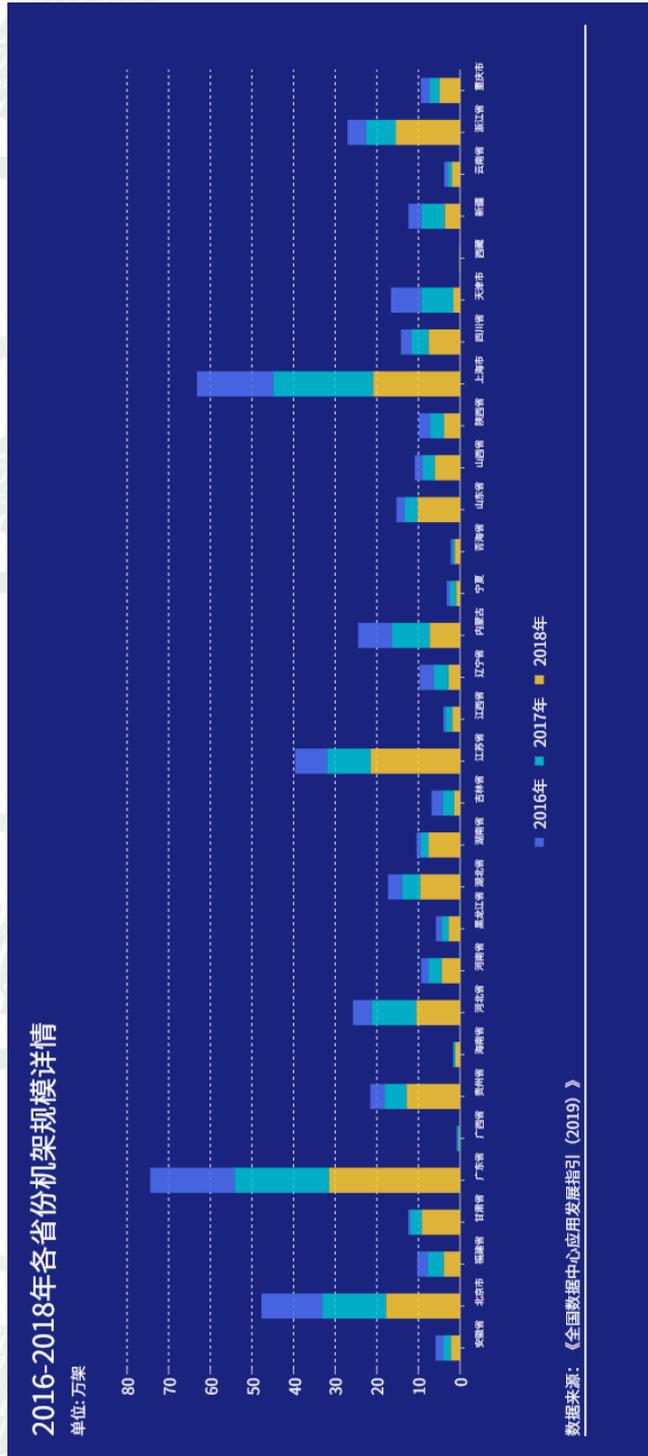


图 6 2016-2018 年各省份机架规模详情（单位，万架）²²

²² 数据来源:《全国数据中心应用发展指引(2019)》

从机架区域分布来看，我国各地区数据中心机架分布存在着很大的不均衡性。图 6 反映了 2016 年至 2018 年我国各省份数据中心机架规模变化，从图中可以看出我国东部省份数据中心机架数较多，中西部省份数据中心机架数较少，由此可以进一步推断出我国东部地区算力充足，中西部地区算力相对贫乏。近年来，我国政府对数据中心能效管控日益严格，降低数据中心 PUE 水平逐渐成为行业共识，贵州、内蒙古等偏远地区在电力、土地价格及自然冷源方面较东部地区具有明显优势，不少数据中心运营商开始将贵州、内蒙古等西部地区作为数据中心建设选址的优选地区。

我国数据中心机架规模区域分布不均衡的根本原因在于我国不同地区用户对算力需求存在较大差异，东部地区用户对各类网络应用需求较大，为了向用户提供更为高效、及时的算力支持，不少数据中心选择在东部地区选址建设，数据中心机架规模增长较快。中西部地区尽管存在一定的资源优势，但是本地用户对算力需求相对较少，另外，数据在远距离传输过程中的网络延时可能会对服务质量产生一定的影响，这使得业务导向型数据中心更加愿意在东部算力需求旺盛的地区选址，中西部地区数据中心数量难以进一步增长，机架规模增长缓慢，中西部地区算力供给也无法得到有效保障。

4.2. 我国上架率情况

数据中心总体算力水平及算力能效不仅与机架总体规模有关，同

时还会受到数据中心在用上架率的影响，数据中心在用上架率能够反映当前数据中心的实际算力水平及算力能耗，在评价社会总体或某地区数据中心实际在用算力时应充分考虑到机架规模及上架率。

截至 2018 年年底，国内数据中心总体平均上架率为 47.9%²³。部分省市的上架率已超过 50%，其中，上海市上架率为 78.9%、北京市上架率为 69.9%、河北省上架率为 72.3%、广东省上架率为 53.6%、浙江省上架率 52.4%。全国超大型数据中心的上架率为 28.2%，大型数据中心的上架率为 54.3%，中小型数据中心的上架率为 59.8%。总体来看，东部发达地区及一些自然资源较为充足的中西部省份上架率相对较高，东部发达地区对时效性较高的“热数据”需求较多，提升上架率有助于进一步满足这种实时的算力需求。中西部等自然资源较为充足的地区在建设能效导向型数据中心方面具有一定优势，也逐渐受到资本加持，一些实效性要求不高的“冷数据”通常可以在这些地区进行远端部署，这将提升数据需求旺盛地区的算力资源利用率。

4.3. 服务器出货量

4.3.1. CPU 服务器

根据 Gartner 的数据来看，中国（不包括港澳台地区）的 CPU 架构服务器出货量在 2015-2019 年基本呈现上升趋势，5 年复合增长率近 8%，2019 年的出货量为 340 万台左右，其中 x86 架构在 CPU 市

²³ 数据来源：《全国数据中心应用发展指引（2019）》

场的占比都在 99%以上。

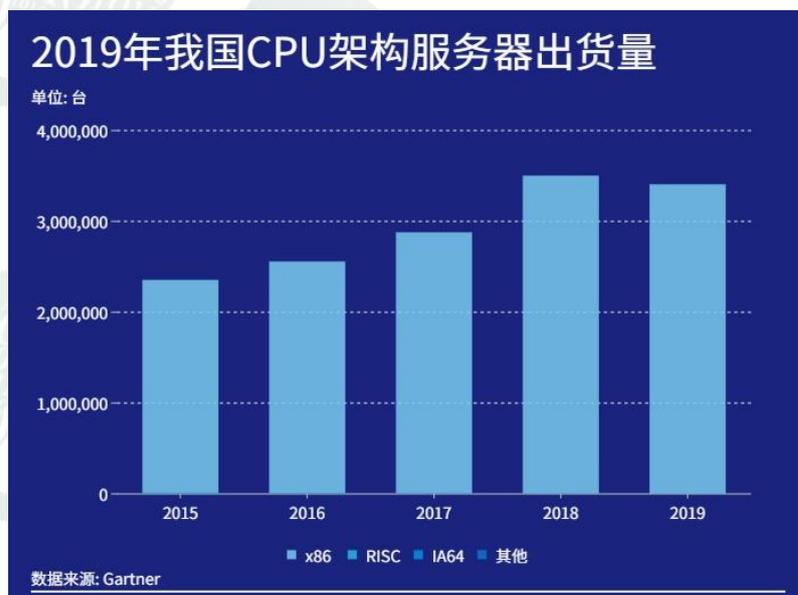
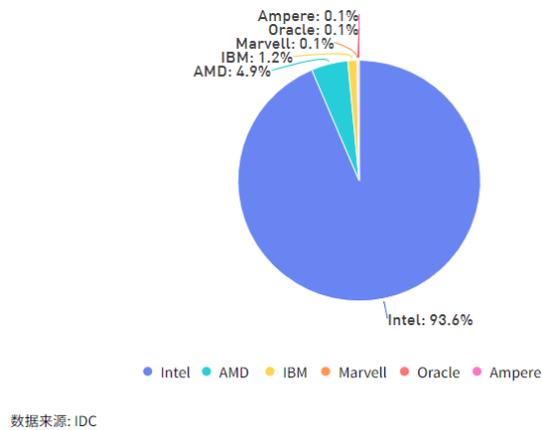


图 7 2015-2019 年我国 CPU 架构服务器出货量²⁴

在厂商市场份额方面，Intel 市场基础庞大，在 CPU 市场市占率基本维持近 95%左右的水平。以 2019 年 Q4 为例，根据 IDC 数据显示，Intel 在全球数据中心 CPU 微处理器市场份额的占比为 93.6%，其次为 AMD 为 4.9%。

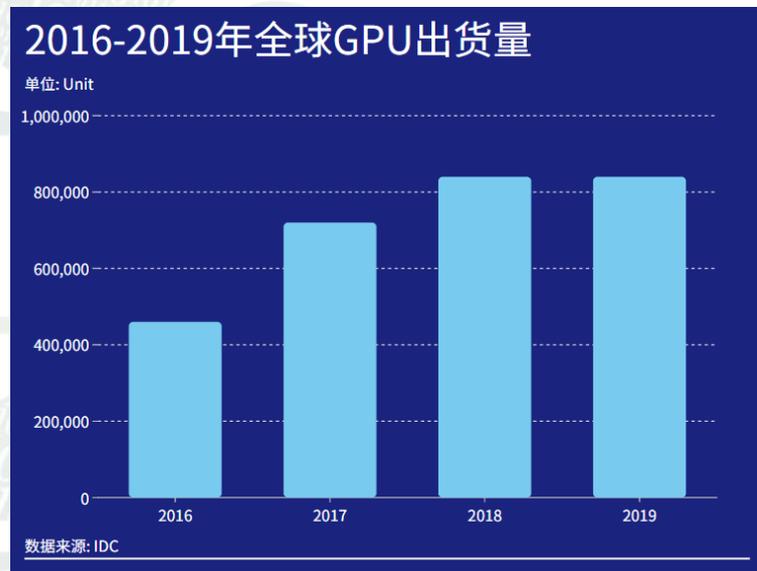
²⁴ 数据来源: Gartner

2019年Q4全球数据中心CPU微处理器市场份额

图 8 2019 年 Q4 全球数据中心 CPU 微处理器市场份额²⁵

4.3.2. GPU 服务器

根据 IDC 的数据显示，全球 GPU 的出货量呈现上升趋势，在 2019 年达到了 840 万 Unit。

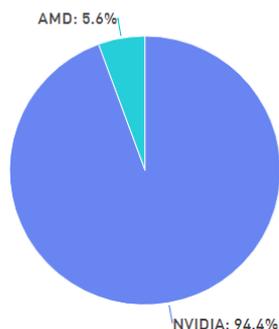
图 9 2016-2019 年全球 GPU 出货量（单位：Unit）²⁶

²⁵ 数据来源: IDC 《Datacenter Processing, 4Q19: Server CPU, GPU, FPGA, AI ASICs, and ASSPs》

²⁶ 数据来源: IDC 《Datacenter Processing, 4Q19: Server CPU, GPU, FPGA, AI ASICs, and ASSPs》

市场份额方面，NVIDIA 为行业龙头企业。以 2019 年 Q4 为例，根据 IDC 数据显示，NVIDIA 在全球数据中心 GPU 服务器市场份额的占比为 94.4%，其余为 AMD，占比 5.6%。

2019年Q4全球数据中心GPU服务器市场份额



数据来源: IDC

图 10 2019 年 Q4 全球数据中心 GPU 服务器市场份额²⁷

4.4. 数据中心能效情况

数据中心在对外提供应用服务的过程中会产生大量的能耗，在数据中心发展早期，数据中心数量较少，规模较小，其能耗问题还没有受到过多关注。21 世纪以来，互联网用户规模的增长及各类网络应用的出现使得网络数据量不断增长，数据中心数量及规模开始有所增长。近年来，云计算、人工智能、物联网、区块链等新兴互联网技术的兴起使得网络数据量激增。为了更好地处理不断增长的网络数据，数据中心数量和规模得到了进一步扩充，与此同时，数据中心能耗也在逐

²⁷ 数据来源：IDC 《Datacenter Processing, 4Q19: Server CPU, GPU, FPGA, AI ASICs, and ASSPs》

渐增加，有关机构调查显示 2014 年全球数据中心能耗达到世界能源使用总量的 1.62% 左右²⁸。

如果把数据中心看成是一个对外提供算力的基础设施，数据中心能够根据给定的算法处理用户数据，为用户提供算力服务，因此，数据中心能耗的增长也就意味着数据中心算力总能耗的增长。

从能源利用率及社会经济运行效率角度来看，进一步提升数据中心算力能效，提升算力资源的利用率将成为数据中心算力资源管控的重点。在满足社会总体算力需求的前提下，数据中心应该进一步降低 IT 设备能耗，降低总体能耗水平，以更少的能耗代价创造更多的算力资源，为全社会提供更高质量的算力服务。

在数据中心算力能效的测算、评估及改善过程中，数据中心功耗测算是基础。数据中心总功耗主要由供配电系统、制冷系统及 IT 设备功耗构成，在很长一段时间内，数据中心制冷系统能耗过高限制了数据中心能效水平的提升，数据中心 PUE 居高不下，近年来，随着新型制冷技术的应用、冷热通道布局方式的优化以及自然冷源的引入，数据中心整体能效水平有所提升。截至 2018 年年底，全国超大型数据中心平均 PUE 为 1.40，大型数据中心平均 PUE 为 1.54，与前两年比有所提升。全国规划在建数据中心平均设计 PUE 为 1.35 左右，超大型、大型数据中心平均设计 PUE 分别为 1.32、1.40²⁹。

²⁸ <http://www.idc311.com/zhongxin/4962.html>

²⁹ 《全国数据中心应用发展指引（2019）》及 ODCC 实测

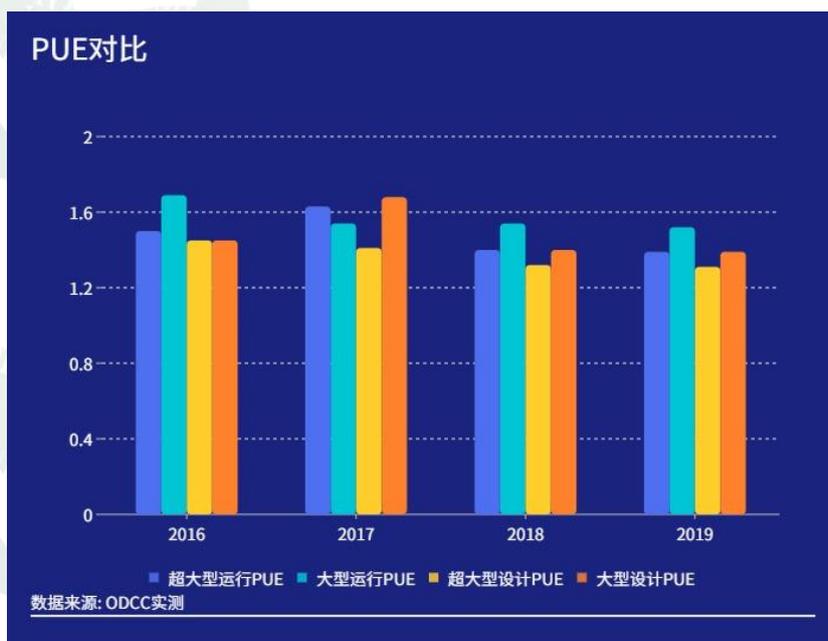


图 11 ODCC 实测 PUE 数据

在数据中心能效管控方面，不少发达国家政府都制定了相应的政策、法规，并提出行业发展倡议，希望能够进一步提升数据中心能效水平，降低 PUE。我国数据中心的能效管理水平依然还有很大的进步空间。

5. 我国数据中心算力情况

5.1. 算力分析

根据前文，我国数据中心内主要使用的服务器类型分为 CPU 服务器和 GPU 服务器，所以本白皮书基于此进行拆解测算。

(1) 通用算力

CPU 类型的服务器几乎部署在所有的数据中心的，根据 4.3.1 的

数据可知 x86 服务几乎占据了 CPU 服务器的全部市场，同时基于我国当前数据中心机架数以及每年服务器的出货量，分别选取 Intel、AMD 等厂商在中国 2015-2019 这五年中每年出货的几款典型产品对数据中心的 CPU 部分算力进行测算。

经过测算，截止到 2019 年底，我国数据中心通用计算能力为 **71.96 EFLOPS³⁰ (FP32)**。

(2) 高性能算力

GPU 更多的使用在 AI 等应用场景中，所以部署于部分数据中心中且规模较小。根据 5.3.2，我们选取 NVIDIA 在 2017-2019 年发布的几款产品（例如：P100、V100 等）对数据中心的 GPU 部分的算力进行测算。

经过测算，截止到 2019 年底，我国数据中心高性能计算能力为 **3.90 EFLOPS (FP64)**，折算为单精度浮点算力为 **7.78 EFLOPS (FP32)**。

(3) 小结

综上，截止到 2019 年底，我国数据中心总算力（含通用算力和高

³⁰ 1 EFLOPS (exaFLOPS) = 10¹⁸ FLOPS

性能算力)，即 CP 为 **79.74 EFLOPS (FP32)**。

5.2. 算效分析

截止到 2019 年底，我国数据中心的通用计算能力的算效为 15.7 GFLOPS/W (FP32)；高性能计算能力的算效为 22.8 GFLOPS/W (FP64)，折算为单精度浮点的算效为 45.5 GFLOPS/W (FP32)。综合通用计算能力和高性能计算能力的算效，全国数据中心的总体算效达到 **18.16 GFLOPS/W (FP32)**。

6. 发展趋势

6.1. 异构加速计算需求旺盛

传统数据中心，在面对机器学习、人工智能、无人驾驶、工业仿真等新兴技术领域的崛起的过程中，遭遇通用 CPU 在处理海量计算、海量数据/图片时越来越多的性能瓶颈。

在数据中心加快步伐部署 48 核以及 64 核心等更高核心 CPU 来应对激增的算力需求的同时，为了应对计算多元化的需求，越来越多的场景开始引入通用加速芯片，如 GPU，专用加速芯片，如 FPGA，ASIC 等异构硬件。加速硬件承担了大部分的新算力需求。

如 AMD Chiplet，Intel Foveros 等技术，正在致力于设计统一的高速互联技术，实现 CPU、GPU、FPGA、ASIC 等计算单元的按需组合，来应对更多样的异构计算需求。

6.2. 大型、边缘两极化发展

随着 DCIM 和智能巡检机器人等技术的发展,远程管理和无人值守数据中心逐渐出现,数据中心运维自动化程度越来越高,大规模数据中心及边缘数据中心的组建在管理上成为可能。

一方面,大型、超大型数据中心,便于服务器的云化和虚拟化,体现规模效应,不管是全球还是国内,超大规模数据中心的占比都在不断提升。

另一方面,爆炸式增长的数据,倒逼计算资源向数据源头靠近。在 5G 带宽大幅增加且天然支持网络切片(更适合传统产业及工业场景)的推动下,离数据产生和处理场景更近的边缘节点也将越来越多发挥中心节点无法替代的作用,在场景需求密集区域,可能成为稀缺资源。

6.3. 计算、存储、网络深度融合

未来,数据除了在计算节点上(如 CPU+GPU)处理,也会在存储和网络流动中处理,例如:智能网卡会承担更多的算力,如数据的加解密、网络协议卸载等;存储端也会承担更多的算力,如数据去重、存储协议卸载等。2020 年 Q4, NVIDIA 首款“网卡”NVIDIA Mellanox 400G InfiniBand 发布,除了提供 400Gbps 的网络互联性能之外,增加了新的 NVIDIA 网络计算引擎,实现了额外的加速。

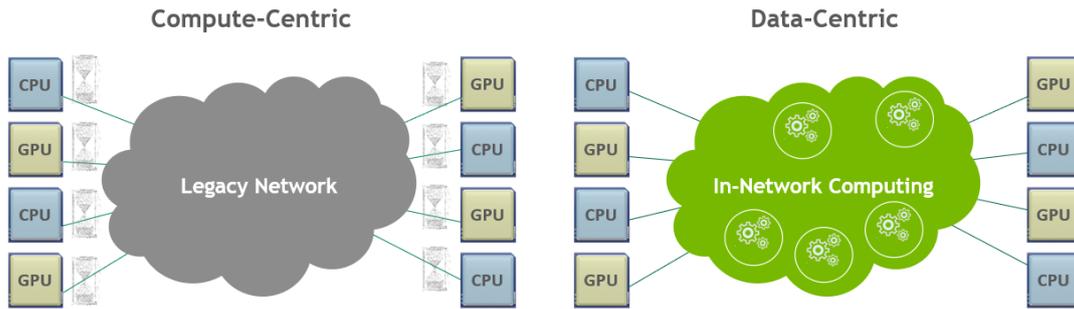


图 12 存储、网络和计算的深度融合

6.4. 对更高计算效能的追逐

数据中心能耗随着单芯片功耗的提升将更加被关注。更高制程工艺的芯片成为数据中心部署的首选，例如 7 纳米 CPU 制程工艺，相比 14 纳米，为数据中心在单位部署空间和单位能耗方面带来最高双倍算力。虚拟化技术将物理硬件基础设施虚拟成按需使用的云化资源，以提高资源的共享率和利用率，将数据中心 IT 环境改造成为更加强大、更具弹性、更富有活力的架构，使得单台服务器的平均利用率大幅提升，通过减少闲置资源，提升计算效能。

6.5. 量子计算大有可为

2020 年 12 月，中国科学技术大学潘建伟研究团队与中科院上海微系统所、国家并行计算机工程技术研究中心合作，通过在量子光源、量子干涉、单光子探测器等领域的自主创新，成功构建了 76 个光子 100 个模式的高斯玻色取样量子计算原型机“九章”。实验结果显示，“九章”处理特定问题的速度比目前世界排名第一的超算“富岳

(Fugaku)”快一百万亿倍，成功实现了我国在量子计算领域的突破。

未来，量子计算在算力领域将发挥更大的价值，量子计算机在原理上具有超快的并行计算能力，提供比传统计算机更强的算力支持。

7. 相关后续工作

本白皮书为数据中心算力提供了一种行之有效的衡量方法，后续开放数据中心委员会（ODCC）将依托 ODCC 实验室，由中国信息通信研究院云计算与大数据研究所开展数据中心算力评估的相关业务。通过评估，各数据中心可以明确自身的“算力”和“算效”，这将有助于精细化明确数据中心的计算能力以及能耗的真正的利用情况，使得数据中心在不断调优 PUE 的同时，通过不断调优“算效”，从而进一步将数据中心作为新型基础设施的杠杆作用发挥到极致。



图 13 数据中心算力和算效的等级评估 logo

同时，下一步我们就不断扩充算力研究的内涵以及多样性，在把

已有研究细化深化之后，继续将算力的成本、经济效益、社会影响等方面也纳入相应的考虑，使得研究的体系更加完善。

8. 总结

数据中心作为新一代信息通信技术的重要载体，是算力输出的底座。异构加速计算的需求日益旺盛，高性能计算能力将大有可为，将逐渐成为数据中心算力的主要力量；计算、存储、网络的深度融合，更加丰富了“算力”的内涵；同时，“大型+边缘”的两极化发展对算力提出了多样性的要求。算力发展的挑战将来自于功耗，应大力推动“绿色算力”的发展，在提高算力的同时降低数据中心的能耗，使得能源在数据中心的利用效益最大化。



开放数据中心委员会（ODCC）

开放数据中心委员会（ODCC）为中国通信标准化协会（CCSA）技术机构之一，由腾讯、阿里巴巴、百度、中国电信、中国移动、中国信息通信研究院以及英特尔等发起成立，围绕服务器、数据中心设施、网络、边缘计算、新技术与测试、智能监控与管理等内容，已发布 150 多项成果，引领技术创新，推动产业合作推广应用，搭建活跃高效的生态圈和开放平台。



在产业界企业和专家们的共同努力下，开放数据中心委员会(ODCC)成果日益丰富、影响力稳步提升，已成为数据中心界知名开放平台。未来，开放数据中心委员会(ODCC)将继续秉承“开放、创新、合作、共赢”的理念，引领我国数据中心产业乘着“新基建”的东风踏上新征程，相信 ODCC 的未来将如星辰大海般璀璨辽阔，欢迎加入我们！

联系方式：ODCC@odcc.org.cn

网 址：www.odcc.org.cn



WWW.ODCC.ORG.CN

本页为规范最后一页