云计算简史

围绕云计算相关的技术领域、技术名词和技术产品令人眼花缭乱。在云计算发端之初，应用开发环境还比较简单，当时还有所谓的全栈工程师存在，意味着如果不考虑开发周期，一个人就能搞定整个应用软件。今天，这个称谓已经名不符实。很少再有一个人，甚至一个企业能够全面掌握和云计算有关的所有技术栈。他们可能会应用旁人完成的一些成果，结合自有的一些专有经验，来形成在某个细分市场有竞争力的产品，或者为客户交付期望的产出。

即使作为纯粹的应用者，要想全面了解和云计算有关的技术，做到合理架构，恰当选型，顺利完成集成开发和部署的全过程，也都比过去难得多，需要的技术人才也比过去昂贵。坦率来说，在当下的人才竞争度下，一般行业的企业，即使拥有信息部门，也都不太可能独立驾驭这样复杂的开发设施，他们将不得不广泛依赖云计算平台提供的服务。这给软件行业的解决方案商带来了新的市场机会。谁能够为数字化转型企业提供友好的应用开发和部署环境，谁就能够获得和保有客户。

# 一、云计算市场的形成和结构

我们今天能够享受经济和便捷的云计算服务，主要来自两大动力，一是计算资源的虚拟化技术，二则是规模经济效应。前者发端于2000年后VMWare推出的Hypervisor虚拟化软件，它不再依赖一个母体操作系统，就允许用户将硬件和网络资源划分成多个单元，从而实现计算资源的池化、共享和按需调度。

2006年，Amazon推出了S3对象存储服务和SQS简单队列服务，开创了公共云计算服务的先河。在此后，微软，IBM，谷歌，中国的阿里，腾讯和华为等都陆续加入了公共云服务的市场，提供的服务也从基础计算资源扩展到数据库、人工智能、物联网等多个技术领域。目前，这个行业已经成长为年收入2500亿美元的巨大市场。

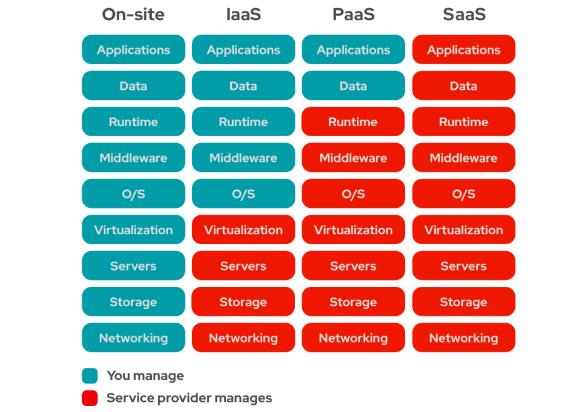
在这十多年的发展过程中，当然出现了很多的公司，产品和服务，但是概括起来这些事物的涌现基本沿着两条明显的路线在进行：

## 1.1趋势一：从基础设施，到应用，再到应用相关的平台服务。

### (1) 基础云 (Infrastructure as a Service)

最早的云计算服务就是最基础的云主机(Virtual Machine)，服务商把裸金属装上Hypervisor，把计算和网络资源分块后就可以卖了。随后，基础服务被拆分为主机、存储、网络、数据库和安全等几个重要的基础云产品，允许用户灵活组合，并实现了弹性计费(目前国外基础云厂商大多都提供按分钟或按秒的计费精度，存储则可以按月计费，比如AWS的S3服务每GB数据的标准存储月费在0.0125美元，而深度归档存储的每GB月费可以低至每GB0.001美元)。

我们一般把主机，存储，网络，数据库和安全相关的计算服务统称为基础云服务。在这些服务之上，开发者需要完成所有的技术栈搭建，构建自己的数据架构，开发编码，部署运维，最终才能实现云端应用。而初代的云计算客户大多数都是互联网公司。他们并非云服务的最终消费者，而是生产者。



### (2) 应用即服务(Software as a Service)

和Amazon Web Services几乎同时起步的另外一家公司Dropbox是一家面向个人和团队提供文件存储和共享服务的创业公司。赶上了AWS起步的时候，Dropbox就直接使用了AWS现成的S3对象存储服务，这让一家团队人数很小的初创公司有机会能够聚焦在应用开发和营销上，让Dropbox通过短短的几年时间发展成市场份额第一的文件共享应用。和Dropbox类似的大票SaaS企业大多也都是在随后的十年内陆续出现，他们千篇一律都使用了云计算平台的服务，而不再自建基础设施。这当中也包括一个超级大用户“奈飞“(Netflix)，他们的下行流量占到整个互联网下行流量的15%之多，也是AWS的客户。

广义来说，最早的云服务在基础云公司之前就出现了。1999年创办的Salesforce，就是一个典型的SaaS公司，只不过当年没有这样的行业术语。2016年，据说Salesforce也已经成为了AWS的客户。因为SaaS服务形式的存在，使得云计算能够间接提供服务给大量的中小企业和非互联网行业企业。今天，几乎所有的企业都或多或少应用一些SaaS服务。

云计算市场发展的第一波主要由互联网企业用户带动。他们具备比较完整的开发和自助运维能力，而且也有日益增长的用量，属于基础云服务最理想的客户群体。直到今天，阿里云和腾讯云的主要客户群体依然是泛互联网行业。

SaaS企业是云计算基础服务的重要推动者，虽然这个门类和2C的网络服务相比，贡献的经济价值要小得多，但是他们深谙企业市场的需求，推动了云计算平台的应用开发环境日益成熟。这就是趋势的下一步：平台即服务。

### (3) (开发)平台即服务(Platform as a Service)

所谓平台即服务，特指开发平台。应用开发工作从本地迁移到云端，自然需要在云计算环境中提供对应的更优解决方案。所以过去传统的中间件市场陆续发生变迁，逐一转换为在云计算平台上的某项服务。比较常见的开发平台服务包括：

通讯：提供音视频通信、消息推送、短信、邮件等服务。

地理信息：提供地图、定位、导航相关的服务。

应用开发框架：提供应用开发环境和运行时环境。

媒体服务：提供图片和音视频等媒体文件的编码、加工和存储服务。

机器学习框架：提供面向AI应用开发者的机器学习数据标注和模型训练平台。

小到发送一条验证码短信也是一项PaaS服务。

作为PaaS服务，主要是为开发者服务的，所以除了功能性服务以外，PaaS厂商也要提供开发友好性相关的周边能力，比如弹性扩展的能力，调试和控制权限的能力等。参与的开发者越多，一项PaaS服务就能够有更多的改进机会和摊低的平均成本。

PaaS服务是不是一定由独立的PaaS厂商来提供呢?不一定。实际上，主流的PaaS服务大多被IaaS公司所覆盖了。如果你打开阿里云的产品列表，在数百个产品中，你会发现基础云服务只是其中一个门类，其他十几个门类都是和开发环境有关的服务。这意味着，一家创业公司想要独立成为一家成功的PaaS厂商，需要相当聚焦地执行，而且产品有明显的技术领先度。一旦做到这一点，也不用担心和基础云公司的竞争，因为我后面会讲到云计算市场的技术发展，其中已经有众多的技术趋势保障了独立性PaaS公司建立跨云服务的独特优势。

以上说的是云计算服务过去十五年发展中的一条脉络，从基础云到应用的共生，再到日益丰富的开发平台即服务。云计算覆盖的用户越来越多，依赖的是这三个层次的服务互为补充。

## 1.2趋势二：从公共云、私有云到混合云，再到多云

第二条脉络有关云计算服务的部署模式(Deployment Model)。当云计算概念被提出时，它显然指的就是公共云服务，客户不需要保有任何基础设施，直接像水电煤一样使用云计算资源就可以。但是商业的现实和技术企业的理想之间总是存在沟壑。云计算到底是技术还是服务，在很长一段时间内是缺乏共识的。

在云计算服务开启之前，很多大型企业和组织都有自己的服务器。2010年，全球服务器市场就有500亿美元的规模，这些服务器大多数都卖给了企业和政府。企业拥有了这些基础设施，难道再花钱买公共云服务吗?既然云计算技术这么好，为什么我自己来实现呢?政府、金融、医药等行业客户更加不可能在公共云计算服务诞生的初期就义无反顾地采纳，他们有各种各样所谓的合规要求。

### (1) 私有云 (Private Cloud)

果不其然，有需求就有供给。2010年Rackspace和NASA公开了一个叫做OpenStack的开源项目组。它包含了一系列用于构筑云计算服务的开源软件。这意味着，所有拥有硬件基础设施的用户都可以用很低的成本来实现和AWS类似的技术架构。Rackspace是一家IDC公司，它这么做的动力显然是很强的。它认为只要帮助客户解决虚拟化问题，自己的主机托管生意一样可以兴旺发达。

虽然软件是开源免费的，但是要实施Open Stack依然需要云计算相关的专业知识。因此，从2010年开始，出现了很多基于OpenStack帮助企业建立私有云的服务商。在国内，公共云服务商甚至都提供过这类服务。十年过去了，这股由OpenStack带起的私有云风潮基本告一段落。除了极少数大型用户在经济上能够承受自己维护独立的云计算平台，绝大多数用户根本无法得到经济上合理的回报。虚拟化只是云计算服务的一个技术前提，但并非所有的价值。私有云方案永远无法利用到资源弹性利用(可大可小)和真正的规模经济效应，除非用户根本不关切经济理性。



在中国市场，重点行业可能至今依然无法使用商业云服务，但是电信运营商和一些国家级的科技企业也在公共云服务商的帮助下建立了各种行业云。比如移动云，联通云和电信天翼云都是这样形成的，他们为金融、政府、交通、教育等重点行业提供公共云服务。

故事到这里，似乎公共云已经大获全胜。但是，商业现实又回来了。在越来越同质化的云计算服务市场，客户难道完全没有议价能力吗?客户的需求如果不能得到满足，总有供应商会愿意创新。于是混合云(Hybrid Cloud)出场了。

### (2) 混合云 (Hybrid Cloud)

其实混合云并不是什么独特的云计算技术，它实质上是一组通讯服务。只要堆上足够好的网络设备和奢侈的专线连接，世界上任何地点的计算设备都可以组成高速专网。即便客户预算有限，只要对安全性和连通性的要求没有那么高，也可以自助搭建经济的虚拟专用网络。围绕通过商业网络连接组建混合云的技术被称之为“SD-WAN”(软件定义广域网)。有了网络连接，就可以把客户自有的计算设施和公共云计算设施连接在一起，称之为“混合云”。

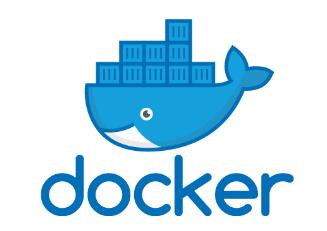
混合云对客户的好处是明显的。首先，每个企业都可能有云计算基础用量，但也可能有短期的激增需求。有了混合云，客户就可以围绕自己的基础用量采购自有IT资产，自己运营私有云，而短期波动的增量则可以通过公共云服务满足，等需求高峰过去，就可以去掉这部分的开支。企业也可以将运维难度比较低的基础云服务保留在自己的设施内，而同时使用公共云提供的复杂计算服务，比如机器学习平台等。Dropbox是一个大规模的SaaS应用，它在2016年做了很大的架构调整，大部分的服务不再使用AWS的公共云，一举节省了7000万美元的年度云计算开销。

混合云策略现在已经得到了厂商和客户的双重支持，它终结了公共云和私有云非黑即白的争议，让整个IT产业更加务实。这其中也诞生了很多的商业机会。微软，亚马逊，IBM，Google等领先的云计算厂商都推出了自己的混合云解决方案。因为混合云方案主流化，云计算厂商的竞争开始从基础云资源的成本向应用开发生态环境迁移。因为在混合云架构下，客户面临如何规划流畅的数据连接，如何快速交付云原生应用的新挑战。所以，云计算的终极竞争不是硬件的竞争，也不是软件的竞争，而是应用开发和部署(AD&D)环境的竞争。

### (3) 多云 (Multi-Cloud)

多云概念是云计算市场最近几年出现的概念。它把所有的云计算平台，客户的私有云设施全部视作一般基础设施。所有的应用在所有的云上都能一致并可靠地运行。多云解决方案不仅是基础设施提供者需要协调的，更重要的是应用开发和部署要面向多云运行目标。

2013年，Y Combinator孵化企业Docker Inc开源了Docker项目。它成为应用跨云部署的重要前提。Docker允许用户将复杂的应用、数据和依赖的环境，包括操作系统本身打包到一个“容器”中，通过标准的Docker引擎，在任何计算环境中都可以一致地运行。有了这项技术，把一个应用系统从阿里云转移到腾讯云就和传输一个文件一样简单，云和云之间已经没有边界。为什么Windows和mac OS的应用永远不兼容，而云计算厂商却眼睁睁地看着这些事情发生呢?很简单，因为整个云计算技术生态都建立在开源软件上，亚马逊再大，它也只是一个服务提供者，收的是租金。而客户方，则越来越看重自主可控性，他们不希望被单一的云计算公司锁定，毕竟自己的客户和交易数据都运行在云计算上，它是所有企业的命脉了。



2015年，Google开源了Kubernates项目，让多云解决方案更胜一筹。K8S能够对容器的创建、扩展等进行自动编排。这意味着无论应用有多么复杂，它都能够在多云环境中进行统一运维。比如自家的某种类型存储用完了，就可以临时购买一些亚马逊的存储。数据过时了，就定期自动地转移到低价格的冷存服务中。

有了多云技术框架和服务，同时意味着云计算平台必须提供广泛支持。阿里云当然希望多卖一些云主机服务，但是如果因为技术框架落后，客户就会流失。所以，全世界的云计算平台目前都义无反顾地支持了多云策略，希望在这个过程中继续以专业服务商的地位存在。

多云策略对应用开发者的影响也很大。首先开发者必须从第一天就按照云计算环境来规划，支持多云部署，自动伸缩，采用微服务架构以实现容器部署。其次，应用开发者也能够从这样的架构中受益。因为它使得客户获得私有软件也像应用SaaS一样简单，唯独不同的是应用和数据运行在客户控制的计算环境中，但是软件本身都是基于单一代码库的(Single Code Base)。

前面我们提到了云计算公司的竞争将向应用开发和部署环境迁移。那么它具体指的是什么呢?它有关于围绕云计算相关的四个技术领域。Tom Siebel把他们概括为云计算本身、大数据、人工智能和物联网。

接下来，我们会逐一介绍这十五年来，伴随云计算发展起来的数字化技术领域。正是因为云计算服务的普及，才催化了这些新兴的技术领域，反过来，这些技术领域的发展也让现代云服务更加完善，当然也更加复杂。正是这些复杂性，让企业数字化转型工作变得阻力重重。相较于更早前的基础信息化工作，企业届要认知和掌握的技术范畴要比宽广得多。因此，我们介绍云计算的发展简史，就必须要把关联技术域的发展也一并介绍。

## 1.3测评过程

本次等级测评分为四个过程：测评准备过程、方案编制过程、测评实施过程、分析与报告编制过程。具体如图 1-1 所示。其中，各阶段的时间安排如下：

（1）YYYY年MM月DD日～MM月DD日，测评准备过程。

（2）YYYY年MM月DD日～MM月DD日，方案编制过程。

（3）YYYY年MM月DD日～MM月DD日，现场实施过程。

（4）YYYY年MM月DD日～MM月DD日，分析与报告编制过程。

# 二、云计算相关的技术领域

## 1. 大数据(Big Data)

在大数据概念出现之前，数据存储、处理和分析的技术早已存在。随着存储成本的下降和云计算提供的弹性计算能力增强，越来越多的数据场景已经不能被传统的数据库技术所处理。这些新场景可以被概括为高数据量(Volume)，高频度(Velocity)和多数据类型(Variety)三个特点。比如在电子商务、金融和物联网领域，系统往往在很短的时间内会产生大量的数据。这些数据甚至在存储的过程中就会产生瓶颈，更不用说实时性很强的计算和分析。所以，从搜索引擎时代开始，大数据相关的技术就开始孕育。

### (1) MapReduce和Hadoop

搜索引擎的霸主Google成立于1998年，几年以后，Google的搜索服务所承载的数据量已经是一个天文数字，而且还在以光速增加。传统的数据处理技术完全依赖硬件算力的铺陈，这会让Google在未来的发展中不堪重负。2004年，Google在内部推出了GFS分布式文件系统和分布式计算框架MapReduce。前者解决了单一硬件资源的限制，后者通过一系列数学原理，将多类型的数据进行切片并分散存储在特定的分区中，这个设计能够让未来的计算和分析大幅提效。MapReduce的技术原理是大数据技术发展的最重要基础。

很快，开源软件领域开始响应这项技术方案，Lucene项目创始人Doug Cutting在2006年正式独立出Hadoop开源项目，在其中包括了分布式文件系统，在集群资源上的调度工具，以及最核心的大数据并行处理开发框架。有了Hadoop以后，那些面对海量数据分析难题行业从此有了更好的解决方案。只是在2006年前后，主要的应用行业还是互联网行业本身。Yahoo，

中国的百度等都很快应用了Hadoop来解决海量数据的存储和检索问题。



### (2) Hive，Spark和流式计算

在随后的几年中，Hadoop相关的大数据处理技术继续得到增强。Facebook开源的Hive分析工具用更高层和抽象的语言来描述算法和数据处理流程，能够用SQL语句进行大数据分析，这大大降低了使用者门槛，也提升了大数据技术的应用效率。不要小看这项改进，它让全世界大多数现有的数据分析人员可以轻易掌握大数据技术。

2009年，加州大学伯克利分校的AMP实验室开发了Spark开源集群计算框架，通过完善API和库，提供更完善的能力和通用性。而且Spark的特色是能够将数据存储在内存中，所以数据处理和查询效率要比利用硬盘存储的MapReduce框架快百倍。目前，Spark已经加入Apache Software Foundation，成为Apache开源项目中的明星项目，被大数据技术领域作为最重要的工具框架。

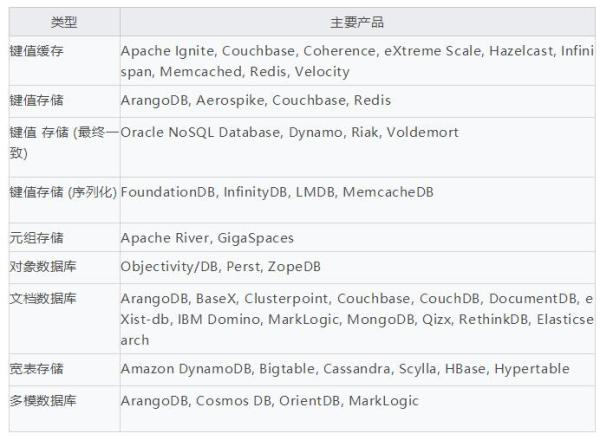
至此为止的技术栈基本解决了针对海量数据批量进行处理和分析的需求。比如零售业企业如果需要研究顾客和交易数据，从而对顾客群进行特征细分，这些技术就足够了。但是，数字化技术的发展总是会刺激出更高级的需求。比如，在线上零售中，商品和顾客的行为数据是永续不断在发生的，我们希望在数据发生的时刻就立即进行计算，及时地给顾客推送一张个性化的优惠券，而不是定时进行某种批量计算，这时候就需要大数据技术的一个分支—流式计算。

流式计算的常用框架包括Storm和Spark Stream和Flink，他们在零售和电子商务行业中的交易分析、金融风控、物联网中的态势监控、车联网中的自动驾驶等领域都被广泛应用。2019年，阿里巴巴用1亿美元收购了Flink，是因为我们用的淘宝天猫中的搜索、商品推荐，包括双11的实时监控大屏数据都是由Flink来驱动的。Flink用几乎无延迟的速度截获双十一最后一秒钟结束后的GMV数值，可见它在实时处理数据方面的性能。

### (3) NoSQL数据库

与大数据技术同步发展的还包括NoSQL(非关系型)数据库市场。在上个世纪，大多数商业数据库都是关系数据库，通过SQL语言进行数据处理和查询。当大数据技术发展起来后，技术专家们发现数据库完全可以用不同的形态来存储数据，这样可以大幅减少数据分析过程中的预处理工作量。所以，从2009前后开始，各种NoSQL数据库开始进入市场。

下图是维基百科上针对NoSQL数据库类型的分类方法：



读者可以忽略其中的细节技术语言，只需要了解不同类型的NoSQL数据库会有利于特定场景的应用开发。比如文档数据库采用JSON格式存储，可以随心所欲定义不同的数据结构，而且横向扩展性很强(数据规模增大后可以保证查询效率)。

NoSQL数据库普遍支持分布式文件系统，所以都具备很强横向扩展性。和关系数据库相比，NoSQL数据库大多不具备事务一致性，但是这个牺牲交换得到数据处理的效率，因而作为大数据技术相关的常见存储方案。

### (4) 云计算平台上的大数据服务

以上我们介绍了大数据技术发展依赖的各个重要技术栈。很显然，和传统的应用开发相比，大数据技术相对更加复杂。它不仅涉及复杂的编程框架，还需要一个专业的运维体系。这使得大部分普通企业用户很难自己来搭建大数据开发环境。所以云计算平台在基础云服务之外，也开始结合云计算资源提供大数据服务。阿里云上的MaxCompute是一个全托管的大数据SaaS服务，用户甚至无需管理主机基础设施，直接按照大数据计算任务量付费。顺便说一下，这种直接将计算服务提供给开发者的模式被称为“无服务器”(Serverless)计算，它的目的是为了简化开发工作中的运维任务，让开发者聚焦在应用开发上。不仅仅是大数据领域，在AI，物联网等其他技术领域，无服务器服务模式正在日益成为主流。E-MapReduce则是一整套大数据相关的PaaS服务，用户可以选择利用现成的服务在自己控制的云主机上完成部署，客户主要支付的是基础云的资源费用。和阿里云类似，亚马逊AWS等其他云计算平台也提供丰富的大数据相关平台服务。

### (5) 应用领域

我们前面提到大数据技术起源于搜索引擎应用。在随后的十多年中，它的主要应用场景依然还是在互联网领域。最常见的应用包括计算广告(依据用户和内容数据动态决定广告投放策略和定价)，内容检索和推荐(百度、头条)，商品推荐和营销活动优化(淘宝、拼多多)。不要小看这几个场景，它们几乎和互联网用户上网过程中的每一分秒都有关系，所以创造了巨大的经济价值。

数据的价值当然不仅仅局限于互联网行业，几乎每个产业都有机会在大数据技术的帮助下发掘出数据的价值，或者改善运营效率，或者发现出新的业务机会。金融行业是较早的受益者。银行贷款业务中的风险控制、零售和结算业务中的欺诈发现、保险业务中的精算和保单个性化定价、证券行业中的期货定价和股价预测等都实实在在在创造出财富。

大数据在研究和开发领域也在大显身手。在生物医药领域，大数据技术在帮助缩短药物研发的周期和提高成功率;合成化学行业也在利用大数据和机器学习技术来加快发现新材料。有人甚至认为数据科学将成为实验、推演和仿真以外的一种新的科学研究方法，成为“第四范式”。

大数据在城市交通、社会治理、能源传输、网络安全、航空航天等领域也都已经有了现实的应用。但在这些资本投入密集的领域之外，大数据在一般行业和企业中的应用依然道路曲折。这不是因为大数据技术不够完善，而是诸多行业尚未能够明确抽象出大数据应用的价值以及可付诸实施的方法论。正如前面提到的，云计算和大数据对于普通中小企业来说依然是一个模糊的技术工具，一般企业也很难雇佣大数据专家，而专业服务企业目前还没有找到利用自己的技术专长提供普遍服务的有效机会。通用领域中的大数据应用还停留在理念阶段。所以，在过去几年出现的大数据技术公司大多都还在服务金融、公安、交通、能源等大客户集中的行业。

突破的关键点可能在两个方面，一是大数据技术栈本身十分复杂，当下的工具还依赖专门训练的计算机专家，产业还没有抽象出一个通用领域的应用模型，也无法提供一个类似SaaS这样友好的应用界面。这值得数据技术领域和企业应用领域中的跨界专家来探索。二是企业数字化建设还刚刚开始，很多企业缺失稳定和可靠的数据采集和记录的过程。如果没有数据流，自然就不会有大数据应用。因此大数据技术被广泛应用可能还需要五到十年的时间。

# 三、人工智能(Artificial Intelligence)

人工智能的概念和基本原理起源早至1950年代。早期的人工智能研究集中在加州大学伯克利分校，麻省理工，斯坦福和南加州大学等计算机实验室中。今天已经商业化的神经网络算法就来自于半个世纪多前麻省理工大学的明斯基教授发表的《感知元》论文，但是计算机算力在当时实在是太弱了，以至于任何计算理论上的假设都很难付诸于现实。因此，在长达五十年的时间内，人工智能技术都停留在理论研究和一部分不成功的实践上。

虽然人工智能领域经历了漫长的冬天，但它所提出的机器向人类学习，并最终在特定领域能够做得比人类更好的假设却是千真万确的。

## 1. 千禧年后的AI复苏

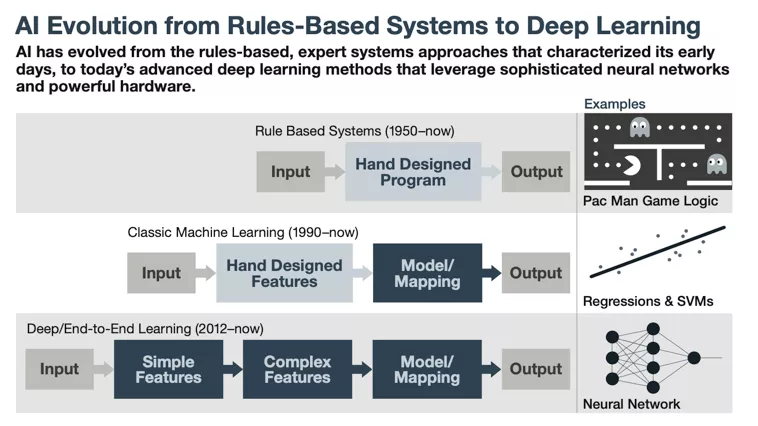
2000年以后，有几大动因推动了人工智能概念的复兴。

首先，因为摩尔定律的存在，计算机的运算速度和单位存储成本均用指数速率发展到一个新的阶段。云计算和大数据技术也允许计算机用很快的速度处理TB甚至PB级的数据。

其次，网络服务的兴起在诸多领域生产出丰富的数据，Google，Netflix和Amazon的业务就像数据机器一样，每分每秒都能产生海量的用户行为数据。

第三，在人工智能的数学方法研究中，AT&T贝尔实验室的三位科学家(Tin Kam Ho, Corinna Cortes, 和 Vladimir Vapnik)在机器学习领域取得了突出的进展。机器学习技术可以将复杂和不确定的非线性问题通过线性的数学公式来解决。在解决不同的问题的过程中，机器学习理论方法和实践被明确验证。最早的一批互联网企业，包括Google，Facebook，Linkedin等在这个过程中既提供了海量数据，也从研究过程中获得了巨大的成果。尤其是Google，它是机器学习及其分支深度学习领域最重要的信奉者和推动者。2010年，Google成立了Google大脑，一个专注人工智能研究的内部组织，后来又收购了英国企业DeepMind。后者在2016年3月击败了人类围棋冠军李世石。

下图是Tom Siebel在Digitlal Transofrmation一书中对AI技术进化史的一张插图，显示了从1950年代开始到现在的主要技术迭代历史。



## 2. 机器学习(Machine Learning)

机器学习是推动AI复苏的最重要动力。它的兴起标志着人工智能很长时间弯路的终结。要想让机器比人做得更好，并不是依靠人来教机器规则，而是让机器从历史数据中学习。比如最常见的机器学习场景——物体识别，要想让机器从各种照片中找出“猫”，只要让机器学习各种各样猫的照片对象。机器学习算法会将训练用的猫图像背后的向量特征总结为一个预测模型，让这个模型预测任何一张新图片中包含猫的概率。同样的道理，语音识别、语言翻译、人脸识别等都是使用的类似的原理。喂养算法的数据量越大，通常预测的准确率就越高。

机器学习应用可以分为有监督学习和无监督学习。前者需要人工参与训练数据的标识，后者则通过数学方法自动聚类出存在相似性的对象。在缺少训练数据的情况，无监督机器学习就会起到更大的作用。

机器学习的一个分支被称为深度神经网络(DNN)，它的设计已经高度参照了人类大脑神经元的连接结构。在深度神经网络中，数据被输送到输入层，结果则从输出层产生，在输入层到输出层之间存在多个隐藏层，每一层会对输入数据的各个特征进行推断，最终能够得到更为准确的预测结果。打败李世石的AlphaGo就是一个基于深度神经网络的算法。但是，DNN对于用户来说依然是一个黑盒子。设计者并不需要也不会知道神经网络中的每一层到底在判断什么具体特征，以及它是如何分解特征的。它背后都是高度抽象的数学方法。不管它有多么玄妙，深度神经网络的确厉害，它不仅具备高超的自学习能力，而且还简化了传统机器学习中大量复杂和耗时的特性工程(Feature Engineering，通过行业专有知识来调优机器学习算法的过程)。

## 3. TensorFlow

2015年，Google开源了内部的TensorFlow框架，开始将人工智能计算框架作为一项云计算服务向外界提供。在核心开源库之后，TensorFlow还陆续推出了Javascript版本，满足在浏览器和Node.js上开发和训练机器学习模型，以及在移动设备和IoT设备上部署的Lite版本。另外，TensorFlow Extended是一个端到端的机器学习生产平台，它连带提供了编程环境和数据处理工具。

当然，TensorFlow并不是唯一的机器学习框架，Caffe，Torch，Keras等都是。它们无一例外都是开源的。在云计算的前沿领域，软件开源是一个普遍的策略。为什么如此复杂和高级的软件都会义无反顾地选择开源呢?一方面因为框架性产品本身并不直接包含商业价值，价值需要开发者进行二次创造，另一方面，在云计算服务的商业模式大前提下，通过API来提供封装好的人工智能服务是一个非常容易实现的商业手段。这些开源产品的运营者没有必要对框架进行收费。

## 4. 人工智能服务

事实上，即便你不使用这些机器学习框架，也能直接使用人工智能服务。国内外云计算平台都已经在通过API提供各色各样的人工智能服务。这些服务已经完全封装成应用开发接口，开发者完全不需要了解和处理复杂的机器学习过程，只要把自己当作用户就可以了。

但是这些服务都非常具体和专向，并不存在任何通用的AI接口，每个接口只能为用户解决一类具体问题。以下是阿里云AI类目下的服务分布。你可以看出这些服务都和用户的某一个具体需求有关。比如语音识别可以让移动开发者开发出让用户直接通过语音来控制功能的应用。人脸识别可以识别出影像中的人脸对象和实现身份对比验证。



提供一次此类服务要收多少钱呢?在云计算平台上，这类AI应用开发接口大多按照次数或者每秒次数级别(QPS)进行收费。比如识别一张身份证上的信息大约要收取1-5分钱，听起来不少吧?

实际上，从事人工智能技术的企业并不仅仅是云计算平台提供商。比如中国市场中，Face++，科大讯飞、商汤科技、寒武纪、优必选等都分别在计算机视觉、语音、机器人等领域有专长。但是它们的专向定位让这些企业很难提供普遍的开发者服务。因为开发者往往希望在一个云计算平台上获得一揽子服务，而且用户的基础云资源也是从云计算平台购买的。作为开发者来说，拥有一个统一和完善的应用开发环境是非常重要的。

所以，在人工智能的商业化中，还有不少企业利用自己的专向技术优势来解决更加细分的问题。比如科大讯飞主要通过自己在语音和自然语言处理方面的技术积累为教育和司法等行业提供解决方案，中国法院的庭审文字记录现在很多都是通过自动化的语音转录而实现的。商汤科技和旷视科技则主要在智慧城市和安防领域提供软硬件一体化方案。还有一组创业企业专注于解决高价值的自动驾驶问题，并从中派生出更细分的AI芯片设计和制造企业。

## 5. 技术栈和人才

AI相关的技术栈是前面介绍的大数据技术的一个扩展。也就是说，没有离得开数据获取和处理的人工智能项目。要把如此众多的开发框架和微服务组合在一起，对于非云计算专业企业来说是非常困难的。除了技术栈的复杂性以外，开发者还需要搞定大规模训练数据的获取和处理过程，这个成本在短时间内一定会成为牵制企业投入的因素。

成本还是相对容易克服的问题，因为只要问题足够值钱，有长期主义价值观的企业总是愿意投入。但是更致命的问题在于AI相关人才的激烈竞争。能够从事AI应用开发的团队需要包含大数据相关的数据库专家，深谙数学建模的算法专家，以及熟练掌握C++或Python等编程语言的高级程序员，同时还离不开有技术素养的业务专家参与。而在当下阶段，云计算巨头企业和专业企业像吸铁石一样吸引走了绝大多数专长人才，让普通企业根本无从获取。

考虑到AI技术的复杂性和专业度，它极可能像云计算服务一样，大多数企业都只会成为用户级别的角色，这就给专业开发者留下了创新的空间，看谁能够进行足够合理的抽象，组合出更加易用，面向通用业务场景的AI服务。

# 四、物联网(Internet of Things)

## 1. 消费产品引爆的物联网普及

云计算服务的普及不仅为用户提供了弹性伸缩的经济性，还提供了一个泛在的可连接性。任何计算设备只要连上互联网，就彼此通过TCP/IP协议能够相互访问。这个互联价值在物联网技术发展之前还仅仅限于传统计算设备，也就是服务器和个人计算终端。在个人、家庭和企业世界，还有大量非传统计算设备并没有联入这个数字化世界。

汽车、家电、个人穿戴设备、工厂的制造设备现在都已经有接入互联网的条件，市场上流通的这些互联智能产品也越来越多。当连接的设备丰富到一定程度的时候，各种智能化场景才能真正实现。IHS Markit预测到2025年，全球联网设备总数量将达到750亿个。万物互联，正是物联网技术所追求的愿景。

具有数字化连接能力的非常规计算设备在上个世纪90年代就已经出现，比如可以无线连接的摄像头。真正具备中长距离连接能力的设备首先出现在零售和工业制造领域，包括西门子，通用电气等企业开发的工业设备互联协议(M2M)。在当时，这些设备已经可以通过低速的无线局域网使用IP协议连接到工厂的控制中心。这样的网络被成为工业以太网。但当时商业互联网并没有开始发展起来，所以M2M的出现只能算作是物联网技术的局部发展。

物联网开始成型还是通过消费产品市场来推动的。2000年代初期，LG率先推出了可以接入互联网的家电产品，一台联网冰箱售价高达20000美元，这显然无法真正带动市场。在随后几年中，像Garmin GPS和Fitbit智能手环这样的消费电子产品开始获得更大的产销量，从而带动相关的低功耗芯片行业发展。到了2011-12年，消费电子领域出现了更多的明星级产品，这其中包括后来被Google收购的家用传感器Nest，Philip推出的Hue智能灯泡等。在中国市场，以小米为代表的智能手机厂商开始扩展到物联网产品领域，推出了一系列围绕个人和家庭的智能设备和家庭网关产品。苹果也于2015年正式进入可穿戴产品市场，推出了Apple Watch，后来还推出了智能音箱HomePod。Google和国内的互联网巨头企业也都加入了这场通过新型个人数字设备争夺用户和数据的竞争。目前，全球可穿戴产品市场已经多年保持了40%以上的年度增长率。

个人和家庭智能设备的量产促进了和物联网相关的协议发展和元器件成本的降低。在这期间，蓝牙5.0，WiFi-6，IPv6，NFC和RFID等关键传输和通信协议得到进一步发展，让设备能耗和连接速率都进一步提升。在同一时间，云计算基础服务和大数据处理技术也起到了关键作用。物联网设备往往在短时间内会产生大量数据，如果没有前文提到的大数据技术栈，传统的数据库工具是无法承载的，同时云计算也是设备数据汇聚的海洋，今天几乎所有的物联网技术平台都架构在云计算平台上，他们是典型的互生行业。

## 2. 物联网的技术栈

物联网相关的技术栈非常综合。它横跨软硬件，既包含感知探测相关的硬件技术，也包含网络传输和应用构筑的软件技术。直至今天，物联网相关的技术栈都没有完全稳定下来，甚至很可能长期保持多元的特征。但是概括起来，整个技术栈还是有一些层次特征。

行业一般把物联网相关的技术架构分解为四层，分别定义为和物理环境相关的设备感测层、和数据传输和通信有关的网络层，IoT相关的平台管理层，以及最终实现用户价值的业务应用层。无论是针对消费市场还是企业市场的物联网系统都会有这四个层次。

设备感测层是由各种类型的传感器和可交互访问的硬件模块及其嵌入式软件而组成的。例如温度湿度传感器、摄像头、电源开关和插座和网关等。感测层设备不仅是单向获取数据，还可能从外界接受指令改变硬件状态(比如智能锁)。行业内一般把这个层次称为“边缘”(Edge)。

设备感测层的技术栈主要由软硬件协同开发的嵌入式系统构成。我们用的智能手机本质上也是一个嵌入式系统，只是它的嵌入度非常完整，甚至不亚于一台标准计算设备。嵌入式系统开发经历了早期的单片机和嵌入式操作系统/CPU阶段，目前最前沿的是SoC(片上系统)，把一个专用系统的所有嵌入式软件完全整合在一个集成电路上。今天的智能手机、智能电视等都是由若干个SoC整合而成的。在嵌入式系统中，固化在硬件上的软件程序甚至依然可以得到更新，而且这种更新目前大多都可以通过连接互联网实现，这种升级被称之为OTA (Over-the-air)更新。

另外，设备感测层还需要解决设备的访问协议问题。物联网系统目前已经广泛支持IPv6协议。IPv6能够提供全球的IP地址总量高达2的128次方，这是一个天文数字，可以确保任何物联网设备都能够拥有独立的IP地址，从而实现在全球的唯一寻址。当全球物联网设备达到千亿，甚至万亿数量级的时候，IPv6功不可没。

网络传输层要解决的是感测设备和计算设备之间，以及最终与平台管理层之间的数据传输问题。根据连接性质不同，又可以分为短距、中距和长距类型，以及有线和无线类型。在这些连接协议中，蓝牙、NFC、Wi-Fi、无线射频(RFID)、4G和5G等是比较常用的。这些传输协议在设备侧一般都是直接设计在板上系统上的，通过IP协议提供可访问地址。开发者需要根据连接的距离、速率、功耗和成本等要求做出合理的选择。

IoT平台层是物联网系统中至关重要的部分，它的出现也标志着基于云计算平台的物联网系统诞生。一个IoT平台的核心作用是管理成千上万的物联网设备，包括他们的状态，数据上报接收，建立对它们的控制，对设备进行运维分组，并能够实现从云端往边缘侧的更新推送(OTA)。同时，物联网平台也要借用上文提到的大数据技术栈，对设备上报数据进行处理，并用各种数据库完成存储，这其中比较重要的数据库类型就是时序数据库。

更完整的IoT平台还包括围绕设备数据建立自动化工作流的能力，数据分析工具以及为更上层的应用开发提供数据开发接口的设计。

目前，主流云计算平台都专门为客户提供了物联网技术平台，结合基础云和大数据相关服务获取增值业务收入。阿里云，AWS，Azure和Google Cloud都有专门的解决方案，国内外也有专门的物联网平台技术公司将自己的解决方案架构在基础云上或者提供跨云服务。

Oracle，Salesforce和微软Azure等企业软件厂商的物联网平台不仅提供了上述基础服务，还结合了自己的企业应用套装优势，提供一站式的物联网应用开发平台。它们更适合企业物联网系统建设。

最上一层的应用层是整个物联网技术架构中最不标准化的部分。应用层最终要将连接的设备和数据用于具体的商业场景。比如共享充电宝就是一个物联网系统，它的应用层包含面向C端的租用和支付系统，面向商户的设备状态报告，收益结算系统，以及面向运营部门的设备运维管理系统。换到另外一个物联网场景，应用层的构成可能完全不一样。

## 3. 边缘计算和AIoT

物联网技术架构的基本思路是分层分工，感测层主要取得数据和建立对物理硬件的控制，数据通过网络层和计算平台连接，计算则在云端完成。但是随着物联网应用场景的丰富，设备规模的扩大，以及芯片技术的发展，边缘计算概念开始得到认可。所谓边缘计算，就是利用设备端和邻近网关的计算能力处理和存储数据，减少和云端的数据传输，从而实现更快速的应用响应。除了速度的提升，边缘计算还能够大幅减少云计算和数据传输成本。比如针对一个大型的视频监控网络，如果摄像头将所有的视频流数据都传输到云端，云端算力要求和成本将会非常高。而如果在摄像头设备内部实现必要的视觉计算(比如识别异常)，整个物联网系统的效率将大大提高。再例如广泛应用的人脸认证和识别系统，如果不能依赖本地的设备计算能力，十亿用户的高频度使用将会让云平台不堪重负。

上面两个例子显示了边缘计算往往和人工智能应用相关，设备侧往往完成的是模式识别类的人工智能算法，因此往往需要专门芯片的助力。Nvidia公司推出的Jetson系列模块就是专门为边缘计算场景而服务的。这些芯片模块被安装在机器人、自动驾驶汽车等边缘设备上，所以这组技术产品也被称为“自主机器”。因为物联网和这些人工智能应用的结合，因此，这套技术方案也常被称为AIoT。

## 4. 应用领域

如果把2012年前后作为基于云计算的物联网平台技术发展开端的年份，仅仅八年左右的时间，物联网应用领域已经发展得非常广泛。只是我们身在其中，享受它带来的便利，并不一定能够感知它的存在。这个高速发展过程很大程度上受益于基础云计算服务和大数据技术栈的同步发展。

在消费应用领域，个人穿戴设备已经从手表、手环发展到耳环戒指这样的杂项。在智能家居领域，我们可见的家电、门锁、照片、开关、音箱等都已经是互联网设备。在个人和家庭领域，物联网技术竞争已经不再重要，竞争的焦点已经迁移到内容生态和用户网络效应。在这些领域，苹果、谷歌和国内的华为、小米等已经牢牢占据了领先地位。小米生态链重点指的就是依托米家体系的一群消费电子产品企业。

在工业、农业、交通运输、能源和社会管理领域，物联网的应用场景更为广泛。我们的街道上已经密布了各种摄像头，这些摄像头通过专门网络构筑了天眼系统;我们头顶上的电力传输网络和用电单位终端也已经完成了智能电网改造;矿场和工地上也都布满了各种安全监控设备。这些都是最近十年来重大的IT投资。预计未来十年这样的建设和更新依然不会结束。

# 五、云计算的未来

本文主要是对云计算过去十五年的一个回顾。技术领域的发展速度是如此之快，以至于我们很难预测未来的十五年会发生什么。在本文的最后，我仅仅对当下已经呈现的云计算市场趋势做一些简单的概括，它们甚至已经在发生，但我们不知道这些技术趋势的发展具体会颠覆哪些巨头，会催生出哪些明星。

存储和计算的成本将进一步下降，但消耗会同步增长。考虑到全球范围内的计算服务还有大量没有转移到云计算环境中，未来十年基础云服务的算力还会大量增加，服务单价会持续下降。 前沿技术将持续融合到云计算平台，包括量子计算、AR/VR、区块链等。尤其是那些依赖海量数据计算能力和弹性计算资源的应用，云计算是成就它们更快成长速度的轨道。 云计算的核心竞争将聚焦在应用开发环境的优越性上。谁能够提供廉价、完善和前沿的开发技术栈环境，谁就能够获得更多的开发者用户。当开发者用户选择了云服务商后，他们其实就为终端用户做出了选择。 多云、或称混合云环境成为长期的企业应用策略，云计算用户将综合使用边缘、各个云计算服务商和自有IT设施。 云计算开发技术栈会越来越复杂，这会让应用开发领域的分工更加明确。面向终端用户的应用开发将变得更加简洁，应用生成的方式会越来越多元，无需编码，依靠普通业务用户建立应用程序的服务商会越来越多。