

基于 SDN 的 Actwill 数据迁移业务探究

赵呈亮¹ 唐宾徽²

四川大学锦城学院 计算机与软件学院 四川 成都 611731

DOI: 10.18686/jsjxt.v1i3.1261

【摘要】SDN(Software Defined Network)即软件定义网络,是一种新兴的网络设计理念,它打破了传统网络的架构,推出了可以集中式软件管理,可编程化,控制转发层面分开的新型网络架构。SDN 是一次技术的革新、一个新的网络思想,改变了传统网络不兼容、信息转发复杂的问题,同时基于这项技术衍生出多种技术种类,本文将介绍基于 SDN 的数据迁移业务技术的研究及分析结论。

【关键词】SDN AWS 灾备 网络模型 数据迁移 应用集群

引言

云计算时代的来临,标志着信息化时代的飞速发展,在这一背景下,网络的革新必不可少。各个企业在使用云服务的同时,会涉及到业务的变动和网络的优化,如何安全高效的进行数据迁移成为一个难题。企业大部分业务是建立在云端并实时运行的,通过虚拟化技术将各类网络资源整合在一起,大量的数据在虚拟化服务器内流通,在进行数据迁移的过程中,要保证迁移的业务系统和数据是完整的,同时在迁移的时候要保证企业的正常业务不受太大的影响。因此,引入 SDN 的设计理念将数据进行规范化控制,保证数据在迁移的过程中受外界因素影响最小。

1 SDN 网络模型介绍

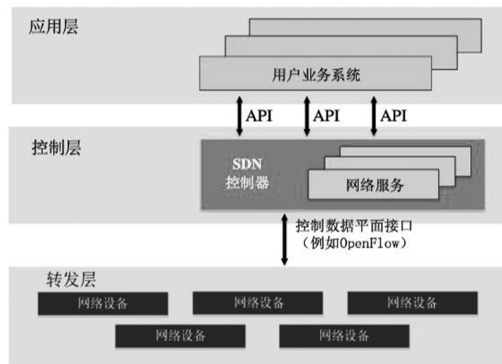
在传统网络中,采用分层结构,并将不同种类的设备设立在各层级结构中,层与层之间依赖于协议进行连接,协议繁冗且实现困难,运维难度比较大,并且数据的转发主要依赖于路由器产生的路由表进行路径计算和选择,导致数据流量的灵活调整度不够。此外,在企业业务不断增长这一状态下,设备的部署和升级成为必不可少的一个阶段,设备一旦出现故障,会影响业务开展,大大降低了工作效率。同时,各大厂商生产的网络设备功能不同,设备与设备之间的兼容性存在很大问题,多个外接的应用服务很少能在一个工作环境下运行,管理困难。

SDN 网络抛开传统网络架构模式,推出应用平面、控制平面、数据平面三大平面,三大平面又通过一个管理的平台进行处理,在一定层面上简化了整个网络环境^[1]。应用平面通过北向接口与控制平面进行交互,将应用平面的应用抽象化并使应用能够直接控制网络,而数据平面通过南向接口与控制平面进行交互,对数据进行转发行为控制,提供设备性能查询,进行统计和通知等功能^[2]。SDN 的三个主要特征是转控分离、集中控制、开放接口,即对网元进行控制,通过协议计算,产生流表,控制器集中管理和下发流表,在对设备进行配置以后可通过开放接口对整个网络进行管理。SDN 最重要的一个环节是对控制平面和数据平面进行解耦合,通过 Openflow 的架构将各类数据变为流表,进行数据的查找与转发,借助于这一项优势,数据可以通过最优路径传递到目的网络设备,理论上可以将数据按照流表的形式对数据进行迁移,既保证了数据的完整性,也保证了传输数据的效率。

现目前,控制平面的开发阶段还不够成熟,初步计划是将数据转变为通用的带标签形式的数据流,但是如何将这种形式的数据流按照特定双方迁移的设备进行设计是一个值得思考的问题,我们将网络设备抽象为通用的转发模型,使得不同厂商的设备可以同时兼容,同时在数据上标记,使得同一类型的数据会按照一定的路径分布到对应的设备中,提高了数据迁移的效率^[3]。因此,设计在开发的控制平

面上对流表进行定向控制,可以支持数据的定向转发,同时也利用虚拟化技术对数据进行处理迁移。接下来将介绍数据迁移的具体实现方法及利用集群应用进行迁移抽象数据的设计流程。

ONF定义SDN基本架构



2 主流数据迁移技术

2.1 数据迁移简介

一个企业的核心资源是数据,大量数据的背后潜藏的是商业价值,而数据往往会有一个生命周期,从数据的诞生,发展,使用再到清除,数据是保存在物理存储设备上的,尽管在云平台上利用的是虚拟存储设备,但是数据最终还是会间接的放在物理存储设备上或是利用虚拟磁盘映射等方式实现,因此数据离不开各类设备的支持^[4]。一般我们所使用的网络设备都要它的质保期,在长时期的使用过程中会消耗其物理性能,使得数据在其中传输的过程中会受到阻碍,从一定意义上来说,数据完全依赖其宿主设备,一旦设备不能满足数据的运行环境,就要考虑更换设备,这也是为什么要进行数据迁移的原因。而选择一个全新的宿主设备意味着数据的运行环境受到改变,这就要求设备的性能及兼容性的问题,往往在上一个设备上能正常运行的设备在新设备上就会出现各种问题。如果是对于企业级的数据迁移,还要考虑到迁移过程中的数据安全、数据可用性及迁移时带来的停机时间。所以在进行数据迁移前,往往要先做好备份措施,一旦出现异常,就要及时进行数据回退操作,重新启动服务,这是一个很难得到完全解决的问题,数据迁移还带有一点运气成分。

2.2 数据迁移方法

数据迁移对企业来说必不可少,目前的主流迁

移技术大致分为应用迁移、卷管理迁移、网络迁移、存储层的迁移。在应用迁移中,分为数据库迁移、文件系统迁移和虚拟机迁移,而在三者迁移过程中,往往需要有一个中转软件工具,即迁移双方所支持的文件格式或者文件类型不同,需要转换成共同的兼容文件,才能继续数据迁移,往往在这个过程中,花费了太多时间。基于卷管理迁移则需要对 SAN 裸设备进行聚合或者拆分提交给上层应用,通过卷管理软件进行数据管理,利用卷管理软件自带的镜像进行迁移。对于网络迁移,则需要将设备虚拟化,将数据通过网关进行传递,然后利用镜像复制到目标存储设备并将相关的配置文件进行修改。存储层数据迁移需要将源存储设备映射到目标设备,通过读写操作将目标设备中的数据从源存储设备中导出来,但一般要求两个存储对象是来自于同一个平台产品,即迁移的局限性。

3 Actwill 迁移方案设计流程

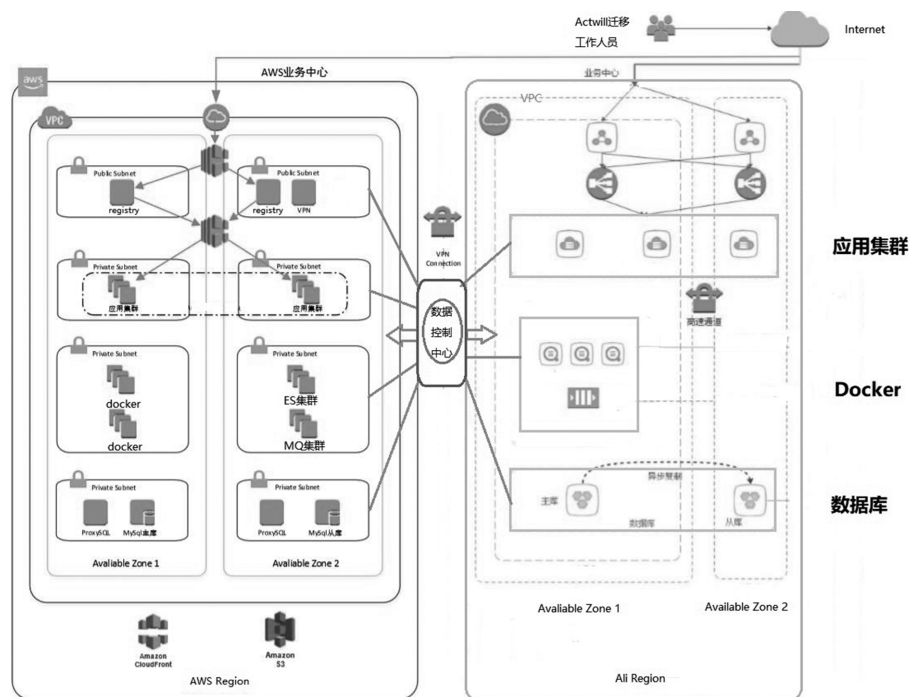
3.1 迁移基本架构

从前面介绍的内容来看,SDN 层面需要解决控制平面的设计问题,怎么统一管理数据并将数据拆解为通用带标签的数据流是一个急需解决的问题,从数据迁移角度来讲,设计将各类设备进行分类处理,同时数据可以定向传递而不是靠中转的方式。Actwill 迁移方案参照的是 AWS 数据迁移架构,具体内容为建立一个站点区域,在这个区域中架设虚拟私有云 VPC,即在公有云平台的基础上再次建立私有云^[5]。在私有云的内部结构中分为公有子网和私有子网,并使用 CIDR 地址块进行子网划分,然后依次在公有子网上堆放接入 vpn 服务,注册信息服务等,在私有子网上堆放各类应用集群,其中也包含 Docker 容器以及各类实例服务,根据对应的服务内容划分到同一个私有子网。一般情况,同一个 VPC 中会放置多个 AZ 可用区,用于放置多项业务并能提供备份容灾,类似于数据库和管理层面的应用都需要在两个站点同时堆放,如果一个站点发生业务中断,将会自动启动另一个站点并保证业务的正常运行。

在本次迁移方案中,涉及到的是从阿里云向 AWS 进行数据迁移,通过 VPN 连接两个数据中心,

在连接接口处设立负载均衡器,以达到负载均衡,选择最近的站点区域作为服务响应区,使得请求的业务和传输的数据能最快到达。最重要的就是设立一个自动数据控制应用集群,使得所有要进行数据迁移的 Docker 容器以及实例服务能够进行实时分类,并将数据分离出来,按照一定的逻辑在新的子网中进行堆放。数据控制中心在工作时,会收集各类资源的信息,将其统一整合并进行数据抽象,并会根据信息流内的标签信息转移至各个对应的子网环境中。本次迁移的实例为 linux 操作系统,每个实例会包含多个属性,主要有实例 ID、镜像 ID 和实例规格

族,在开发数据控制集群应用时,将按照属性的分布规则编制数据的数据流。在进行层级迁移的时候,会将数据流相似的实例转移至另一个可用区内的数据控制集群应用中,再将数据流解析出来还原至原有的实例属性,此时只需要配置好对应可用区的配置信息,就能实现从一个可用区到另一个可用区的迁移,不需要再经历导入导出再导入的这个过程。此种集群应用在开发时还应注意数据迁移时的时间戳,有可能在数据迁移过程中某一数据的属性不对应而导致错误,因此在实现完美迁移上还存在难度,数据可能依赖于原有的资源环境。



3.2 迁移方案实施

针对不同的业务层级进行分类,分别进行业务切换,在应用级层面上配置相关的网络文件,手动切换 DNS 服务,重建镜像仓库,使得从阿里云中的镜像数据能导入 AWS 的镜像仓库,并且横跨三个 AZ 可用区部署,保证数据迁移后应用层服务的正常运行^[6]。对于中间所堆积的 MQ 消息队列服务集群和 ES 集群进行开关操作,在 AWS 上的 EC2 上重新部署 ES 实例节点,在完成数据验证后,逐步清理原有阿里云节点,同时重建 MQ 集群。对于数据库迁移,将使用 DTS 数据转换服务将数据抽取并转换合并到目标位置,数据库割接以后,在 AWS 的 EC2 上重

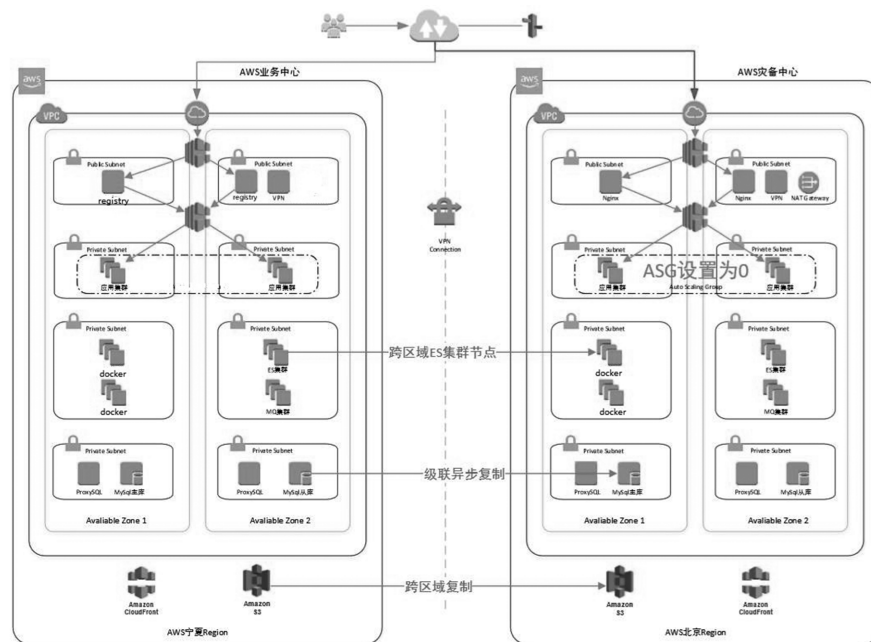
建数据从节点。对象存储将使用 CDN 网络构建模式对服务器进行加速,收集网络中的节点连接、负载状况信息,并及时将服务导向最近服务节点,提高站点的稳定性,在迁移时,将原有的 CDN 网络用全新构建的 CDN 网络代替。在完成这一系列的工作部署以后,需要进行测试并添加回退预案,如果发生迁移意外的错误,将进行回退操作回退到原有的设备运行状态。整个迁移过程使用 VPN 专网搭建,并配置专有的数据中心进行数据传输,保障整个传输过程中的传输安全。

3.3 同步迁移灾备

AWS 在中国有两个 Region,分别为宁夏区域和

北京区域,为了使迁移后的业务正常运行,还需要借助这两个地区的连接来实现迁移灾备^[7]。两个 VPC 之前通过网络及 VPN 相连,将可用区的所有业务资源按照一比一的比例完全复制,使得所有业务能同时拥有两个业务访问区域,可以增加业务的可靠性,最大限度减少延迟。此外,一旦发生数据丢失造成业务中断,两个可用区将会相互反馈并将数据互相复制,及时启用另一个区域的业务,保证所有业务正

常运行。针对不同应用集群,将采用跨区域节点复制,首先预先复制资源池,将里面的存储服务进行配置,再对应用集群进行跨节点复制。数据库复制采用级联异步复制, master 服务器记录数据的改变并写入一个二进制日志, slave 服务器会对日志进行探测,一旦发生改变,将会触发事件,读取日志信息进行复制。而存储设施等则可以依靠 AWS 上的应用进行实时部署,实现跨区域复制。



4 结语

对于开发的数据应用控制中心在实现上依赖于 SDN 控制平面的发展,在多层级的数据迁移上不能完全做到数据的分离,在对数据进行抽象化时,可能会因为标签形式的数据冲突导致数据传输发生不可预估的错误。在数据迁移的过程中,将参考主流的

迁移技术并以 AWS 私有云为核心进行数据迁移,做好备份操作以及回退预案,将整个迁移过程中可能带来的业务损失降低到最小,目前将会对所迁移的业务和数据进行整合处理,采取多层级技术进行整个业务的迁移工作。

【参考文献】

- [1]软件定义网络研究综述[J]. 张顺森,邹复民. 计算机应用研究. 2013(08)
- [2]对 SDN 技术的研究与思考[J]. 刘露,陈清金,张云勇. 互联网天地. 2013(03)
- [3]SDN 的特征、发展现状及趋势[J]. 郑毅,华一强,何晓峰. 电信科学. 2013(09)
- [4]数据迁移的研究[J]. 史晓燕. 浙江工商职业技术学院学报. 2007(03)
- [5]亚马逊 AWS:云上构筑大数据框架[J]. 赵艳秋. IT 经理世界. 2017(10)
- [6]技术与服务成就 AWS[J]. 朱琨. 软件和集成电路. 2017(09)
- [7]亚马逊 AWS 云服务分析[J]. 杜忠岩,张伟强,鲁华伟. 中国新通信. 2015(17)