ICS 35.240.40
CCS A 11

团 体 标 准

T/BFIA 013—2022

面向新型金融应用的全 IP 数据中心网络技术要求

All-IP data center network technical requirements for new financial applications

2022－08－16 发布　　2022－08－16 实施

北京金融科技产业联盟　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由北京金融科技产业联盟归口。

本文件起草单位：中国银行股份有限公司、北京金融科技产业联盟、中国工商银行股份有限公司、华为技术有限公司。

本文件主要起草人：潘润红、许泓、聂丽琴、吴仲阳、刘新儒、赵文烜、张继东、王磊、李露伟、赵力、黄本涛、李明艳、周豫齐、李璐、宋新超、崔洪斌、徐晓宇、林艺宏。

引　言

金融机构的网络主要分为数据中心网络、广域网络、分支园区网络，其中数据中心网络作为业务运行和交换承载体，是金融业务高效、稳定运行的基石。

新技术的飞速发展驱动传统金融行业进行深刻的业务变革。新型金融应用如智慧营销、智慧风控、智慧经营的规模部署及金融行业向分布式架构转型，对ICT基础设施提出更高的要求。RoCEv2技术应用和全IP数据中心网络架构成为金融数据中心重要的演进方向，不仅实现了技术产业代际升级，提升海量用户的并发处理能力，缩短新型智慧金融应用系统的实时响应时间，同时为金融行业IPv6规模化部署给出新的演进方案。

本文件提出面向新型金融应用的全IP数据中心网络技术要求，对于指导金融机构的全IP数据中心网络技术架构、关键技术能力等核心问题给出了可行的技术参照体系。

面向新型金融应用的全 IP 数据中心网络技术要求

1 范围

本文件规定了面向新型金融应用的全IP数据中心网络部署架构和关键技术要求。

本文件适用于银行业机构金融数据中心网络建设，可供保险、证券等其他金融机构参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件，不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

YD/T 3096—2016 数据中心接入以太网交换机设备技术要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

新型金融应用 new financial applications

为满足智慧营销、智慧风控、智慧经营等新兴业务，采用大数据平台、人工智能、分布式数据库等技术的金融应用系统。

3.2

数据中心 data center

由计算机场地（机房）、机房基础设施、信息系统硬件（物理和虚拟资源）、信息系统软件、信息资源（数据）和人员以及相应的规章制度组成的组织。

[来源：GB/T 33136-2016,3.1.1]

3.3

全IP数据中心 all-IP data center

数据中心的网络承载技术统一采用IP技术栈实现。

3.4

同城双中心 dual-data center

根据金融业务等级的恢复时间目标（RTO）和恢复点目标（RPO）灾备指标要求，在同城或邻近城市可独立承担业务的两个数据中心，两个数据中心的物理距离通常在100公里以内。

3.5

基于融合以太网的RDMA（第2版） RDMA over Converged Ethernet version 2

一种基于以太网络的网络层协议，允许在以太网上使用RDMA（远程直接内存访问）网络技术，可减少CPU开销，提高数据吞吐，降低网络延时和抖动。

3.6

网络遥测 Network Telemetry

新一代从网络设备上远程高速采集数据的网络监控技术，设备通过“推模式（Push Mode）”周期性地主动向采集器上送设备信息，提供更实时、更高速、更精确的网络监控功能。

[来源：IETF RFC 9232]

4 缩略语

下列缩略语适用于本文件。

CPU：中央处理器（Central Processing Unit）

DSCP：区分服务编码点（Differentiated Services Code Point）

ECN：明确拥塞通告（Explicit Congestion Notification）

ICT：信息和通信技术（Information and Communications Technology）

IP：网际协议（Internet Protocol）

I/O：输入输出（Input/Output）

PFC：基于优先级的流量控制（Priority-Based Flow Control）

QoS：服务质量（Quality of Service）

RDMA：远程直接内存访问（Remote Direct Memory Access）

RoCEv2：基于融合以太网的RDMA（第2版）（RDMA over Converged Ethernet version 2）

TCP：传输控制协议（Transmission Control Protocol）

UDP：用户数据报协议（User Datagram Protocol）

5 网络部署架构

5.1 流量特征

新型金融应用如智慧营销、智慧风控、智慧经营的规模部署及金融行业向分布式架构转型。各业务场景典型应用流量特征如表1：

表 1 新型金融应用流量特征分析

场景	典型应用	主要流量特征和要求
计算类	大数据应用	I/O 密集，多对一流量模型，大带宽、时延敏感。
	人工智能	I/O 密集，多对一、多对多流量模型，大带宽。
存储类	分布式存储	I/O 密集，多对一流量模型，大带宽。
	集中式存储	I/O 密集，时延敏感，高可靠。

5.2 网络架构

为满足新型金融应用各业务场景流量特征要求，根据金融业务系统对数据中心内部的通用计算、大数据人工智能等密集计算、后端数据存储等流量特征维度进行网络类型划分，全IP数据中心网络部署参考架构如图1所示：

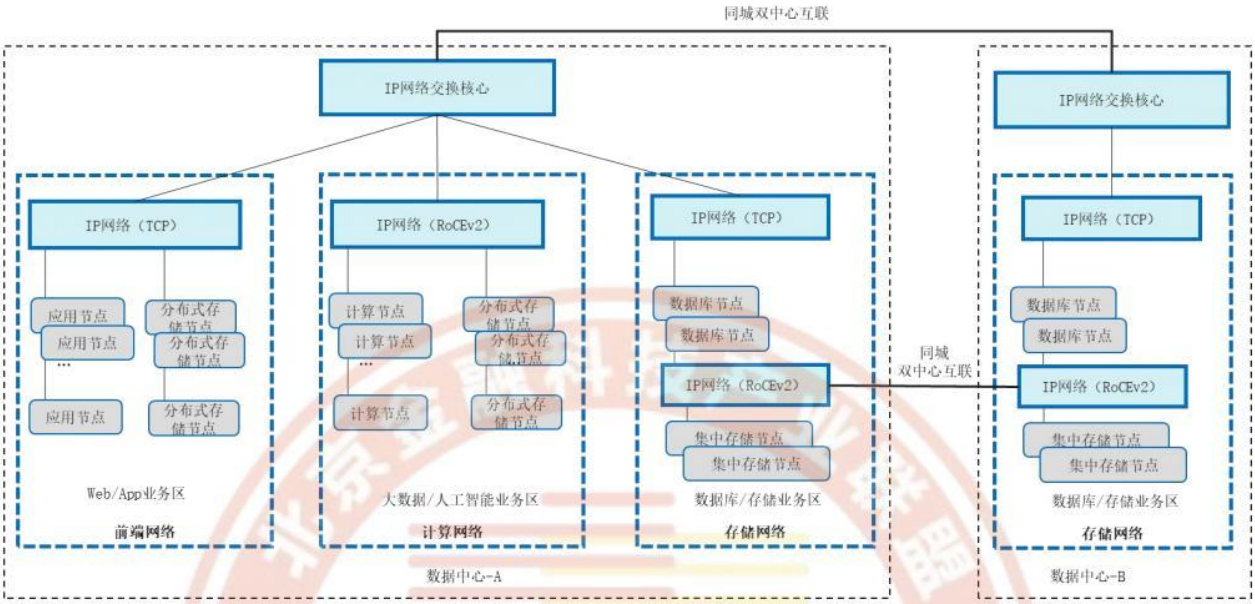

图1　全IP数据中心网络部署架构

前端网络：用于连接前置类应用的服务器节点、存储节点，以及对外访问的互联。服务器要求实现虚拟化、容器化、云化管理部署，要求网络满足业务弹性扩容和自动化部署的要求，采用主流TCP/IP协议栈相关技术，满足Web/APP类应用节点互联和对外互联的诉求。

计算网络：用于大数据分析、人工智能业务区的计算类节点和存储节点。该类业务区要求计算节点提供高算力，满足海量数据的处理和分析，涉及金融实时类业务要求提供分钟级、秒级等处理响应能力，要求网络提供高并发计算、大流量的低时延连接能力。业界RMDA技术栈已经成熟应用，相比TCP/IP技术栈，在吞吐量、CPU开销等方面具备更大的优势，采用RoCEv2网络承载技术，满足大数据、人工智能、分布式存储等应用的低时延、I/O密集型的通信诉求。计算网络支持的计算节点部署规模和应用场景相关。

存储网络：用于连接金融核心业务系统的数据库节点、存储节点。该类业务要求提供高并发I/O吞吐、以及低时延响应、高可靠要求。金融核心存储业务要求同城实时复制，以及异地灾备复制能力，为避免不受其他业务流量的冲击，存储网络要求单独组网。当前RDMA技术已验证具备更好的I/O吞吐和低时延，本地数据中心和同城双中心满足实时性要求，采用RoCEv2网络承载技术，满足低时延、高并发，以及同城复制场景下的低时延长距诉求。

IP网络交换核心：用于多个计算网络以及前端网络等交换，能够提供对TCP/UDP/RoCEv2混合流量的SLA保障，支持同城双中心互联。

5.3　网络协议栈

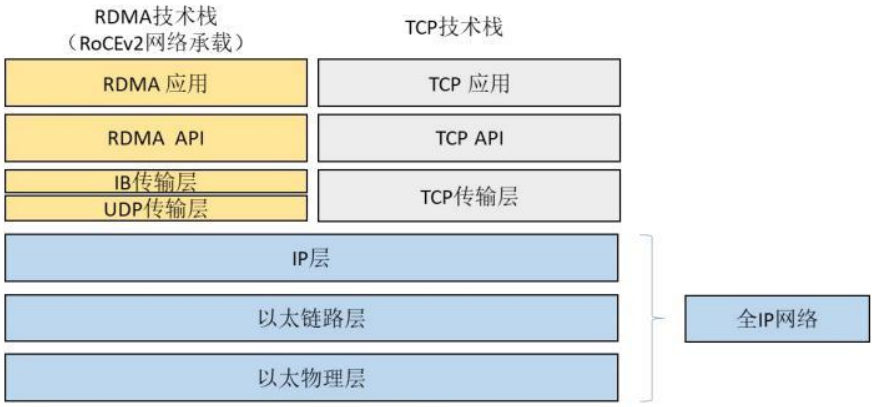

图2　全IP网络技术协议栈

金融分布式架构下ICT基础设施的服务器、存储、网络要求采用开放互联技术，根据金融业务应用的不同要求可以选择采用TCP技术栈或RDMA技术栈，如图2所示。

采用RDMA技术栈，应用系统、操作系统、网卡、网络设备需要配套支持RoCEv2网络技术；存储网络场景下，存储系统需要支持NVMe-oF（NVMe over Fabrics）接口协议标准。RoCEv2技术是基于IP网络实现（UDP协议），与TCP/IP技术栈在IP层/以太链路层/以太物理层实现了统一，能够充分满足新型金融应用的发展需要。

全IP统一承载网络架构在网络的端口带宽演进（1GE/10GE/25GE/100GE/400GE等）、运维、云平台协同、IPv6应用升级改造等方面，产业生态发展更健康，符合金融分布式改造的演进方向。

6　关键技术要求

6.1　总体要求

新型金融应用的部署区域网络涉及计算网络、存储网络和IP网络交换核心三类，可根据业务需求采用RoCEv2网络技术承载。

关键网络技术要求见表2。

表 2 关键网络技术要求

技术项	技术要求	计算网络	存储网络	IP 网络交换核心
网络基础要求	交换和路由功能	●	●	●
	QoS 功能	●	●	●
	可靠性组网	●	●	●
网络增强要求	拥塞控制技术	●	●	●
	长距低时延网络控制优化技术	NA	●	NA
	故障快速切换	NA	●	NA
网络运维要求	设备 Telemetry 采集技术	●	●	●
	网络故障告警	●	●	●
	网络设备关键资源和状态的可视	●	●	●
注：●表示必备要求，NA 表示不涉及。				

网络基础要求是为了满足金融应用系统运行的基本网络诉求，各类区域网络均应达到基础要求。

网络增强要求是指对应用系统的I/O吞吐能力以及对时延和可靠性有高要求的场景（如同城双中心存储实时复制场景），对网络提出的更高技术要求，满足此类场景业务要求的区域网络建设应达到特定的增强要求。

网络运维要求是面向各种业务场景下的通用网络技术要求，各类区域网络均应达到运维要求。

6.2 网络基础要求

6.2.1 交换和路由功能

支持的链路层数据交换和路由协议功能，应符合 YD/T 3096—2016。

6.2.2 QoS 功能

支持 QoS 功能要求，应符合如下技术要求：

a) 支持多队列技术，基于流量优先级（具体内容参考 IEEE802.1P）或 DSCP 实现基于优先级的多队列功能，不同优先级的业务流量进入不同队列；
b) 支持基于严格优先级和加权循环优先级调度；
c) 支持队列与带宽保障功能；
d) 支持显示拥塞通知 ECN（具体内容参考 IEEE 802.1Qau 拥塞通知），支持通过配置修改高低门限；
e) 支持 PFC 优先级流量控制（具体参考 IEEE 802.1Qbb 基于优先级的流量控制），支持基于队列灵活配置 PFC，调整 PFC 水线；
f) 支持 PFC 死锁检测技术，支持在指定检测周期内检测到死锁时，暂停响应或者关闭 PFC；
g) 支持 PFC 风暴扩散预防技术；
h) 支持流量整形功能，限制流量与突发，支持均匀的速率向外发送。

6.2.3 网络可靠性组网

支持网络设备可靠性组网，应符合如下技术要求：

a) 支持网络设备双上联跨设备链路捆绑；
b) 支持网络设备 VRRP 路由冗余技术；
c) 支持服务器双上联接入交换机。

6.3 网络增强要求

6.3.1 拥塞控制

拥塞控制技术要求如下：

a) 应支持 RoCE 和 TCP 流量混跑 SLA 保障协议；
b) 应支持 PFC 死锁预防技术，避免死锁引起的流量传输大幅下降或停滞；
c) 应支持 PFC 风暴隔离技术，支持针对出现 PFC 风暴的端口进行自动关闭，实现故障隔离；
d) 应支持存储转发和直通转发模式。如果出现网络拥塞，则自动改为存储转发模式转发报文，网络拥塞消除后，自动恢复为直通模式转发报文；
e) 应支持在网络间歇性出现拥塞时，自动调整 ECN 高低门限和标记概率，来满足低时延、高吞吐的要求。

6.3.2 长距低时延网络控制优化

长距低时延网络控制优化技术要求如下：

a) 应支持同城双数据中心的长距流量突发时，支持根据拥塞情况进行调度,通过网络基于流进行快速降速通告，满足跨数据中心的低时延、高吞吐的要求；
b) 应支持同城双数据中心的长距离网络时延检测，分配缓存，满足无丢包的要求；
c) 应支持不低于 100km 的同城长距无损网络转发能力，端口速率应支持 10/25/40/100G。

6.3.3 故障快速切换

故障快速切换技术要求如下：

a) 应支持 IP 域管理，管理接入主机/存储设备的信息，监控网络状态；网络设备之间需要实现 IP 业务域与主机/存储设备状态信息同步，实现主机和存储设备之间自动建立连接；
b) 应支持与主机/存储设备的协同，实现主机/存储的端到端故障路径切换时间<2 秒；
c) 应支持同城双数据中心的长距离无损链路故障切换时间<2 秒，避免影响同城存储复制类 I/O 吞吐量；
d) 应支持数据中心网络自身故障的快速切换能力，网络设备单点、链路故障场景，要求网转发路径切换时间<2 秒。

6.4 网络运维要求

网络运维要求如下：

a) 应支持设备 Telemetry 采集技术，间隔达到 100 毫秒级，满足设备实时状态监控；
b) 应支持如下网络故障告警：
   1) 网络设备转发表项资源不足；
   2) 网络设备丢包；
   3) 网络与主机 IP 地址冲突；
   4) PFC 触发死锁或风暴；
   5) 光模块异常或亚健康状态。
c) 应支持如下网络设备关键资源和状态的可视：
   1) 设备/单板级的 CPU、内存利用率；
   2) 接口级的收发字节数、收发丢包数、收发错包数等；
   3) 队列级 buffer 缓存的资源监控；
   4) 基于网络设备队列的吞吐可视；
   5) 网络设备队列 KPI 可视：PFC 计数、ECN 统计、死锁统计等；
   6) 异常事件可视，如发生异常丢包、设备转发超时延。

参 考 文 献

[1] GB/T 33136—2016 信息技术服务 数据中心服务能力成熟度模型

[2] YD/T 1693—2007 基于光纤通道的IP存储交换机技术要求

[3] IEEE 802.1P 有关流量优先级的局域网第二层 QoS/CoS 协议（LAN Layer 2 QoS/CoS Protocol for Traffic Prioritization）

[4] IEEE 802.1Qau 拥塞通知（Congestion Notification）

[5] IEEE 802.1Qbb 基于优先级的流量控制（Priority-based Flow Control）

[6] IETF RFC 3168 TCP/IP显式拥塞通知机制（The Addition of Explicit Congestion Notification (ECN) to IP）

[7] IETF RFC 8257 数据中心TCP拥塞控制机制（TCP Congestion Control for Data Centers）

[8] IETF RFC 9232 网络遥测框架（Network Telemetry Framework）

[9] Liang Guo, Paul Congdon. IEEE SA Industry Connections - IEEE 802 Nendica Report: Intelligent Lossless Data Center Networks[J/OL]. IEEE SA Industry Connections, 2021 [2021-06-22]. https://ieeexplore.ieee.org/document/9457238

[10] NVM Express. NVM Express™ over Fabrics Revision 1.1a[S/OL]. NVM Express Base specification.[2021-07-12].https://nvmexpress.org/wp-content/uploads/NVMe-over-Fabrics-1.1a-2021.07.12-Ratified.pdf