

第5章

网络盛宴里的新生代

网络时代,新生事物源源不断涌入到人们的生活中,为人们的生活提供了极大的便利。本章介绍了云计算、大数据和物联网的概念、原理及用途,为我们认知这个世界又打开了一个新的窗口。

5.1 云计算

5.1.1 云计算的概念



云计算

云计算是一种基于互联网的分布式计算,在虚拟的数据中心里,成千上万台计算机和服务器的连接成一片“云”,形成超级强大的运算服务能力,甚至可以体验到每秒钟 10 万亿次的运算。拥有如此强大的计算能力,可以模拟核爆炸、预测气候变化和市场发展趋势。用户通过计算机、笔记本电脑、手机等终端设备接入数据中心,按自己的需求共享这种超级运算。云计算有狭义云计算和广义云计算之分。

1. 狭义云计算

提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的,并且可以随时获取,按需使用,随时扩展,按使用付费。这种特性经常被称为像水电一样使用 IT 基础设施。

2. 广义云计算

这种云计算服务可以是 IT 和软件、互联网相关的,也可以是任意其他的服务。

这种资源池称为“云”。“云”是一些可以自我维护 and 管理的虚拟计算资源,通常为一些大型服务器集群,包括计算服务器、存储服务器、宽带资源等。云计算将所有的计算资源集中起来,并由软件实现自动管理,无须人为参与。这使得应用提供者无须为烦琐的细节而烦恼,能够更加专注于自己的业务,有利于创新和降低成本。

云计算是并行计算(Parallel Computing)、分布式计算(Distributed Computing)和网格计算(Grid Computing)的发展,或者说是这些计算机科学概念的商业实现。云计算是虚拟化(Virtualization)、效用计算(Utility Computing)、IaaS(基础设施即服务)、PaaS(平台即服务)、SaaS(软件即服务)等概念混合演进并跃升的结果。

总的来说,云计算可以算作网格计算的一个商业演化版。

3. 私有云

私有云(Private Clouds)是为一个客户单独使用而构建的,因而提供对数据、安全性和



私有云

服务质量的最有效控制。该公司拥有基础设施,并可以控制在此基础设施上部署应用程序的方式。私有云可部署在企业数据中心的防火墙内,也可以将它们部署在一个安全的主机托管场所。私有云极大地保障了安全问题,目前有些企业已经开始构建自己的私有云。

优点:提供了更高的安全性,因为单个公司是唯一可以访问它的指定实体。这也使组织更容易定制其资源以满足特定的IT要求。

缺点:安装成本很高。此外,企业仅限于合同中规定的云计算基础设施资源。私有云的高度安全性可能会使得从远程位置访问也变得很困难。

4. 公有云

公有云通常指第三方提供商用户能够使用的云,公有云一般可通过Internet使用,可能是免费或成本低廉的。这种云有许多实例,可在当今整个开放的公有网络中提供服务。公有云的最大意义是能够以低廉的价格,提供有吸引力的服务给最终用户,创造新的业务价值。公有云作为一个支撑平台,还能够整合上游的服务(如增值业务、广告)提供者和下游的最终用户,打造新的价值链和生态系统。它使客户能够访问和共享基本的计算机基础设施,其中包括硬件、存储和带宽等资源。

优点:除了通过网络提供服务外,客户只需为他们使用的资源支付费用。此外,由于组织可以访问服务提供商的云计算基础设施,因此他们无须担心自己安装和维护的问题。

缺点:与安全有关。公共云通常不能满足许多安全法规遵从性要求,因为不同的服务器驻留在多个国家,并具有各种安全法规。而且,网络问题可能发生在在线流量峰值期间。虽然公共云模型通过提供按需付费的定价方式通常具有成本效益,但在移动大量数据时,其费用会迅速增加。

5. 混合云

混合云是公有云和私有云两种服务方式的结合。由于安全和控制原因,并非所有的企业信息都能放置在公有云上,这样大部分已经应用云计算的企业将会使用混合云模式。很多企业选择同时使用公有云和私有云,有一些也会同时建立公众云。因为公有云只会向用户使用的资源收费,所以集中云将会变成处理需求高峰的一个非常便宜的方式。比如对一些零售商来说,他们的操作需求会随着假日的到来而剧增,或者是有些业务会有季节性的上扬。同时混合云也为其他目的的弹性需求提供了一个很好的基础,比如灾难恢复。这意味着私有云把公有云作为灾难转移的平台,并在需要的时候去使用它。这是一个极具成本效应的理念。另一个好的理念是,使用公有云作为一个选择性的平台,同时选择其他的公有云作为灾难转移平台。

优点:允许用户利用公共云和私有云的优势。还为应用程序在多云环境中的移动提供了极大的灵活性。此外,混合云模式具有成本效益,因为企业可以根据需要决定使用成本更昂贵的云计算资源。

缺点:因为设置更加复杂而难以维护和保护。此外,由于混合云是不同的云平台、数据和应用程序的组合,因此整合可能是一项挑战。在开发混合云时,基础设施之间也会出现主要的兼容性问题。



公有云

5.1.2 云计算的原理及特点

1. 云计算的原理

云计算的基本原理是,通过使计算分布在大量的分布式计算机上,而非本地计算机或远程服务器中,企业数据中心的运行将更与互联网相似。这使得企业能够将资源切换到需要的应用上,根据需求访问计算机和存储系统。

这是一种革命性的举措,打个比方,这就好比是从古老的单台发电机模式转向了电厂集中供电的模式。它意味着计算能力也可以作为一种商品进行流通,就像煤气、水电一样,取用方便,费用低廉。最大的不同在于,它是通过互联网进行传输的。

云计算的蓝图已经呼之欲出:在未来,只需要一台笔记本电脑或者一个手机,就可以通过网络服务来实现我们需要的一切,甚至包括超级计算这样的任务,如图 5-1 所示。从这个角度而言,最终用户才是云计算的真正拥有者。

云计算的应用包含这样的一种思想,把力量联合起来,给其中的每一个成员使用。

2. 云计算的特点

1) 数据安全可靠

首先,云计算提供了最可靠、最安全的数据存储中心,用户不用再担心数据丢失、病毒入侵等麻烦。

很多人觉得数据只有保存在自己看得见、摸得着的计算机里才最安全,其实不然。你的计算机可能会因为自己不小心而被损坏,或者被病毒攻击,导致硬盘上的数据无法恢复,而有机会接触你的计算机的不法之徒则可能利用各种机会窃取你的数据。

反之,当你的文档保存在类似 Google Docs 的网络服务上,当你把自己的照片上传到类似 Google Picasa Web 的网络相册里,你就再也不用担心数据的丢失或损坏。因为在“云”的另一端,有全世界最专业的团队来帮你管理信息,有全世界最先进的数据中心来帮你保存数据。同时,严格的权限管理策略可以帮助你放心地与你指定的人共享数据。这样,你不用花钱就可以享受到最好、最安全的服务,甚至比在银行里存钱还方便。

2) 客户端需求低

其次,云计算对用户端的设备要求最低,使用起来也最方便。

大家都有过维护个人计算机上种类繁多的应用软件的经历。为了使用某个最新的操作系统,或使用某个软件的最新版本,必须不断升级自己的计算机硬件;为了打开朋友发来的某种格式的文档,不得不疯狂寻找并下载某个应用软件;为了防止在下载时引入病毒,不得不反复安装杀毒和防火墙软件。所有这些麻烦事加在一起,对于一个刚刚接触计算机,刚刚接触网络的新手来说不啻一场噩梦!如果你再也无法忍受这样的计算机使用体验,云计算也许是你的最好选择。只要有一台可以上网的电脑,有一个喜欢的浏览器,你要做的就是

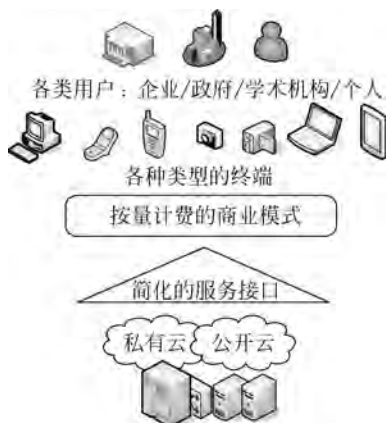


图 5-1 云计算模型

浏览器中输入 URL,然后尽情享受云计算带来的无限乐趣。

你可以在浏览器中直接编辑存储在“云”的另一端的文档,你可以随时与朋友分享信息,再也不用担心软件是否是最新版本,再也不用为软件或文档染上病毒而发愁,因为在“云”的另一端,有专业的 IT 人员帮你维护硬件,帮你安装和升级软件,帮你防范病毒和各类网络攻击,帮你做你以前在个人计算机上所做的一切。

3) 轻松共享数据

此外,云计算可以轻松实现不同设备间的数据与应用共享。

不妨回想一下,你自己的联系人信息是如何保存的。一个最常见的情形是,你的手机里存储了几百个联系人的电话号码,你的个人计算机或笔记本电脑里则存储了几百个电子邮件地址。为了方便在出差时发邮件,你不得不在个人计算机和笔记本电脑之间定期同步联系人信息。买了新的手机后,你不得不在旧手机和新手机之间同步电话号码。

考虑到不同设备的数据同步方法种类繁多,操作复杂,要在这许多不同的设备之间保存和维护最新的一份联系人信息,你必须为此付出难以计数的时间和精力。这时,你需要用云计算来让一切都变得更简单。在云计算的网络应用模式中,数据只有一份,保存在“云”的另一端,你的所有电子设备只需要连接互联网,就可以同时访问和使用同一份数据。

仍然以联系人信息的管理为例,当你使用网络服务来管理所有联系人的信息后,你可以在任何地方用任何一台计算机找到某个朋友的电子邮件地址,可以在任何一部手机上直接拨通朋友的电话号码,也可以把某个联系人的电子名片快速分享给好几个朋友。当然,这一切都是在严格的安全管理机制下进行的,只有对数据拥有访问权限的人,才可以使用或与他人分享这份数据。

4) 可能无限多

最后,云计算为人们使用网络提供了几乎无限多的可能。

为存储和管理数据提供了几乎无限多的空间,也为人们完成各类应用提供了几乎无限强大的计算能力。想像一下,当你驾车出游的时候,只要用手机联入网络,就可以直接看到自己所在地区的卫星地图和实时的交通状况,可以快速查询自己预设的行车路线,可以请网络上的好友推荐附近最好的景区和餐馆,可以快速预订目的地的宾馆,还可以把自己刚刚拍摄的照片或视频剪辑分享给远方的亲友……

离开了云计算,单单使用个人计算机或手机上的客户端应用,我们是无法享受这些便捷的。个人计算机或其他电子设备不可能提供无限量的存储空间和计算能力,但在“云”的另一端,由数千台、数万台甚至更多服务器组成的庞大的集群却可以轻易地做到这一点。个人和单个设备的能力是有限的,但云计算的潜力却几乎是无限的。当你把最常用的数据和最重要的功能都放在“云”上时,我们相信,你对计算机、应用软件乃至网络的认识会有翻天覆地的变化,你的生活也会因此而改变。

互联网的精神实质是自由、平等和分享。作为一种最能体现互联网精神的计算模型,云计算必将在不远的将来展示出强大的生命力,并将从多个方面改变人们的工作和生活。无论是普通网络用户,还是企业员工,无论是 IT 管理者,还是软件开发人员,他们都能亲身体验到这种改变。

5.1.3 云计算的服务类型

云计算主要有 3 种服务类型：基础架构即服务、软件即服务和面向广大互联网开发者的平台即服务。

1. 基础架构即服务

基础架构即服务 (Infrastructure as a Service, IaaS) 一般面向的是企业用户, 它的代表有 Amazon 的 AWS (Amazon Web Service), 还有国内的 PPPOCloud 等。

这种云计算最大的特征在于, 它并不像传统的服务器租赁商一样出租具体的服务器实体, 它出租的是服务器的计算能力和存储能力。AWS 将 Amazon 计算中心的所有的服务器的计算能力和存储能力整合成一个整体, 然后将其划分为一个个虚拟的实例, 每一个实例代表着一定的计算能力和存储能力。购买 AWS 云计算服务的公司就以这些实例作为计量单位。

基础架构即服务与平台即服务有显著的区别, 基础架构即服务提供的只有计算能力和存储能力的服务, 平台即服务除了提供计算能力和存储能力的服务, 还提供给开发者完备的开发工具包和配套的开发环境。也就是说, 开发者使用平台即服务时, 可以直接开始进行开发工作。而使用基础架构即服务时, 则必须先进行如安装操作系统、搭建开发环境等准备工作。基础架构即服务是云计算的基石, 平台即服务和软件即服务构建在它的上面, 分别为开发者和消费者提供服务, 而它本身则为大数据服务。

2. 软件即服务

软件即服务 (Software as a Service, SaaS) 是普通消费者可以感知到的云计算, 它的代表有 Dropbox, 还有国内用户熟悉的百度云、腾讯微云等。这种云计算最大的特征就是消费者并不购买任何实体的产品, 而是购买具有与实体产品同等功能的服务。

以前, 人们花钱购买的是服务器上的存储空间。现在, 人们花钱购买的是 Dropbox 的存储服务。表面上看, 两者没有实际的区别。但是换一个角度来看, 两者却完全不同。以前花钱购买服务器上的存储空间, 假设空间容量是 10GB, 我们是真正地买到了服务器上的 10GB 空间。如果不上传文件的话, 那么服务器上的这 10GB 空间就是空的。现在, 我们购买 Dropbox 的存储服务, 假设空间容量还是 10GB, 却并没有真正地买到 Dropbox 服务器上 10GB 的空间, 而买到的是 10GB 空间的服务。也就是说, 如果上传文件, Dropbox 会将文件分开放在任何地方的任何服务器上, 如果不上传文件, Dropbox 的服务器上就根本没有属于我们的任何空间。

3. 平台即服务

与软件即服务不同, 平台即服务 (Platform as a Service, PaaS) 是面向开发者的云计算。这种云计算最大的特征是它自带开发环境, 并向开发者提供开发工具包。它的代表有 Google 的 GAE (Google App Engine), 还有国内的百度的 BAE、新浪的 SAE 等。

平台即服务与软件即服务之间可以相互转换。如果是消费者, 购买 Dropbox 的服务, 那 Dropbox 就是软件即服务。如果是开发者, 利用 Dropbox 提供的开发包借助 Dropbox 的服务开发自己的服务, 那么 Dropbox 本身就是平台即服务, 构筑在 Dropbox 之上的开发者的服务就是软件即服务。

以前,开发者如果要搭建一个网站,需要做很多准备工作,比如购买服务器,安装操作系统,搭建开发环境等。现在,开发者如果购买平台即服务云计算,就可以省去上面费时费力的准备工作,直接进行网站的开发。不仅如此,开发者还可以使用各种现成的服务,比如 GAE 会向开发者提供 Google 内部使用的先进的开发工具和领先的大数据技术。这一切都使得网站开发变得比以前轻松很多,这也是云计算时代互联网更加繁荣的原因之一。

5.1.4 云计算应用案例

1. 应用背景

H3C 提供大量的网络设备和网络机房运维服务,拥有完善的售后服务体系,大量的售后技术人员为用户提供技术支撑服务,H3C 业务的特性决定了售后技术支持人员需要到客户现场处理技术、软硬件故障等一系列问题。同时,由于 H3C 产品和解决方案众多,客户数量庞大,对售后支持有着严格的审批和处理流程,然而售后技术服务管理的系统位于 H3C 内部网络,这就造成了 H3C 技术支持人员在客户现场无法根据现场情况实时派发售后工单,而只能在现场了解客户情况后,回到内部网络在开始售后技术支持工单的审批工作,或者在客户现场通过 VPN 访问内部网络系统派发工单,这无疑严重影响了售后技术支持人员的工作效率,也降低了 H3C 带给客户良好的售后服务体验。

2. 技术方案

H3C 经过对微软 Azure 云平台的评估和测试,决定分步骤分阶段地将业务系统迁移到 Azure 上。H3C 在杭州有自建的数据中心,拥有丰富的计算资源和大量关键的业务系统和核心数据,在使用 Azure 的同时,需要确保与自有数据中心的互联互通,对系统做出平滑的迁移,而不影响服务的可用性,在业务上相互协作。

H3C Learning 是迁移到 Azure 上的第一个业务系统,Azure 中虚拟机(VM)的应用是整个业务系统中的重点。H3C 通过在 Azure 中快速产生和部署虚拟机,将 H3C Learning 整个业务系统通过虚拟机镜像的方式快速迁移到了 Azure 中。H3C 部署了两台 CentOS 的虚拟机,通过 Apache 搭建 Web Server 服务器,用于用户访问 H3C Learning 的前端 Web 展示页面,使用 Azure LB 实现了负载均衡。同时使用了一台 Windows Server 虚拟机,部署了文件服务器,用于 H3C Learning 的各种类型的文件数据存储。H3C Learning 系统使用了 Oracle 的数据库服务,通过在 Azure 上生成一台新的虚拟机,将自由数据中心的 Oracle 数据库整体迁移到了 Azure 的虚拟机中,实现了业务系统的平滑迁移。

H3C 使用 Azure 提供点到点的 VPN 服务,通过配置和管理 Azure 中的虚拟专用网络,将其与原有数据中心的 IT 基础设施的 VPN 进行安全连接。在使用 Azure 的 VPN 服务时,由于 H3C 自身的产品就有 VPN 客户端,但是 H3C 的客户端并不在 Azure 的支持列表中,通过对实际环境的评估以及提供的适配参数,H3C 选择了一款产品并与 Azure 成功适配,满足了业务需求。通过虚拟网络,H3C 将原有数据中心扩展到云端,使得部署在 Azure 中的应用程序安全连接到原有数据中心的本地系统中。

为了解决售后技术支持人员在客户现场的业务需求,H3C 基于移动设备开发了移动端的 App,以便售后技术支持人员能随时随地地访问内部办公系统,并生成工单,快速进入业务的审批流程。H3C 将移动 App 的前端整体迁移到了 Azure 中,通过生成新的虚拟机部

署移动端 App 接口服务,接受来自客户端的请求,而后端服务仍在自有数据中心。用户通过移动端 App 访问 Azure 中的前端服务器,对于用户简单的业务请求直接返回用户请求数据,而对于需要与自有数据中心的后端服务器交互的数据,通过 VPN 将请求发送至后端服务器处理,并返回数据到 Azure 中的前端服务器,完成用户不同业务层面的需求,整个过程属于无缝连接与切换,用户完全感受不到,如图 5-2 所示。

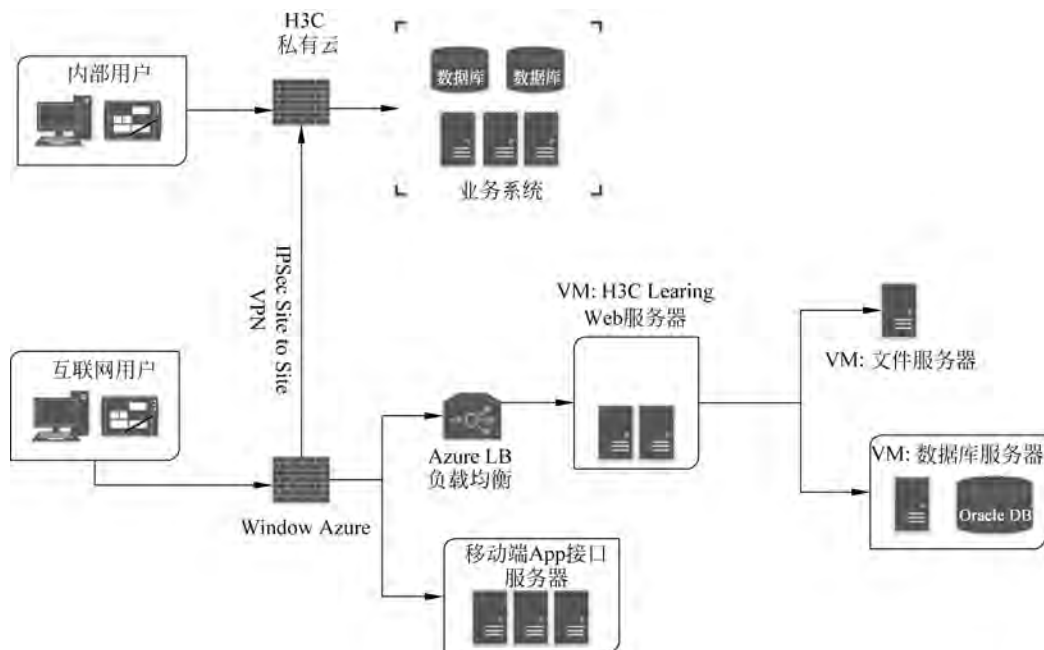


图 5-2 H3C 云计算系统架构

5.2 大数据

5.2.1 大数据的概念

大数据(Big Data)指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合,是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产;是目前存储模式与能力、计算模式与能力不能满足存储与处理现有数据集规模产生的相对概念。麦肯锡全球研究所给出的定义是:一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合,具有海量的数据规模(Volume)、快速的数据流转(Velocity)、多样的数据类型(Variety)和价值密度低(Value)四大特征。

5.2.2 大数据的特点

大数据带来的信息风暴正在变革人们的生活、工作和思维,大数据开启了一次重大的时代转型;大数据时代最大的转变就是,放弃对因果关系的渴求,而取而代之关注相关关系。也就是说,只要知道“是什么”,而不需要知道“为什么”。这就颠覆了千百年来人类的思维惯

例,对人类的认知和与世界交流的方式提出了全新的挑战。

大数据是指“无法用现有的软件工具提取、存储、搜索、共享、分析和处理的海量的、复杂的数据集合。”业界通常用4个V(Volume、Velocity、Variety、Value)来概括大数据的特征。

(1) 数据体量巨大(Volume)。截至目前,人类生产的所有印刷材料的数据量是200PB($1\text{PB}=2^{10}\text{TB}$),而历史上全人类说过的所有的话的数据量大约是5EB($1\text{EB}=2^{10}\text{PB}$)。当前,典型个人计算机硬盘的容量为TB量级,而一些大企业的数据量已经接近EB量级。

(2) 处理速度快(Velocity)。这是大数据区别于传统数据挖掘的最显著特征。根据IDC的“数字宇宙”的报告,到2020年,全球数据使用量达到35.2ZB。在如此海量的数据面前,处理数据的效率就是企业的生命。

(3) 数据类型繁多(Variety)。这种类型的多样性也让数据被分为结构化数据和非结构化数据。相对于以往便于存储的以文本为主的结构化数据,非结构化数据越来越多,包括网络日志、音频、视频、图片、地理位置信息等,这些多类型的数据对数据的处理能力提出了更高要求。

(4) 价值密度低(Value)。价值密度的高低与数据总量的大小成反比。以视频为例,一部1小时的视频,在连续不间断的监控中,有用数据可能仅有一两秒。如何通过强大的机器学习更迅速地完成任务“提纯”成为目前大数据背景下亟待解决的难题。

5.2.3 大数据的主要问题

计算机中存在不断变大的数据集,不存在绝对的大数据,计算机中的所有数据集都是有限集合。大数据主要有以下一些问题。

(1) 大数据采样——把大数据变小、找到与算法相适应的极小样本集、采样对算法误差的影响。

(2) 大数据表示——表示决定存储、表示影响算法效率。

(3) 大数据不一致问题——导致算法失效和无解、如何消解不一致。

(4) 大数据中的超高维问题——超高维导致数据稀疏、算法复杂度增加。

(5) 大数据中的不确定维问题——多维度数据并存、按任务定维难。

(6) 大数据中的不适宜性问题——高维导致问题的解太多难以抉择。

5.2.4 大数据的处理过程

大数据处理分为四步,分别是数据获取、数据抽取、统计分析、数据挖掘。大数据处理的普遍流程至少应该满足这四个方面的步骤,才能算得上是一个比较完整的大数据处理。具体过程如图5-3所示。

1. 数据获取

大数据的采集是指利用多个数据库接收发自客户端的数据,并且用户可以通过这些数据库进行简单的查询和处理工作。在大数据的采集过程中,其主要特点和挑战是并发数高,因为有可能会有成千上万的用户同时进行访问和操作,比如火车票售票网站和淘宝,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,并且如何在这些数据库之间进行负载均衡和分片也需要深入的思考和设计。

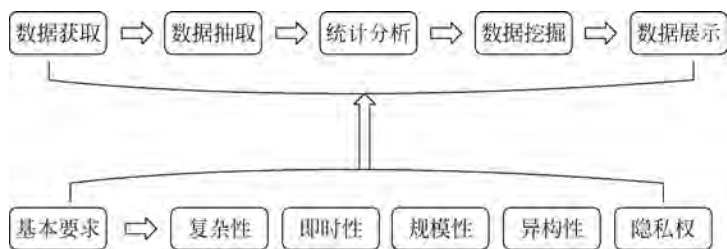


图 5-3 大数据处理过程

2. 数据抽取

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,还是应该将这些来自前端的数据导入到一个集中的大型分布式数据库,或者分布式存储集群,并且可以在导入基础上做一些简单的清洗和预处理工作。导入与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,或者分布式计算集群来对存储于其内的海量数据进行普通的分析和分类汇总等,以满足大多数常见的分析需求。统计分析这部分的主要特点和挑战是分析涉及的数据量大,其对系统资源会有极大的占用。

4. 数据挖掘

与前面统计分析过程不同的是,数据挖掘一般没有什么预先设定好的主题,主要是在现有数据上面进行基于各种算法的计算,从而起到预测的效果,从而实现一些高级别数据分析的需求。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,常用数据挖掘算法都以多线程为主。数据挖掘主要体现在以下几个过程,如图 5-4 所示。

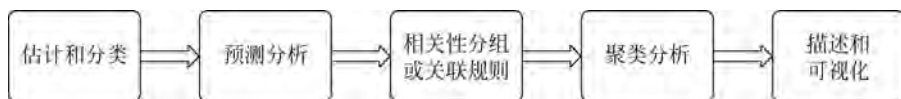


图 5-4 数据挖掘主要步骤

1) 估计和分类

根据所采集的数据,通过估值得到未知的连续变量的值,然后根据预先设定的阈值比对,运用分类技术对数据进行分类,并建立分类模型。

2) 预测分析

通过估值和分类得出模型后,将模型用于对未知变量的预测,并形成阶段性分析结论。

3) 相关性分组或关联规则

决定各预测结果间的联系性,找出关联规则,推导关联过程。

4) 聚类分析

聚类是对挖掘记录或预测结果进行分组,把相似的记录放在一个聚集里做研究。

5) 描述和可视化

描述和可视化是对数据挖掘结果的表示方式。通过数据可视化工具进行数据展现、分析、钻取,将数据挖掘的分析结果以更形象、更深刻、更可视化的方式展现出来。

5.2.5 大数据的发展趋势

趋势一：数据的资源化。

资源化,是指大数据成为企业和社会关注的重要战略资源,并已成为大家争相抢夺的新焦点。因而,企业必须提前制定大数据营销战略计划,抢占市场先机。

趋势二：与云计算的深度结合。

大数据离不开云处理,云处理为大数据提供了弹性可扩展的基础设备,是产生大数据的平台之一。自2013年开始,大数据技术已开始和云计算技术紧密结合,预计未来两者关系将更为密切。除此之外,物联网、移动互联网等新兴计算形态,也将一齐助力大数据革命,让大数据营销发挥出更大的影响力。

趋势三：科学理论的突破。

随着大数据的快速发展,就像计算机和互联网一样,大数据很有可能是新一轮的技术革命。随之兴起的数据挖掘、机器学习和人工智能等相关技术,可能会改变数据世界里的很多算法和基础理论,实现科学技术上的突破。

趋势四：数据科学和数据联盟的成立。

未来,数据科学将成为一门专门的学科,被越来越多的人所认知。各大高校将设立专门的数据科学类专业,也会催生一批与之相关的新的就业岗位。与此同时,基于数据这个基础平台,也将建立起跨领域的数据共享平台,之后,数据共享将扩展到企业层面,并且成为未来产业的核心一环。

趋势五：数据泄漏泛滥。

未来几年数据泄漏事件的增长率也许 would 达到100%,除非数据在其源头就能够得到安全保障。可以说,在未来,每个财富500强企业都会面临数据攻击,无论他们是否已经做好安全防范。而所有企业,无论规模大小,都需要重新审视今天的安全定义。在财富500强企业中,超过50%将会设置首席信息安全官这一职位。企业需要从新的角度来确保自身以及客户数据,所有数据在创建之初便需要获得安全保障,而并非在数据保存的最后一个环节,仅仅加强后者的安全措施已被证明于事无补。

趋势六：数据管理成为核心竞争力。

数据管理成为核心竞争力,直接影响财务表现。当“数据资产是企业核心资产”的概念深入人心之后,企业对于数据管理便有了更清晰的界定,将数据管理作为企业核心竞争力,持续发展,战略性规划与运用数据资产,成为企业数据管理的核心。数据资产管理效率与主营业务收入增长率、销售收入增长率显著正相关。此外,对于具有互联网思维的企业而言,数据资产竞争力所占比重为36.8%,数据资产的管理效果将直接影响企业的财务表现。

趋势七：数据质量是BI(商业智能)成功的关键。

采用自助式商业智能工具进行大数据处理的企业将会脱颖而出。其中要面临的一个挑战是,很多数据源会带来大量低质量数据。想要成功,企业需要理解原始数据与数据分析之间的差距,从而消除低质量数据并通过BI获得更佳决策。

趋势八：数据生态系统复合化程度加强。

大数据的世界不只是一个单一的、巨大的计算机网络,而是一个由大量活动构件与多元参与者元素所构成的生态系统,终端设备提供商、基础设施提供商、网络服务提供商、网络接

入服务提供商、数据服务使能者、数据服务提供商、触点服务、数据服务零售商等一系列的参与者共同构建的生态系统。而今,这样一套数据生态系统的基本雏形已然形成,接下来的发展将趋向于系统内部角色的细分,也就是市场的细分;系统机制的调整,也就是商业模式的创新;系统结构的调整,也就是竞争环境的调整等,从而使得数据生态系统复合化程度逐渐增强。

5.2.6 大数据的应用方向

传统产业人士通常认为大数据是大型互联网公司的“专利”。当龙头企业通过互联网平台将大数据应用连接到传统企业,并带来巨大的商业价值后,通信行业、金融行业、服务零售业以及传统的装备制造业等,都纷纷开始进军大数据。

根据调查显示,32.5%的公司正在搭建大数据平台,处于测试阶段;29.5%的公司已经在生产环节实践大数据,并有成功的产品。总体来看,目前正在开发和已经使用的大数据应用平台占比超过60%,而准备开发的占24.52%,并且这个比例还会日渐上升,说明企业对大数据的需求明显加大。

下面给出几个大数据应用方向。

1. 对顾客群体细分

大数据可以对顾客群体细分,然后对每个群体量体裁衣般地区分服务。瞄准特定的顾客群体来进行营销和服务是商家一直以来的追求。云存储的海量数据和大数据的分析技术使得对消费者的实时和极端的细分有了成本效率极高的可能。

2. 个性化精准推荐

在企业运营商内部,根据用户喜好推荐各类业务及应用是常见的,比如应用商店软件推荐、IPTV 视频节目推荐等,而通过关联算法、文本摘要抽取、情感分析等智能分析算法后,可以将之延伸到商用化服务,利用数据挖掘技术帮客户进行精准营销,今后的盈利可以来自于客户增值部分的分成。

很多人讨厌广告的原因,在于它推送的是无用的信息。互联网的出现更是放大了这一特点,而如今人们发现自己搜索过的或者买过的商品都能被针对性地推荐,出现在浏览的网页广告中。这便是随着信息数量的持续增加,大数据的到来。在这些数据中,隐藏了消费者的消费习惯、市场的变化、产品的趋势以及大量的历史记录,这些关键数据对于企业和组织的后续运营和发展起到了至关重要的作用。更准确的营销手段已经成为一种广告工具,这种个性化的广告推广,主要是为了缩小范围来针对某一类人群。

3. 数据精准搜索

数据搜索是一个并不新鲜的应用,随着大数据时代的到来,实时性、全范围搜索的需求也就变得越来越强烈。我们需要能搜索各种社交网络、用户行为等数据。

其商业应用价值是将实时的数据处理分析和广告联系起来,即实时广告业务和应用内移动广告的社交服务。运营商掌握的用户网上行为信息,使得所获取的数据具备更全面的维度,更具商业价值。

4. 数据存储空间出租

企业和个人有着海量信息存储的需求,只有将数据妥善存储,才有可能进一步挖掘其潜在价值。具体而言,这块业务模式又可以细分为针对个人文件存储和针对企业用户两大类。

主要是通过易于使用的 API,用户可以方便地将各种数据对象放在云端,然后再像使用水、电一样按用量收费。

5.2.7 大数据应用案例：业务分析

1. 应用背景

随着余额宝用户数持续呈指数级增长,数据量也成倍增长。在这种情况下,已经无法通过简单的 Hadoop 集群进行数据的管理工作,而业务端面临需要通过数据了解用户、分析行为进而对业务决策和用户行为进行精准预测。基于这些业务的需求驱动需要一个大数据平台来承载,我们在对稳定性、成本、自身能力和复杂度等进行综合考量后,决定采用当前最流行和最成熟的云平台阿里云 MaxCompute。

2. 技术方案

从技术指标的角度搭建大数据平台是达到数据存储和数据计算两大目标,而从各个业务环节的角度看是数据采集、数据清洗、在线/离线分析与预测、实时/非实时查询。而业务目标是为了能够快速响应业务需求,能够为业务分析提供稳定的开发和建模平台,为业务提供逻辑清晰和灵活便捷的可视化平台,从而实现从数据支持业务到数据驱动业务的逐步升级。

整个架构都是搭建在阿里云上的,该架构是成熟的三层架构:采集层+整合层+应用层,如图 5-5 所示。

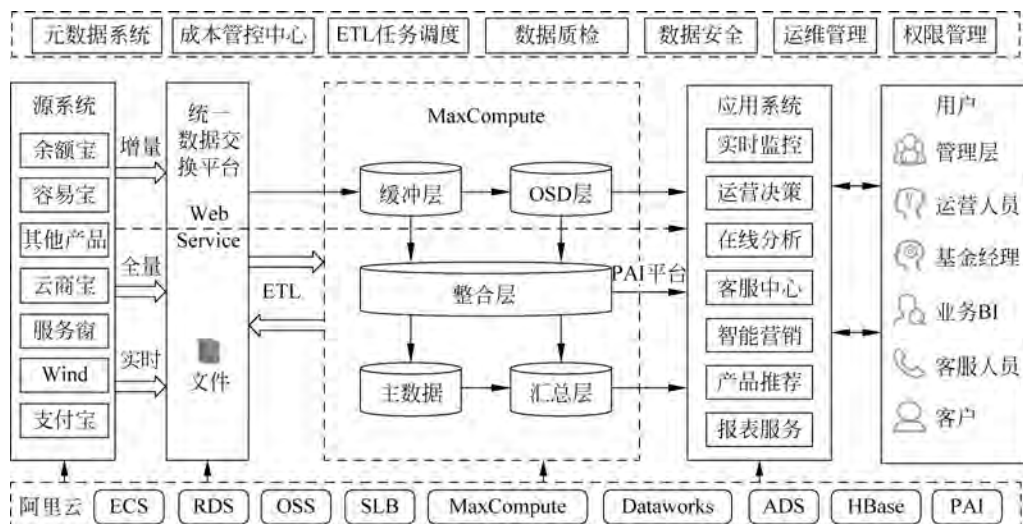


图 5-5 大数据架构案例

采集层对接了几乎所有的业务,采集数据的频率有实时的、分钟级、小时级、日级、月级,支持不同的采集频率,而且这些都是灵活可配置的。将采集的数据通过企业级的数据交换平台进行存储和交换,该平台使用 OSS 实现。通过 OSS 可以实现数据的中转、分发和备份存储。

在 MaxCompute 整个整合层包含五大区:缓冲区、ODS 区、整合区、主数据和汇总区。不同的区域实现不同的功能。缓冲区是为了在正式进入数仓应用数据模块之前进行数据质

检,满足质检后方可进行真正的加工处理,避免因为数据错误污染整个数仓的数据;ODS区是为了保留源系统格式的数据模块,一方面能够在有问题时追根溯源,另一方面能够满足部分业务的需要;整合区是数据仓库的核心区域,通过主题建模的方式进行数据的模型化处理,使得数据的解释口径具有统一性;主数据则是与业务结合比较紧密的主题数据,这样更方便业务方的使用;汇总区则是提前将需要预加工统计的数据进行统计计算,避免多次开发计算带来的时间成本、开发成本和计算成本等。

应用层主要是通过监控、管理看板、报表等可视化系统给业务提供直观的数据呈现,而为业务的决策提供更加有力的数据支撑。在应用层通过 RDS、ADS、HBase 等不同的产品满足了不同的需求。

数据仓库是一个比较复杂的系统,需要很多配套的系统辅助才能做好这样的项目。而其中很多系统在 MaxCompute、DataWorks 中都已经产品化,大大简化了大数据平台的搭建和运维,提供了一站式的解决方案,而且通过阿里云 MaxCompute、Dataworks 能够实现敏捷开发、快速响应、轻量化运维,低成本地实现大数据平台架构。其中包括最核心的调度系统、权限管理、元数据管控、数据安全保护伞等一系列功能。而在使用中,数据分析师能够快速上手完成数据的加工和分析。

5.2.8 大数据应用案例:病虫害检测

1. 应用背景

传统方法处理茶树或者茶叶信息,都是在实验室里面完成的,是静止的且数据量比较小,而基于大数据处理的茶园病虫害检测系统所采取的方法是对整个茶园可视的每一株每一叶进行处理,是一个海量数据库的系统,所以把它归到大数据处理方法。

茶园种植面积普遍有上亿平方米,茶园病虫害检测系统数据的采集分为两种,第一种是在茶园里布置很多地面传感结点,采集茶园的湿度、湿度、风向、风速、光照度和降水量等信息;第二种方法是用无人机在整个茶园的上空采集视频信息,通过相关软件,比如在 PhotoScan 中呈现三维效果,管理者和消费者将来都会通过这个模型,看到地面以及高空的一些信息。这样,大家都能够通过这个平台很好地掌握整个茶园茶树的生长过程。茶园传感结点搜集的数据,会有一个专门开发的服务器进行收集,大量服务器可以放在 DC(Data Center)机房里,为了能更长时间、更加清晰地保存这些信息,服务器容量也会很大。

采用无人机采集茶园图像信息的过程:控制无人机起飞到达高空中,然后在中间点悬停,以 365°定点旋转拍摄全景照片,通过 4G 无线传输的方式将采集到的图像传回到后台服务器中,然后再通过 PhotoScan 做三维建模,再采用地面信息补充模型。例如,茶叶上很多都有红色的斑,是炭疽病的一种表现,通过航拍的方式以及地面的一些信息采集来识别这一种病害。

由于采用的数据量很大,所以可以采用大数据处理算法来识别这些病害。茶叶上面的这些红斑,如果是通过图像处理的方法,是要识别它的轮廓,也可以采用光谱分析方法,不识别它的轮廓,就可以识别出来了。

2. 技术方案

茶园病虫害预警系统有很多关键模块,主要有以下几部分。

1) 茶园信息采集模块

该模块包括宏观的卫星图像,也有用四旋翼无人机对整个茶园进行低空拍摄的微观图

像,另外还有采集茶园的湿度、风向、风速、光照度和降水量等信息的传感网络。

2) 处理模块

系统需要有病虫害分析算法的计算服务器对这些数据进行处理,把处理结果通过专业设备展现给生产者和消费者,也需要有一个庞大的 DC 机房来存储这些数据。

3) 展示模块

数据展示还需要基于安卓开发 App 应用程序,生产者可以通过数据显示结果来调控执行部件,消费者可以通过数据显示结果对茶叶产品的安全放心。

茶园病虫害预警系统结构框架图见图 5-6。

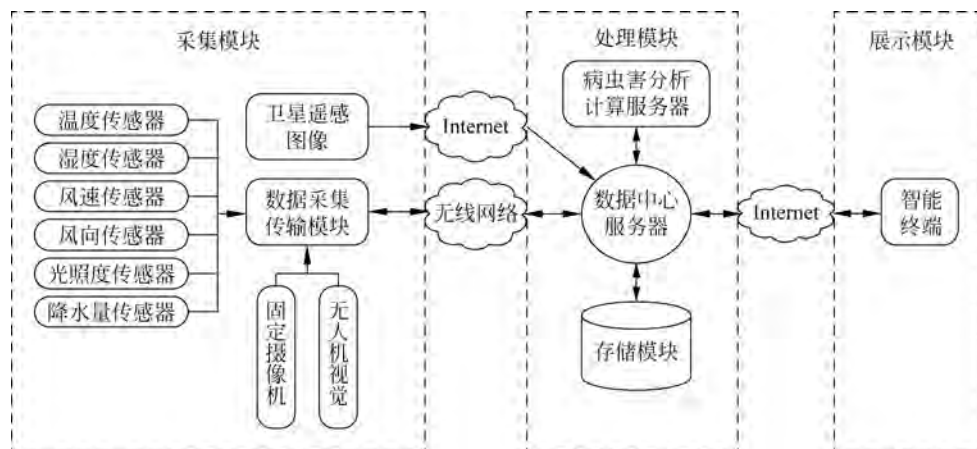


图 5-6 茶园病虫害预警系统结构框架图

通过在茶园里布置的大量传感结点,连续采集地面的温度、湿度、风向、风速、光照度和降水量等微气候要素大规模数据,这些气象数据也可以通过机器学习和深度学习的方法找到与病虫害的相关性。

采用传统的卫星遥感图像,分辨率达不到分析要求,而固定摄像机采集图像的面积是有限的,采用配高清摄像机的低空四旋翼无人机采集图像,涉及的面积是比较大的,整个茶园只需要一架四旋翼无人机就可以采集完整茶园图像,并且用 4G 模块可以实时将数据上传到服务器。这其中涉及的大数据首先体现在图像数据量会比较大,而且还要连续从图像中提取有价值的信息,并把这些信息实时地发送给服务器。

对下载的卫星遥感图像 MODIS 数据采用 ENVI 软件处理,首先进行校正、裁剪、拼接等预处理操作,然后计算 NDVI 值,归一化后得到 MODIS 图像。因为 MODIS 数据选择在出现病虫害发生之后的一段时间,这个值在一定程度上反映了病虫害对茶叶的破坏作用。

项目用 PhotoScan 软件基于无人机低空遥感图像对南京市六合区平山茶园部分区域构建 3D 模型,见图 5-7,该 3D 模型便于茶园大数据可视化分析和展示。

这个系统大数据不但能够体现在对病虫害的研究上,还可以研究茶叶的生长过程,以及茶叶销售的整个销售链过程。使得消费者不仅可以事后溯源,也可以实时同步观测生产者是如何管理茶叶的生长,确保茶叶质量,与智慧农业有效地结合起来,可见大数据处理技术在农业领域有着非常广阔的应用前景。

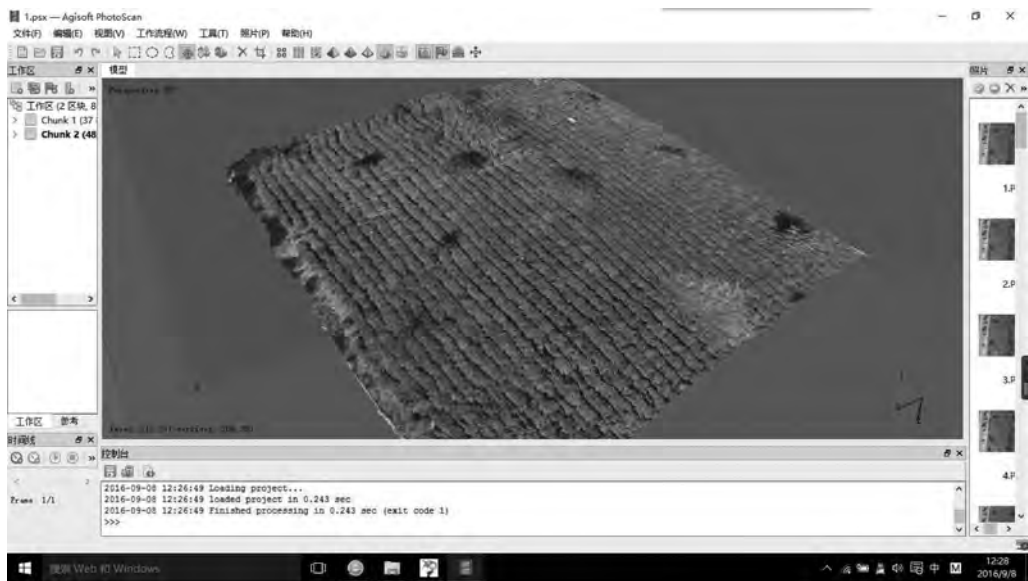


图 5-7 基于 PhotoScan 的茶园 3D 建模

5.2.9 大数据应用案例：内涝监测

1. 应用背景

南京夏季经常下暴雨,可以通过城市内涝监测系统来实时观测城市淹水的情况。城市内涝监测系统由宏观和微观两个系统组成,宏观系统是通过卫星遥感图像来对整个城市的宏观内涝情况进行分析;微观系统是通过每一个小区、每一条道路布置的监控摄像机来解析微观积水情况。

从大数据层面上来研究到现有的局部布点系统有了一个很大的提升。之前我们国家可能做了一些防汛抗灾系统,针对某一个点历史数据进行预测,是静态的、非实时性的。而从大数据层面来讲,研究的系统是动态的、连续的和实时的。降雨时雨量在不断地增加,暴雨导致城市的积水量在不断变化,这些都需要实时监测,而且是基于视频监控,因此这个数据量是巨大的,就是依据这样的大数据来对整个城市的内涝进行实时解析,来通过实时监测平台展示给防灾减灾部门,甚至个人都可以通过实时监测平台来进行监测,决策自己的出行。基于大数据的城市内涝检测系统是一个时间等待是秒级的实时系统。

2. 技术方案

城市内涝监测系统分为三个子系统,具体如下。

1) 采集模块

采集模块通过宏观和微观手段来采集宏观图像和微观图像。宏观采集系统是通过卫星遥感获取遥感图像,微观采集模块是通过在每一个小区、每一条道路布置的监控摄像机来采集图像。

2) 处理模块

大规模服务器集群作为处理中心,对宏观和微观图像进行解析,尤其是微观的,要通过

积水的面积以及积水的深度来判定受灾的情况。同时用存储模块对有研究价值的图像和处理结果进行保存。

3) 展示模块

数据展示同样还需要基于安卓开发 App 应用程序,需要将结果通过 App 的方式,或者通过其他的一些方式,呈现给使用者,及时地了解灾情。

具体系统结构框架图见图 5-8。

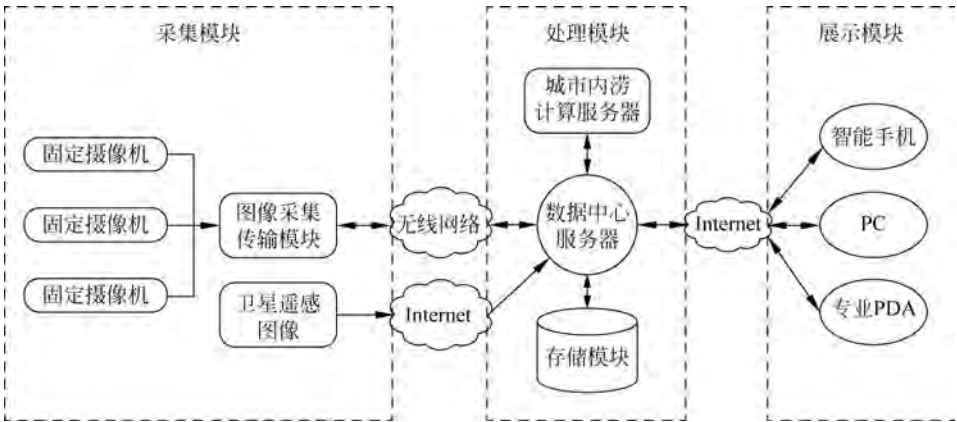


图 5-8 城市内涝监测系统结构框架图

系统主要采用图像处理手段,图像的信息量是非常巨大的,而且采用实时方法一天 24 小时连续采集图像,一个点一天可以达到几十 GB 的数据量,整个城市布置几千个点,这个信息量是非常巨大的,需要用专业的大数据处理软件来对它进行处理,而且要有非常高效的硬件平台支撑,所以这是一个“规模巨大”的大数据处理。

对于数据存放的问题是很多大数据处理所面临的共同问题,通常解决方法有两种,第一种是租用服务器提供商的服务器,但这是非常有限的。对于城市内涝监测系统,需要建设自己专门的 DC 机房,也就是建立服务器集群。根据数据量的需求,DC 机房的规模,取决于数据量的大小,确保有足够的空间来存放海量的实时数据。项目需要存储的数据量,完全是由它本身的量来决定的,当然这也需要有强大的资金来作为支持。

网上曾经有一个非常热门的话题,一个小伙子家里突然受到了水灾,他的一边是他的父亲,另一边是母亲,而这两个地方他到底该先救谁? 因为他不知道哪边受灾更严重,从而产生了一个非常复杂、焦虑的情况。如果用到以上大数据处理的的城市内涝监测系统,就可以从容应对。

在布置监控摄像机的时候会布置其所在位置的对应的二维码,二维码下面会有一个标尺,标记积水深度。利用大数据图像处理的方法,很容易地分析得到整个小区内部的积水情况以及积水深度。以此为依据来判定到底先救谁。所以这个小伙子只要拿着手机看一下,到底是父亲那边淹水严重,还是母亲那边淹水严重,就可以做出非常准确和迅速的判断。从个体来看,可以判别先救谁,而从整个城市宏观来看,可以判定先救哪个区域。

5.3 物联网

2009 年年初,“智慧地球”这一概念由 IBM 公司首先提出。智慧地球也称为智能地球,就是把感应器嵌入和装备到电网、铁路、桥梁、隧道、公路、建筑、供水系统、大坝、油气管道等各种物体中,并且被普遍连接,形成“物联网”,然后将“物联网”与现有的互联网整合起来,实现人类社会与物理系统的整合。

5.3.1 物联网的概念

物联网的英文名称为“The Internet of Things”,简称 IOT。物联网是在互联网、移动通信网等通信网络的基础上,针对不同应用领域的需求,利用具有感知、通信与计算能力的智能物体自动获取物理世界的各种信息,将所有能够独立寻址的物理对象互联起来,实现全面感知、可靠传输、智能处理,构建人与物、物与物互联的网络智能信息服务系统,如图 5-9 所示。

理解物联网的定义与技术特征,需要注意以下 4 个基本的问题。

(1) 物联网是在互联网基础上发展起来的,它与互联网在基础设施上有一定程度的重合,但它不是互联网概念、技术与应用的简单扩展。

(2) 互联网扩大了人与人之间信息共享的深度与广度,而物联网更加强调它在人类社会生活的各个方面、国民经济的各个领域广泛与深入的应用。

(3) 物联网的主要特征是全面感知、可靠传输、智能处理。物联网的基础是感知技术,支撑环境是计算机网络、移动通信网及其他可以用于物联网数据传输的网络,核心价值应该体现在对自动感知的海量数据的智能处理,利用所产生的知识形成反馈控制指令,通过人与执行机制的结合,“智慧”地处理物理世界的问题。

(4) 物联网是一种新的计算模式。它使人类对客观世界具有更透彻的感知能力、更全面的认知能力、更智慧的处理能力,可以在提高人类的生产力、效率、效益的同时,改善人类社会发展与地球生态环境的和谐性以及可持续发展的关系。

物联网是通过智能感知、识别技术与普适计算、泛在网络的融合应用,被称为继计算机、互联网之后世界信息产业发展的第三次浪潮。与其说物联网是网络,不如说物联网是业务和应用,物联网也被视为互联网的应用拓展。因此应用创新是物联网发展的核心,以用户体验为核心的创新是物联网发展的灵魂。

这里的“物”要满足以下条件才能够被纳入“物联网”的范围。

- (1) 要有相应信息的接收器。
- (2) 要有数据传输通路。
- (3) 要有一定的存储功能。

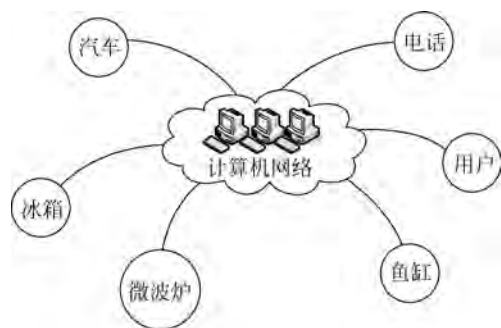


图 5-9 物联网工作示意图

- (4) 要有 CPU。
- (5) 要有操作系统。
- (6) 要有专门的应用程序。
- (7) 要有数据发送器。
- (8) 遵循物联网的通信协议。
- (9) 在世界网络中有可被识别的唯一编号。

5.3.2 物联网的体系结构

物联网是在互联网和移动通信网等网络通信基础上,针对不同领域的需求,利用具有感知、通信和计算的智能物体自动获取现实世界的信息,将这些对象互联,实现全面感知、可靠传输、智能处理,构建人与物、物与物互联的智能信息服务系统。

物联网体系结构主要由三个层次组成:感知层、网络层和应用层,模型如图 5-10 所示。



图 5-10 物联网层次结构模型

1. 感知层

感知层设备主要分为两类:自动感知设备与人工生成信息设备。

一类为能够自动感知外部物理物体与物理环境信息的 RFID 和传感器。传感器根据它能够感知的参数可以分为物理传感器、化学传感器与生物传感器。另一类是用来人工生成信息的智能电子设备,如智能手机、个人数字助理(PDA)、GPS、智能家用电器、智能测控设备,它是自动感知技术的辅助手段。

人们将 RFID 形容成能够让物体“开口”的技术。RFID 标签中存储了物体的信息,通过无线信道将它们储存的数据传送到 RFID 应用系统中。一般的传感器只具有感知周围环境参数的能力。例如,在环境监测系统中,一个温度传感器可以实时地传输它所测量到的环境温度,但是它对环境温度不具备控制能力。而一个精准农业物联网应用系统中的植物定点浇灌传感器结点,系统设计者希望它能够在监测到土地湿度低于某一个设定的数值时,就自动打开开关,给果树或蔬菜浇水,这种感知结点同时具有控制能力。在物联网突发事件应急处理的应用系统中,核泄漏现场处理的机器人可以根据指令进入指定的位置,通过传感器将周边的核泄漏相关参数测量出来,传送给指挥中心。根据指挥中心的指令,机器人需要打开某个开关或关闭某个开关。从这个例子可以看出,作为具有智能处理能力的传感器结点,它必须同时具备感知和控制能力,还应具备适应周边环境的运动能力。因此,从一块简单的 RFID 芯片、一个温度传感器到一个复杂的测控装置和智能机器人,它们之间最重要的区别表现在智能物体是不是需要同时具备感知能力和控制、执行能力,以及需要什么样的控制、执行能力。

2. 网络层

网络层包括接入层、汇聚层和核心交换层。很多文献将这一层叫作“传输层”。需要注意的是:物联网的“网络层”或“传输层”与计算机网络体系结构七层协议中的网络层、传输层内涵是不相同的。物联网的“网络层”或“传输层”相当于计算机网络中传输网的概念。

1) 接入层

物联网接入层相当于计算机网络 OSI 参考模型中的物理层与数据链路层。传感器与接入层设备构成了物联网感知网络的基本单元。

接入层网络技术类型可以分为两类:无线接入与有线接入。无线接入主要有 IEEE 802.15.4 协议、6LoWPAN 协议、蓝牙协议、ZigBee 协议与无线移动通信网的 M2M 协议。有线接入主要有现场总线网接入、电力线接入与电话线接入。从接入层可以选择的各种网络技术的特点可以看出,接入层技术对应于计算机网络体系结构中的物理层与数据链路层,与网络层的 IP 协议无关。因此,可以不必担心目前互联网中 IP 协议安全问题对物联网接入层的影响。

2) 汇聚层

一个实际的公路、铁路、输油管线的安全监控物联网系统由分布在很长的线路上的多个无线传感器网络组成。实际的智能电网由从发电厂、输变电电路到用户智能电表多种感知单元与数据处理单元组成。要将这些系统中的多个感知单元的数据准确、实时、有序地汇集起来,传送到高层数据处理中心,在整体物联网网络拓扑结构设计时也必须考虑在接入层之上加入汇聚层,将汇集、整理后的数据流通过核心交换层传送到高层数据处理中心。

汇聚层位于接入层与核心交换层之间,它的基本功能包括:汇接接入层的用户流量,进行数据分组传输的汇聚、转发与交换;根据接入层的用户流量,进行本地路由、过滤、流量均衡、QoS 优先级管理,以及安全控制、地址转换、流量整形等处理;根据处理结果把用户流量转发到核心交换层或在本地进行路由处理。

汇聚层网络技术可以分为无线与有线两类。无线网络技术主要有符合 IEEE 802.15.4 协议标准的无线个人区域网(WPAN),符合 IEEE 802.11 协议标准的无线局域网

(WLAN),符合 IEEE 802.16 协议标准的无线城域网(WMAN),符合 M2M 协议标准的无线移动通信网(4G/5G),以及专用无线通信网。有线网络技术主要有符合 IEEE 802.3、IEEE 802.4 或 IEEE 802.5 协议标准的局域网(LAN),工业现场总线网标准,以及电话交换网(PSTN)。

3) 核心交换层

核心交换层为物联网提供高速、安全与具有服务质量保障能力的数据传输环境。目前,物联网核心交换层分为 3 种基本的结构:IP 网、非 IP 网和混合结构。目前,非 IP 网的研究主要是基于移动通信网(4G/5G)传输网与专用无线通信网技术。同时也存在 IP 网和非 IP 网互联的混合结构

3. 应用层

物联网的“应用层”又称为“应用管理层”。物联网的应用层非常复杂,它既包括各种行业性应用的应用层协议,又包括支持这些应用实现的各种软件技术,因此需要进一步将应用层分为两个子层:管理服务层与行业应用层。管理服务层通过中间件软件实现了感知硬件与应用软件物理上的隔离与逻辑上的无缝连接,提供海量数据的高效、可靠地汇聚、整合与存储,通过数据挖掘、智能数据处理与智能决策计算,为行业应用层提供安全的网络管理与智能服务。

1) 管理服务层

管理服务层位于传输层与行业应用层之间。当感知层产生了大量数据经过传输层传送到应用层时,如果不经有效地整合、分析和利用,那么物联网就不可能发挥应有的作用。在提供数据存储、检索、分析、利用服务功能的同时,管理服务层还要提供信息安全、隐私保护与网络管理功能,在管理之中也体现出服务的目的。

2) 行业应用层

物联网的特点是多样化、规模化与行业化。物联网可以用于智能电网、智能交通、智能物流、智能数字制造、智能建筑、智能农业、智能家居、智能环境监控、智能医疗保健、智慧城市等领域。

物联网体系结构的行业应用层由多样化、规模化的行业应用系统构成。为了保证物联网中人与人、人与物、物与物之间有条不紊地交换数据,就必须制定一系列的信息交互协议。行业应用层的主要组成部分是应用层协议(Application Layer Protocol)。应用层协议是由语法、语义与时序组成。语法规定了智能服务过程中的数据与控制信息的结构与格式。语义规定了需要发出何种控制信息,以及完成的动作与响应。时序规定了事件实现的顺序。不同的物联网应用系统需要制定不同的应用层协议。例如,智能电网的应用层协议与智能交通的协议不可能相同。为了实现复杂的智能电网的功能,人们必须为智能电网的工作过程制定一组协议。为了保证物联网中大量的智能物体之间有条不紊地交换信息、协同工作,人们必须制定大量的协议,构成一套完整的协议体系。

物联网网络体系结构是物联网网络层次结构模型与各层协议的集合。物联网体系结构将对物联网应该实现的功能进行精确定义。物联网体系结构是抽象的,而实现协议的技术是具体的。目前我们在研究不同领域物联网应用系统设计方法时,会发现很多针对不同应用场景的新的协议标准的研究和出台。

5.3.3 物联网应用案例

1. 应用背景

丰赏农业总部位于江苏省南通市,拥有数千万平方米的种植土地,作物包括蔬菜、花卉、香料等不同的品种,面向上海市提供产品销售服务,主要客户群涵盖超市、邮轮等高端客户。对于丰赏农业来说,企业不仅要实现管理规模庞大的农场,更需要通过精益化的管理来提升农作物的产量和品质,避免食品安全风险。而农业物联网能够通过对环境的监测,数据的采集来提供相应的数据支撑,一方面可以帮助企业及时根据农作物的需求调节环境温湿度、光照等促进作物的生长;另一方面也能够积累作物生长全周期的数据,为农业科研、食品安全溯源等应用领域提供可靠的数据支持。

2. 技术方案

该公司通过在微软 Azure 云平台上使用 IoT Hub 连接包括 HTTP、高级消息队列协议(AMQP)和 MQ 遥测传输(MQTT)的各类传感器,用于监测 pH 值、溶解氧、电导率、温度、湿度、CO₂、Lux 流明、土壤湿度、水温、PIR 等指标,并采用 Stream Analytics 对数据进行流式处理,实时接收来自农场的各类传感器数据。

为了让农场的管理者能够及时了解农作物的生长环境,采用了 Power BI 实现数据的可视化,以具有丰富交互式图标的仪表板进行展现,让农场管理者和一线员工都能够直观地了解所需的信息。此外,该平台还通过使用 Azure 通知中心面向工作人员的移动电话和 App 推送特定通知,让工作人员在任何时间和任何地点及时掌握农场运营信息,如图 5-11 所示。

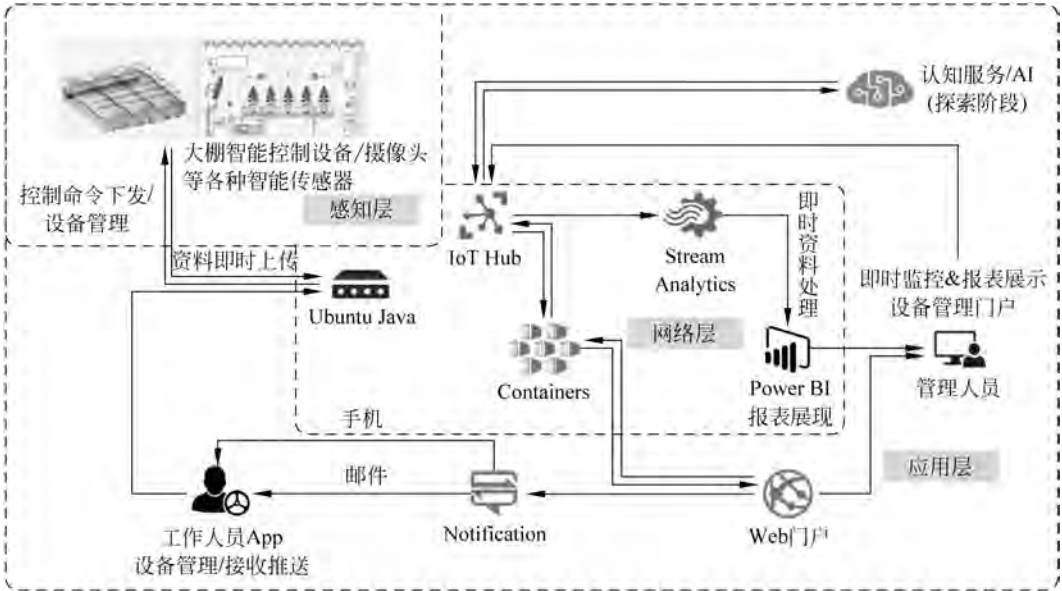


图 5-11 物联网应用案例

感知层包含大棚智能控制设备、摄像头和各类传感器。网络层包含各类传输和信息处理装置,如 IoT Hub、Ubuntu Java 主机、交换机、路由器等。应用层包含 Web 应用、通知推送、监控和认知服务。

习 题

- (1) 什么是云计算?
- (2) 简述广义云计算与狭义云计算的区别。
- (3) 什么是大数据?
- (4) 大数据有哪些特点?
- (5) 什么是物联网?
- (6) 物联网的原理和用途有哪些?