

边缘计算基础资源架构

作为一种新型的服务模型,边缘计算将数据或任务放在靠近数据源头的网络边缘侧进行处理。网络边缘侧可以是从数据源到云计算中心之间的任意功能实体,这些实体搭载着融合网络、计算、存储等核心能力于一体的边缘计算平台,为终端用户提供实时、动态和智能的服务计算。同时,数据就近处理的理念也为数据安全和隐私保护提供了更好的结构化支撑。

边缘计算模型的总体架构主要包括核心基础设施、边缘数据中心、边缘网络和边缘设备。从架构功能角度划分,边缘计算包括基础资源(计算、存储、网络)、边缘管理、边缘安全以及边缘计算业务应用,如图 5-1 所示。

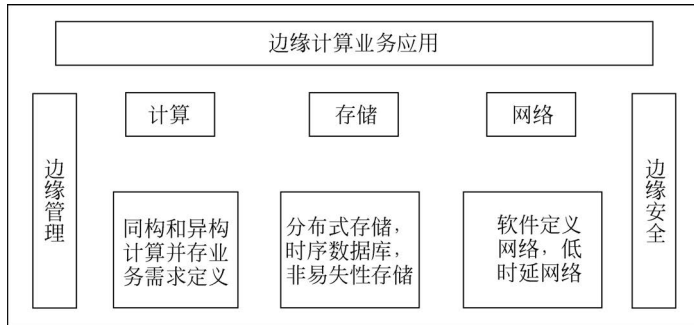


图 5-1 边缘计算功能划分模块

边缘计算的业务执行离不开通信网络的支持,其网络既要满足与控制相关业务传输时间的确定性和数据完整性,又要能够支持业务的灵活部署和实施。时间敏感网络(TSN)和软件定义网络(SDN)技术是边缘计算网络部分的重要基础资源。异构计算支持是边缘计算模块的技术关键。随着物联网和人工智能的蓬勃发展,业务应用对于计算能力提出了更高的要求。计算需要处理的数据种类也日趋多样化,边缘设备既要处理结构化数据,又要处理非结构化数据。为此,边缘计算架构需要解决不同指令集和不同芯片体系架构的计算单元协同起来的异构计算,满足不同业务应用的需求,同时实现性能、成本、功耗、可移植性等的优化均衡。

目前,业界以云服务提供商为典型案例,已经实现部署了云上 AI 模型训练和推理预测的功能服务。将推理预测作为边缘计算工程应用的热点,既满足了实时性要求,又大幅减少了占用云端资源的无效数据。边缘存储以时序数据库(包含数据的时间戳等信息)等分布式存储技术为支撑,按照时间序列存储完整的历史数据,需要支持记录物联网时序数据的快速写入、持久化、多维度的聚合等查询功能。

本章首先介绍边缘计算与前沿技术的关联和融合,然后详细介绍边缘计算网络、存储、计算三大基础资源架构技术。

第 1 节 边缘计算与前沿技术的关联和融合

1. 边缘计算和云计算

边缘计算的出现不是替代云计算,而是互补协同,也可以说边缘计算是云计算的一部分。边缘计算和云计算的关系可以比喻为集团公司的地方办事处与集团总公司的关系。二者各有所长,云计算擅长把握整体,聚焦非实时、长周期数据的大数据分析,能够在长周期维护、业务决策支撑等领域发挥优势;边缘计算则专注于局部,聚焦实时、短周期数据的分析,能更好地支撑本地业务的实时智能化处理与执行。云边协同将放大边缘计算与云计算的应用价值:边缘计算既靠近执行单元,又是云端所需的高价值数据的采集单元,可以更好地支撑云端应用的大数据分析;反之,云计算通过大数据分析,优化输出的业务规则或模型可以下发到边缘侧,边缘计算基于新的业务规则进行业务执行的优化处理。

边缘计算不是单一的部件,也不是单一的层次,而是涉及边缘 IaaS、边缘 PaaS 和边缘 SaaS 的端到端开放平台。如图 5-2 所示为云边协同框架,清晰地阐明了云计算和边缘计算的互补协同关系。边缘 IaaS 与云端 IaaS 实现资源协同;边缘 PaaS 和云端 PaaS 实现数据协同、智能协同、应用管理协同、业务编排协同;边缘 SaaS 与云端 SaaS 实现服务协同。

2018 年年底,阿里云和中国电子技术标准化研究院等单位发表了《边缘云计算技术及标准化白皮书(2018)》,提出了边缘云的概念。现阶段被广为接受的云计算定义是《信息技术云计算概览与词汇》中给出的定义:云计算是一种将可伸缩、弹性的共享物理和虚拟资源池以按需自服务的方式供应和管理的模式。云计算模式由关键特征、云计算角色和活动、云能力类型和云服务类别、云部署模型、云计算共同关注点组成。

但是,目前对云计算的概念都是基于集中式的资源管控提出的,即使采用多个数据中心互联互通的形式,依然将所有的软硬件资源视为统一的资源进行管理、调度和售卖。随着 5G、物联网时代的到来以及云计算应用的逐渐增加,集中式的云

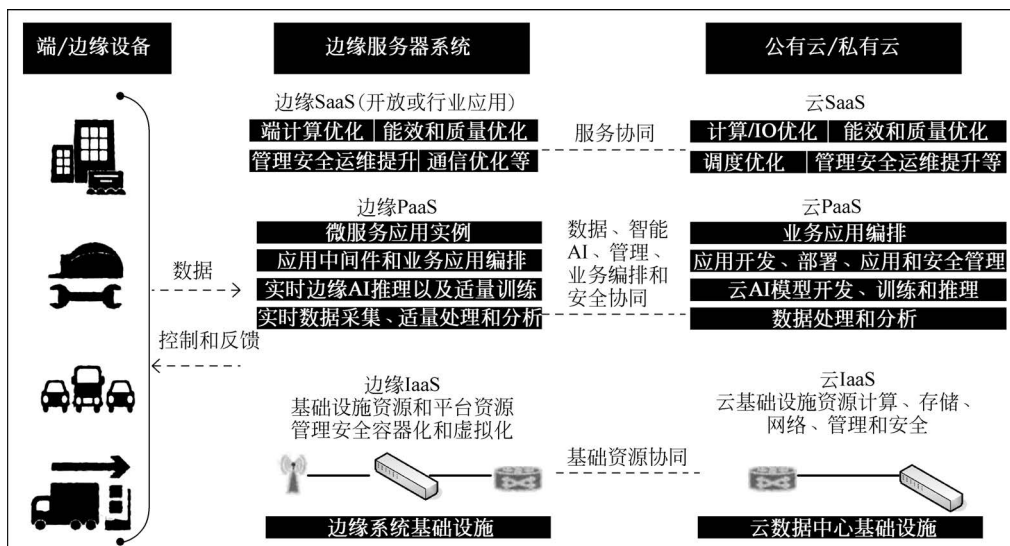


图 5-2 云边协同框架

已经无法满足终端侧“大连接、低时延、大带宽”的资源需求。结合边缘计算的概念,云计算将必然发展到下一个技术阶段:将云计算的能力拓展至距离终端更近的边缘侧,并通过“云-边-端”的统一管控实现云计算服务的下沉,提供端到端的云服务。边缘云计算的概念也随之产生。

《边缘云计算技术及标准化白皮书(2018)》把边缘云计算定义为:基于云计算技术的核心和边缘计算的能力,构筑在边缘基础设施之上的云计算平台。同时,边缘云计算也是形成边缘位置的计算、网络、存储、安全等能力的全面的弹性云平台,并与中心云和物联网终端形成“云边端三体协同”的端到端的技术架构。通过将网络转发、存储、计算、智能化数据分析等工作放在边缘处理,可以降低响应时延、减轻云端压力、降低带宽成本,并提供全网调度、算力分发等云服务。

边缘云计算的基础设施包括但不限于分布式 IDC、运营商通信网络边缘基础设施、边缘侧客户节点(如边缘网关、家庭网关等)等边缘设备及其对应的网络环境。图 5-3 描述了中心云和边缘云协同的基本概念。边缘云作为中心云的延伸,将云的部分服务或者能力(包括但不限于存储、计算、网络、AI、大数据、安全等)扩展到边缘基础设施之上。中心云和边缘云相互配合,实现中心—边缘协同、全网算力调度、全网统一管控等能力,真正实现“无处不在”的云。

边缘云计算在本质上是基于云计算技术的,为“万物互联”的终端提供低时延、自组织、可定义、可调度、高安全、标准开放的分布式云服务。边缘云可以最大限度地与中心云采用统一架构、统一接口、统一管理,这样能够降低用户的开发成本和运维成本,真正实现将云计算的范畴拓展至距离产生数据源更近的地方,弥补传统

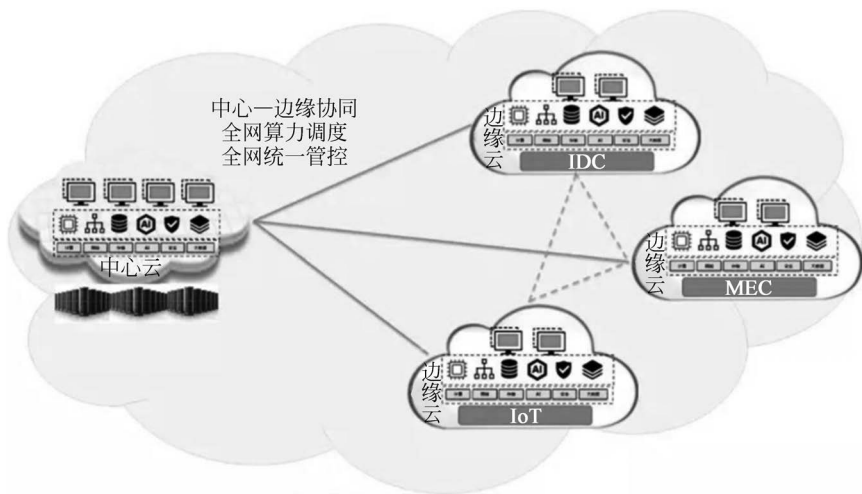


图 5-3 中心云和边缘云协同

架构的云计算在某些应用场景中的不足。

根据所选择的边缘云计算基础设施的不同以及网络环境的差异,边缘云计算技术适用于以下场景:将云的计算能力延展到距离“万物”10km 的位置,例如将服务覆盖到乡镇,街道级“10km 范围圈”的计算场景。“物联网云计算平台”能够将云的计算能力延展到“万物”的身边,可称为“1km 范围圈”,工厂、楼宇等都是这类覆盖的计算场景。

除了网络能够覆盖到的“10km 计算场景”和“1km 计算场景”,边缘云计算还可以在无法覆盖的地域,即通常被称为“网络黑洞”的区域提供“边缘云计算服务”,例如“山海洞天”(深山、远海航船、矿井、飞机)等需要计算的场景。在需要时将处理的数据进行实时处理,联网之后再与中心云协同处理。边缘云计算具备网络低时延、支持海量数据访问、弹性基础设施等特点。同时,空间距离的缩短带来的好处不只是降低了传输时延,还降低了复杂网络中各种路由转发和网络设备处理的时延。此外,由于网络链路被争抢的概率大大减小,能够明显降低整体时延。边缘云计算给传统云中心增加了分布式能力,在边缘侧部署部分业务逻辑并完成相关的数据处理,可以大大缓解将数据传回中心云的压力。边缘云计算还能够提供基于边缘位置的计算、网络、存储等弹性虚拟化的能力,并能够真正实现“云边协同”。

2. 边缘计算和大数据

大数据是指无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合。大数据技术是指从各种各样的数据中快速获得有价值信息的能

力,适用于大数据的技术包括大规模并行处理(MPP)数据库、数据挖掘网络、分布式文件系统、分布式数据库、云计算平台、互联网和可扩展的存储系统。大数据具有以下4个基本特征。

① 数据体量巨大。

百度资料表明,其新首页导航每天需要提供的数据超过1.5PB(1PB=1024TB),这些数据如果用A4纸打印出来,将超过5000亿张。有资料证实,到目前为止,人类生产的所有印刷材料的数据量仅为200PB。

② 数据类型多样。

现在的数据类型不仅是文本形式,更多的是图片、视频、音频、地理位置信息等多种类型的数据,个性化数据占绝大多数。

③ 数据处理速度快。

数据处理遵循“1秒定律”,可从各种类型的数据中快速获得高价值的信息。

④ 数据价值密度低。

以视频为例,在不间断的监控过程中,时长为1小时的视频中可能有用的数据仅有一两秒。

(1) 大数据分析理论。

只有通过大数据进行分析才能获取很多智能的、深入的、有价值的信息。如今,越来越多的应用涉及大数据,而这些大数据的属性(包括数量、速度、多样性等)呈现了大数据不断增长的复杂性。所以,大数据分析的方法在大数据领域就显得尤为重要,可以说是判断最终信息是否有价值的决定性因素。基于此,大数据分析普遍存在的方法理论如下。

① 可视化分析。

大数据分析的使用者有大数据分析专家和普通用户,但是二者对于大数据分析最基本的要求都是可视化分析。因为可视化分析能够直观地呈现大数据的特点,同时非常容易被读者接受,就如同看图说话一样简单明了。

② 数据挖掘算法。

大数据分析的理论核心就是数据挖掘算法,各种数据挖掘算法基于不同的数据类型和格式,才能更加科学地呈现出数据本身具备的特点。也正是因为这些统计方法,我们才能深入数据内部,挖掘出公认的价值。另外,也正因为有了这些数据挖掘的算法,才能更快速地处理大数据。

③ 预测性分析。

大数据分析最重要的应用领域之一是预测性分析。预测性分析是从大数据中挖掘出信息的特点与联系,并科学地建立模型,之后通过模型导入新的数据,从而预测未来的数据。

④ 语义引擎。

非结构化数据的多元化给数据分析带来新的挑战,我们需要一套工具系统地分析和提炼数据。语义引擎需要具备人工智能,以便从数据中主动地提取信息。

⑤ 数据质量和数据管理。

大数据分析离不开数据质量和数据管理,有了高质量的数据和有效的数据管理,无论是在学术研究还是在商业应用领域,都能够保证分析结果的真实性和价值。

(2) 大数据的处理方法。

对大数据的处理有采集、导入和预处理、统计分析、挖掘四种方法。

① 采集。

大数据的采集是指利用多个数据库接收客户端(Web、App 或传感器形式等)的数据,并且用户可以利用这些数据库进行简单的查询和处理。例如,电商会使用传统的关系数据库存储每一笔事务数据;除此之外,非关系数据库也常用于数据的采集。

在大数据的采集过程中,其主要特点和挑战是并发数高,因为同时会有成千上万个用户进行访问和操作。例如,火车票售票网站和淘宝网,它们并发的访问量在峰值时达到上百万,所以需要在采集端部署大量数据库才能支撑。如何在这些数据库之间进行负载均衡和分片,需要深入地思考和设计。

② 导入和预处理。

虽然采集端本身会有很多数据库,但是如果要对这些海量数据进行有效的分析,应将这些数据导入一个集中的大型分布式数据库或者分布式存储集群中,并且在导入的基础上做一些简单的数据清洗和预处理工作。也有一些用户会在导入时使用 Twitter 的 Storm 对数据进行流式计算,来满足部分业务的实时计算需求。

导入和预处理过程的特点和挑战主要是导入的数据量大,每秒的导入量经常会达到百兆甚至千兆级别。

③ 统计分析。

统计分析主要利用分布式数据库或分布式计算集群对海量数据进行分析 and 分类汇总等操作,以满足大多数常见的分析需求。在这方面,一些实时性需求会用到 EMC 的 Green Plum、Oracle 的 Exadata,以及基于 MySQL 的列式存储数据库 Infobright 等。而一些批处理或者基于半结构化数据的需求,可以使用 Hadoop 来满足。

统计分析的主要特点和挑战是涉及的数据量大,其对系统资源,特别是 I/O 会有极大的占用。

④ 挖掘。

与统计分析过程不同的是,数据挖掘一般没有预先设定好的主题,主要是在现

有数据上面进行基于各种算法的计算,从而达到预测的效果,实现一些高级别数据分析的需求。比较典型的算法有用于聚类的 K-means、用于统计学习的 SVM 和用于分类的 NaiveBayes,使用的工具主要有 Hadoop 的 Mahout 等。该过程的特点和挑战主要是用于挖掘的算法很复杂,并且计算涉及的数据量和计算量都很大,常用的数据挖掘算法都以单线程为主。

现在,大多数请求被大规模离线系统处理,云服务商也正开发新的技术以便适应这种趋势。持续的大数据处理不仅缩短了磁盘的使用寿命,还会降低云服务器的整体工作寿命。常规 Web 服务器硬件组件的使用寿命达到 4~5 年,而与大数据相关的组件和云服务器的生命周期不超过 2 年。引入边缘计算将帮助解决这个问题,在采集端将信息过滤,在边缘做预处理和统计分析,仅把有用的待挖掘信息提交给云端。

基于云的大数据分析非常强大,给系统提供的有用信息越多,系统就越能对问题提供更好的答案。例如,在零售环境中,面部识别系统收集的消费者画像统计数据可以添加更详细的信息,让商家不仅知道销售了什么,还知道谁在购买这些商品。此外,在制造过程中,测量温度、湿度和波动等信息的物联网传感器有助于构建运维配置信息,预测机器何时可能发生故障,以便提前维护。

以上情景的困难在于,在大多数情况下,物联网设备生成的数据数量非常惊人,而且并非所有数据都是有用的。以消费者画像统计信息为例,它基于公有云的系统,物联网摄像机必须先收集视频,再将其发送到中央服务器,然后提取必要的信息。而借助边缘计算,连接到摄像机的计算设备可直接提取消费者画像统计信息,并将其发送到云中进行存储和处理。这大大减少了收集的数据量,并且可以仅提取有用的信息。

同样使用物联网传感器,是否有必要每秒发送一次测量数据进行存储呢?通过在本地图存数据和计算能力,边缘设备可以帮助减少噪声、过滤数据。最重要的是,在人们担心安全和隐私的时代,边缘计算提供了一种负责任和安全的方式来收集数据。例如,消费者画像统计信息案例中,没有私人视频或面部数据被发送到服务器,而是仅仅发送有用的非个性化数据。

大数据分析有两种主要的实现模式:数据建模和实时处理。数据建模有助于进行业务洞察和全局分析,实时数据可让用户对当前发生的事情做出反应。边缘人工智能提供了最有价值的实时处理。例如在面部识别和消费者画像统计方面,零售商可以根据屏幕前客户的喜好定制显示内容或者调整报价,吸引更多的观看者,从而提升广告关注度和购买转化率。传统的方式会将视频流发送到云,对其进行处理,然后显示正确的商品,这样非常耗时。使用边缘计算,本地可以解码人物画像统计信息,然后在短时间内调整显示内容或商品报价。

3. 边缘计算和人工智能

人工智能革命是从弱人工智能,通过强人工智能,最终达到超人工智能的过程。现在,人类已经掌握了弱人工智能。

2018年,华为发布的《GIV2025: 打开智能世界产业版图》白皮书也指出:到2025年,全球物联数量将达1000亿,全球智能终端数量将达400亿。边缘计算将提供AI能力,边缘智能成为智能设备的支撑体,人类将被基于ICT网络、以人工智能为引擎的第四次技术革命带入一个万物感知、万物互联、万物智能的智能世界。

全球研究和预测机构认为,到2023年,IoT将推动数字业务创新。2019年有142亿个互联设备被使用,2021年达到250亿个,这一过程会产生大量的数据。人工智能将应用于各种物联网信息,包括视频、静止图像、语音、网络流量活动和传感器数据。因此,公司必须在物联网战略中建立一个充分利用AI工具和技能的企业组织。目前,大多数物联网端设备使用传统处理器芯片,但是传统的指令集和内存架构并不适合于端设备需要执行的所有任务。例如,深度神经网络(DNN)的性能通常受到内存带宽的限制,而并非受到处理能力的限制。

到2023年,预计新的专用芯片将降低运行DNN所需的功耗,并在低功耗物联网端点中实现新的边缘架构和嵌入式DNN功能。这将支持新功能,例如与传感器集成的数据分析,以及低成本电池供电设备中所设置的语音识别。Gartner建议人们注意这一趋势,因为支持嵌入式AI等功能的芯片将使企业能够开发出高度创新的产品和服务。

边缘计算可以加速实现人工智能就近服务于数据源或使用者。尽管目前企业不断将数据传送到云端进行处理,但随着边缘计算设备的逐渐应用,本地化管理变得越来越普遍,企业上云的需求或将面临瓶颈。由于人们需要实时地与他们的数字辅助设备交互,因此等待数千米(或数十千米)以外的数据中心是行不通的。以沃尔玛为例,沃尔玛零售应用程序将在本地处理来自商店相机或传感器网络的数据,而云计算带来的数据时延对沃尔玛来说太高了。

人工智能仍旧面临优秀项目不足、场景落地缺乏的问题。另外,随着人工智能在边缘计算平台中的应用,加上边缘计算与物联网“云-边-端”协同推进应用落地的需求不断增加,边缘智能成为边缘计算的新形态,打通物联网应用的“最后一公里”。

(1) 边缘智能应用领域。

① 自动驾驶领域。

在汽车行业,安全性是最重要的问题。在高速驾驶情况下,实时性是保证安全

性的首要前提。由于网络终端机时延的问题,云端计算无法保证实时性。车载终端计算平台是自动驾驶计算发展的未来。另外,随着电动化的发展,低功耗对于汽车行业变得越来越重要。天然能够满足实时性与低功耗的 ASIC 芯片将是车载计算平台未来的发展趋势。目前,地平线机器人与 Mobileye 是 OEM 与 Tier1 的主要合作者。

② 安防、无人机领域。

相比于传统视频监控,AI+视频监控最主要的变化是把被动监控变为主动分析与预警,解决了需要人工处理海量监控数据的问题。安防、无人机等终端设备对算力及成本有很高的要求。随着图像识别与硬件技术的发展,在终端完成智能安防的条件日益成熟。海康威视、大疆公司已经在智能摄像头上使用了 Movidius 的 Myriad 系列芯片。

③ 消费电子领域。

搭载麒麟 980 芯片的华为 Mate 20 与同样嵌入 AI 芯片的 iPhone XS 将手机产业带入智能时代。另外,亚马逊的 Echo 引爆了智能家居市场。对于包括手机、家居电子产品在内的消费电子行业,实现智能的前提是要解决功耗、安全隐私等问题。据市场调研表明,搭载 ASIC 芯片的智能家电、智能手机、AR/VR 设备等智能消费电子产品已经处在爆发的前夜。

(2) 边缘智能产业生态。

边缘智能产业生态架构已经形成,主要有以下三类玩家。

① 第一类:算法玩家。

从算法切入,如提供计算机视觉算法、NLP 算法等,典型的公司有商汤科技和旷视科技。2017 年 10 月,商汤科技同美国高通公司宣布将展开“算法+硬件”形式的合作,将商汤科技机器学习模型与算法整合到高通面向移动终端、IoT 设备的芯片产品中,为终端设备带来更优的边缘计算能力。而旷视科技为了满足实战场景中不同程度的需求,也在持续优化算法以适配边缘计算的要求。

② 第二类:终端玩家。

从硬件切入,如提供手机、PC 等智能硬件。拥有众多终端设备的海康威视在安防领域深耕多年,是以视频为核心的物联网解决方案提供商。其在发展过程中,将边缘计算和云计算加以融合,更好地解决物联网现实问题。

③ 第三类:算力玩家。

从终端芯片切入,例如开发用于边缘计算的 AI 芯片等。对于边缘计算芯片领域,华为在 2018 年发布昇腾系列芯片——昇腾 310,面向边缘计算产品。

国际上,谷歌云推出 TPU 的轻量级版本——Edge TPU,用于边缘计算,并开放给商家。亚马逊也被曝光开发 AI 芯片,主要用来支持亚马逊的 Echo 及其他移

动设备。不过单独占据一类的参与者不是终极玩家,边缘智能需要企业同时具备终端设备、算法和芯片的能力。

4. 边缘计算和 5G

5G 技术以“大容量、大带宽、大连接、低时延、低功耗”为诉求。联合国国际电信联盟(ITU-R)对 5G 定义的关键指标包括:峰值吞吐率 10Gb/s、时延 1ms、连接数 100 万、移动速度 500km/h。

(1) 高速度。

相对于 4G,5G 要解决的第一个问题就是高速度。只有提升网络速度,用户体验与感受才会有较大提高,网络才能在面对 VR 和超高清业务时不受限制,对网络速度要求很高的业务才能被广泛推广和使用。因此,5G 的第一个特点就定义了速度的提升。

其实和每一代通信技术一样,很难确切说出 5G 的速度到底是多少。一方面,峰值速度和用户的实际体验速度不一样;另一方面,不同的技术在不同的时期速率也会不同。5G 的基站峰值要求不低于 20Gb/s,随着新技术的使用,还有提升的空间。

(2) 泛在网。

随着业务的发展,网络业务需要无所不包,广泛存在。只有这样才能支持更加丰富的业务,才能在复杂的场景上使用。泛在网有两个层面的含义:广泛覆盖和纵深覆盖。广泛是指在社会生活的各个地方需要广覆盖。高山峡谷如果能覆盖 5G,可以大量部署传感器,进行环境、空气质量,甚至地貌变化、地震的监测。纵深覆盖是指虽然已经有网络部署,但是需要进入更高品质的深度覆盖。5G 的到来,可把以前网络品质不好的卫生间、地下车库等环境都用 5G 网络广泛覆盖。

在一定程度上,泛在网比高速度还重要。只建一个少数地方覆盖、速度很高的网络,并不能保证 5G 的服务与体验,而泛在网才是 5G 体验的一个根本保证。

(3) 低功耗。

5G 要支持大规模物联网应用,就必须有功耗的要求。如果能把功耗降下来,让大部分物联网产品一周充一次电,甚至一个月充一次电,就能大大改善用户体验,促进物联网产品的快速普及。eMTC 基于 LTE 协议演进而来,为了更加适合物与物之间的通信,对 LTE 协议进行了裁剪和优化。eMTC 基于蜂窝网络进行部署,其用户设备通过支持 1.4MHz 的射频和基带带宽,可以直接接入现有的 LTE 网络。eMTC 支持上下行最大 1Mb/s 的峰值速率。而 NB-IoT 构建于蜂窝网络,只消耗大约 180kHz 的带宽,可直接部署于 GSM 网络、UMTS 网络或 LTE 网络,以降低部署成本、实现平滑升级。

(4) 低时延。

5G 的新场景是无人驾驶、工业自动化的高可靠连接。要满足低时延的要求,需要在 5G 网络建构中找到各种办法,降低时延。边缘计算技术也会被采用到 5G 的网络架构中。

(5) 万物互联。

在传统通信中,终端是非常有限的,在固定电话时代,电话是以人群定义的。而手机时代,终端数量有了巨大爆发,手机是按个人应用定义的。到了 5G 时代,终端不是按人来定义,因为每人可能拥有数个终端,每个家庭也可能拥有数个终端。

在社会生活中,大量以前不可能联网的设备也会进行联网工作,更加智能。井盖、电线杆、垃圾桶这些公共设施以前管理起来非常难,也很难做到智能化,而 5G 可以让这些设备都成为智能设备,利于管理。

(6) 重构安全。

传统的互联网要解决的是信息高速、无障碍的传输,自由、开放、共享是互联网的基本精神,但是在 5G 基础上建立的是智能互联网。智能互联网不仅要实现信息传输,还要建立起一个社会和生活的新机制与新体系。智能互联网的基本精神是安全、管理、高效、方便。安全是 5G 之后的智能互联网第一位的要求,如果 5G 无法重新构建安全体系,那么会产生巨大的破坏力。

在 5G 的网络构建中,在底层就应该解决安全问题。从网络建设之初,就应该加入安全机制,信息应该加密,网络并不应该是开放的,对于特殊的服务需要建立起专门的安全机制。网络不是完全中立、公平的。

如图 5-4 所示,在目前的网络架构中,由于核心网的高位部署传输时延比较高,不能满足超低时延业务需求;此外,业务完全在云端终结并非完全有效,尤其一些区域性业务不在本地终结,既浪费带宽,也增加时延,因此,时延指标和连接数指标决定了 5G 业务的终结点不可能全部在核心网后端的云平台,移动边缘计算正好契合该需求。一方面,移动边缘计算部署在边缘位置,边缘服务在终端设备上运行,反馈更迅速,解决了时延问题;另一方面,移动边缘计算将内容与计算能力下沉,提供智能化的流量调度,将业务本地化,内容本地缓存,让部分区域性业务不必大费周章地在云端终结。

5G 三大应用场景之一的“低功耗大连接”要求能够提供具备超千亿网络连接的支持能力,满足每平方千米 100 万个的连接密度指标要求,在这样的海量数据以及高连接密度指标的要求下,保证低时延和低功耗是非常重要的。5G 甚至提出 1ms 端到端时延的业务目标,以支持工业控制等业务的需求。要实现低时延以及低功耗,需要大幅度降低空口传输时延,尽可能减少转发节点,缩短从源到目的节

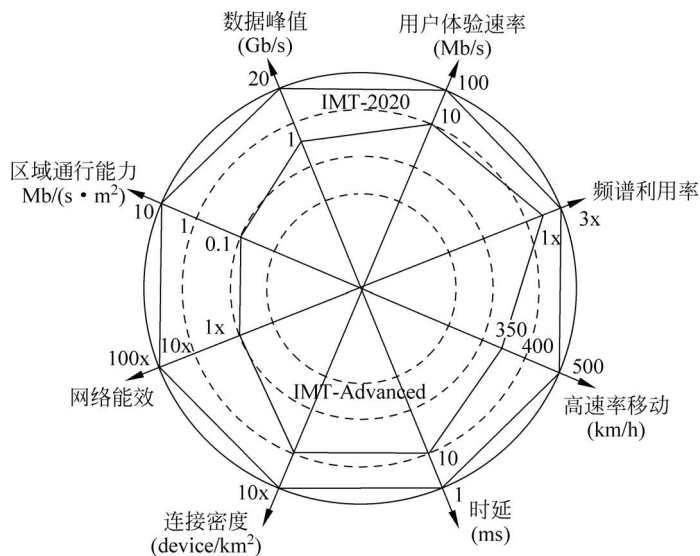


图 5-4 5G 网络架构需求驱动边缘计算发展

点之间的“距离”。

目前的移动技术对时延优化并不充分, LTE 技术可以将空口吞吐率提升 10 倍, 但对端到端的时延只能缩短到原来的 1/3。其原因在于, 当大幅提升空口效率以后, 网络架构并没有充分优化反而成了业务时延的瓶颈。LTE 网络虽然实现了 2 跳的扁平架构, 但基站到核心网往往距离数百千米, 途经多重会聚、转发设备, 再加上不可预知的拥塞和抖动, 根本无法实现低时延的保障。

移动边缘计算部署在移动边缘,将无线网络和互联网技术有效地融合在一起,并在无线网络侧增加计算、存储、处理等功能,构建移动边缘云,提供信息技术服务环境和云计算能力。由于应用服务和内容部署在移动边缘,可以缩短数据传输中的转发时间和处理时间,降低端到端时延,满足低时延要求。因此在网络拥堵严重影响移动视频观感的情况下,移动边缘计算是一个好的解决方法。

① 本地缓存。

移动边缘计算服务器是一个靠近无线侧的存储器,可以事先将内容缓存至移动边缘计算服务器上。在有观看移动视频的需求时,即用户发起内容请求,移动边缘计算服务器立刻检查本地缓存中是否有用户请求的内容,如果有就直接提供服务;如果没有,则去网络服务提供商处获取,并缓存至本地。在其他用户下次有该类需求时,可以直接提供服务。这样便降低了请求时间,也解决了网络堵塞问题。

② 跨层视频优化。

此处的跨层是指“上下层”信息的交互反馈。移动边缘计算服务器通过感知下层无线物理层的吞吐率,服务器(上层)决定为用户发送不同质量、清晰度的视频,

在减少网络堵塞的同时提高线路利用率,从而提升用户体验。

③ 用户感知。

根据移动边缘计算的业务和用户感知特征,可以区分不同需求的客户,确定不同的服务等级,实现对用户差异化的无线资源分配和数据包时延保证,合理分配网络资源,以提升整体的用户体验。

5. 边缘计算和物联网

由无数类型的设备生成的大量数据需要推送到集中式云以保留(数据管理)、分析和决策。然后,将分析的数据结果传回设备。这种数据的往返消耗了大量网络基础设施和云基础设施资源,进一步增加了时延和带宽消耗问题,从而影响关键任务的物联网使用。例如,在自动驾驶的连接车中,每小时产生了大量数据,数据必须上传到云端进行分析,并将指令发送回汽车。低时延或资源拥塞可能会延迟对汽车的响应,严重时可能导致交通事故。

这就是边缘计算的用武之地。边缘计算体系结构可用于优化云计算系统,以便在网络边缘执行数据处理和分析,更接近数据源。通过这种方法,可以在设备本身附近收集和处理数据,而不是将数据发送到云或数据中心。边缘计算驱动物联网发展的优势包括以下方面。

① 边缘计算可以降低传感器和中央云之间所需的网络带宽(即更低的时延),并减轻整个 IT 基础架构的负担。

② 在边缘设备处存储和处理数据,而不需要网络连接来进行云计算,这消除了高带宽的持续网络连接。

③ 通过边缘计算,端点设备仅发送云计算所需的信息而不是原始数据,有助于降低云基础架构的连接和冗余资源的成本。当在边缘分析由工业机械生成大量数据并且仅将过滤的数据推送到云时,这是有益的,从而显著节省 IT 基础设施。

④ 利用计算能力使边缘设备的行为类似于云类操作。应用程序可以快速执行,并与端点建立可靠且高度响应的通信。

⑤ 通过边缘计算实现数据的安全性和隐私性:敏感数据在边缘设备上生成、处理和保存,而不是通过不安全的网络传输,并有可能破坏集中式数据中心。边缘计算生态系统可以为每个边缘提供共同的策略,以实现数据完整性和隐私保护。

⑥ 边缘计算的出现并不能取代对传统数据中心或云计算基础设施的需求。相反,它与云共存,因为云的计算能力被分配到了端点。

第2节 边缘计算优势、覆盖范围和基础资源架构准则

1. 边缘计算优势

在实际应用中,边缘计算可以独立部署,但大多数情况下都是与云计算协作部署。云计算适合非实时的数据处理分析、大量数据的长期保存、通过大数据技术进行商业决策等应用场景;而边缘计算在实时和短周期数据的处理和分析,以及需要本地决策的场景下起着不可替代的作用,例如无人驾驶汽车、智能工厂等。它们都需要在边缘就能进行实时的分析和决策,并保证在确定的时间内响应,否则将导致严重的后果。

边缘计算具备一些云计算没有的优势,除低时延之外,还包括如下优势。

(1) 数据过滤和压缩。

通过边缘计算节点的本地分析能力,可以大大降低需要上传的数据量,从而降低上传网络的带宽压力。

(2) 环境感知能力。

由于边缘计算节点可以访问无线网络,例如 Wi-Fi 热点、5G 的无线接入单元 RRU 等,因此可以给边缘应用提供更多的信息,包括地理位置、订阅者 ID、流量信息和连接质量等,从而具备环境感知能力,为动态地进行业务应用优化提供了基础。

(3) 符合法规。

边缘计算节点可以将敏感信息在边缘侧处理并终结,而不传输到公有云中,从而符合隐私和数据定位信息等相关法律法规。

(4) 网络安全性。

可以通过边缘计算节点来保护云服务提供商的网络不受攻击。

如图 5-5 所示,边缘计算和云计算在数字安防中协同工作,网络摄像头在地理上分散部署,如果将所有视频流和相关元数据都上传到云端进行分析和存储,将消耗大量的网络带宽和成本。通过添加边缘计算节点网络硬盘录像机(NVR),可以在网络边缘侧进行视频流的保存和分析,只将分析结果和感兴趣的视频数据上传到云端,然后进行进一步的分析和长期保存,可以大大降低对网络带宽的要求及由此产生的流量成本,同时降低了响应时间并提高了服务质量。同时,由于边缘计算节点更靠近设备端,因此可以获得更多网络摄像头的位置等环境信息,为进一步提高边缘智能提供了基础。

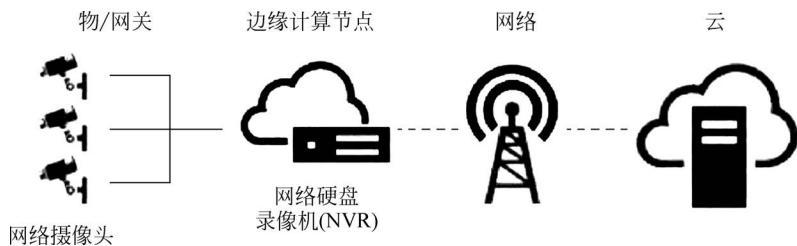


图 5-5 边缘计算和云计算在数字安防中协同工作

2. 边缘计算覆盖范围

如图 5-6 所示,从企业、网络运营商和云服务提供商的角度,边缘计算覆盖的范围不同。对于企业而言,边缘计算由最靠近设备和用户现场的计算节点组成,例如办公室和家庭的智能网关设备,智能工厂内的智能控制器、边缘服务器等;对于运营商而言,边缘计算包括从接入网到核心网之间的基站机房和中心机房内的边缘服务器等。

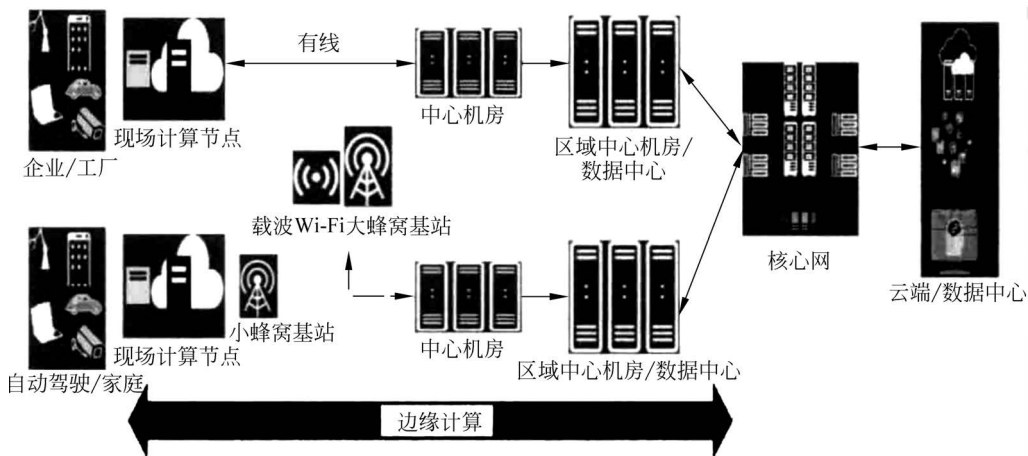


图 5-6 边缘计算覆盖范围

3. 边缘计算基础资源架构准则

(1) 边缘时延要求。

为了应对市场压力,企业变得越来越敏捷。在这样的趋势下,信息技术领域面临着越来越大的压力,因为它需要确保企业能对越来越快的业务速度做出响应。云计算彻底提升了企业可用的后端敏捷性,能够非常快速地为任何企业提供海量的计算和存储能力。敏捷性的下一阶段是前端敏捷,需要重点减少由网络和距离

导致的时延。不同业务对时延的要求差异巨大。在工厂自动化中,微秒之差也是至关重要的。例如,运动控制应用需要几十微秒的周期时间,而在 $10\mu\text{s}$ 内,光在一根典型的光纤中能够传输约 3000ft ($1\text{ft}=0.3048\text{m}$)。在这种情况下,即便是缩短 1m 的距离也可能极为重要。

边缘计算的整体架构设计和部署与实际应用场景是分不开的。如图 5-7 所示,不同的应用对于最大允许时延的要求也有很大不同。例如,对于智能电网控制、无人驾驶、AR 或 VR 应用等,时延需要控制在几十毫秒以内;一些工业控制、高频交易等应用甚至需要控制在 1ms 以内。这些应用场合一般都需要边缘计算来提高响应速度,在确定的时间内完成任务。对于 4K 高清视频流媒体、网页加载、网络聊天等应用,虽然它们对时延的敏感度没有那么高,可允许的最高时延一般在 1~4s,但过高的时延也会影响用户体验和服务质量。因此,也需要 CDN 来进行边缘侧的内容缓存和分发,从而降低由于网络和距离导致的时延。

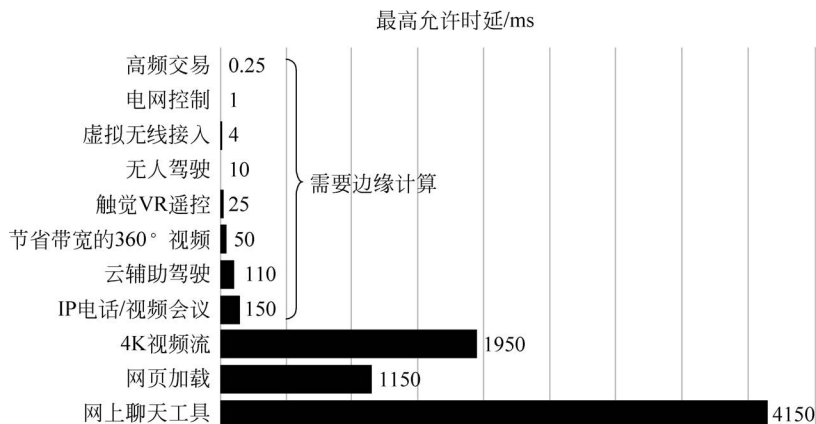


图 5-7 实际应用场景的最大允许时延要求

为了达到边缘应用所需要的高性能和低时延要求,可以从多个方面进行优化。

① 对于虚拟化场景下的网络功能。

可以借助 SR、IOV、直通访问、DPDK、高速网卡(50G/100G)和 NUMA 等来提升性能。

② 对于存储功能。

可以借助分层的存储结构,包括 Memory、SATA、NVMe 等,以及选择合适的内存数据库和数据处理框架来实现。

③ 对于计算功能。

特别是在处理深度学习的推理算法、对称加密或非对称加解密等计算密集型业务时,标准的 CPU 平台是没有太大优势的。因此需要异构的计算平台,例如基于 FPGA、GPU 或者 NPU 的加速卡来卸载这些操作,以缩短计算时间,提升响应速度。

(2) 异构计算。

随着 AI 技术的快速发展,基于机器学习或深度学习的 AI 技术越来越多地被引入边缘计算节点甚至边缘设备中。如图 5-8 所示,同样是数字安防的例子,在智能摄像头中可以集成人脸识别或跟踪的算法,而在分布式的边缘计算节点中,可以进行人脸对齐或特征提取。同时,在带有本地存储的边缘计算节点中进行人脸匹配或特征存储,并周期性地将聚合的数据同步到云端服务器,进行更大范围的人脸匹配或特征存储。在这个过程中,边缘计算节点除了运行本身的业务和应用外,还需要能够执行边缘的模型推理,或根据收集的带标签的数据进行模型的更新和优化。这就需要在边缘计算节点中增加更多算力来更有效率地执行这些算法,例如,基于 FPGA、GPU、ASIC 等的加速器来卸载这部分业务负载。

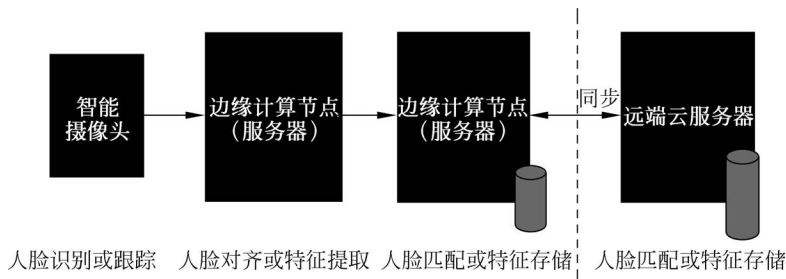


图 5-8 AI 在人脸识别边缘服务器的应用

(3) 负载整合和业务编排。

在对边缘计算提出更多功能需求的同时,用户往往需要简化系统结构,以降低成本,这就需要将单一功能的设备用多功能的设备取代。随着处理器计算能力的提高,以及虚拟化技术的成熟,基于虚拟化和容器实现多负载整合成为业界发展的趋势。

由于边缘计算节点的分布式特征,既有南北向的节点,也有东西向的节点,并且不同应用对于硬件配置、实时性、网络带宽等需求不尽相同,所以如何在边缘计算节点间进行合理的业务编排是关键。目前,流行的如 Kubernetes 或 Apache Mesos 等容器管理和业务编排器,也正通过用户定制调度器来应对边缘场景下复杂调度的问题。同时,通过业务编排器在分布的边缘节点间实现容灾备份,可提高系统可靠性。

(4) 本地互动性。

互动性是指系统协作的速率——本地人与物的“健谈程度”,即确定行动所需的传感器和顺序交互数量。例如,一个人购物的过程包括以下步骤:定位感兴趣的物品,试用这些物品,更换物品,最终做出决定。这是一系列为最终做决策而连续进行的交互。与实时交互的、移动中的人和物组成的复杂多变的系统相比,传感器和制动器对计算能力和时延的要求截然不同。例如,对于自动驾驶车辆在自身

系统内、与其他车辆以及周边环境之间进行的交互而言,迅速且果断的决策可以拯救生命。即使往复一次的时延很低,但一个协作的系统会将时延放大多倍,从而需要更短的时延才能满足要求。高水平的本地互动性除了要求解决方案的物理部署位置更接近于边缘,还需要更强大的信息处理能力、多输入关联能力和数据分析能力,而且可能还需要机器学习功能。

(5) 数据和带宽。

可以说,今天的互联网是围绕涌向边缘的数据而设计的。而物联网的发展趋势正在打开边缘数据爆炸性增长的大门,这和早期云计算的数据流向是相反的。边缘数据的价值特点是:只在边缘对本地决策有价值,对中心总量分析有价值,时间敏感程度高、半衰期短且很快失去价值。

某一些数据可能比另一些数据更有价值,例如捕捉到物体移动的一帧镜头比空帧或仅记录风吹草动的一帧更有价值。有些数据可能需要归档,有些则不用。一些传感器会产生大量复杂数据,而其他传感器只会产生极少量的数据流。带宽的可用性和成本需要与数据价值、生命周期以及是否需要存储和归档相平衡,排列本地优先级、实行数据过滤和智能化有助于减少数据流量。

当数据仅在本地有价值时,边缘计算能够更近距离地处理甚至储存和归档原始数据,从而节约成本。数据存储和远程数据管理将至关重要。当需要处理海量数据时,本地分析和过滤能够减少需要进行维护或送往云端或企业数据中心的数据量。这降低了组网成本,并为更重要的流量处理保留了有限的网络带宽。

因此,应用在云端服务中心的大数据分析技术在边缘计算节点上应用得也越来越多。而随着边缘侧大数据的 4V 特性的显著增长,数据更快、更大、更多样,不可能像传统的 MapReduce 那样将数据先存储下来,然后进行处理和分析。此外,企业对于边缘侧的大数据处理也提出了更高的诉求,要求更快、更精准地捕获数据价值,高性能的流处理将是解决这些问题的关键之一,在大数据处理中也将扮演越来越重要的角色。例如,通过 Spark Streaming、Flink 流处理框架提供内存计算,并在此之上发展出数据处理、高级分析和关系查询等能力。

(6) 隐私和安全。

隐私、安全和监管要求可能需要边缘计算解决方案来满足。对于运营商的网络,一般认为核心网机房处于相对封闭的环境,受运营商控制,安全性有一定保证。而接入网相对更容易被用户接触,处于不安全的环境。由于边缘计算的本地业务处理特性,使得大多数数据在核心网之外得到终结,运营商的控制力减弱,攻击者可以通过边缘计算平台或应用攻击核心网,造成敏感数据泄露、DDoS 攻击等。

边缘计算中的一些数据是公共的,但很多数据是企业保密信息、个人隐私或受到监管的信息。一些边缘计算架构和拓扑将根据数据需要在何地进行安全合法的

存储和分析来决定。边缘的场景可能是工作场所、工厂或家庭,边缘计算可以与人和物在一起“就地部署”。或者边缘本身可能并不安全,例如位于公共空间。在这种情况下,边缘计算需要远离人和物部署才能保障安全。监管要求可能因地理位置而异,因此,不同位置的应用有着不同的网络拓扑和数据归档要求。

隐私和安全问题将推动边缘计算拓扑、数据管理、归档策略和位置以及数据分析方案的形成。为满足不同边缘位置的地理和监管要求,不同的边缘计算解决方案之间可能大不相同。

(7) 有限的自主性。

虽然边缘计算是中央数据中心或云服务的一部分或与之相连,位于边缘的用例可能需要一定程度的独立性和自主性,这包括自组织和自发现(处理新连接的人和物),或当一条连接断开时能够继续操作。边缘计算解决方案还可能依赖于云端或中央数据中心的某些功能或协调能力,而后续这种依赖将减弱。自主性要求还与用例如何确保自我恢复能力、如何处理后端的不一致和不确定的时延有关,也可能与用例如何包含边缘机器学习有关。

不依赖于连接后端的边缘计算解决方案需要更广泛的处理能力和数据缓存能力,也就是自我恢复能力。一旦重新建立连接,这些边缘计算解决方案将需要与它的云端或企业数据中心重新同步。它们需要足够灵活,以根据连接是否可用来动态变更计算能力。它们可能需要更丰富的机器学习能力来自我组织和自我发现,而非依赖核心系统的协调。

(8) 边缘部署环境。

对于靠近现场设备端部署的边缘计算节点,一般需要考虑环境的要求。例如,在智能工厂应用中,边缘计算节点可能直接部署在车间的设备旁。因此,为了保证节点长时间稳定运行,需要支持宽温设计、防尘、无风扇运行,具备加固耐用的外壳或者机箱。

边缘计算服务器通常部署在靠近设备端的办公室内或网络边缘等,边缘计算服务器与 BBU 部署在同一个站址,因此其运行环境必须符合 NEBS 要求。NEBS 要求包括:服务器的工作温度通常为 $-40\sim 50^{\circ}\text{C}$,工作湿度为 $5\%\sim 100\%$,并需要具有良好的防水、防腐、防火性能,以及设备可操作性、抗震性等特性。同时,边缘服务器可能会在机架外进行操作和使用,因此外壳尺寸相对于数据中心要小些,并能够灵活地支持各种固定方式,例如固定在墙上、桌子上或者柜子里。

第3节 边缘计算架构

1. 边缘计算架构的组成

(1) 服务器。

服务器是构建边缘计算架构的核心。相对于传统的数据中心服务器,边缘服

务器应能够提供高密度计算及存储能力。这主要是由于在实际的边缘部署环境中,边缘服务器能够得到的工作空间十分局促,通常不足传统单个数据中心机架(约 40U)的 10%。为了尽可能多地容纳业务部署,边缘服务器需要采用高密度组件,例如多核 CPU(多于 20 核)、预留至少两个半高半长规格的 PCIe 插槽,支持 M.2 E-key 的 Wi-Fi、卸载模块或 M-key 的存储模块、大容量 ECC 内存,以及大容量固态存储器等。

在供电和功耗方面,考虑到深度学习模型推理的场景需要使用卸载卡,总功耗至少在 300W 以上,使用直流电或者交流电供电。对于 5G 基站内部署的服务器,需要支持 48V 直流供电,并支持无风扇散热能力,降低对部署环境的散热要求。

带外管理可以帮助用户在远端管理边缘服务器平台,例如升级系统或诊断故障,是可选的特性。

(2) 异构计算。

随着物联网应用数据的爆炸性增长以及 AI 应用的普及,异构计算在边缘计算架构中也越来越重要。它能够将不同指令集的计算单元集成在一起,从而发挥它们各自最大的优势,实现性能、成本和功耗等的平衡。例如,GPU 具有很强的浮点和向量计算能力,可以很好地完成矩阵和向量的并行计算,适用于视频流的硬件编解码、深度学习模型的训练等;FPGA 具有硬件可编程能力及低时延等特性;而 ASIC 具有高性能、低功耗等特点,可用于边缘侧的深度学习模型推理、压缩或加解密等卸载操作。异构计算在带来优势的同时,也增加了边缘计算架构的复杂度。因此,需要虚拟化和软件抽象层来提供给开发者统一的 SDK 和 API 接口,从而屏蔽硬件的差异,使开发者和用户能够在异构平台上方便地开发和安装。

(3) 虚拟机和容器。

借助虚拟机和容器,系统能够更方便地对计算平台上的业务负载进行整合、编排和管理。虚拟机和容器的主要区别如表 5-1 所示。

表 5-1 虚拟机和容器的主要区别

对比项目	虚拟机	容器
虚拟化位置	硬件	操作系统
抽象目标	从硬件抽象 OS	从 OS 抽象应用
资源管理	每个虚拟机有自己的 OS 内核、二进制和库	容器有同样的主机 OS 和需要的二进制和库
密度	几 GB,服务器能够运行有限的虚拟机	几 MB,服务器上可以运行很多容器
启动时间	秒级	毫秒级
安全隔离度	高	低
性能	接近原生	弱于原生
系统支持	单机支持上千个容器	一般支持几十个容器

虚拟机和容器的选择主要依赖业务需要。若业务之间需要达到更强的安全隔离,虚拟机是较好的选择;如果更看重节省资源、轻量化和高性能,则容器更好。容器可以单独运行在主机 OS 之上,也可以运行在虚拟机中。Docker 等容器技术在多数应用中更适合边缘计算的场景。但是,依然有些边缘场景需要使用传统虚拟机(VM),包括同时需要支持多个不同 OS 的场景,例如 Linux、Windows 或者 VxWorks;以及业务间相差较大并对相互隔离需求更高的时候,例如在一个边缘计算节点中同时运行工业上的 PLC 实时控制、机器视觉和人机界面等。

由于容器具有轻量化、启动时间短等特点,所以能够在需要的时候及时安装和部署,并在不需要的时候立即消失,释放系统资源。同时,一个应用的所有功能再也不需要放在一个单独的容器内,而是可以通过微服务的方式将应用分割成多个模块并分布在不同的容器内,这样更容易进行细粒度的管理。在需要对应用进行修改的时候,不需要重新编译整个应用,而只要改变单个模块即可。

容器管理器用于管理边缘端多个主机上的容器化的应用,例如 Kubernetes 支持自动化部署、大规模可伸缩、应用容器化管理。在生产环境中部署一个应用程序时,通常要部署该应用的多个实例,以便对应应用请求进行负载均衡。在 Kubernetes 中,我们可以创建多个容器,每个容器里面运行一个应用实例,然后通过内置的负载均衡策略,实现对这一组应用实例的管理、发现、访问,而这些细节都不需要运维人员去进行复杂的手工配置和处理。

在边缘计算中,终端节点不再是完全不负责计算,而是做一定量的计算和数据处理,之后把处理过的数据再传递到云端。这样一来可以解决时延和带宽的问题,因为计算在本地,而且处理过的一定是从原始数据中进行过精炼的数据,所以数据量会小很多。当然,具体要在边缘做多少计算也取决于计算功耗和无线传输功耗的折中——终端计算越多,计算功耗越大,无线传输功耗通常就可以更小,对于不同的系统存在不同的最优值。

百度 AI 边缘计算架构作为一个典型、完整的边缘计算技术体系,包括基础设施、性能加速、平台资源管理、PaaS、AI 算法框架和开放应用。基础设施主要包括智能终端、接入网技术、移动边缘站点、云边缘站点和 PoP 站点,根据不同资源程度来分配计算任务;性能加速完成边缘计算节点的计算、存储、I/O 优化和节点连接的加速优化;平台资源管理实现对 CPU、GPU、存储及网络的虚拟化和容器化功能,满足资源的弹性调度和集群管理要求;PaaS 提供应用设计及开发阶段的微服务化、运行态环境、通信框架和管理面运行状态监控等支持;AI 算法框架从时延、内存占用量和能效等方面,实现边缘计算节点上 AI 推理加速和多节点间 AI 训练算法的联动;开放应用凭借 AI 算法框架完成强交互的人机交互、编解码、加解密等信息预处理和算法建模,同时需要在数据源带宽低收敛比、低时延响应的物理

资源环境中满足数据传输和交互需求。

2. 边缘计算平台架构

边缘计算的基础资源包括计算、网络和存储三个基础模块,以及虚拟化服务。

(1) 计算。

异构计算是边缘计算侧的计算硬件架构。近年来,摩尔定律仍然在推动芯片技术不断取得突破,但物联网应用的普及带来了信息量的爆炸式增长,AI 技术应用也增加了计算的复杂度,这些对计算能力都提出了更高的要求。计算要处理的数据种类也日趋多样化,边缘设备既要处理结构化数据,也要处理非结构化数据。同时,随着边缘计算节点包含了更多种类和数量的计算单元,成本成为关注重点。

为此,业界提出将不同指令集和不同体系架构计算单元协同起来的新计算架构,即异构计算,以充分发挥各种计算单元的优势,实现性能、成本、功耗、可移植性等方面的均衡。

同时,以深度学习为代表的新一代 AI 在边缘侧应用还需要新的技术优化。当前,即使在推理阶段,对一幅图片的处理也往往需要超过 10 亿次的计算量,标准的深度学习算法显然不适合边缘侧的嵌入式计算环境。业界正在进行的优化方向包括自顶向下的优化,即把训练完的深度学习模型进行压缩来降低推理阶段的计算负载;同时,也在尝试自底向上的优化,即重新定义一套面向边缘侧嵌入系统环境的算法架构。

(2) 网络。

边缘计算的业务执行离不开通信网络的支持。边缘计算的网络既要满足与控制相关业务传输时间的确定性和数据完整性,又要能够支持业务的灵活部署和实施。时间敏感网络 and 软件定义网络技术会是边缘计算网络部分的重要基础资源。

为了提供网络连接需要的传输时间确定性与数据完整性,国际标准化组织 IEEE 执行了 TSN(Time-Sensitive Networking)系列标准,针对实时优先级、时钟等关键服务定义了统一的技术标准,是工业以太网连接的发展方向。

SDN(Soft Defined Network)逐步成为网络技术的发展的主流,其设计理念是将网络的控制平面与数据转发平面进行分离,并实现可编程化控制。将 SDN 应用于边缘计算,可支持百万级海量设备的接入与灵活扩展,提供高效低成本的自动化运维管理,实现网络与安全的策略协同与融合。

(3) 存储。

数字世界需要实时跟踪物理世界的动态变化,并按照时间序列存储完整的历史数据。新一代时序数据库 TSDB(Time Series DataBase)是存放时序数据(包含数据的时间戳等信息)的数据库,并且需要支持时序数据的快速写入、持久化、多纬

度的聚合查询等基本功能。为了确保数据的准确性和完整性,时序数据库需要不断插入新的时序数据,而不是更新原有数据。

(4) 虚拟化。

虚拟化技术降低了系统开发成本和部署成本,已经开始从服务器应用场景向嵌入式系统应用场景渗透。典型的虚拟化技术包括裸金属(Bare Metal)架构和将主机(Host)等功能直接运行在系统硬件平台上,然后再运行系统和虚拟化功能。前者有更好的实时性,智能资产和智能网关一般采用该方式。后者是虚拟化功能运行在主机操作系统上。

对于边缘计算系统,处理器、算法和存储器是整个系统中最关键的三个要素,以下进行详细分析。

① 用于边缘计算的处理器。

常规物联网终端节点的处理器是一块简单的 MCU,以控制目的为主,运算能力相对较弱。如果要在终端节点加边缘计算能力有两种做法:第一种是把这块 MCU 做强,例如使用新的指令集增加对矢量计算的支持,使用多核做类似 SIMD 的架构等;第二种是依照异构计算的思路,MCU 还是保持简单的控制目的,计算部分则交给专门的加速器 IP 来完成,AI 芯片大部分做的其实就是这样的一个专用人工智能算法加速器 IP。显然,按前一种思路做出来通用性好,第二种思路则计算效率更高。未来预计两种思路会并行存在,平台型的产品会使用第一种通用化思路,而针对某种大规模应用做的定制化产品则会走专用加速器 IP 的思路。然而,因为内存的限制,IoT 终端专用加速器 IP 的设计会和其他领域的专用加速器有所不同。

② 算法与存储器。

众所周知,目前主流的深度神经网络模型的大小通常在几 MB 甚至几百 MB,这给在物联网节点端的部署带来了挑战。物联网节点端出于成本和体积的考量不能加 DRAM,一般用 Flash(同时用于存储操作系统等)作为系统存储器。我们可以考虑用 Flash 存储模型权重信息,但是缓存必须在处理器芯片上完成,因为 Flash 的写入速度比较慢。由于缓存大小一般都是在几百 KB 到 1MB,限制了模型的大小,因此算法必须能把模型做到很小,这也是最近“模型压缩”话题受关注的原因。

如果算法无法把模型做到很小,就需要考虑内存内计算。内存内计算(In-Memory Computing)是一种与传统冯·诺伊曼架构不同的计算方式。冯·诺伊曼架构的做法是把处理器计算单元和存储器分开,在需要的时候处理器从存储器读数据,在处理器处理完数据之后再写回存储器。因此,传统使用冯·诺伊曼架构的专用加速器也需要配合 DRAM 内存使用,使得这样的方案在没法加 DRAM 的物

联网节点端难以部署。内存内计算则是直接在内存内做计算,而无须把数据读取到处理器里,节省了内存存取的额外开销。一块内存内计算的加速器的主体就是一块大容量 SRAM 或者 Flash,然后在内存中再加一些计算电路,从而直接在内存内做计算,理想情况下,能在没有 DRAM 的条件下运行相关算法。

当然,内存内计算也面临一定的挑战。一方面,除了编程模型需要仔细考虑,内存内计算目前的实现方案本质上都是做模拟计算,因此计算精度有限。需要人工智能模型和算法做相应配合,对于低精度计算有很好的支持,避免在低精度计算下损失太多正确率。目前,已经有不少 BNN(Binary Neural Network)出现,即计算的时候只有 1 位精度 0 或者 1,并且仍然能保持合理的分类准确率。

另一方面,目前 IoT 节点终端内存不够的问题除了可以用模型压缩解决外,另一种方式是使用新存储器解决方案来实现高密度片上内存,或者加速片外非易失性存储器的读写速度,并降低读写功耗。因此,边缘计算也将催生新内存器件,例如 MRAM、ReRAM 等。

3. 边缘计算平台架构选型

(1) 英特尔至强 D 平台。

随着 5G 网络等新技术的崛起,终端的数量以及生成、消费的数据量正以指数级别增长,依赖于云端的数据中心处理和分析数据可能会具有较高的时延,并占用大量的带宽。大量终端需要近距离的数据处理能力,并且还要兼顾成本、空间和能耗。全新推出的英特尔至强 D-2100 处理器可以完美满足这些要求,通过集成强大的 Intel Skylake 计算核心、I/O 能力,以及独特的 Intel QAT 加速器和 iWARP RDMA 以太网控制器,它提供了数据中心级别的能力:强大的性能以及极高的可靠性。同时,英特尔至强 D-2100 的热设计功耗维持在 100W 以下,在性能、成本、空间、功耗上取得了平衡。在绝大多数边缘计算场景中,至强 D 系列处理器都可适用,它提供的性能足以应付边缘 AI 和数据分析的工作。

(2) 华为发布面向边缘计算场景的 AI 芯片昇腾 310。

在 HC2018 上,华为正式发布全栈全场景 AI 解决方案。其中,昇腾 310 芯片是面向边缘计算场景的 AI SoC。当前,最典型的几种边缘计算场景是安防、自动驾驶和智能制造。无论哪一种边缘计算场景,都对空间、功耗、算力提出了苛刻的条件。一颗昇腾 310 芯片可以实现高达 16TOPS 的现场算力,支持同时识别包括人、车、障碍物、交通标识在内的 200 个不同的物体;一秒钟内可处理上千张图片。无论是在急速行驶的汽车上,还是在高速运转的生产线上,无论是复杂的科学研究,还是日常的教育活动,昇腾 310 都可以为各行各业提供触手可及的高效算力。昇腾系列 AI 芯片的另一个独特优势是采用了华为开创性的统一、可扩展的架构,

即“达·芬奇架构”，它实现了从低功耗到大算力场景的全覆盖。“达·芬奇架构”能一次开发适用于所有场景的部署、迁移和协同，大大提升了软件开发的效率，加速 AI 在各行业的切实应用。

(3) ARM 的机器学习处理器。

机器学习处理器是专门为移动和相邻市场推出的全新设计，性能为 4.6TOPS，能效为 3TOPS/W。计算能力和内存的进一步优化大大提高了它们在不同网络中的性能，其架构包括用于执行卷积层的固定功能引擎以及用于执行非卷积层和实现选定原语和算子的可编程层引擎。网络控制单元管理网络的整体执行和网络的遍历，DMA 负责将数据移入、移出主内存。板载内存可以对重量和特征图进行中央存储，减少流入外部存储器的流量，从而降低功耗。有了固定功能和可编程引擎，机器学习处理器变得非常强大、高效和灵活，不仅保留了原始性能，还具备多功能性，能够有效运行各种神经网络。

为满足不同的性能需求，从物联网的每秒几 GOP 到服务器的每秒数十 TOP，机器学习处理器采用了全新的可扩展架构。对于物联网或嵌入式应用，该架构的性能可降低至约 2GOPS，而对于 ADAS、5G 或服务型应用，性能可提高至 150TOPS，这些多重配置的效率可达到现有解决方案的数倍。由于与现有的 ARM CPU、GPU 和其他 IP 兼容，且能提供完整的异构系统，该架构还可通过 TensorFlow、TensorFlow Lite、Caffe 和 Caffe2 等常用的机器学习框架来获取。

随着机器学习的工作负载不断增大，计算需求将呈现出多种形式。ARM 已经开始采用拥有不同性能和效率等级的增强型 CPU 和 GPU，运行多种机器学习用例。推出 ARM 机器学习平台的目的在于扩大选择范围，提供异构环境，满足每种用例的选择和灵活性需求，开发出边缘智能系统。

(4) 霍尼韦尔 Mobility Edge 平台。

Mobility Edge 平台是一款统一、通用的移动计算平台解决方案，它以统一的内核支持三个系列共 9 款不同形态与等级的移动数据终端产品，帮助交通运输、仓储物流、医疗及零售等领域的企业提高移动作业效率，并节约成本。Mobility Edge 平台整合了霍尼韦尔在移动终端领域的技术与经验，为移动数据终端产品提供高度的一致性、可重用性和可扩展性，从而实现对整个终端方案快速安全的开发、部署、性能与生命周期管理。霍尼韦尔已推出第一款基于 Mobility Edge 平台架构的产品——Dolphin CT60 移动计算机。霍尼韦尔 Dolphin CT60 移动计算机专为企业移动化设计，具备网络连接性、扫描性能、坚固的产品设计与贴心的使用体验，能够随时随地为关键业务应用和快速数据输入提供实时连接。Mobility Edge 平台可帮助企业加速配置、认证和部署流程，实现投资回报率最大化，降低总拥有成本，并简化高重复性任务。无论是轻型仓库、制造业还是现场服务，霍尼韦尔 Dolphin

CT60 移动计算机都会是企业绝佳的移动化工作伙伴。

(5) NI 发布 IP67 工业控制器支持 IoT 边缘应用。

美国国家仪器公司(National Instruments, NI)发布首款 IP67 级工业控制器 IC-3173T。全新的控制器非常适合在恶劣的环境中作为工业物联网边缘节点使用,包括喷涂制造环境、测试单元和户外环境,而且无需保护外壳。IP67 防护等级可以确保机器在粉尘和潮湿环境下严格按照 IEC 60529 标准稳定运行。NI 正在不断研发可支持时间敏感型网络(TSN)的新产品,工业控制器就属于其中的一部分。TSN 是 IEEE 802.1 以太网标准的演进版,提供了分布式时间同步、低时延和时间关键及网络流量收敛。除使用 TSN 进行控制器之间的通信外,工程师还可使用 NI 基于 TSN 的 CompactDAQ 机箱来集成高度同步的传感器测量。

(6) AWS Greengrass。

AWS Greengrass 立足于 AWS 公司现有的物联网和 Lambda(Serverless 计算)产品,旨在将 AWS 扩展到间歇连接的边缘设备,如图 5-9 所示。借助 AWS Greengrass,开发人员可以从 AWS 管理控制台将 AWS Lambda 函数添加到联网设备,而设备在本地执行代码,以便设备可以响应事件,并近乎实时地执行操作。AWS Greengrass 还包括 AWS 物联网消息传递和同步功能,设备可以在不连回到云的情况下向其他设备发送消息。AWS Greengrass 可以灵活地让设备在必要时依赖云、自行执行任务和相互联系,这一切都在一个无缝的环境中进行。Greengrass 需要至少 1GHz 的计算芯片(ARM 或 x86)、128MB 内存,还有操作系统、消息吞吐量和 AWS Lambda 执行所需的额外资源。Greengrass Core 可以在从 Raspberry Pi 到服务器级设备的多种设备上运行。

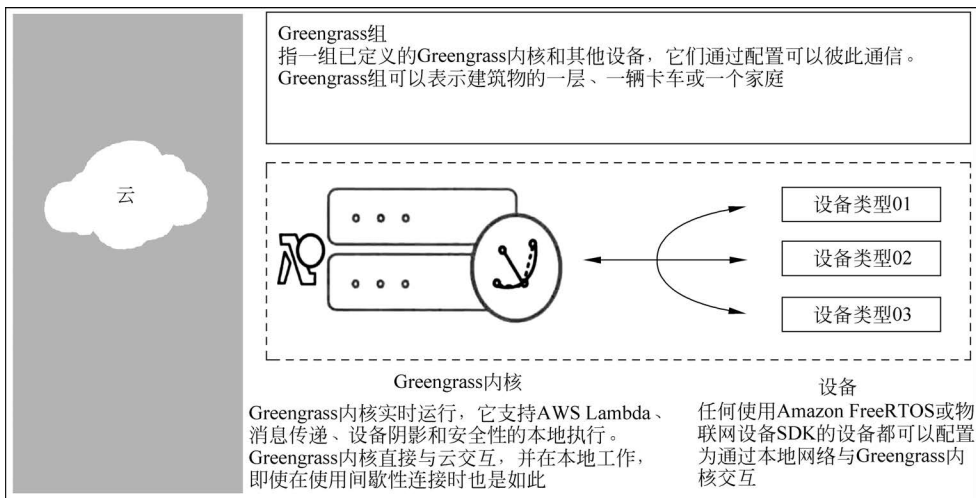


图 5-9 AWS Greengrass

(7) Edge TPU。

谷歌公司推出了能让传感器和其他设备高效处理数据的芯片 Edge TPU,并先投入工业制造领域进行“实验性运行”,其主要用途是检测屏幕的玻璃是否存在制造缺陷。消费电子产品制造商 LG 也将开始对这个芯片进行一系列的测试。据悉,Edge TPU 比训练模型的计算强度要小得多,而且在脱离多台强大计算机相连的基础上进行独立运行计算,效率非常高。

2018 年 7 月,谷歌公司宣布推出两款大规模开发和部署智能联网设备的产品: Edge TPU 和 Cloud IoT Edge。Edge TPU 是一种专用的小型 ASIC 芯片,旨在边缘设备上运行 TensorFlow Lite 机器学习模型。Cloud IoT Edge 是软件堆栈,负责将谷歌公司的云服务扩展到物联网网关和边缘设备上。

如图 5-10 所示,Cloud IoT Edge 有三个主要组件:便于网关级设备(至少有一个 CPU)存储、转换和处理边缘数据,并从中提取信息的 Cloud Dataflow 运行时环境,同时与谷歌云 IoT 平台的其余组件协同操作;Edge IoT Core 运行时环境可将边缘设备安全地连接到云;Cloud ML Engine 运行时环境基于 TensorFlow Lite,使用预先训练的模型执行机器学习推理。

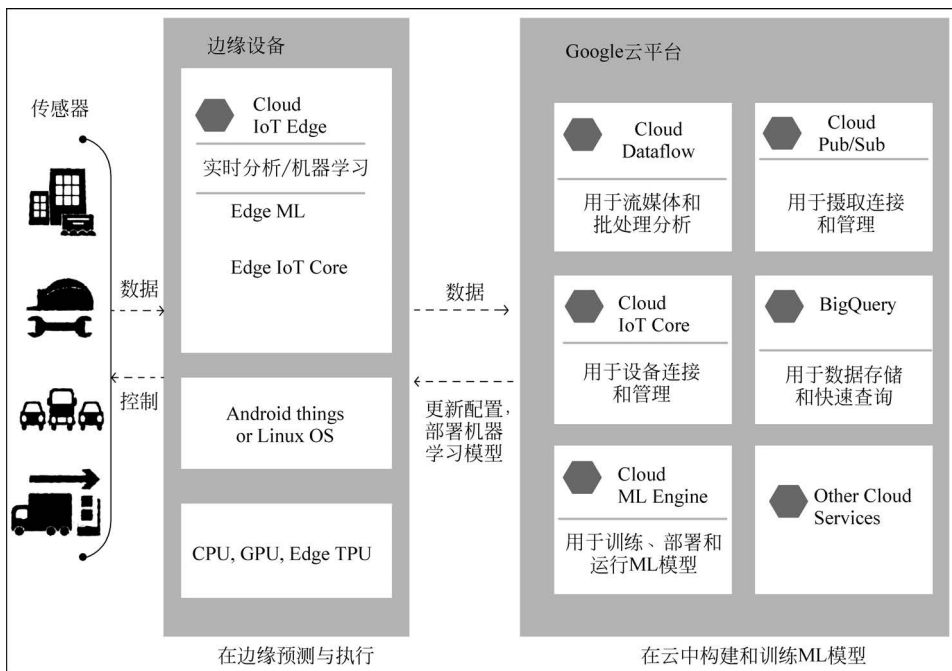


图 5-10 谷歌边缘服务架构

(8) 百度 DuEdge。

百度 DuEdge 借助边缘网络计算的力量,破局云与端之间的数据传输和网络

流量难题,提升业务灵活性和运行效率。使用 DuEdge 服务网站将使访问速度更快,通过智能路由技术解决不同运营商之间的跨网问题;借助缓存减少设备回源请求,释放带宽资源,提升响应速度。DuEdge 将包括云端设备消息收发、函数计算、安全防护在内的一系列能力拓展到边缘节点,使其成为可编程化的智能节点。此外,百度安全 DuEdge 依靠边缘网络计算的分布式计算原理及在物理上更靠近设备端的特性,能够更好地支持本地数据任务的高效处理和运行,减缓了由设备端到云端中枢的网络流量压力。同时,DuEdge 根据用户的实际使用量计费,可有效减少资源占用开支,节省源站的带宽成本和计算成本。而基于百度安全的一站式服务,用户可依照自身需求选择网站可用性监控和 SEO 等多种增值服务,通过按需配置与资源整合,实现产品整体开发和运维成本的下降。

(9) 阿里云 Link IoT Edge。

阿里云推出的 IoT 边缘计算产品 Link IoT Edge,将阿里云在云计算、大数据、人工智能的优势拓宽到更靠近端的边缘计算上,打造“云-边-端”一体化的协同计算体系。借助 Link IoT Edge,开发者能够轻松地将阿里云的边缘计算能力部署在各种智能设备和计算节点上,例如车载中控、工业流水线控制台、路由器等。此外,Link IoT Edge 支持包括 Linux、Windows、Raspberry Pi 等在内的多种环境。

(10) Azure IoT Edge。

微软的 Azure IoT Edge 技术旨在让边缘设备能够实时地处理数据。Moby 容器管理系统也提供了支持,这是 Docker 构建的开源平台,允许微软将容器化和管理功能从 Azure 云扩展到边缘设备。Azure IoT Edge 包含三个部分:IoT Edge 模块、IoT Edge 运行时环境和 IoT 中心。IoT Edge 模块是运行 Azure 服务、第三方服务或自定义代码的容器,它们部署到 IoT Edge 设备上,并在本地执行。IoT Edge 运行时环境在每个 IoT Edge 设备上运行,管理已部署的模块。而 IoT 中心是基于云的界面,用于远程监控和管理 IoT Edge 设备。

微软 Azure 边缘服务架构如图 5-11 所示。

(11) Oracle 与风河。

Oracle 与风河正在携手合作,提供一个集成化的物联网解决方案,将企业应用系统的信息处理能力扩展到边缘设备。通过实现 Oracle IoT Cloud Service 与风河 Wind River Helix Device 的整合,让企业应用系统自动采集边缘设备传感器中的数据并实现情景化,工业企业就可以在设备的网络互连、管理和安全性等方面节省大量的时间,获取更大的效益。这套集成化的解决方案使设备中的数据快速进入企业后端的 ERP、CRM、资产管理和各种特定目标领域的应用系统中,而且为企业客户提供了简洁明了的配置和部署经验,甚至可以直接远程启动设备,并将其中的数据安全地导入企业应用系统。Wind River Helix Device Cloud 是对 Oracle

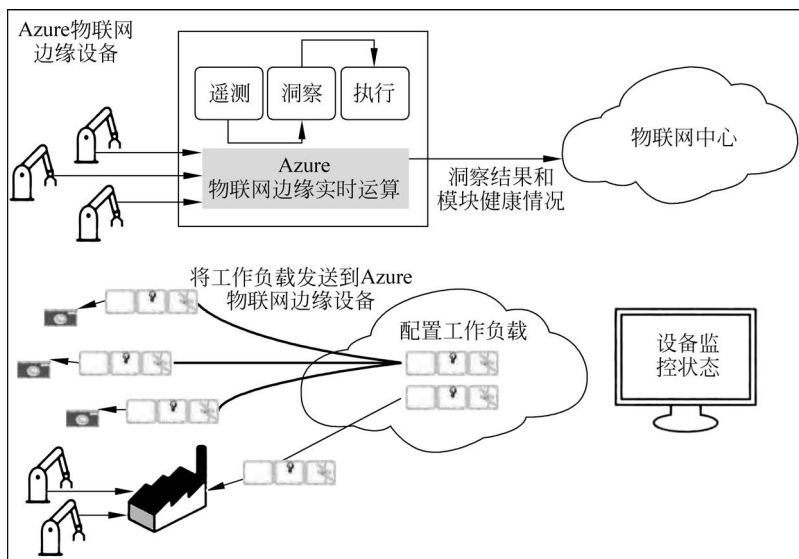


图 5-11 微软 Azure 边缘服务架构

IoT Cloud Service 的扩展,为工业物联网中的设备提供了集中化的设备生命周期管理服务,涵盖安全部署、监视、服务、管理、更新和退役。

4. 机器学习在边缘计算架构中的演进

由于深度学习模型的高准确率与高可靠性,深度学习技术已在计算机视觉、语音识别与自然语言处理领域得到了广泛的应用。

(1) 不同的应用场景,不同的精度需求。

AI 系统通常涉及训练和推断两个过程。训练过程对计算精度、计算量、内存数量、访问内存的带宽和内存管理方法的要求都非常高。而对于推断,更注重速度、能效、安全和硬件成本,模型的准确度和数据精度则可酌情降低。

人工智能工作负载多属于数据密集型,需要大量的存储和各层次存储器间的数据搬移,导致“内存墙”问题非常突出。为了弥补计算单元和存储器之间的差距,学术界和工业界正在以下两个方向进行探索。

① 富内存的处理单元,增加片上存储器的容量并使其更靠近计算单元。

② 创建具备计算能力的存内计算(Process-in-Memory, PIM),直接在存储器内部(或更近)实现计算。

(2) 低精度、可重构的芯片设计是趋势。

关于 AI 芯片的定义并没有一个严格和公认的标准,一般认为面向人工智能应用的芯片都可以称为 AI 芯片。低精度设计是 AI 芯片的一个趋势,在针对推断的

芯片中更加明显。同时,针对特定领域而非特定应用的可重构能力的 AI 芯片,将是未来 AI 芯片设计的一个指导原则。

TensorFlow 和 PyTorch 等 AI 算法开发框架在 AI 应用研发中起着至关重要的作用。通过软件工具构建一个集成化的流程,将 AI 模型的开发和训练、硬件无关和硬件相关的代码优化、自动化指令翻译等功能无缝地结合在一起,将是成功部署的关键。

人工智能芯片技术白皮书指出,从 2015 年开始,AI 芯片的相关研发逐渐成为热点。在云端和终端已经有很多专门为 AI 应用设计的芯片和硬件系统。在云端,通用 GPU,特别是 NVIDIA 系列 GPU 被广泛应用于深度神经网络训练和推理,其最新的 Tesla V100 能够提供 120TFLOPS(每秒 120 万亿次浮点指令)的处理能力。很多公司也开始尝试设计专用芯片,以达到更高的效率,其中最著名的例子是 Google TPU。谷歌公司还通过云服务把 TPU 开放商用,处理能力达到 180TFLOPS,提供 64GB 的高带宽内存(HBM)、2400GB/s 的存储带宽。

不光芯片巨头,很多初创公司也看准了云端芯片市场,如 Graphcore、Cerebras、Wave Computing、寒武纪及比特大陆等公司也加入了竞争行列。

此外,FPGA 也逐渐在云端的推断应用中占有一席之地。目前,FPGA 的主要厂商如 Xilinx、英特尔都推出了专门针对 AI 应用的 FPGA 硬件。亚马逊、微软及阿里云等公司也推出了专门的云端 FPGA 实例来支持 AI 应用。一些初创公司,例如深鉴科技等,也在开发专门支持 FPGA 的 AI 开发工具。

(3) 边缘 AI 计算让传统终端设备焕发青春。

随着人工智能应用生态的爆发,越来越多的 AI 应用开始在端设备上开发和部署。智能手机是目前应用最为广泛的边缘计算设备,包括苹果、华为、高通、联发科和三星在内的手机芯片厂商纷纷开始研发或推出专门适合 AI 应用的芯片产品。

(4) 云+端相互配合,优势互补。

总体来说,云侧 AI 处理主要强调精度、处理能力、内存容量和带宽,同时追求低时延和低功耗;边缘设备中的 AI 处理则主要关注功耗、响应时间、体积、成本和隐私安全等问题。

云和边缘设备在各种 AI 应用中往往是配合工作。最普遍的方式是在云端训练神经网络,然后在云端(由边缘设备采集数据)或者边缘设备上进行推理。

在执行深度学习模型推理的时候,移动端设备将输入数据发送至云端数据中心,云端推理完成后将结果发回移动设备。然而,在这种基于云数据中心的推理方式下,大量的数据通过高时延、带宽波动的广域网传输到远端云数据中心,造成了较大的端到端时延以及移动设备较高的能量消耗。相比面临性能与能耗瓶颈的基于云数据中心的深度学习模型部署方法,更好的方式是结合新兴的边缘计算技术,

充分运用从云端下沉到网络边缘端的计算能力,从而在具有适当计算能力的边缘计算设备上实现低时延与低能耗的深度学习模型推理。通用处理器完全可以胜任推理需求,一般不需要额外的 GPU 或者 FPGA 等专用加速芯片。

边缘侧的负载整合则为人工智能在边缘计算的应用找到了突破口。“物”连上网将产生庞大的数据量,数据将成为新的石油,人工智能为数据采集、分析和增值提供了全新的驱动力,也为整个物联网的发展提供了新动能。虚拟化技术将在不同设备上独立地负载整合资源到统一的高性能计算平台上,实现各个子系统在保持一定独立性的同时还能有效分享计算、存储、网络等资源。边缘侧经过负载整合,产生的节点既是数据的一个汇总节点,同时也是一个控制中心。人工智能可以在节点处采集分析数据,也能在节点处提取洞察做出决策。

如何将人工智能应用到边缘侧?网络优化将是关键性技术之一。英特尔认为,可以通过低比特、剪枝和参数量化进行网络优化。低比特指在不影响最终识别的情况下,通过降低精度来降低存储和计算负荷。剪枝指剪除不必要的计算需求,从而降低计算复杂度。参数量化指可以根据参数的特征做聚类,用相对比较简单符号或数字来表述,从而降低人工智能对存储的需求。

第4节 边缘计算相关网络

1. 通信网络

边缘计算的研究主要针对与传统通信运营商和云服务商业关系密切的 5G 移动网络和部分固网接入业务,下边主要介绍移动网络和固定网络。

移动网络一般可以分为接入网、承载网和核心网等。无线接入网也就是常说的 RAN(Radio Access Network),以 4G 为例,主要包括天线、馈线、无限远端单元(Radio Remote Unit,RRU)和基带处理单元(Base Band Unit,BBU)等功能。天线负责把无线终端的无线信号转换成有线信号,通过馈线进入 RRU 进行射频信号和中频信号的转换,之后由 BBU 完成包括编码、复用、调制和扩频等信号处理,然后通过背传(Backhaul)网络接入移动承载网。

承载网早期主要使用时分复用(Time-Division Multiplexing,TDM)的 T1/E1 技术建立准同步数字系统(Plesiochronous Digital Hierarchy,PDH)网络来支持电话业务,由于语音的带宽是固定的 64kHz,正好对应 TDM 的一个时隙,所以 TDM 特别适合语音业务。但是,由于 PDH 只是准同步,不同级汇聚的设备很复杂,缺乏统一标准,而且维护困难。随着光传输技术的发展,诞生了以光纤为介质的完全同步的光网络 SDH,具有标准统一、支持大容量传输、具备电信级自愈能力等优势,所以一度统一了传输网。然而,随着数据业务开始兴起,需要承载更多类型的接入

业务,特别是净荷大小不固定的 IP 数据。于是,在 SDH 的基础上衍生出了多业务传输平台(Multi-Service Transmission Platform,MSTP),即把多业务通过 SDH 传输,实质还是基于 TDM 的网络,虽然可以支持数据业务,但是带宽利用率和灵活性大打折扣。

随着后续技术的演进,产生了分组传送网(Packet Transport Network,PTN),PTN 主要借鉴了多协议标签交换(Multi-Protocol Label Switching,MPLS)的数据交换技术,又增加了管理维护(Operation Administration and Maintenance,OAM)等功能,形成了 MPLS-TP(Transport Profile)协议。通过给数据包打上标签,为报文建立虚拟的转发通道,每台设备对报文只需根据标签进行交换即可。与固定时隙的 SDH 不同,PTN 的传送单元是大小可变的 IP 报文,可以灵活地支持蓬勃发展的数据业务。另外,随着软交换技术的兴起,例如 VoIP、Volte 等语音业务也可以通过数据包传输,TDM 网络越来越边缘化。

随着技术的演进,相比于 2G 和 3G 网络,LTE 网络架构将原控制平面的部分功能和用户数据流转传送的部分功能向基站转移。这要求回传网络架构必须具备智能路由功能,于是希望把所有移动业务都通过三层 IP 传输的 IP RAN 应运而生。相比 PTN 提供二层以太网业务,IP RAN 以 IP 和 MPLS 技术为基础,直接承载 L3 的 IP 业务,侧重于用路由器和交换机构建整个网络,支持丰富的路由协议,业务调度灵活,开放性好。但是,由于 L3 层网络复杂,硬件设备成本高,安全和管控也有待提高。传统无线网络的业务和数据处理都是在核心网进行的,包括用户认证和授权、会话建立和管理、数据路由和转发等。

固网可以分为接入网、汇聚网和城域网。与移动接入网络不同,早期家庭客户的接入侧使用有线连接,例如电话线和同轴电缆的各种 xDSL(Digital Subscriber Loop)。国内大城市已经普及无源光网络 PON(Passive Optical Network),实现光纤大带宽的接入;接入机房一般通过 OLT、BRAS、交换机和路由器等设备实现业务数据处理,然后一级一级汇聚接入城域网。政企用户对于带宽可靠性有特殊要求,可以在运营商处申请专线业务,通过业务路由器或 PTN 网络接入城域核心网络。在城域网之上,通过基于波分复用的光传送网 OTN(Optical Transport Network)在全国互联形成骨干网。

(1) 数据中心网络。

在早期的大型数据center里,网络通常包括接入层、汇聚层和核心层的三层架构,如图 5-12 所示。接入层交换机通常位于机架顶部,也称为 ToR(Top of Rack),连接服务器和存储设备等。汇聚交换机把接入层的数据汇聚起来,同时提供防火墙、入侵检测、网络分析等服务。核心交换机连接多个汇聚交换机,并为数据包提供高速转发。

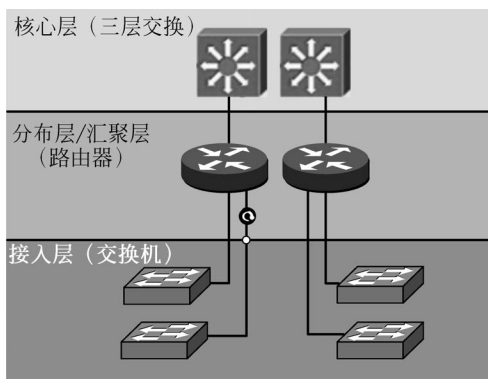


图 5-12 传统三层网络架构

由于核心交换机价格昂贵，所以在实际部署中，接入、汇聚和核心交换机之间有一定的超占比，即：接入层交换机总带宽 $>$ 汇聚层总带宽 $>$ 核心层总带宽。以太网本身是一种尽力而为的网络设计，允许丢包，而且早期的数据中心以南北向流量居多，即数据请求主要来自外部的设备，由数据中心的部分服务器或者存储设备处理，所以这种网络设计是满足需求的。

随着技术的发展，数据的内容和形式发生了变化。分层的软件架构导致软件功能的解耦合模块化等，使一个系统的软件功能通常分布在多个 VM 或者容器中；新应用的出现，如分布式计算、大数据、人工智能等，使一项业务需要多台甚至上千台服务器合作完成，导致数据中心内的东西向流量急剧增加。对于增加的东西向流量如果都在同一个接入层或者汇聚层交换机中，那么数据转发能力尚可。但是，如果东西向流量需要占用大量宝贵的核心交换机转发带宽，由于超占比的存在，整个网络的数据转发能力会明显下降，并且影响正常的南北向流量。如果单纯依靠提高超占比来提升网络性能，势必要大幅增加昂贵的核心交换机的带宽，网络设备的支出会大幅上升。

为了应对大量涌现的东西向流量，新的数据中心主要采用的是 Spine/Leaf 网络架构，如图 5-13 所示。Leaf 交换机相当于传统三层架构中的接入交换机，作为 ToR(Top of Rack)直接连接物理服务器，Spine 交换机相当于核心交换机。Spine 交换机和 Leaf 交换机之间通过 ECMP(Equal Cost Multi Path)动态选择多条路径。每个 Leaf 交换机的上行链路数等于 Spine 交换机的数量，从而缓解了超占比的问题，能够更好地处理增加的东西向流量。Spine/Leaf 网络架构本质上是使用中型的交换机替代传统昂贵的核心交换机，成本低，架构更加扁平化，适应云计算的发展趋势，有利于计算等硬件资源的虚拟化和容器化。随着网络规模和东西向流量的增加，可以增加 Spine 交换机的数量，扩展性好；同时，更多的 Spine 交换机可以更好地实现容错能力。Spine/Leaf 架构最具有代表性的就是 Facebook 公开

的数据中心架构。

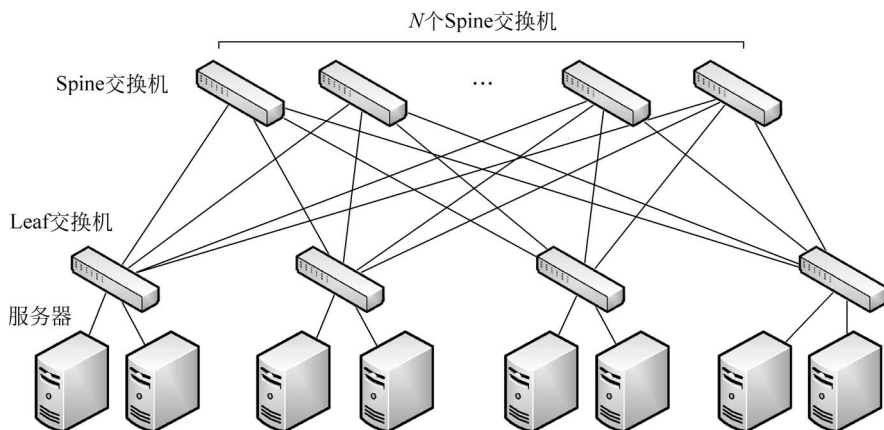


图 5-13 Spine/Leaf 网络架构

在 Spine/Leaf 架构中, L2/L3 层的分界点在 Leaf 交换机上, Leaf 交换机之上是三层网络。这样设计能分隔 L2 广播域, 适用的网络规模更大, 但也导致服务器的部署和 VM 的迁移局限在 L2 网络内。一种解决方案是 VXLAN, 即通过在物理 L2 网络之上构建一个虚拟 L3 网络来解决上述问题。VXLAN 定义了 8 字节的 VXLAN Header, 其中的 24bit 用来标识不同的二层网络, 这样总共可以标识 1600 多万个不同的二层网络。这个 Header 不属于 L2 MAC 包, 而是借用了 L3 的 UDP 包头, 将 Ethernet Frame 封装在 UDP 包里面, 物理网络的二层边界还存在。但是, VM 的创建迁移的网络数据在三层网络传输, 从而跨越了物理二层网络的限制。VXLAN 的优点是突破了 L2/L3 层网络的限制, 缺点是实际上模糊了 L2 和 L3 层的网络分层, 这点是不符合典型的网络分层架构理念的。

(2) SDN 和 NFV。

SDN 和 NFV 是目前网络发展的热门技术, 在使用中经常被混淆。SDN (Software Defined Network) 指软件定义网络, 其背景是随着网络规模的急剧增加, 传统的网络设备交换机和路由器等设备通过专用硬件实现, 数据交换面和控制管理面集成在一起, 主要通过命令行接口控制, 配置部署复杂, 对运维人员要求高。一旦部署, 后续由于业务需求需要改变网络拓扑非常麻烦, 发生问题也不容易定位, 维护管理不方便。SDN 的核心思想是把网络的控制面从数据面剥离, 主要分为三层, 网络基础设施层由专用硬件的交换机负责数据面的转发, 简化硬件交换机的设计复杂度并提高数据带宽, 从而降低网络成本。另外, 由于所有硬件交换机的功能一致, 网络可扩展性非常好。控制层具体负责转发控制策略, 可以集中运行在通用的服务器上, 网络部署只需要改变控制面即可, 简单方便。应用层主要通过控制层提供的开放的 API 接口, 灵活支持各种不同的业务和应用。一般把应用层和

控制层的接口叫作北向接口,控制层和网络基础设施层的接口称为南向接口,目前 OpenFlow 已经成为事实上的南向接口的标准,但是北向接口的标准目前还没有统一。

NFV(Network Function Virtualization)是指网络功能虚拟化。首先,有必要提一下网络虚拟化(Network Virtualization)。网络虚拟化泛指在一个物理的网络上构造多个虚拟的网络。典型的有 VLAN,在一个物理 L2 网络上通过 Tag 隔离出多个虚拟的 VLAN,来实现数据的隔离和保护,VxLAN 也是一种网络虚拟化。但是 NFV 主要指随着通用处理器处理能力的增强和虚拟化技术的发展,把之前运行在网关、交换机和路由器等专用硬件设备上的网络功能通过软件在通用处理器上实现。同时,利用虚拟化将硬件资源池化,根据需求灵活配置资源,从而实现基于实际业务的灵活部署,并降低昂贵的专用网络设备成本。NFVI(NFV Infrastructure)指的是 NFV 的硬件基础设备,其中的两个重要技术是网卡和交换机的虚拟化。

NFV 和 SDN 侧重点不同,但是也有交集,例如 SDN 的控制和数据分离的策略也可以应用在 NFV 的设计中。SDN 和 NFV 是未来网络灵活性和开放性的主要推动力量。

(3) 网卡虚拟化 VMDq 和 SR-IOV。

随着通用处理器核数的增加和虚拟技术的流行,一台物理主机最多可以支持数十个 CPU 核,每个核可以运行一台虚拟机。但是,一台主机实际的 I/O 设备,特别是物理网卡数量有限,必须通过虚拟化更多的 I/O 接口来支持。由于纯软件的虚拟化方案对 CPU 的资源消耗太大,实际上更多的是通过部分硬件来协助实现 I/O 的虚拟化,特别是通过物理网卡虚拟化更多的逻辑网口来支持实现。其中 VMDq(Virtual Machine Device Queues)和 SR-IOV(Single-Root Input/Output Virtualization)是针对 I/O 设备虚拟化提出来的两个重要技术。

VMDq 技术使得网络适配器具备了数据分类功能,从而使虚拟化主机的 I/O 访问性能接近线速吞吐量,并降低了 CPU 占用率。虚拟机管理器 VMM 在网络适配器中为每个虚拟机主机分配一个独立的队列,使得该虚拟机的数据流量可直接发送到指定队列上,虚拟交换机无须进行排序和路由操作。但 VMM 和虚拟交换机仍然需要将数据流量在网络适配器和虚拟机之间进行复制。

SR-IOV 技术为每个 VM 提供无须软件模拟即可直接访问网络适配器的能力,实现了接近物理宿主机的 I/O 性能。SR-IOV 允许 VM 之间高效分享 PCIe 设备,通过在网络适配器内创建出不同虚拟功能 VF 的方式,呈现给虚拟机独立的网卡设备,VF 和 VM 之间通过直接内存访问 DMA 进行高速数据传输。SR-IOV 的性能优于 VMDq,但需要硬件(包括网络适配器和主板等)的支持。具有 SR-IOV

功能的设备具有以下优点。

① 提供了一个共享任何特定 I/O 设备容量,实现虚拟系统资源有效利用的标准方法。

② 在一个物理服务器上,每个虚拟机接近本地的性能。

③ 在同一个物理服务器上,虚拟机之间的数据保护。

④ 物理服务器之间更平滑的虚拟机迁移,由此实现 I/O 环境的动态配置。

(4) 虚拟交换机 OVS 和 DPDK。

数据中心主要通过虚拟交换机,即 OpenvSwitch(OVS)来实现交换功能的虚拟化。OVS 通过软件扩展,支持网络的自动化运维,具有标准的管理接口和协议,可以在多个物理服务器中运行。与传统的物理交换机相比,虚拟交换机具备配置灵活的优点。一台服务器可灵活支持数十台甚至上百台虚拟交换机。

① 交换虚拟化软件主要具备以下能力。

数据分组的快速导入及导出。

不同物理宿主机和虚拟机之间的高速数据转发。

OVS 的主要瓶颈在于传统 CPU 的 I/O 带宽太小,为了提升转发性能,可将英特尔开发的数据平面开发套件 DPDK(Data Plane Development Kit)和 OVS 结合起来,通过提供高性能数据分组处理库和用户空间的驱动程序,替代 Linux 下的网络数据分组处理功能,实现转发路径的优化。

② 与传统的数据包处理相比,DPDK 具有以下特点。

轮询:降低数据处理时上下文切换的开销。

用户态驱动:减小多余的内存复制和系统调用,加快迭代优化速度。

亲和性与独占:在特定核上绑定特定任务,减小线程核间的切换开销,提高缓存命中率。

降低访存开销:利用内存大页技术降低 TLB 丢失率,利用内存 Lock Step 等多通道技术降低内存时延,并提高内存有效带宽。

软件调优:实现缓存的行对齐、数据预取、批量操作等。

结合 DPDK 的虚拟交换机已经接近中端物理交换机的性能,目前已普遍支持 10Gb/s 的交换速度。

2. 边缘计算网络需求

边缘计算的边缘主要是对应于数据中心,边缘计算的网路可以理解为数据中心之外的网路,即移动网路和政企家庭的固网接入等。

随着高清 2K、4K 甚至 8K 视频点播,虚拟现实,增强现实,人工智能和车联网等高带宽、低时延应用的出现,已有的 4G 网路已经无法满足新兴业务的需求。通

过对移动网络的研究,国际电信联盟 ITU M. 2083 定义了 5G 的三大应用场景,如图 5-14 所示。

① 增强型移动宽带,随着用户对多媒体内容、服务和数据的访问持续增长,对移动宽带的要求将快速增长。

② 大规模机器类型通信,对应海量设备连接场景,数据量和时延性要求都不高,但是需要低成本和超高的电池续航能力,主要包括智能电表、智慧城市等。

③ 超可靠低时延通信,对吞吐量、时延和可靠性等要求十分严格,主要包括远程医疗、自动驾驶等。

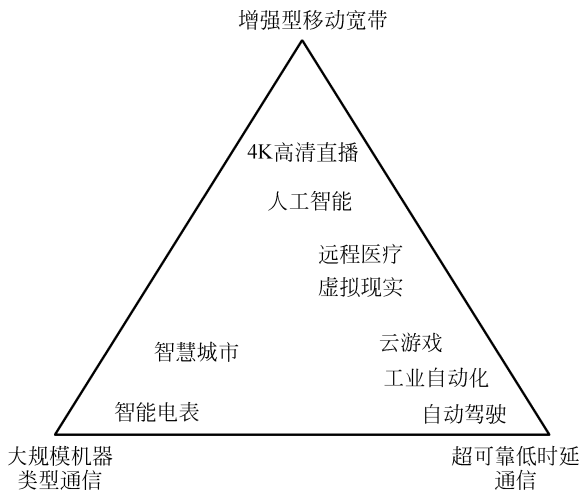


图 5-14 5G 典型应用

欧盟启动了 METIS(Mobile and Wireless Communications Enablers for the Twenty-twenty Information Society)的 5G 研究,并于 2015 年发布“ICT-317669 scenarios, requirements and KPIs for 5G mobile and wireless system with recommendations for future investigations”,其中定义了 5 种典型应用和 12 种测试用例。几个典型例子如下:

第一,以虚拟现实办公为例,要求办公区 95% 的地方需要达到 1Gb/s 以上的带宽,20% 以上的区域甚至要达到 5Gb/s。

第二,以智能电网的远程保护为例,1521 字节信息的传送必须满足小于 8ms 的时延和 99.999% 的可靠性。

第三,对于超过万人的体育场馆,必须保证每人 20Mb/s 以上的带宽和小于 5ms 的空口时延。

第四,对于自动驾驶场景,端到端的时延必须小于 5ms,并保证 99.999% 的可靠性。

我国成立了 IMT-2020(5G)推进组,并发布了《5G 愿景与需求》白皮书,其中提到,移动互联网和物联网是未来移动通信的两大重要推动力,也为 5G 提供了广阔的市场前景。

虽然 5G 定义了大规模机器类型通信 mMTC,希望能把物联网统一到 5G 网络,但 5G 实际上是下一代的移动网络,而智能家电、智能台灯、智能电表等物联网设备实际上是很少移动的。而且 5G 的带宽高、功耗高,对于极少数据量的低功耗物联网终端设备,即使配置一个 SIM 卡也会导致成本增加。除了 5G 移动网络,物联网还包含各种远距离无线传输技术,如窄带物联网(Narrow Band Internet of Things, NB-IoT)、LoRa(Long Range)、Wi-Fi、蓝牙、ZigBee 等近距离无线传输技术。低功耗、低速率的广域网传输技术适合远程设备运行状态的数据传输、工业智能设备及终端的数据传输等。高功耗、高速率的近距离传输技术,适合智能家居、可穿戴设备以及 M2M 之间的连接及数据传输。低功耗、低速率的近距离传输技术,适合局域网设备的灵活组网应用,如热点共享等。但是,无论使用何种无线接入技术,除了个别巨头公司不计成本自行组网,绝大部分物联网的海量数据汇聚后还是会通过无线或者有线接入运营商的已有网络传输,其海量的设备数量和复杂的业务类型也对现有边缘网络提出了很大的挑战。

随着边缘计算和业务的下沉,原来处于安全的数据中心和核心网的业务,数据和算法等需要下沉到边缘接入机房,出于隐私考虑,安全要求越来越高。另外,对于高铁、高速汽车等超高速移动场景,如何保证移动连接和业务也是一大挑战。

另外,随着光纤入户带来的带宽增加,高清视频点播等大带宽消耗应用的快速增长,企业业务向云迁移带来的大量数据迁移等,进一步对现有网络带宽带来挑战,各种高可靠性业务也对网络质量提出了更高的要求。移动通信系统的现实网络逐步形成了包含多种无线制式、频谱利用和覆盖范围的复杂现状,多种接入技术长期共存成为突出特征。在 5G 时代,同一运营商拥有多张不同制式网络的状况将长期存在,多制式网络将至少包括 4G、5G 以及 WLAN 等。如何高效地运行和维护多张不同制式的网络,不断降低运维成本,提高竞争力,是每个运营商都要面临的问题。

综上,5G 移动网络、物联网、光纤大带宽固网接入和云业务的兴起等,对边缘网络的主要需求包括超高速带宽、毫秒级超低端到端时延、超高密度链接、超高速移动链接、多网络融合和安全隐私保障。

3. 边缘计算网络发展趋势

从网络运营商的角度来看,如何在保证投资收益的前提下更好地满足边缘网络的主要需求成为一项巨大挑战。实际上,没有任何一个网络可以支持所有的功

能,以上总结的一些核心需求之间也是互相矛盾的。例如,追求高带宽必然会牺牲时延和可靠性,就好比固定车道的高速公路,大车流必然容易导致意外事故的发生,从而降低数据包的可靠性和时延;而可靠性和低时延的追求,希望尽量减少高速路上不必要的车流量,最好平时没有流量,一旦有数据包传送时,传送过程不会被时延和损坏,从而提高了可靠性和低时延。但是从成本考虑,不可能为不同的业务建造不同的物理网络。

那么如何通过一张物理网络实现多样的网络需求呢?首先,应该通过合理利用边缘业务平台的存储、计算、网络、加速等资源,将部分关键业务应用下沉到接入网络边缘,极大地降低网络时延。然后,把大部分的本地流量终结到边缘侧,仅把必需的经过压缩的数据传送到核心网和数据中心,也可以极大地缓解核心网和骨干网的带宽压力,提升客户的使用体验。另外,毫米波、大规模多天线技术、新型多载波技术、高阶编码调制约束、全双工、小区加密技术、波束赋形和设备到设备等新的无线技术可以极大地改善低时延,提高带宽利用率,支持超高速移动连接。通过引入新的 25G、50G 甚至 100G 大带宽网络提升管道带宽,降低运营成本。

除此以外,还可以通过全新的无线接入技术、CU/DU 分离部署、端到端网络切片、数控分离的核心网、计算能力下沉和软件定义广域网(SD-WAN)来满足网络需求。

(1) CU 和 DU 分离部署。

4G 无线接入网主要通过基带处理单元(Base Band Unit, BBU)+射频拉远单元(Radio Remote Unit, RRU)实现,BBU 主要采用专用芯片实现,网络拓扑固定,扩展性差,难以实现虚拟化和云化。为了应对高带宽和海量的设备连接,以及由此引起的数据业务类型的复杂化,5G 接入网也引入了全新的有源天线单元(Active Antenna Unit, AAU)、集中单元(Centralized Unit, CU)和分布单元(Distributed Unit, DU)灵活部署的方案。BBU 的部分物理层处理功能与原 RRU 及无源天线合并为 AAU,原 BBU 的非实时部分将分割出来,重新定义为 CU,负责处理非实时协议和服务。BBU 的剩余功能重新定义为 DU,负责处理物理层协议和实时服务,CU 通过交换网络连接远端的 DU。基于 5G RAN 架构的变化,5G 承载网由前传(Fronthaul)、中传(Middlehaul)和后传(Backhaul)三部分构成。其中 AAU 和 DU 之间的数据传输是前传,DU 和 CU 之间是中传,CU 和核心网之间是回传。

无线接入网 CU-DU 分离架构的优点在于能够实现数据集中处理和小区间协作;实现密集组网下的集中控制,获得资源的池化增益,有利于 NFV/SDN 的实现,以及满足运营商某些 5G 场景的部署需求。通过不同的 CU 和 DU 的部署模式,支持不同的网络应用场景,当传送网资源充足时,可集中化部署 DU 功能单元,实现物理层协作化技术,而在传送网资源不足时,也可分布式部署 DU 处理单元。

CU 功能的存在实现了原属 BBU 的部分功能的集中,既兼容了完全的集中化部署,也支持分布式的 DU 部署。可在最大化保证协作化能力的同时,兼容不同的传送网能力。

从 4G 单点 BBU 到 5G CU/DU 的两级架构如图 5-15 所示。

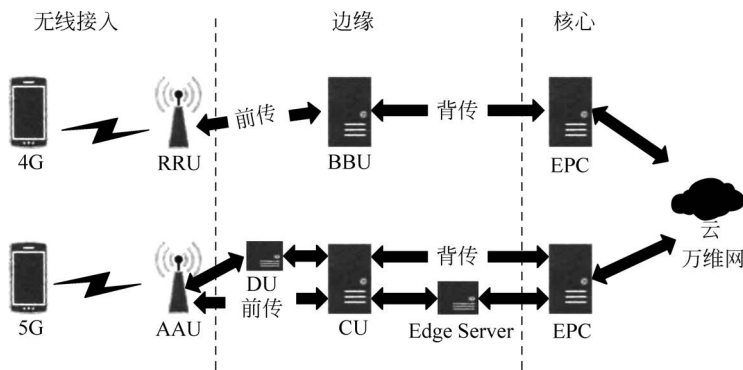


图 5-15 从 4G 单点 BBU 到 5G CU/DU 的两级架构

BBU—基带处理单元；RRU—射频拉远单元；EPC—核心网；

AAU—有源天线单元；DU—分布单元；CU—集中单元

在具体的实现方案上, CU 设备采用通用平台实现, 不仅可支持无线网功能, 也具备支持核心网和边缘应用的能力。 DU 设备可采用专用设备平台或“通用+专用”混合平台实现, 通过引入 NFV 和 SDN 架构, 在管理编排器(MANO)的统一管理和编排下, 实现包括 CU 和 DU 在内的网络配置和资源调度能力, 满足运营商业务部署需求。

以中国联通通信云化网络三层 DC 和一层综合接入机房的四层架构为例, 接入机房的部署和 4G 类似, 使用 CU/DU 合设方案, CU 管理和其同框的 DU 通过机框背板通信, 时延在 2~5ms。边缘 DC 的 CU 大约管理几十个到上百个 DU, 时延小于 10ms; 本地 DC 的 CU 可能管理几百个 DU, 时延小于 20ms; 区域 DC 的 CU 可能管理上千个 DU, 时延小于 50ms。 CU 部署在核心机房, 可以直接运行在使用标准的数据中心服务器上, 便于虚拟化和池化。本地 DC 机房环境如果不能支持使用标准 19in 服务器, 可以采用新兴的边缘服务器支持 CU 的虚拟化和池化, 但是范围会比较小。区域 DC 机房条件恶劣, 前期可以借用已有的 OLT 和 BRAS 等机架设备的空间, 采用类似中兴通讯股份有限公司发布的轻量云内置刀片设计支持 CU 的功能, 后续逐步引入 CU/DU 等合设部署。

对时延极其敏感的 ULRRRC 业务(低时延、高可靠业务), 例如车联网 5ms 的时延、电网远程保护的 8ms 时延, CUR 要部署在接入机房才能满足时延要求。对增强移动宽带的业务而言, 为了保证 5G 的无线性能和时延要求, CU 部署在接入机

房、边缘 DC 和本地 DC 是能满足时延要求的,但是部署在本地 DC 会导致大量的数据占用汇聚网络的带宽。所以,后期随着高带宽、低时延业务的快速发展,需要平衡边缘侧在网络和计算成本上的投入,例如改造边缘 DC 增加计算能力的下沉,或升级边缘和本地 DC 之间的网络带宽,具体需要每家运营商根据已有的基础光纤、网络 and 机房情况进行评估。

联通通信云网络架构如图 5-16 所示。

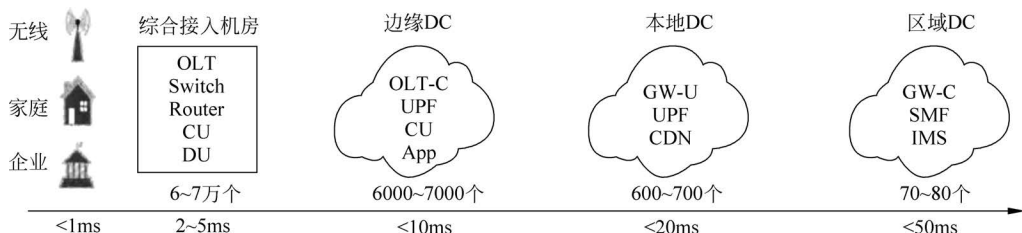


图 5-16 联通通信云网络架构

(2) 端到端网络切片。

5G 承载的业务种类繁多,业务特征各不相同,如果使用一张物理网络支撑三大业务场景,并满足网络的速率、移动性、安全性、时延、可靠性、计费等不同需求,将是巨大的挑战。而端到端的网络切片是 5G 网络支撑多业务的基础,也是 5G 网络架构演进的关键技术。

根据不同的业务需求,端到端的网络切片技术使运营商能够在同一个硬件基础设施上切分出多个虚拟的端到端网络,类似于 VLAN,不同的虚拟网络天然隔离,为不同类型的 5G 场景提供不同的通信业务和网络能力。每一个网络切片在物理上源自统一的网络基础设施,而且不同的网络切片可以按需建立,用完解除,弹性敏捷,大大降低了运营商运营多个不同业务类型的建网成本。网络切片本质上是一种按需组网的方式,可以让运营商在统一的基础设施上切出多个虚拟的端到端网络,每个网络切片从无线接入网到承载网再到核心网在逻辑上隔离,适配各种类型的业务应用。5G 主流的三大应用场景就是根据网络对用户数、QoS、带宽等不同的要求定义的不同通信服务类型,可以对应三个切片。eMBB、uRLLC 以及 mMTC 三个独立的切片承载于同一套底层物理设施之上。在 eMBB 切片中,对带宽有很高的要求,可以在城域机房的移动云引擎中部署缓存服务器,使业务更贴近用户,降低骨干网的带宽需求,提高用户体验。在 uRLLC 切片中,自动驾驶、辅助驾驶、远程控制等场景对网络有着极为苛刻的时延要求,因此需要将 RAN 的实时处理及非实时处理功能单元部署在更靠近用户的站点侧,并在 CO 机房的移动云引擎中部署相应的服务器(V2X Server)及业务网关,城域及区域数据中心中只部署控制面相关的功能。在 mMTC 切片中,对于大多数 MTC 场景,网络中交互的

数据量较小,信令交互的频率也较低,因此可以将移动云引擎部署在城域数据中心,将其他功能以及应用服务器部署在区域数据中心,释放 CO 机房的资源,降低开销。当然,也可以定义更多数量类型的网络切片,但是更多的切片数量会带来设计管理的复杂化,所以切片数量需要根据网络应用和基础设施部署来确定,并不是越多越好。

网络切片的基础首先就是网络本身要能动态支持不同的切分粒度,以 1Gb/s 带宽网卡或者交换机为例,可以很容易切分成 10 个 100Mb/s,甚至 100 个 10Mb/s。但是,传统的基于专用硬件平台的网络通信设备软件和硬件紧耦合,传统物理千兆网卡支持的是一个完整的 1Gb/s 开端,是无法动态切分成不同粒度的。要支持网络切片,最基本的两项技术是网络功能虚拟化和软件定义网络。

NFV 把所有的硬件抽象为计算、存储和网络三类资源并进行统一管理分配,给不同的切片不同大小的资源,且使其完全隔离互不干扰。SDN 实现网络的控制和转发分离,并在逻辑上统一管理和灵活切割。以核心网为例,NFV 从传统网元设备中分解出软硬件的部分。硬件由通用服务器统一部署,软件部分由不同的 NF(网络功能)承担,从而满足灵活组装业务的需求,于是切分粒度的逻辑概念就变成了对资源的重组。重组是根据 SLA(服务等级协议)为特定的通信服务类型选择需要的虚拟资源和物理资源。SLA 包括用户数、QoS、带宽等参数,不同的 SLA 定义了不同的通信服务类型。

一个网络切片包括无线子切片、承载子切片和核心网子切片。核心网目前基于通用处理器的虚拟化程度已经比较高,所以网络切片的实现相对容易。基于服务化架构(Service Based Architecture,SBA),以前所有的网元都被打散,重构为一个实现基本功能集合的微服务,再由这些微服务像搭积木一样按需拼装成网络切片。汇聚层的承载网切片主要基于 SDN 的统一管理,把网络资源抽象成资源池来进行灵活的分配,从而切割成网络切片。网络切片的最大难点在于接入网的虚拟化和云化,5G 引入的 CU 也更易于支持虚拟化。目前,设备商基于 5G 的接入网方案已经基于通用处理器设计,具备相当的虚拟化能力。但是,DU 和 AAU 目前主要通过专用硬件支持,虚拟化的支持有限。英特尔提出的纯软件方案 FlexRAN 是很好的接入网云化方案,但是目前性能还不如专用硬件,适应的场景比较有限。具体部署还要看后续 5G 新方案的研发和实现。

(3) 数控分离的核心网。

传统 4G 核心网架构如图 5-17 所示。移动管理实体(Mobile Management Entity,MME)负责网络连通性的管理,主要包括用户终端的认证和授权、会话建立以及移动性管理。归属用户服务器(Home Subscriber Server,HSS)作为用户数据集,为 MME 提供用户相关的数据,以此来协助 MME 的管理工作。服务网关

(Serving Gateway, SGW)负责数据包路由和转发,将接收到的用户数据转发给指定的 PDN 网关(PDN Gateway, PGW),并将返回的数据交付给基站侧 eNB (Evolved Node B)。PGW 负责为接入的用户分配 IP 地址以及管理用户平面 QoS,并且是 PDN 网络的进入点。MME 仅承担控制平面功能,但是 SGW 和 PGW 既承担大部分用户平面功能,又承担一部分控制平面功能,这就使得用户平面和控制平面严重耦合,不利于网络可编程化。而且,由于历史兼容原因,很多设备运行在专用硬件平台上,软件和硬件紧耦合,网络部署和维护成本高,对四新业务的支持也有限,可扩展性和通用性都较差。

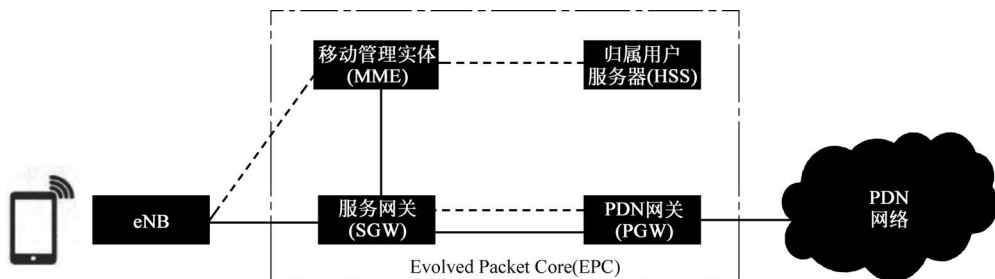


图 5-17 传统 4G 核心网架构

在 5G 时代,众多对时延要求较高的业务需要将网关下移到城域甚至是中心机房,会导致网关节点数量增长 20~30 倍。如果保持现有的网关架构,必然会由于网关业务配置的复杂性,显著增加运营商网络的 CAPEX 和 OPEX。同时,如果控制面订阅了位置、RAT 信息的时间上报,则会产生较多信令在站点、分布式网关、网络控制面功能之间迂回。数量众多的分布式网关会对集中化的控制面网元带来明显的接口链路负担和切换更新信令负荷。

因此,需要对网关的控制面及用户面进行分离,通过剥离网关复杂的控制逻辑,将控制逻辑功能集成到融合的控制面。不仅可以有效降低分布式部署带来的成本压力,还能化解信令路由迂回和接口负担的问题。此外,控制面和用户面分离还能够支持转发面和控制面独立伸缩,进一步提升了网络架构的弹性和灵活性,方便控制逻辑集中,更加容易定制网络切片,服务多样化的行业应用:控制转发演进解耦,避免控制面的演进带来转发面的频繁升级。控制面和用户面分离首先要做到功能轻量化,剥离复杂控制逻辑功能并集中到模块化的控制面中。要对保留的核心基本转发面功能进行建模,定义通用转发面模型和对象化的接口,以实现转发面可编程,支持良好的扩展性。

相比于传统 4G EPC 核心网,5G 核心网采用原生适配云平台的设计思路,基于服务的架构和功能设计提供更泛在的接入、更灵活的控制和转发以及更友好的能力开放。引入了 SDN 的思想以实现控制面和转发面的进一步分离,5G 核心网

与 NFV 基础设施结合,为普通消费者、应用提供商和垂直行业需求方提供网络切片、边缘计算等新型业务能力。

5G 核心网实现了网络功能模块化以及控制功能与转发功能的完全分离。控制面可以集中部署,对转发资源进行全局调度;用户面则可按需集中或分布式灵活部署,当用户面下沉靠近网络边缘部署时,可实现本地流量分流,支持端到端的毫秒级时延。由于未来业务高带宽、低时延、本地管理的发展需求,3GPP、4G CUPS 与 5G New Core 移动网将控制面和转发面分离,使网络架构扁平化。转发面网关可下沉到无线侧,分布式按需部署,由控制平面集中调度。网关锚点与边缘计算技术结合,实现端到端低时延、高带宽,均衡负载海量业务,从而在根本上解决传统移动网络竖井化单一业务流向造成的传输问题,以及核心网负荷过重、时延瓶颈问题。

伴随控制面与转发面分离的网络架构演进需求,传统专有封闭的设备体系也正在逐步瓦解,转向基于“通用硬件+SDN/NFV”的云化开放体系。基于虚拟机以及容器技术承载电信网络的功能,使用 MANO 与云管协同统一编排业务与资源,并通过构建 DevOps 一体化能力,大幅缩短新业务面市周期,从而实现 IT 与 CT 技术深度融合,提高网络运营商的业务竞争力。因此,从中心机房到边缘机房,借助云化技术重构电信基础设施将是必然选择。

(4) SD-WAN 优化网络带宽和时延。

长期来看,移动 2G、3G、4G 网络的长期演进属于 5G 网的一部分。同时,Wi-Fi 作为移动网络的补充也会长期存在。另外,各种有线网络的接入会导致带宽快速增长,企业业务迁移到云上并直接转向企业云,应用程序带来了大量的数据带宽要求,物联网的海量数据连接使接入网的业务更加复杂化。为了在接入网有效地支持多种复杂网络接入,并快速识别不同客户设备的网络需求,最终满足带宽、可靠性和时延的要求,SD-WAN(软件定义的广域网)对网络带宽和数据业务的识别是至关重要的一环。

SD-WAN 使用网络虚拟化来利用、管理和保护互联网宽带,为各种应用构建更强大的网络。SD-WAN 在 WAN 连接的基础上,将提供尽可能多的、开放的和基于软件的技术。SD-WAN 的主要功能是综合利用多条网络链路,根据现网情况及配置的策略,自动选择最佳路径,实现负载均衡,从而保证网络质量,降低流量成本。例如,SD-WAN 如果同时连接了 MPLS 和 Internet,那么可以将一些重要的应用流量分流到 MPLS,以保证应用的可用性。对于一些对带宽或者稳定性不太敏感的应用流量,例如文件传输,可以分流到 Internet 上,这样减轻了企业对 MPLS 的依赖。Internet 也可以作为 MPLS 的备份连接,当 MPLS 出故障时,企业的 WAN 网络不至于受牵连。

SD-WAN 继承了“转控分离”的思想,通过统一的中央控制器实现各 CPE 设备的统一管控。新开 CPE 可通过 Call Home 自动连接控制器下载配置和策略,真正做到零接触快速开通,且配置变更时只需要在控制器上修改一次,所有分支站点自动同步。通过 CPE 对流量的深度识别和基于探针的链路状况检测,使得流量选路策略更加精细灵活,实现策略与网络质量随动,是真正意义上的动态流量调度。

支持云计算的骨干网 SD-WAN 架构提供了 SD-WAN 盒子,可以将企业的站点连接到最近的网络节点(POP),企业的流量在 SD-WAN 提供商的私人、光纤、网络骨干网中传输,可以实现低时延、低丢包率和低抖动。这种方式能够有效提高所有网络流量的性能,特别是语音、视频和虚拟桌面等实时流量。骨干网也与主要的云应用提供商和数据中心等直接相连,这些应用能够提高性能和可靠性。

SD-WAN 可以优化 WAN 管理接口。一般硬件专用设备网络都存在管理和故障排查等较为复杂的问题,WAN 也不例外。SD-WAN 通常也会提供一个集中的控制器来管理 WAN 连接,设置应用流量策略和优先级,监测 WAN 连接可用性等。基于集中控制器,可以再提供 CLI 或者 GUI,从而达到简化 WAN 管理和故障排查的目的。SD-WAN 业务模型如图 5-18 所示。

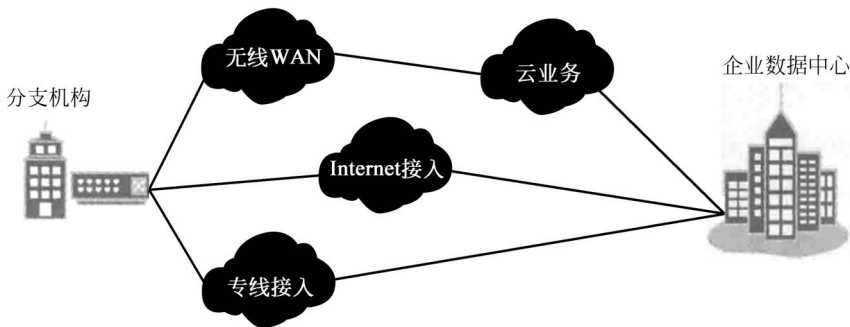


图 5-18 SD-WAN 业务模型

家庭客户或者政企侧一般通过客户端设备 CPE 接入运营商网络,目前的趋势是使用通用客户端设备 uCPE 来替代传统的专用网关。

4. 国内运营商网络演进

由于通信网络越来越 IP 化,很自然地会和本来就是 IP 网络的 IT 网进行融合,也就是 ICT 融合。ICT 融合主要指网络的融合和设备的融合,网络的融合主要指运营商侧的一张 IP 网支撑 IT 和 CT 的业务。例如,传统通信网络的一些设备,包括服务 GPRS 支持节点(Serving GPRS Support Node,SGSN)、网关 GPRS 支持节点(Gateway GPRS Support Node, GGSN)和移动管理入口(Mobile Management Entry,MME)等实际已经运行在通用的处理器上,只是由于历史兼

容原因,在实际部署中仍然使用之前基于 ATCA 的老式专用设备机箱。随着 5G 架构的提出和部署、SDN 和 NFV 技术的发展、网络云化和边缘计算的演进,在下一代网络部署中,越来越多的通信设备会基于通用边缘和传统服务器的硬件平台实现。

考虑未来 5G 和固网业务的需求,以及 SDN、NFV 等技术的发展,基于自身已有的网络基础设施和自身业务特点,结合 ICT 融合的趋势,在下一代网络部署中,中国三大运营商都提出了自己的技术演进策略。

(1) 中国移动的 C-RAN。

2009 年,中国移动首次提出 C-RAN (Centralized, Cooperative, Cloud and Clean RAN) 的概念。2016 年,中国移动发布了《迈向 5G C-RAN: 需求、架构与挑战》白皮书,C-RAN 是集中化、协作化、云化、绿色化概念的集合体。4G 的 C-RAN 主要是基于 BBU 集中化,提高无线协作化和抗干扰能力,从而提高设备利用率,优化机房利用率和降低能耗等。经过长期运维,中国移动证明了 C-RAN 组网方式在综合成本、抗干扰、降低能耗、运维和机房优化等方面的优势很明显。所以,将 C-RAN 作为优选建设方式在全网进行推广。

随着 5G 的发展,为了支持 eMBB、mMTC 和 uRLLC 等多种业务,无线空口资源与具体的无线业务需要解耦并实现按需分配,这就需要一定程度的软硬件解耦,以便于构造资源池,实现空口资源的灵活管理。随着 5G 频率的上升,单个基站的覆盖范围缩小,需要提高基站密度。随着 5G 业务的下沉,接入网络必须支持计算能力和控制面 UPF 的下沉。另外,2G、3G、4G、5G 和固网接入等多网络协作和管理也是一项极大的挑战。传统 4G BBU 堆叠集中部署规模,还不能完全满足 5G 组网的多种要求。

为了应对 5G 网络的挑战,通过引入 NFV 架构,云接入网 C-RAN 基于集中和分布单元 CU/DU 以及下一代前传接口 (Next Generation Fronthaul Interface, NGFI),将部分基带资源集中成资源池并统一管理与动态分配,从而提升资源利用率、降低能耗,进而提升网络性能。5G C-RAN 网络的具体特征如下。

① 集中部署。

首先是 BBU 的集中部署,减少机房数量实现设备共享。其次是 CU/DU 无线高层协议栈功能的集中。

② 协作能力。

相比集中部署,首先是小规模物理层集中,主要实现多小区或多数据发送点间的无线联合发送和接收,提升小区的边缘频谱效率和平均吞吐量。其次是大规模的无线高层协议栈功能的集中,实现多连接、无缝移动性管理、频谱资源高效协调等。

③ 无线云化。

随着 CU/DU 架构的引入, CU 设备具备了云化和虚拟化的基础。一方面是指资源的池化形成了资源共享, 从而实现了灵活部署, 降低整个系统的成本。另一方面是无线空口的云化实现了无线资源与无线空口技术的解耦, 从而灵活地支持无线网络能力调整, 满足特定的定制需求。

④ 绿色节能。

通过集中化、协作化、无线云化等, 优化无线机房的部署, 降低配套设备和综合能耗, 从而优化全系统的整体效能比。

(2) 中国电信的三朵云。

为满足 5G 的不同组网需求, 构建更加灵活开放的网络, 中国电信基于 SDN 和 NFV 的理念及技术, 提出了基于接入云、控制云和转发云的“三朵云”总体网络架构, 如图 5-19 所示。其中, 控制云完成整体网络管理和控制, 通过虚拟化、网络资源容器化和网络切片化等技术, 完成业务需求的定制化, 并实现网络功能的灵活部署, 提供网络开放和可拓展能力。接入云主要实现多网络的多种业务场景, 例如, 3G、4G、5G 网络, 物联网, 车联网等的智能接入和灵活组网, 并具备边缘计算下沉的能力。转发云负责数据的高速转发和处理, 以及各种业务使能单元。基于不同业务的带宽和时延等需求, 转发云在控制云的协调调度下, 通过端到端的网络切片构建不同的虚拟网络, 从而满足 eMBB、mMTC、uRLLC 等业务需求。

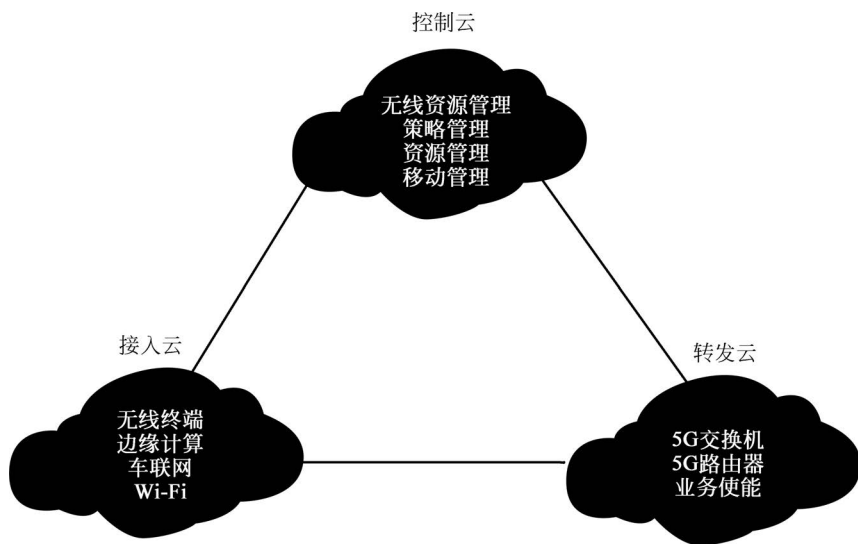


图 5-19 中国电信“三朵云”架构

“三朵云”是电信的远期目标, 具体技术现在不是特别成熟, 而且初期的实际业务需求也不是特别迫切, 对于具体网络的改造必然需要分阶段进行。其中, 重点是

要充分利用已有的网络资源,注重新旧业务的平滑过渡和支持、多网融合和协作,尽量避免网络大规模和频繁的升级改造。

考虑 5G 独立组网方案通过核心网实现与 4G 网络的协作,具有支持网络切片和 MEC 等新特性,对现网改造小,并且业务能力更强,终端设计简单,所以中国电信将优选 SA 组网方案。考虑 5G 初期的主要支持场景是 eMMB 的大带宽需求,短期内无线接入网的 CU 基于通用处理器的虚拟化性能与专用硬件相比偏低,导致成本高,建议使用经济实用的 CU/DU 合设方案。中远期随着 uRLLC 和 mMTC 业务需求的提升和 NFV 技术的发展,适时引入 CU/DU 分离的虚拟化架构。

承载网的演进遵循固移融合的原则实现资源共享,具备支持网络切片、支持 5G 等多种业务需求。回传网络前期采用 IP-RAN,后期按需引入高速率 25Gb/s 接口,业务大且集中的区域采用 OTN 组网,中远期建成基于 SDN 的高速率、超低时延的自动智能回传网络。

由于网络云化及边缘计算的引入,5G 核心网主要部署在省中心的区域 DC,且部分功能将下沉到城域网,甚至接入局所。基于独立组网方案,5G 核心网必须以 SDN、NFV、云计算为基础,具备控制与承载分离的特征。控制面采用通用处理器平台,使用虚拟化技术,构建统一的 NFV1 资源池,实现业务的云化部署和弹性扩容等。基于通用 CPU 和专用硬件联合实现用户面功能,通用 CPU 适用于业务多样化的场景,专用硬件实现纯粹的数据转发。另外,还需考虑多种硬件加速技术并实现一定的标准化和虚拟化,实现不同厂商间的混合部署、网络资源的分布式灵活部署和全局调用,更好地支持端到端的网络切片和边缘计算的按需下沉,以满足未来网络的多样化业务需求。

(3) 中国联通 Edge-Cloud 网络。

为了应对 5G 网络的多样化业务和固网融合等需求,中国联通基于已有的大量边缘 DC 优势资源,在加强运营商管道能力的基础之上,依托 SDN、NFV 和网络云化等技术,通过虚拟技术承载电信网络功能,使用 MANO 与云管协同实现资源的统一管理和业务编排。中国联通将基于 CORD(Central Office Re-architected as a Data Center)将传统局端改造成数据中心,从传统的专用硬件设备,向“基于通用硬件+SDN/NFV 的资源虚拟化和池化”开放的 Edge-Cloud 平台转变。不仅支持边缘计算和网络灵活部署的能力,还可以拓展新型增值业务,将平台存储、计算、网络与安全能力等基础设施通过虚拟池化后开放给第三方。同时,进一步通过统一的 API 接口提供丰富的网络服务,例如无线信息服务、位置服务、带宽管理服务,从而变成综合性的端到端业务提供商,提高竞争力。

未来,组网将采用边缘、本地、区域 DC 的三级布局,基于虚拟技术构建云资源

池,实现统一接入的多网融合网络。区域 DC 类似于核心数据中心,通过骨干网链接,服务全国、大区或者全省的业务和控制面网元。本地 DC 部署于地级市和省内重点县级市,主要承载城域网控制面网元和集中化的媒体面网元,服务本地网的业务、控制面及部分用户面网元。边缘 DC 以终结媒体流功能和转发为主,主要部署接入层以及边缘计算类网元。综合接入机房负责公众、政企、移动、固网等用户的统一接入。考虑边缘计算的下沉和超低时延的需求,综合接入机房和边缘 DC 是边缘业务平台部署关注的重点。

中国联通 Edge-Cloud 平台主要在边缘 DC、本地 DC、区域 DC 实现三级云化,所以也可以称为“三级云”,如图 5-20 所示。每级云都通过虚拟化形成资源的池化,通过基于 OpenStack 管理虚拟资源的平台 VIM(Virtualized Infrastructure Manager)管理一个云的 NFVI。基于分布式部署,节点数较少的边缘 DC 只部署计算节点,控制节点部署在区域 DC 或者节点数较多的边缘 DC 上,对各个边缘 DC

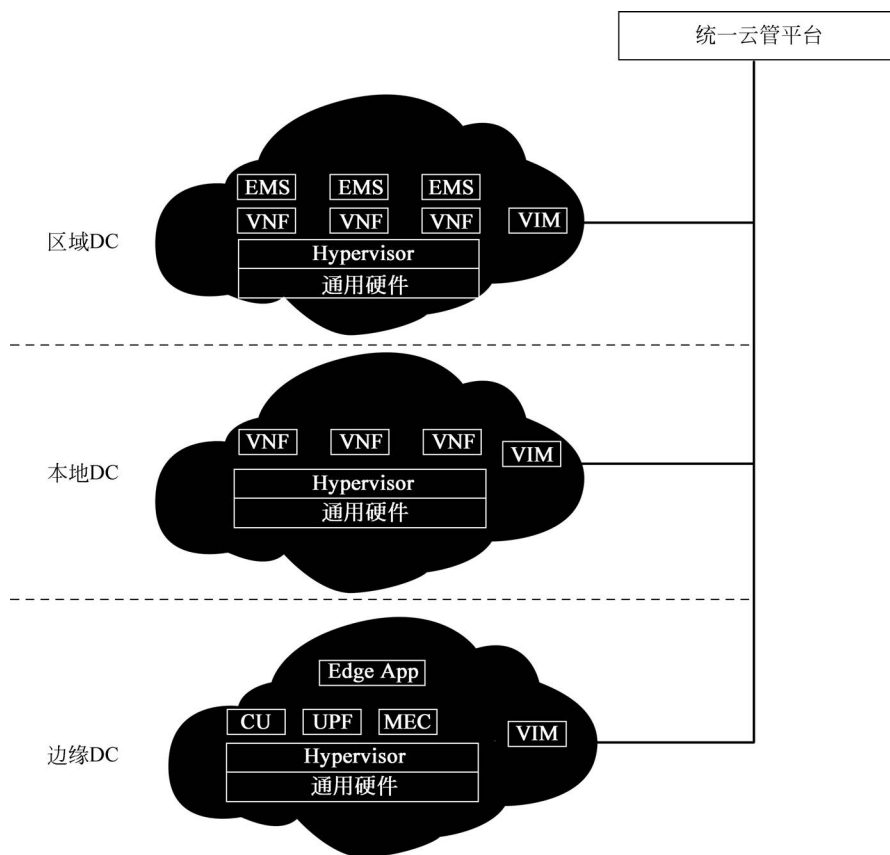


图 5-20 中国联通 Edge-Cloud 云管与 MANO 架构图

进行统一管理。统一云管理平台主要分为两层,可依实际情况部署在边缘数据中心或区域数据中心并对接多个 VIM,提供接入点汇聚、属地化异构资源统一纳管、基于位置的资源调度等能力。然后在区域 DC 实现统一云管平台的部署管理,包括用户授权、VNF 管理等,从而实现对整个通信云基础设施的管理。

除了支持 VNF,中国联通特别强调支持和编排管理第三方的边缘应用(Edge-App),包含应用程序包的软件镜像、格式、应用描述、签名所需的最小计算资源,以及虚拟存储资源、虚拟网络资源等需求。

自 2017 年 9 月起,中国联通启动 Edge-Cloud 规模试点工作,实现 COTS 和 Cloud OS 软硬解耦,Cloud OS 与 MEP(Multi-access Edge Platform)同厂家部署。2018 年,Edge-Cloud 平台已经实现 COTS、Cloud OS、MEP、Edge-App 四层解耦,并以共 COTS 异 Cloud OS 的方式实现通信网元功能的虚拟化。2019 年,启动边缘 DC 云资源池的规模建设,努力实现 5G 网络试商用。从 2021 年到 2025 年,加快 5G 和规模商用节点建设,同期,中国联通通信网络云化架构也将基本成型,预计 2025 年将实现 100% 云化部署和 Edge-Cloud 的彻底四层解耦。

国内运营商面临话音低值化的困境,各种新型的视频直播、社交、在线零售和导航等服务催生了新的行业巨头,但由此产生的高速增长的数据流量并未给运营商带来可观的收益,反而对其原有基础网络设施造成了额外的挑战。传统通信运营商网络基础设施主要基于专用的硬件设备,软件和硬件紧耦合,控制面和数据面集成。

随着日益新增的各种高清视频直播、虚拟现实、AI 等业务需求的兴起,5G 定义的等大带宽、低时延和高可靠性等业务场景的支持,物联网引入的海量设备和传统固网宽带大带宽的飞速增长,运营商急需借助 SDN、NFV、处理能力,以及快速提高的通用处理器、边缘计算、SD-WAN 和 5G 等新兴网络技术,对现有网络基础设施和架构进行升级,支持 CU/DU 分离设计、网络切片和虚拟化等新特性,构造软硬解耦、数控分离的自动智能网络,推动接入网、承载网和核心网的云化,优化基础设施投入,降低运营成本,满足多样的业务需求。同时,借助数以千计的优质边缘 DC 和海量用户资源,在加强管道能力的基础上,打造统一的边缘业务平台,提供开放式的 API 接口。例如,在边缘 DC,由于空间和成本的限制,每家云厂商很难自建边缘 DC。运营商可以在机房自建边缘服务器和存储,租用给云运营商来提供除管道之外的增值业务,从而提升整体的竞争力。

中国移动是 C-RAN 方案的领导者,主导了 CU/DU 分离的设计方案,并积极联合芯片厂商和设备厂商推进接入边缘网络的云化和虚拟化。中国电信三朵云的远景主要是基于 SDN 和 NFV 技术理念,完美地实现数控分离的自动灵活网络。但在实际部署中,前期采取经济务实的分布式改造策略,基于 eMBB 的需求,采取 CU/DU 合体的专用硬件方案部署,后续根据业务需求进一步引入 CU/DU 分离的

虚拟化方案。中国联通基于已有的大量边缘 DC 优势资源以及异厂商共平台的策略,稳步推动 Edge-Cloud 的发展,特别是对 Edge-App 的支持,打造了四层解耦的开放基础服务平台。

传统运营商希望借助 SDN、NFV 和 5G 等新技术拓展管道之上的新增业务,而新兴的云服务商希望凭借已有的应用业务优势,通过应用来定义网络和计算的底层基础架构,只希望传统通信运营商提供最基础的通道业务,甚至只是租用运营商的机房来放置自己的计算、存储和网络等设备。已经有云服务商通过租用传统通信运营商的内部网络接入点,构建基于自身云的虚拟网络,来提升云服务体验。

另外,ODM 热切拥抱软硬分离的设备方案,并已经从白盒服务器和网络设备领域分得了一杯羹。但是,传统通信设备商出于自身利益考虑,对软硬解耦的态度模糊,而且专用硬件加速模块的性能和成本在某些特定应用上高于通用处理器,所以处于领导地位的设备商仍然提供软硬耦合的方案。还有一些设备商倾向于提供基于“通用处理器+专用硬件”加速的软耦合方案,同时传统通信运营商也在考虑针对硬件加速模块的标准化方案,以适应新技术的演化。

下一代新型网络的发展需要通信运营商、云服务商、设备商、芯片商和 ODM 等全行业的共同参与和努力。在新技术演进的滚滚洪流中,是已有的行业巨头借东风更上一层楼,还是会涌现出新的行业弄潮儿,让我们拭目以待。

第 5 节 边缘存储架构

1. 边缘存储

边缘存储就是把数据直接存储在数据采集点或者靠近的边缘计算节点中,例如 MEC 服务器或 CDN 服务器,而不需要将数据通过网络即时传输到中心服务器(或云存储)的数据存储方式。边缘存储一般采用分布式存储,也称为去中心化存储。下面通过几个案例来说明。

第一,在安防监控领域,智能摄像头或网络视频录像机(NVR)直接保存数据,即时处理,不需要将所有数据传输至中心机房再处理。

第二,家庭网络存储服务器,用户更偏向将私人数据存储在自己家中,而不是通过网络上传到提供存储服务的第三方公司,这样第三方公司不会接触到敏感数据,以保证隐私和安全性。

第三,自动驾驶采集的数据往往可以在车载单元或路侧单元中进行预处理,再将处理后的少量数据传输给后台服务中心或云。

为什么目前主要使用的还是中心存储,而不是边缘存储呢? 一个很重要的原因是数据处理在中心,边缘设备的处理能力还不够。另外一个原因是,缺乏成熟可

行的技术方案连接和同步边缘节点,无法使边缘端更多地承担数据采集、处理和存储的任务。

随着芯片技术的发展,边缘端设备的运算能力和处理速度都得到了大幅度提升,使设备成本大大降低,在靠近数据的边缘端已经可以进行较好的数据处理。同时,随着去中心化存储技术的飞速发展,例如 IPFS 采用的 Libp2p,能够很好地解决端设备的局部互联问题,可以在边缘进行连接和处理。以车联网为例,在自动驾驶车辆中,传感器和摄像头采集的数据完全可以存放在本地和路侧单元中,由于在同一个街道或区域运行的汽车很多,它们会采集大量重复的数据,但也有一些数据可以相互补充。当把数据存储在本地图,同一个街道上的汽车能够相互连接,并对数据进行即时聚合,这样需要上传的数据就大大减少了。

边缘存储的主要特点如下。

- ① 低时延,通常小于 5ms。
- ② 分布式查看,隔离操作,同一个网络操作不会影响到其他的网络。
- ③ 本地保存和转发能力,能够降低和优化节点间的带宽占用。
- ④ 能够聚合并传送给中心节点,从而减少网络中冗余数据的传输。
- ⑤ 数据移动性,允许边缘设备在不同的边缘网络中移动,而不影响数据同步和完整性。

2. 边缘存储的优势

(1) 网络带宽和资源优化。

对于以云为核心的存储架构,所有的数据都需要传输到云数据中心,带宽的需求是极大的。同时,并不是所有数据都需要长期保存。例如,对于电子监控的视频数据仅保存数天、数周或数月。而对于智能工厂中机器采集的原始数据,特点是数据频率高、规模大,但有价值的数据相对较少。如果将这些数据存储在网上,会带来网络带宽资源浪费、访问瓶颈以及成本上升等一系列问题。

在某些情况下,由于带宽限制,数据传输质量会受到影响,出现丢包或超时等问题。针对这种情况,可以利用边缘存储来缓存数据,直到网络状况改善后再回传信息。此外,还可以利用边缘存储动态优化带宽和传输内容质量。例如,在边缘录制高质量的视频,然后在远程查看标准质量的视频,甚至可以在网络带宽不受限制时,将录制的高质量视频同步到后端系统或云存储中。

通过边缘存储和云存储的有机结合,可以将一部分数据的存储需求从中心转移到边缘,更加合理有效地利用宝贵的网络带宽,并根据网络带宽的情况灵活优化资源的传输,使得现有网络可以支撑更多边缘计算节点的接入并降低总体拥有成本(Total Cost of Ownership, TCO)。

(2) 分布式网络分发。

由于边缘节点分布式的特点,可以利用边缘存储建立分发网络,分发加速的效果将好于当前站点有限的 CDN 网络。例如,分享一部分存储用于分发,那么观看的热门电视剧就可以被邻居直接下载使用,极大地节省网络带宽,也可以通过分享存储资源获取部分收益。

当边缘存储进入实用阶段时,去中心化的应用也更容易建立。基于地域的社区将可以不通过中心服务器或服务商进行交互,也更容易建立基于社区的私有网络。

同样,由于基于边缘存储的点对点网络的建立,应用或服务商之间的数据共享变得更加容易和便捷。在理想情况下,服务商或应用提供商完全可以不拥有数据,而数据本身属于数据的生产者。这样一来,数据的拥有者就完全可以把这部分数据分享给不同的应用或服务,令其产生超额价值。例如,远程医疗可以让病人把自己的检查结果存储在本地,而病人可以支配自己的检查报告,用于提供给不同的医生或医院进行诊断。同时,如果病人愿意,也可以匿名分享给研究机构作为科研数据。

(3) 可靠性更强。

当数据存储和处理完全是中心化的时候,任何的网络问题或数据中心本身的问题都会导致服务中断,影响巨大。当边缘计算节点具备一定的处理能力,且数据存储于端或边缘之后,对网络的要求大大降低,一部分的网络中断只会影响小部分功能,因为很多处理运算同样可以在本地进行。同时,当边缘的点对点网络建立起来后,网络的冗余性会进一步解决部分网络中断的问题,容错性得到极大的加强。

而对于需要具备高可信度的企业级应用(例如银行、政府和城市监控系统),利用边缘存储可以降低数据丢失的风险。如果主网络中的存储发生问题,边缘存储可以保留数据的备用副本,并在需要的时候同步到后台,从而提高整个系统的可靠性。

为了进一步提高边缘存储的可靠性,边缘存储介质的选型也非常重要。由于边缘节点需要长期全天候运行,工作环境多样,传统的针对消费级的存储介质已经不能满足边缘存储高可靠性的要求。典型的如 IP 摄像头或 NVR 中使用的消费级 Micro SD,这种卡的固件并未针对全天候录制需求进行优化,因此会丢帧,在许多情况下,帧捕捉率下降 20%~30%后,会导致大量数据丢失。因此需要选择工业级的专用存储卡,提高平均无故障时间(MTTF)、降低年度故障率(AFR),并提供可监控运行状况的智能工具,以减少系统停机时间,降低维护成本。

(4) 安全与隐私兼顾。

目前,虽然云计算极大地方便了我们的生活,能够让我们随时随地访问自己的私人数据,但也出现了一些关于数据安全和隐私的担忧。这也是家庭安防、智能家居发展缓慢的原因之一。边缘存储结合点对点网络技术,可以帮助解决这个问题。

在新的解决方案里,用户不需要把数据存储到网上,而是保存在家庭的 NAS 中,所有数据都是可以加密存储的。通过 P2P 网络,也可以建立端设备和家庭数据服务器的点到点连接,让数据私密传送,同时兼顾安全与隐私。

除此之外,边缘存储还有一个优势是与边缘计算相结合。考虑到家庭安防的情况,用户可以对家庭的摄像报警系统进行配置,设置报警的条件,多数情况下采集的视频数据是不需要上传的,只有在出现异常情况时才需要占用网络带宽和外部资源。此外,边缘存储和网络传输可以使用不同格式的视频流,例如本地存储高清晰、高解析度的视频流,而在网络上传输的可以是低码率、低清晰度、占用有限带宽的数据,既可以解决实时监控的问题,也可以进一步分析。

3. 边缘数据和存储类型

(1) 边缘数据类型。

根据数据的访问频率,可以将数据分为三类:热数据、温数据和冷数据。这三类数据与总数据量的比例分别约为 5%、15% 和 80%。同时数据的存储和访问策略也迅速分化,数据分层加剧:热数据一般放在内存或基于 3D XPoint 的持久化内存中;温数据放在基于 PCIeNVM 或者 SATA 接口的 SSD 中;冷数据则放在低转速的 HDD 硬盘中。

根据数据格式的不同,数据也可以分为三类。

① 结构化数据。

结构化数据是表现为二维形式的数据,可以通过固有键值获取相应信息。其数据以行为单位,一行数据表示一个实体的信息,每一行数据的属性是相同的。结构化数据的存储和排列是很有规律的,这对查询和修改等操作很有帮助,一般可以使用关系数据库表示和存储。主要应用的关系数据库有 Oracle、SQL Server、MySQL 和 MariaDB 等。

② 半结构化数据。

严格来说,结构化数据与半结构化数据都是有基本固定结构的数据;而半结构化数据可以通过灵活的键值调整获取相应的信息,且数据的格式不固定。同一键值下存储的信息可能是数值型的,也可能是文本型的,还可能是字典列表型的。半结构化数据属于同一类实体,可以有不同的属性,即使它们被组合在一起,这些属性的顺序也并不重要。常见的半结构化数据有日志文件、XML 文档、JSON 文档、电子邮箱等。

③ 非结构化数据。

非结构化数据简单来说就是没有固定结构的数据,是应用最广的数据结构,也是最常见的数据结构。例如,办公文档、文本、图片、XML、HTML、报表、图像、音

频和视频信息等。这类数据一般直接整体进行存储,而且一般存储为二进制的数据格式。

随着边缘端基于视觉的 AI 应用越来越多,对于图片和视频等非结构化数据进行分析和存储的需求也越来越强:在智能工厂内,边缘计算节点需要对工业摄像头采集的图像进行实时分析并反馈结果;在智能物流的物流仓储中心中,需要对传送带上的包裹标签进行实时识别和确认;在安防中,NVR 对于摄像头传送的视频流进行实时的基于深度学习的算法推理,并将原视频内容(包括元数据)和分析结果存储起来。2020 年,全球非结构化数据量达到 35ZB,相当于 80 亿块 4TB 硬盘,非结构化数据在存储系统中所占据的比例已接近 80%,数据结构变化给存储系统带来了新的挑战。

此外,时序数据是边缘计算节点中保存和处理最多的数据类型。智能制造、交通、能源、智慧城市、自动驾驶、人工智能等行业都产生了巨量的时序数据。例如,无人驾驶汽车在运行时需要监控各种状态,包括坐标、速度、方向、温度、湿度等,每辆车每天就会采集将近 8TB 的时序数据。这些数据在边缘计算节点中不仅要进行快速的保存,同时也需要实现实时的分析和多维度的查询操作,以揭示其趋势性、规律性、异常性等,甚至需要通过大数据分析、机器学习等实现预测和预警。

在这种情况下,使用传统数据库进行时序数据保存和查询的效率非常低,而时序数据库能够实现时序数据的快速写入,以及持久化、多纬度的聚合查询等基本功能。例如,开源的 InfluxDB 能够手动或使用脚本函数创建标签,并使用标签进行数据的索引,优化分析过程中的查询操作。

时序数据库面向的是海量数据的写入、存储和读取,单机很难解决,一般需要采用多机分布式存储。边缘节点处理时序数据的特点如下。

- ① 普遍都是写操作,占比为 95%~99%。
- ② 写操作基本都是顺序添加数据。
- ③ 很少更新。
- ④ 使用块擦除。

由于边缘端对于时序数据有实时处理的需求,因此时序数据库也面临一些挑战,如下所示。

- ① 如何支持每秒上千万甚至上亿数据点的写入。
- ② 如何支持对上亿数据的秒级分组聚合运算。
- ③ 如何进行数据压缩并以更低的成本存储这些数据。

(2) 边缘存储类型。

从存储介质角度来看,边缘存储分为机械硬盘和固态硬盘。机械硬盘泛指采用磁头寻址的磁盘设备,包括 SATA 硬盘和 SAS 硬盘。由于采用磁头寻址,机械

硬盘性能一般,随机 IOPS 一般在 200 左右,顺序带宽在 150Mb/s 左右。固态硬盘是指采用“Flash/DRAM 芯片+控制器”组成的设备,根据协议的不同,又分为 SATASSD、SASSSD、PCIe SSD 和 NVMe SSD。

边缘计算节点数据存储包括持久性和非持久性数据存储。对于非持久性数据存储,考虑到高级分析功能的需求,至少需要 32GB 的 DRAM,以保证在分析过程中不会将内存页面调度到持久性存储介质中。而对于持久性存储,需要考虑不同应用环境的需求。例如,基于 3D XPoint 的 SSD 不能部署在有宽温需求的环境中;对于靠近设备部署的边缘计算节点,由于尺寸限制且不支持热插拔,一般选择使用 M.2 SSD;而对于边缘服务器或部署在中心机房的机架服务器,一般可以选择基于 PCIe NVMe 协议的 U.2 SSD 或 EDSFF SSD,以保证 IOPS、吞吐率等读写性能,并增加存储密度。

随着 NVMe SSD 的出现,使用 Linux 文件系统,I/O 栈已经无法发挥出 NVMe 的性能。为了更好地发挥 SSD 固态硬盘的性能,英特尔开发了一套基于 NVMe SSD 的开发套件 SPDK(Storage Performance Development Kit),它的目标是将固态存储介质的功效发挥到极致。相对于传统 I/O 方式,SPDK 采用用户态驱动和轮询方式来避免内核上下文切换和中断,这将会节省大量的处理开销,提高吞吐量,降低时延,减少抖动。相对于 Linux 内核,SPDK 对于 NVMe SSD 的 IOPS/core 可以提升大约 8 倍,在虚拟存储情况下,可以提升大约 3 倍。

从流程上来看,SPDK 主要包括网络前端、处理框架和存储后端。网络前端由 DPDK、网卡驱动、用户态网络服务构件等组成。DPDK 给网卡提供一个高性能的包处理框架,提供一个从网卡到用户态空间的数据快速通道。用户态网络服务则解析 TCP/IP 包并生成 i-SCSI 命令。处理框架得到包的内容,并将 i-SCSI 命令翻译为 SCSI 块级命令,不过在将这些命令发送给后端驱动之前,SPDK 需要提供一个 API 框架以加入用户指定的功能,例如缓存、去冗、数据压缩、加密、RAID 和纠删码计算等,这些功能都包含在 SPDK 中。数据到达后端驱动,在这一层中与物理块设备发生交互,即读与写等。

从产品定义角度来讲,存储分为本地存储(Direct Attached Storage,DAS)、网络存储(Network Area Storage,NAS)和存储局域网(Storage Area Network,SAN)。本地存储就是本地盘,直接插到服务器上。

网络存储是指提供 NFS 协议的网络存储设备,通常采用“磁盘阵列+协议网关”的方式。

存储局域网与网络存储类似,提供 SCSI/i-SCSI 协议,后端是磁盘阵列。

在计算和存储一体化的存储架构中,一般设置四层数据存储,基于分布式存储软件引擎完全水平拉通,且支持基于强一致的跨服务器数据可靠性。

第一层存储：内存，时延 100ns，作为缓存，通过缓存算法管理。

第二层存储：PCIe SSD，时延 10(本地)~300 μ s(远地)，作为缓存或最终存储。

第三层存储：本地存储，时延 5(本地)~10ms(远地)，作为第四层存储的补充或替代。

第四层存储：存储局域网，时延 5(本地)~10ms(远地)。

将上述各层存储的热点数据读写推至更上一层存储，实现数据 I/O 吞吐及整个系统性能的大幅提升。

从应用场景角度来讲，存储分为文件存储、块存储和对象存储三大类。

4. 边缘分布式存储

(1) 集中式存储。

集中式存储就是整个存储是集中在一个系统中的。但集中式存储并不是一个单独的设备，而是集中在一套系统中的多个设备。目前，企业级的存储设备大都是集中式存储。在这个存储系统中包含很多组件，除了核心的机头(控制器)、磁盘阵列(Just a Bunch of Disks, JBOD)和交换机等设备，还有管理设备等辅助设备。

在集中式存储中通常包含一个机头，这是存储系统中最核心的部件。通常，在机头中包含两个控制器，这两个控制器实现互备的作用，以避免硬件故障导致整个存储系统不可用。在该机头中通常包含前端端口和后端端口，前端端口用于为服务器提供存储服务，后端端口用于扩充存储系统的容量。通过后端端口，机头可以连接更多的存储设备，从而形成一个非常大的存储资源池。可以看出，集中式存储最大的特点是有一个统一的入口，所有数据都要经过这个入口，也就是存储系统的机头。

在集中式存储中，机头可能成为制约系统扩展性的单点瓶颈和单点故障风险点。在虚拟化服务器整合环境中，成百上千个 VM 共享同一个存储资源池。一旦磁盘阵列控制器发生故障，将导致整体存储资源池不可用。尽管 SAN 控制机头自身有主备机制，但依然存在异常条件下主备同时出现故障的可能性。另外，在集群组网环境下，各计算节点的内存、SSD 作为分层存储的缓存彼此孤立，只能依赖集中存储机头内的缓存实现 I/O 加速；共享存储的集群内各节点缓存容量有限，但不同节点缓存无法协同，且存在可靠性问题，导致本可作为集群共享缓存资源的容量白白浪费。

(2) 分布式存储。

分布式存储(Hadoop Distributed File System, HDFS)是相对于集中式存储来说的，它除了传统意义上的分布式文件系统、分布式块存储和分布式对象存储，还包括分布式数据库和分布式缓存等。

分布式存储最早是由谷歌公司提出的,其目的是通过廉价的服务器解决大规模、高并发场景下的 Web 访问问题。面对信息化程度不断提高带来的 PB 级海量数据存储需求,以及非结构数据的快速增长,传统的存储系统在容量和性能的扩展上出现瓶颈。SAN 存储成本高,不适合 PB 级大规模存储系统。DAS 数据共享性不好,无法支持多用户文件共享。NAS 存储共享网络带宽,并发性能差,随着系统扩展,性能会进一步下降。分布式文件系统和分布式存储以其扩展性强、性价比高、容错性好等优势得到了业界的广泛认同。

分布式存储系统具有以下几个特点。

① 高性能。

分布式散列数据路由,数据分散存放,实现全局负载均衡,不存在集中的数据热点和大容量分布式缓存。

② 高可靠。

采用集群管理方式,不存在单点故障,灵活配置多数据副本,不同数据副本存放在不同的集群、服务器和硬盘上,单个物理设备故障不影响业务的使用,系统检测到设备出现故障后可以自动重建数据副本。

③ 高可扩展性。

没有集中式机头,支持平滑扩容,容量几乎不受限制。

④ 易管理。

存储软件直接部署在服务器上,没有单独的存储专用硬件设备,通过 Web UI 的方式进行软件管理,配置简单。

如图 5-21 所示,是分布式存储的简化架构图。在该系统的整个架构中,将服务器分为两种类型:一种名为 namenode,这种类型的节点负责元数据的管理;另外一种名为 datanode,这种类型的服务器负责实际数据的管理。

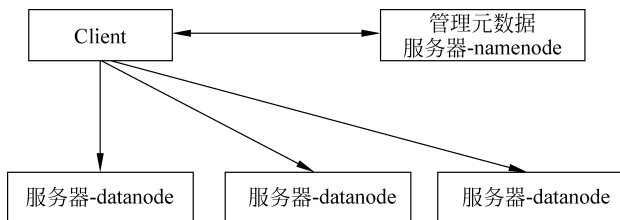


图 5-21 分布式存储的简化架构图

在图 5-21 中,如果客户端需要从某个文件读取数据,首先从 namenode 获取该文件的位置(具体在哪个 datanode),然后从该位置获取具体的数据。在该架构中,namenode 通常采用主备部署方式,而 datanode 则是由大量节点构成的一个集群。由于元数据的访问频度和访问量相对数据都要小很多,因此 namenode 通常不会

成为性能瓶颈,而 datanode 集群可以分散客户端的请求。因此,采用这种分布式存储架构,可以通过横向扩展 datanode 的数量来增加承载能力,也即实现了动态横向扩展的能力。

⑤ 完全无中心架构——计算模式(Ceph)。

Ceph 无中心架构与 HDFS 不同的地方在于它没有中心节点,客户端通过一个设备映射关系计算其写入数据的位置。这样客户端可以直接与存储节点通信,从而避免中心节点出现性能瓶颈。

在 Ceph 存储系统架构中,核心组件有 Mon 服务、OSD 服务和 MDS 服务等。对于块存储类型,只需要 Mon 服务、OSD 服务和客户端的软件即可。其中,Mon 服务用于维护存储系统的硬件逻辑关系,主要是服务器和硬盘等在线信息,Mon 服务通过集群的方式保证其服务的可用性。OSD 服务用于实现对磁盘的管理,实现真正的数据读写,通常一个磁盘对应一个 OSD 服务。

客户端访问存储的大致流程是,客户端在启动后首先从 Mon 服务拉取存储资源布局信息,然后根据该布局信息和写入数据的名称等信息计算出期望数据的位置(包含具体的物理服务器信息和磁盘信息),然后与该位置信息直接通信,读取或者写入数据。

⑥ 完全无中心架构——一致性散列(Swift)。

与 Ceph 通过计算获得数据位置的方式不同,另外一种方式是通过一致性散列获得数据位置。例如,对于 Swift 的 ring 是将设备做成一个散列环,然后根据数据名称计算出的散列值映射到散列环的某个位置,从而实现数据的定位。

数据分布算法是分布式存储的核心技术之一,不仅要考虑数据分布的均匀性、寻址的效率,还要考虑扩充和减少容量时数据迁移的开销,兼顾副本的一致性和可用性。一致性散列算法因其不需要查表或通信即可定位数据,计算复杂度不随数据量的增长而改变,且效率高、均匀性好、增加或减少节点时数据迁移量小,受到开发者的喜爱。但具体到实际应用中,这种算法也因其自身局限性而遇到了诸多挑战,如在存储区块链场景下,几乎不可能获取全局视图,甚至没有一刻是稳定的;在企业级 IT 场景下,存在多副本可靠存储问题,数据迁移开销巨大。