DeepFlow®

混合云全网流量采集与分发方案

云杉网络

目录

1. 前言

经过十多年的发展，企业在 IT 基础设施以及云原生的业务应用上稳步推进。上云业务规模增加，混合云中网络变得更为复杂，企业对业务安全的诉求、行业主管部门监管的要求有增无减。本方案介绍如何在企业混合云中建设统一的全网流量采集平台。

2. 为什么混合云需要全网流量

企业 IT 基础设施部门对于网络监控并不陌生，在传统 IT 环境中，物理网络是主要部分，获取网络流量主要在网络设备及物理链路上，汇聚分流和镜像（*SPAN：Switched Port Analyzer*）是成熟的方案选择。

如今企业中的混合云环境，同样面临网络性能分析、网络问题定位及排障、网络安全管理、合规审计、网络扩展等问题。在解决以上问题时，有能力获取完整的网络流量，是一个前提。混合云包括本地部署的私有云以及使用云服务商所提供的云基础设施服务，这本身就是一个涉及多资源池信息汇总的难题。在本地部署的私有云环境中，通常涉及到多数据中心中的各类资源池，包括 OpenStack、VMware、裸金属、容器等；从网络区域中划分，涉及到业务区、互联网接入区、外联区、DMZ 区等；在云计算转型比较深入的企业中，会涉及到更多的网络功能服务链。

网络的保障涉及到配置、日志及现网流量或流数据等元素。在混合云环境中获取并管理好现网监控流量并不是一件轻松的事情。客户的业务运行在逻辑网络中，而逻辑网络是通过网络虚拟化技术，在物理交换机、虚拟交换机基础上实现的，所以通过传统的汇聚分流、物理交换机镜像方案，不能完整地描绘逻辑网络的全部流量视图，以致所熟悉的应用端到端性能分析、网络数据钻取、网络异常发现、安全分析等网络分析功能都遇到了阻碍。

在云环境下，选择网络流量采集方案需要考虑以下几个方面：

- 流量获取的方式
- 环境中的流量模型
- 规模及可管理性
- 对现网环境的影响
- 平台开放性

流量获取的方式

在云环境中获取到虚拟交换机上的流量，是完整绘制虚拟机或容器之间访问关系的必要组成部分。仅仅是获取虚拟交换机的流量，通过在交换机上设置镜像策略就很容易达到。但在生产环境中，这并不是最优的选择。主要的两个突出原因，其一是侵入生产网络的转发平面，存在镜像流表与转发流表配置冲突的风险；其二是镜像功能影响虚拟交换机的处理性能。

在目前的技术方案下，通常有以下几种方案

1) 在虚拟机或工作负载（*Workload*）中安装采集探针，从操作系统层面获取需要的信息，包括各个接口的流量。此方案由于安装基础在虚拟机，安装规模涉及数量多，并且需要获取虚拟机根（*Root*）权限。

2) 通过在虚拟交换机（*OVS：Open vSwitch*、*VDS：*vSphere Distributed Switch、*VSS：Virtual Standard Switch*）上配置镜像或广播策略，将所需流量引出。这种方案下，通常是将流量通过交换机端口引至一台虚拟机或服务器进行集中处理或分析，需要对生产平面的虚拟交换机进行配置。

3) 在宿主机 Hypervisor（*如 Openstack Hypervisor*）上通过安装采集探针，以用户态进程形式独立获取虚拟交换机上的流量，不需要对生产平面的虚拟交换机进行配置。

具体选择哪种采集方式最优，需要根据 IT 网络及资源池的实际环境情况进行选择配置或者组合。

环境中的流量模型

规划网络流量采集方案时，现网中的流量模型、主要业务的流量特征是方案选择的重要依据，基础特征包括 IP 分配、流量、包长、协议、端口、TCP、Http 信息等，同时也需要考虑组合特征，尤其是可能出现的渗透、异常等因素。

规模及可管理性

混合云环境中，网络规模宏大且资源池类型繁多，需要考虑多数据中心的整体方案，避免针对不同需求重复安装探针，分散建设分散管理的情况。虚拟交换机不再是物理网络设备，其数量相等于计算节点数量，与物理链路的采集点相比，数量是几个

数量级的增长。此外，虚拟化及容器资源池动态性很强，尤其是容器，其资源随应用需求变化频繁发生迁移、切换或回收，流量采集策略、流量分发策略也要随着变化进行迁移或释放。

在构建整体采集方案时，应充分考虑需要监控、优化的业务，分布在哪些链路、区域以及资源池，采集平台可以分阶段进行部署，但要具备扩展和统一管理能力。

对现网环境的影响

应尽可能地避免对现有云环境的影响，在已经投入生产的环境中，可能存在未规划独立的流量监控平面；逻辑 CPU 已按用途完全划分；已经部署应用不同的网络虚拟化产品方案等情况。在进行流量采集部署时，需要满足平滑部署且保证业务不间断，同时，有机制保障对计算资源的消耗限制。

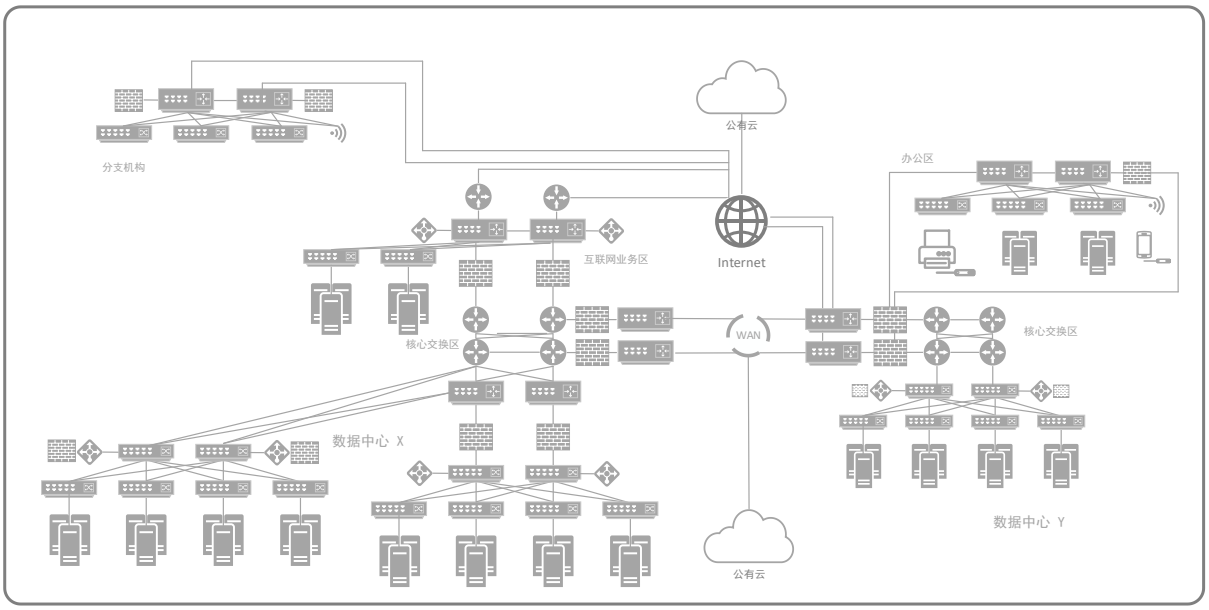
此外，流量采集系统的部署也要保证对已有的物理网络分流镜像有能力进行兼容或平滑切换，并可以对接已有的分析工具。

平台开放性

首先采集平台本身应具备开放性，避免采集端与消费端绑定，导致在现网中不断部署垂直竖井式的流量采集系统，对于流量数据应具备一次采集，可按需多处进行分析消费的能力。此外，还考虑具备数据开放性，针对原始流量数据进行处理，得到流日志、统计、特征等数据，有能力提供高性能存储写入、检索查询、API 输出等数据服务。

3. 全网流量采集与分发方案

多数大型企业目前都存在多数据中心、混合云的 IT 设施资源，从网络的角度看如下图所示，自有的数据中心通过专有网络互联，并划分业务区，并且有可能存在多个分支机构网络。为保障资源弹性，业务快速上线等，也大量使用公有云资源，选择多个云服务商。企业从运维排障、运营管理、业务性能等方面都需要对网络有全面清晰的画像。



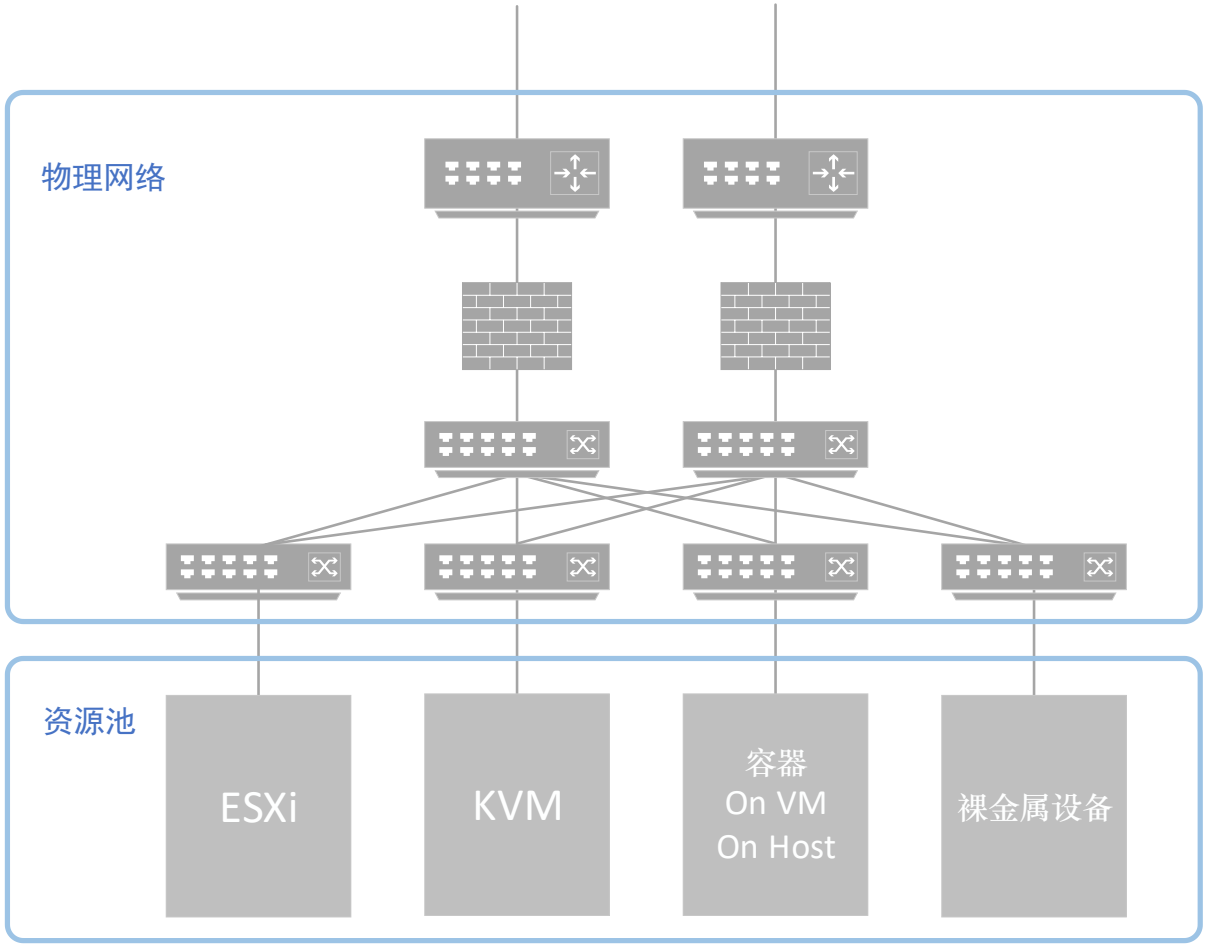
本方案的目标是为企业混合云建立统一高效的网络流量采集及处理平台，面对各类资源池实现统一的流量采集抽象层，支持 IPv4、IPv6 协议环境，并且能对流量实现过滤、去重、压缩、截短等处理功能，能为网络运营中心（*NoC：Network Operation Center*）、安全运营中心（*SoC：Security Operation Center*）、大数据分析平台等多方流量消费端提供数据供给。

实现全网流量采集及处理，可以从区域以及资源池来规划，以下分别以数据中心侧、公有云侧及整体控制管理侧来阐述方案。

3.1 数据中心侧

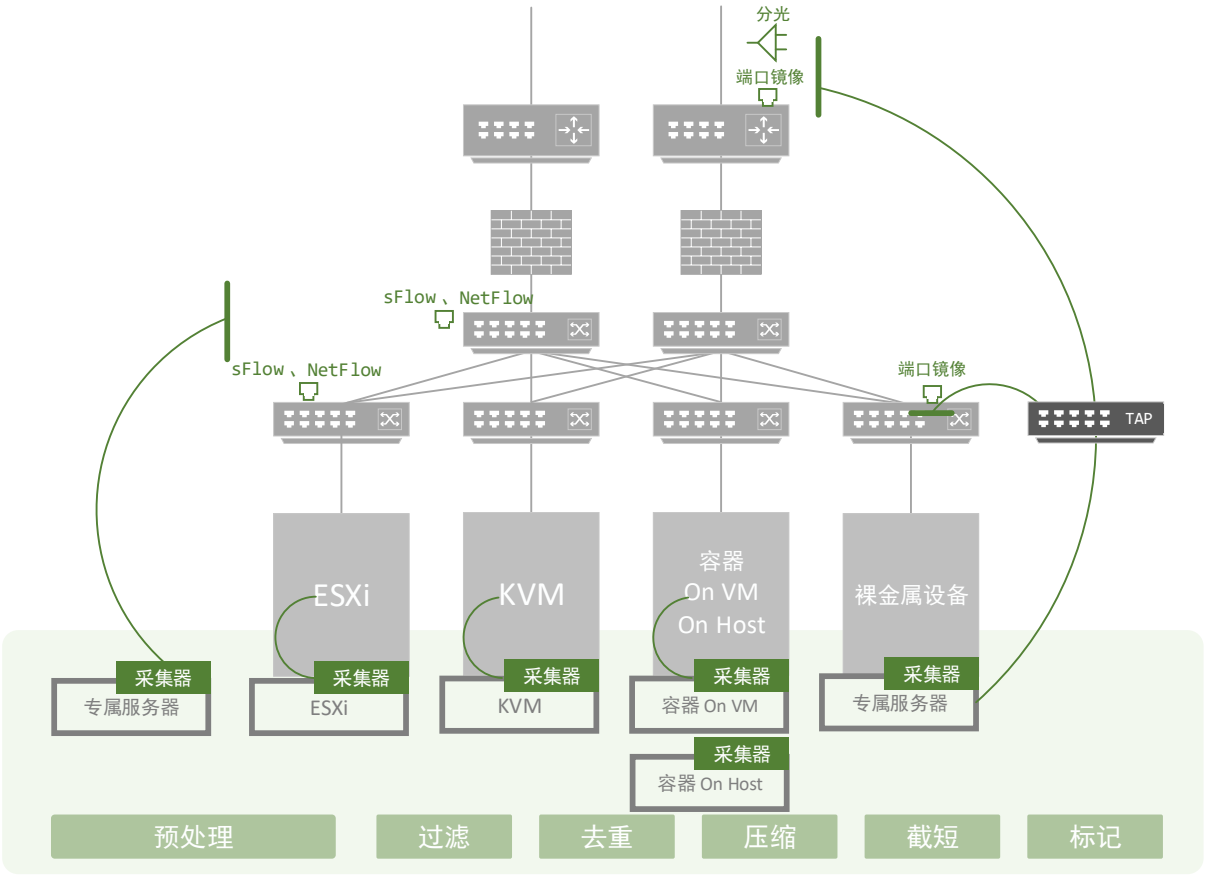
在数据中心侧，以通常所划分的网络分区为例，如互联网业务区、外联业务区、核心业务区等，对于整个平台的部署，可将数据中心按区域（*Region*）来定义，区域内可包含多个可用区（*AZ：Available Zone*）。

要获取区域内的网络流量，可包含可用区内的物理网络和资源池内的网络数据流量，如图典型区域所示。



在物理网络涉及的范围，除可用区内部网络外，还包括各类链路，如专线、互联网（*ISP：Internet Service Provider*）链路等。获取流量或流信息可通过镜像、分光、sFlow、NetFlow/IPFIX 等方式。在混合云环境中，方案挑战性比较大的是在资源池内所涉及的范围，网络边界主要由各类虚机交换机所构成，数量多、波动大，同时新技术也涉及多。

各类型号的 DeepFlow® Trident 流量采集器为全网流量采集方案提供基础捕获能力。



资源池内网络流量采集

对不同的资源池配备不同形态的采集器，提供最优的网络流量捕获能力，包括 VMware ESXi 采集器、KVM 采集器、KVM-DPDK 采集器、HyperV 采集器、容器 OnVM 采集器、容器 OnHost 采集器，避免配置池内虚拟交换机，采集器以进程形态独立运行，减少对现网的影响以及避免可能的干扰生产配置风险。

对于裸金属设备资源池，获取其池内网络流量可通过 Leaf 交换机、接入交换机的端口镜像，汇总至 TAP 设备后交由专属服务器类型采集器实现对数据包处理操作，也可以选择将采集器安装在每一台需要采集的裸金属设备系统上。

物理网络流量采集

在物理网络中，流量获取主要通过端口镜像、分光等方式获得，采集器主要对网络数据包进行过滤、去重等处理，分布的采集点主要有互联网业务区中的 ISP 线路、外联区域的专线线路、各区出口线路以及防火墙、负载均衡设备前后线路。

在对物理网络管理中，需要对物理网络交换矩阵（*Fabric*）的转发路径、端口统计、遥测数据（*Telemetry*）等信息收集展示，通常由物理设备厂商的监控方案来提

供并解决。DeepFlow®采集器可对接交换矩阵（*Fabric*）中网络设备 sFlow、NetFlow/IPFIX 等标准数据输出。

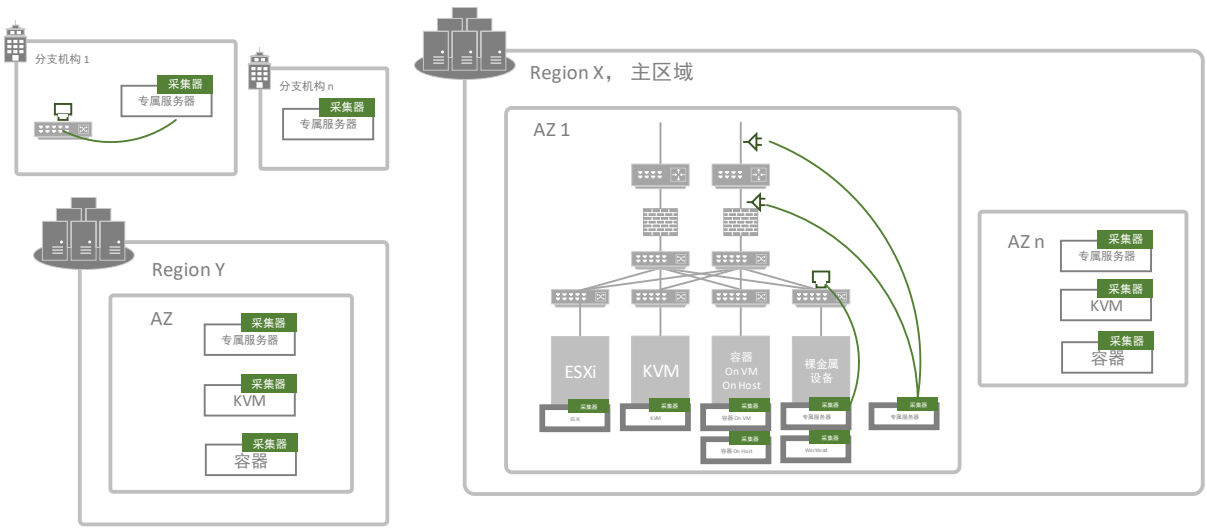
DPDK 环境下的支持

在运营商 CT（*Communications Technology*）网络中，已经应用 NFV 技术方案，在其虚拟网络实现中，虚拟交换机，如 OvS（*Open vSwitch*），通过使用数据平面开发套件（*DPDK：Data Plane Development Kit*）提升数据包处理性能。在 CT 环境中，虚拟网元（*VNF：Virtual Network Feature*）间的通信流量，尤其是控制信令流量，同样面临网络虚拟化后的采集难题。

在企业环境中，如果存在采用 DPDK 套件的资源池，方案中可以选择 KVM-DPDK 采集器进行资源池内流量采集。

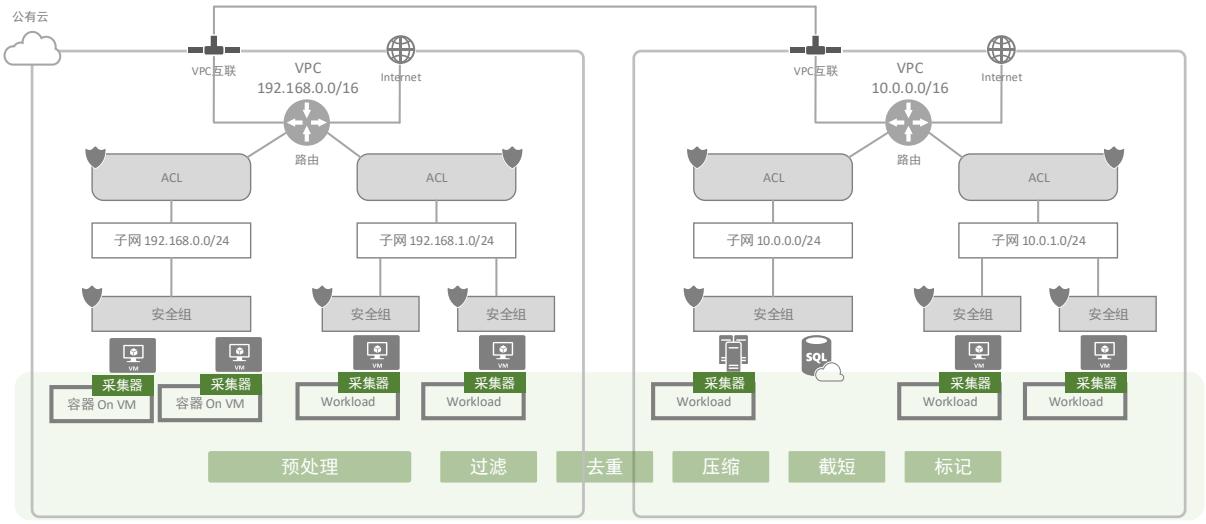
多区域支持

多数考虑统一流量采集平台的企业，IT 资源都存在于多个数据中心，而且存在众多分支机构。如下图所示，各地数据中心区域、各类资源池，网络流量采集需求都由相应型号的采集器完成。



3.2 公有云侧

公有云为租户提供 VPC 网络，Workload 采集器以用户态的软件形式部署在虚拟机、容器、裸金属设备等 Workload 上，支持 Linux、Windows 等主流操作系统，实现 VPC 内各类资源的网络流量采集。

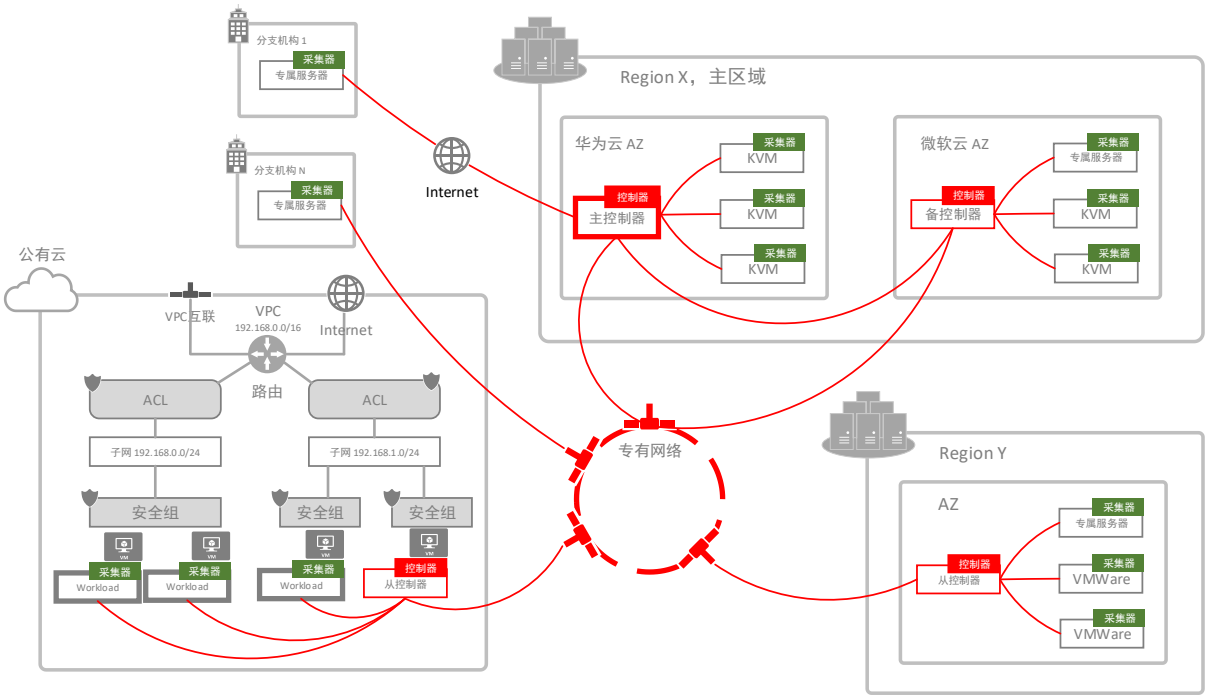


在公有云侧，Underlay 网络由云运营商维护提供，采集器以用户态进程方式安装在 Workload 操作系统上，完成网络流量获取。同时在虚拟机部署容器的环境中，容器采集器可以实现容器 POD 的网络流量获取。

由于部署安装在 Workload 操作系统上，采集器数量多，可以通过镜像进行预装。

3.3 控制管理侧

前面两章节主要介绍了采集获取各类型资源池流量的能力，既然采集器数量大，策略维度多，波动突出，对其的管理控制是方案能力评估的重点。



面对多数据中心、多云异构的混合云基础设施，统一建设网络流量的管理调度平台，解决规模大及可管理性的问题，控制面的设计是核心点。控制器是管理控制采集器及策略下发的控制中枢，可分为主控制器、备控制器、从控制器，可按照部署要求进行选择。

主控制器：整个 DeepFlow®平台的控制中枢和提供对外交互、服务的接口。部署后的 DeepFlow®平台中只有一台主控制器，主控制器所在的区域称之为主区域。

备控制器：与主控制器的功能完全一致，当主控制器出现宕机或不能提供服务的其他故障时，自动切换为主控制器。在没有备控制器的情况下，DeepFlow®控制器集群没有高可用能力。整个 DeepFlow®中只有一台备控制器，且必须和主控制器在同一个区域中，并共享一个用于提供外部服务的虚 IP 地址。

从控制器：负责控制所在区域（*Region*）或可用区（*AZ：Available Zone*）中的采集器及数据节点，将主控制器的策略和云平台资源信息同步至所有的采集器和数据节点。除主、备控制器所属的区域，每个区域中至少部署一台从控制器，同一个区域的多台从控制器之间可以实现负载均衡和高可用。

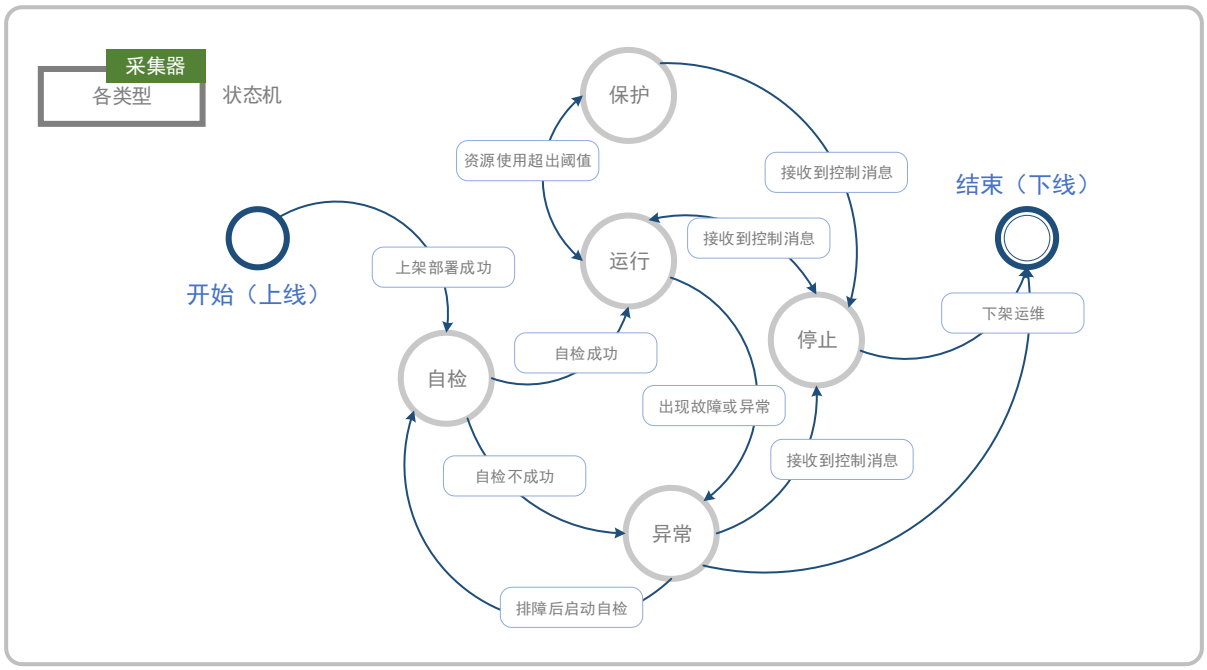
在多点的部署环境中，首先指定主区域（*Region*），主控制器存在于主区域中，当启动主控制器高可用功能，主区域内应部署多台控制器，通过心跳保证控制器间的状态同步，及时启动主、备控制器选举。选举产生主控制器后，为整体流量管理平台提供控制入口。除主区域外的其他区域控制器为从控制器，不参与主控制器选举。

在区域中可以划分多个可用区（*AZ：Available Zone*），通常以可用区为单元，由单一控制器独立控制可用区内的各类型采集器，对本地采集器进行采集策略、分发策略、预处理策略下发。多区域间可通过专线网络进行控制通信，主要包括管理、策略等通信。

通常在有分支机构的环境下，数量相对数据中心较多，主要是请求服务的流量，其区域内没有服务端，需要流量数据主要是构建网络整体状况以及业务端到端网络性能分析。不需要独立部署控制器，可以按实际情况，将采集器划分在附近区域的控制器管理下。

公有云环境中，控制器部署在虚拟机中，管理范围内的采集器。

控制器完全控制采集器状态，各类采集器具备相同状态机机制，如下图：



各类型的采集器可能处于自检、运行、停止、异常、保护等几种状中，其中保护状态，是确保采集器工作时，平台能对其使用 CPU、内存资源使用上限的限定。当配置采集器资源限制在 1vCPU1G 内存时，运行过程中，如果出现压力过大，采集器状态将由“运行”切换至“保护”状态，对所采集、处理的数据包进行丢弃，以确保不对生产环境产生影响，直至重新调整资源配置或处理压力下降，切回至“运行”状态。

另外，控制器通过对接虚拟化资源池控制器、配置管理数据库（*CMDB：Configuration Management Data Base*）、公有云开放 API 等，实现多粒度下发采集、分发策略，在云环境、容器环境中，更灵活、更贴近业务应用。

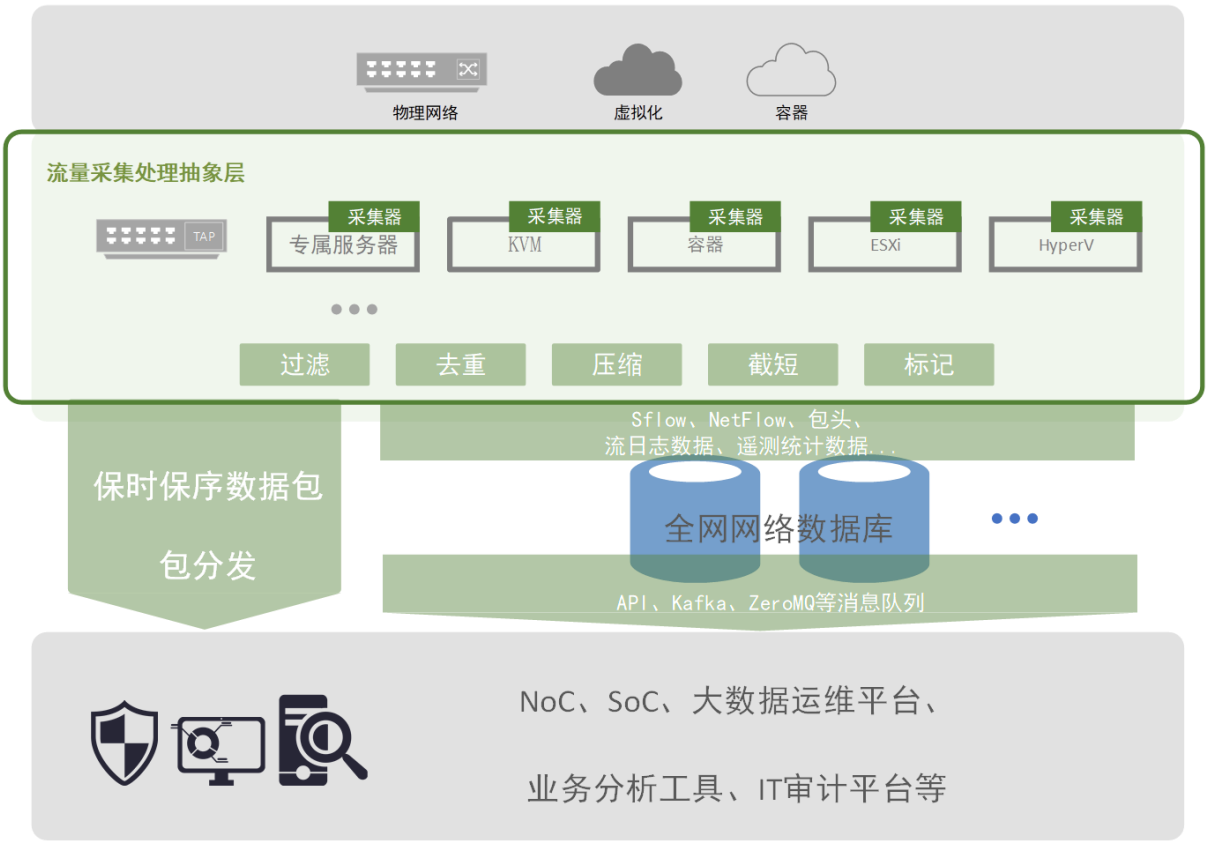
单一控制器可支持 2000 个采集器工作，这通常是一个可用区涉及的采集器规模。主、备控制器与从控制器协同工作，控制器规模最大支持 50 台，并在主区域内实现选举机制。方案整体可满足 10 万台采集器规模，具备大型企业私有 IT、公有云、容器等对网络流量采集要求。

3.4 基于分布式的监控流量处理

采集器不再是简单地获取网络流量管道，是具备对本地采集的网络流量进行处理的计算单元，众多采集器以及控制器构建成一个与云网规模一致的分布式流量处理系统。

全网网络整体状况、应用服务间访问、负载均衡、安全策略应用情况等需要对现网捕获到的流量进行处理后存储供分析展示，集中后处理海量流量需要大量扩展计算资源，在此方案中，采集器具备专利算法的前置计算能力，分布在资源池中按需对流量进行处理，有效减少分发数据对监控网络和后端分析工具的压力。

通过各类型的采集器实现流量采集处理抽象层，主要对数据包处理能力进行抽象，包括过滤、去重、数据包截短、压缩、特征标记等功能。



3.4.1 过滤

过滤能力是高效进行流量获取、以及精准实现网络流量数据价值的基础，对所有数据包无区分的处理、存储可不是一个好的可执行的方案。有过滤后，那条件维度就是下一个重要因素，对采集策略、分发策略、处理策略仅仅基于网络五元组设置过滤条件是远远不够的，在池化、多租户、容器环境中，仅有这些可以说是过时的。更丰富的过滤条件，如业务、主机、服务、POD 等维度都是需要加入的。

3.4.2 去重、截短、流日志、压缩、标记

去重能力是保证获取流量数据后的准确性，采集器存在于网络流量的两端，有可能分布在不同资源池，不同区域，同时捕获后，进行统计、区分源、目的端等，为分析、可视化的准确性提供支撑。

截短能力是对获取数据包后，针对性应答数据消费端需求的能力，同时，也是压缩能力的基础之一。可提供数据包包头，包头后指定偏移长度的截短能力。

流日志能力是对流量数据包获取网络元数据的能力，是对全网整体绘图、回溯查询的基础。支持 80 种类型元数据获取，除了基础的源、目的 MAC、IP 地址、端口外，按需求可获取更多 TCP 以及性能的日志信息。

压缩能力是保证获取流量数据后，有效利用传输带宽、存储资源。如果后端消费端需要数据包包头或流日志数据，可获得的压缩比 100:1，如果仅需要遥测统计数据，压缩比达到 10000:1，这也保证在整体方案中，对于一些分支机构，仅需要获取流日志信息，不需要专属网络线路，可以通过 Internet VPN 搭建方案。

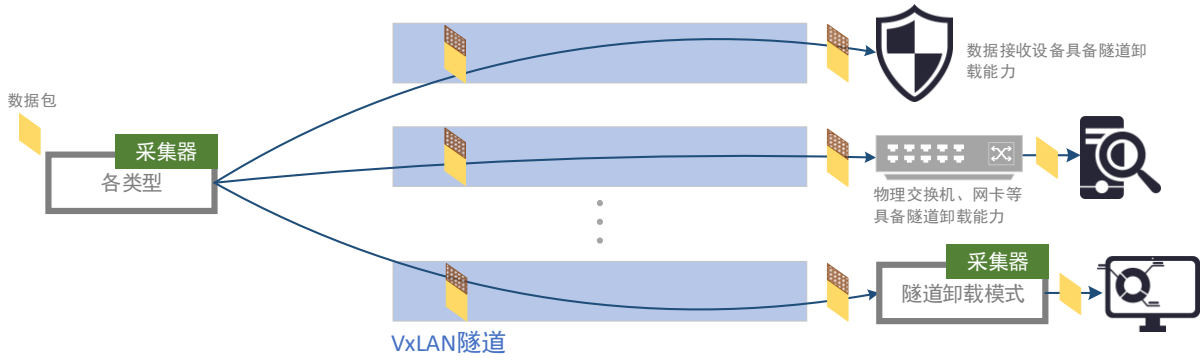
特征标记能力是对数据包分发过程中，分发起点在封装的隧道包头保留字段中，标记特征值，卸载封装时可以识别。供后端分析工具、运维平台针对特殊数据包识别，可用于网络诊断、采集点定位等场景。

3.4.3 包分发

包分发功能是解决对完整的数据包分析、安全保证的需求，核心需要保障数据包的原始性，包括数据包内容、顺序等，同时可以针对单一数据包进行多目的地的分发。

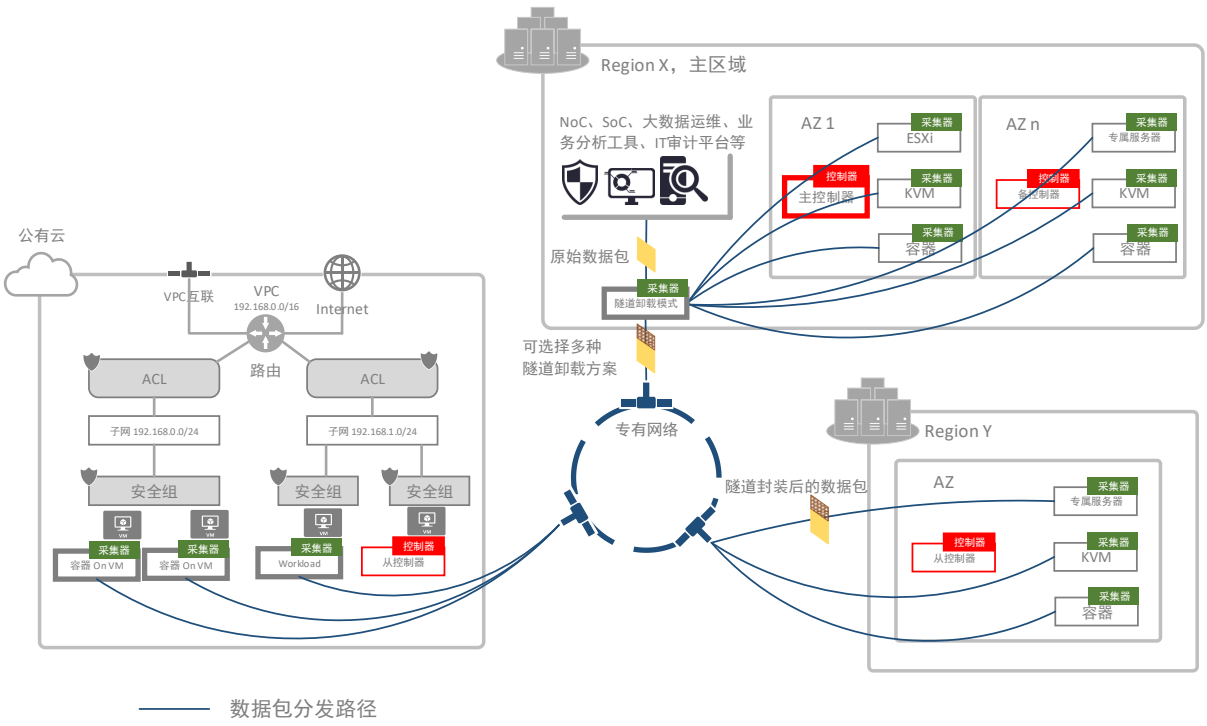
包分发功能是通过三层隧道实现（ERSPAN、VxLAN），控制器统一下发分发策略，由涉及的采集器端直接进行数据包封装，并发往目的端，支持多目的端发送。

在混合云数据包分发的方案中，需要考虑分发的网络平面，如果分发流量较大，可以考虑预留独立的网络监控平面。如果仅针对少量核心业务，可以复用已有的物理网络。



在分发的目的端，需要考虑对封装隧道的卸载方案。通常如果目的端是 NoC、SoC、大数据平台等大型平台，可以由物理交换机的 VxLAN 卸载功能进行隧道解封装。同时如果针对一些传统的分析工具，没有隧道卸载能力，可以使用专属服务器采集器，运行隧道卸载模式，部署在分析工具前端，解封装后，将数据包送至分析工具。当然，如果分析工具具备 VxLAN 卸载能力，可以直接接收隧道数据包。

整体分发方案应用如下图所示。在传统物理网络环境中，NPB 设备及方案很好地完成了数据包分发的需求，但在混合云环境中，资源池数量多，种类不同，在混合云环境中实现监控流量按需分发能力，并将分发操作以分布式架构分布在各个采集点实现，避免单点性能瓶颈及适配逻辑网络跨多资源的场景。

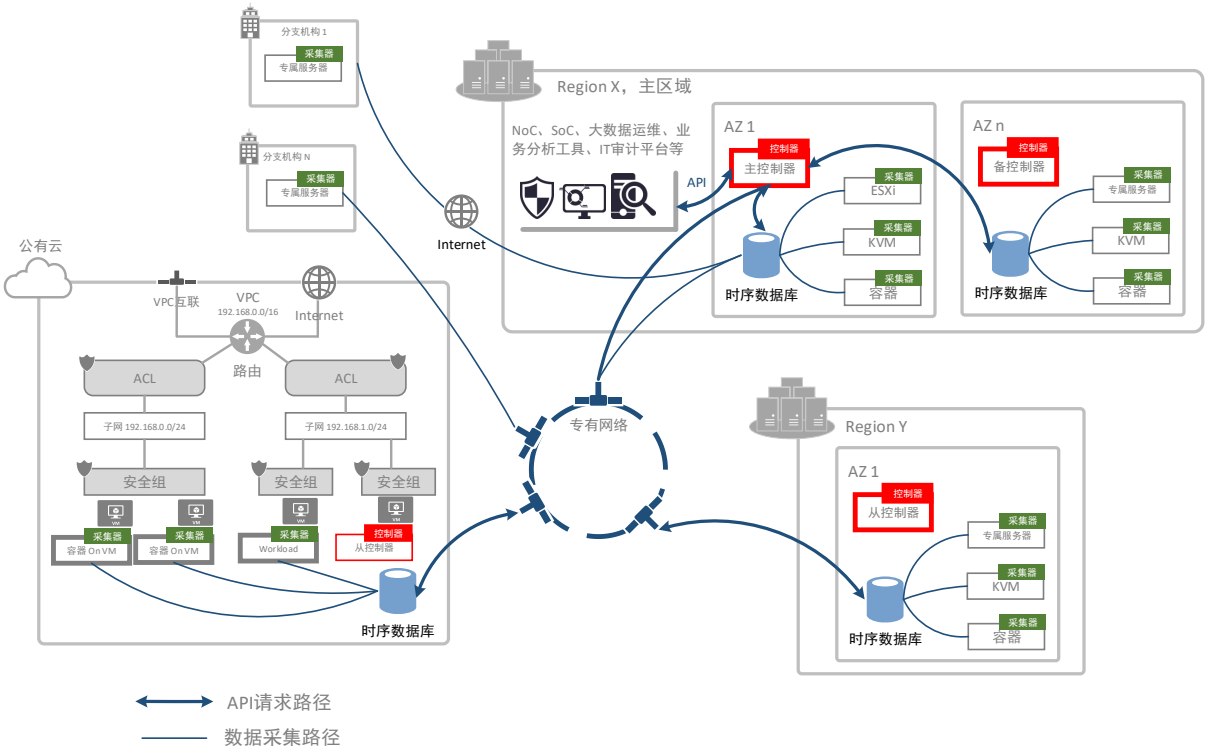


3.4.4 数据服务

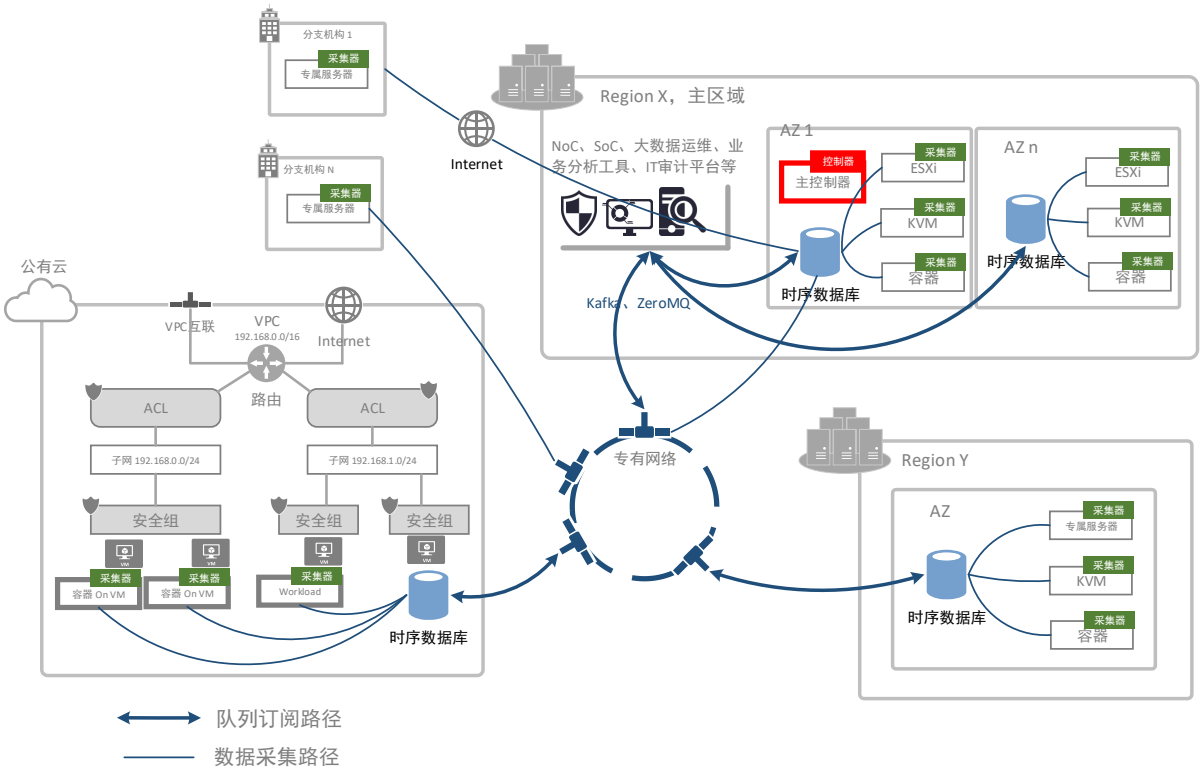
对于非原始数据包的数据消费需求，平台提供开放的数据订阅方式。处理后的包头，网络元数据、遥测统计数据通过网络平面汇总至高性能时序数据库中，可通过API，消息队列为其他数据消费平台调用。

在每个区域、可用区都可以配置高性能时序数据库，通常在分支机构环境下，不需要部署时序数据库，其数据通过压缩后写入纳管区域内的数据库。

主控制器直接响应对网络数据的 API 调用，由主控制器查询本地时序数据库，或收集区域内数据库 API 返回的结果并回复请求端。



数据订阅可通过ZeroMQ等消息队列提供，由数据需求平台向数据库发起消息队列请求后，就可执行订阅服务。



数据订阅消息队列示例，Python：

```python
#!/usr/bin/python

import traceback

import zmq

from google.protobuf.text_format import MessageToString

import dfi_pb2

STREAM_PORT = 20204

def get_socket():
    ctx = zmq.Context.instance()
    socket = ctx.socket(zmq.SUB)
    socket.set(zmq.LINGER, 0)
    socket.setsockopt(zmq.SUBSCRIBE, "")
    socket.connect('tcp://127.0.0.1:{}'.format(STREAM_PORT))
    return socket

def main():
    socket = get_socket()
    flow = dfi_pb2.Flow()
    try:
        while True:
            bs = socket.recv()
            flow.ParseFromString(bs)
            print(MessageToString(flow, as_one_line=True))
    except KeyboardInterrupt:
        pass
    except:
        traceback.print_exc()
    socket.close()

if __name__ == '__main__':
    main()
```

3.5 部署

整体方案主要涉及采集器、控制器、高性能时序数据库三部分，在完成规划整体方案后，可分区域、分资源池按阶段投入建设，最终为企业混合云 IT 基础设施环境构建统一的流量监控管理平台。

由于传统物理网络已具备完整的监控方案，通常选择以 KVM、容器资源池进行第一步部署实施，解决虚拟网络环境流量“黑盒”不可见的问题，满足对虚拟网络流量合规审计的要求；采集流量对接已存在的监控分析工具，闭合私有云、容器环境中的运维、业务分析工具链。

第二步纳入更多资源池，与新建扩容的资源池同步部署，接入物理网络中交换机 sFlow 数据，接入专线等分光流量数据，实现对整体数据中心的流量采集能力，存在统一的流量监控平面；对接网络中心、安全中心、智能运维等平台，提供数据包、流数据服务，满足各平台对现网流量数据的展示、分析需求。

第三步可以对存在公有云上所运行的 Workload 或实例流量进行采集，完成对混合云 IT 环境整体监控流量管理，具备整体网络画像、流量分发、支持对多平台流量数据分发服务能力。

对于已经运行的混合云环境，可以在不影响生产环境运行的情况下部署实施，网络规划上将 DeepFlow®平台所涉及的管理、监控分发平面复用在已有的网络平面中，通常可以复用已经存在的网络管理平面。

对于整体规划的方案，建议对整体混合云规划独立的网络监控平面，对于混合云的监管流量统一、独立地进行管理。另外，采集器对计算能力的要求，可以根据处理流量、资源情况进行整体规划，对单一采集器最低可配置 1vCPU 128M 的资源使用。

3.6 方案优势

流量采集先进

DeepFlow®全网采集方案中，主要是以采集器技术实现流量采集，采集器支持 KVM、VMWare、容器等型号，以进程形态部署安装，最大程度上避免对生产交换平面的干扰，不存在与生产平面交换机流表冲突的风险，同时在操作系统上继承进程级保护优势，实现整体系统稳定。

分布式处理系统

采集到数据包后避免集中处理，采用分布式架构，采集点分布处理控制器集中管理。

场景全规模大

整体方案是基于分布式设计模型以及多地域管理，可以充分保障资源池规模弹性扩展，整体系统可管理 10 万台采集器，涵盖虚拟机、容器、公有云资源池。

可管理性

平台主控制器是管理员统一集中的采集、分发策略配置入口，同时具备对所有采集器状态管理能力。各类操作贴近资源池特性，支持虚拟机名称、子网、集群、容器 POD 等多维度进行。资源存在迁移、回收、重新部署等场景，策略跟随保障采集能力在动态环境下的持续执行。平台管理的粒度是单一采集器，对采集器的管理控制以及运行状态，历史记录都可回溯跟踪。

数据包、流数据服务

数据服务是将流量采集与后端平台对接的重要环节，完整流量数据包多目的地分发，高性能网络时序数据库通过 API、ZeroMQ、Kafka 等消息队列提供流数据服务。同时也将采集与后端各类分析工具解耦，避免流量采集器局限在仅为单一工具服务的竖井中。

4. 总结

DeepFlow®混合云全网流量监控采集与分发解决方案为企业在混合云、云原生等新型 IT 基础设施环境演进过程中，提供完整地、可持续的平台级流量管理，避免重复投入，重复安装，解决实际网络监管难题，也为企业规划整体运维、安全平台补齐现网流量、流日志这一板块。本方案已应用于金融、运营商等客户 IT 环境中。

了解更多信息

专业的售前技术支持及商务合作，协助您选择最合适的解决方案

详询：400-9696-121

网址：www.yunshan.net

北京云杉世纪网络科技有限公司

北京市海淀区成府路 28 号优盛大厦 A 座 1209