

# 全新英特尔® 数据中心 GPU Flex 系列， 助力火山引擎打造高画质、低时延 云游戏体验



“借助云服务、边缘计算等技术的加持，我们的云游戏解决方案正帮助更多玩家不受终端配置束缚，体验到更多的高品质游戏大作。而英特尔® 数据中心 GPU Flex 系列产品在音视频渲染与编码等方面的卓越表现，使新方案能在更低的成本下为玩家提供更高画质的视觉效果和低时延的操作互动，也获得了玩家的高度认可。”

梁宇  
云游戏架构师  
火山引擎

## 前言概述

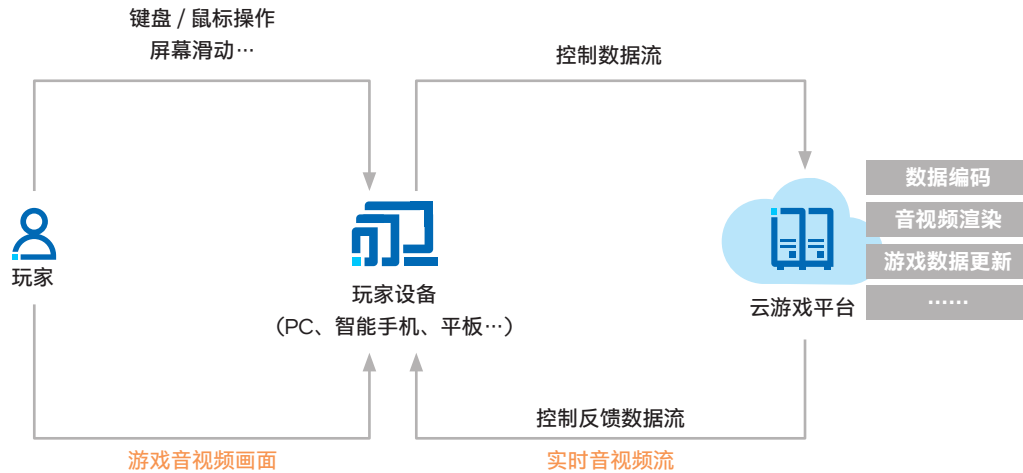
5G 网络以及边缘计算 (Mobile Edge Computing, MEC)、服务器 GPU、音视频编解码等技术的持续发展，正为云游戏产业带来新的契机。这一大潮中，作为字节跳动旗下的云服务品牌，火山引擎也正借助其丰富的互联网运营经验和雄厚的技术积累，为云游戏服务提供高品质方案。

云游戏的本质是结合云计算和低时延实时音视频流技术的云服务。高品质的云游戏解决方案，需要在成本可控的前提下构建起高效能的音视频渲染和编码能力，并可根据网络状况将编码后的音视频流高效地推送到玩家终端上。而以上能力的实现，则有赖于是否为云端平台配备了强有力的 GPU 产品，并进行了端到端的全链路优化。

把握云游戏产业发展新契机，火山引擎与英特尔展开合作，引入英特尔® 数据中心 GPU Flex 系列，作为云端渲染和编码工作的核心引擎，结合自研的 RTC (Real Time Communication, 实时通信) 产品 ByteRTC、智能调度技术等，打造全新的火山引擎云游戏解决方案。通过在《航海王热血航线》等云游戏的成功部署，新方案已被证明具有强劲的渲染和编码能力，并能以低时延保证操作的流畅性，为玩家提供了可媲美本地游戏的良好体验。

## 背景概述

作为一种不依赖终端配置、适用于各类系统平台的全新游戏体验方式，云游戏能让玩家在不购买高配置终端的前提下，免安装即可进入酣畅淋漓的高品质游戏世界，受到了众多游戏玩家的青睐，并被视为游戏产业的重要发展趋势。有研究报告预测，2022 年至 2025 年，中国云游戏市场复合增长率 (CAGR) 达到 76.6%，2025 年中国云游戏市场收入有望达 248 亿元，月活用户超 1.8 亿人<sup>1</sup>。



图一 云游戏基本架构

从本质上而言，云游戏结合了云计算和低时延实时音视频流技术，如图一所示，玩家在不同的设备上通过客户端进入游戏后，可通过 5G 等高速网络将鼠标操作、屏幕滑动等控制数据传送到部署在云端或边缘节点的云游戏平台中；平台收到控制数据后，会根据游戏场景、人物和行为的变化，更新游戏数据并重新渲染游戏画面，并将新的音视频进行编码后以实时音视频流的方式反馈到玩家设备上。

众所周知，今天的游戏追求融合各类画面、声音和交互操作的沉浸式体验。相对于本地游戏，非本地化配置的高性能游戏处理能力既是云游戏的优势，也是其面临的巨大技术挑战。长距离数据传输、音视频渲染、编码都会增加游戏运行时的时延，并由此带来画面卡顿、操作延迟等问题，影响游戏体验。要有效应对这些挑战，云游戏平台需要根据玩家接入的密度与游戏场景的变化，对整个链路开展一系列优化，使其接近于本地游戏的时延水准。

而在游戏品质日趋精良，且玩家活跃度不断提升的今天，这一优化并非易事。目前，许多电竞级云游戏已达 144FPS (Frames Per Second, 每秒帧数) 的刷新率和 2K 级的分辨率，同时业界编码格式也逐渐从 H.264 向 H.265 迁移。此外，云游戏对渲染和编码能力的需求与游戏画面的分辨率、刷新率等参数正相关，且会随着游戏并发路数的增加而同步增加。与此同时，为降低网络带宽压力，云端的游戏画面通常需经过编码后再进行传输。因此当海量玩家在线时，需要实时渲染和编码的工作负载就会达到一个惊人的数量。

这些工作负载需要云游戏平台部署大量的 GPU 来承载，而传统的 GPU 由于并未特别考虑云游戏场景下的需求，在渲染、硬件编码能力等方面往往遇到瓶颈。用户通常只能通过堆叠更多硬件来予以解决，而由此也带来了巨大的部署成本。与此同时，面向复杂多变的网络传输情况，云游戏平台还需要通过有效的优化方案来应对网路抖动、网络丢包等问题，解决因网络问题引发的画面丢帧、卡顿等现象，提升用户体验。

为有效应对以上挑战，字节跳动旗下的火山引擎与合作伙伴英特尔一起，借助其强大的云服务技术积累以及丰厚的互联网服务经验，开发新一代火山引擎云游戏解决方案，在保证其支撑的云游戏大作具有更华丽炫酷的画面品质和更加流畅顺滑的操作体验的同时，具备以下效能表现：

- **更快的端到端响应速度：**为了让玩家体验云游戏时，不因响应时延过长而影响游戏体验，端到端响应时延压降至 70 毫秒以内<sup>2</sup>；
- **更优的网络状态调节能力：**为了使云游戏在不同移动网络环境下依然能保持良好的流畅感，新方案计划加入网络状态动态调节能力来确保网络延迟一致性；
- **更低的单路云游戏实例成本：**计划通过提升 GPU 处理效能，降低硬件和网络带宽成本。



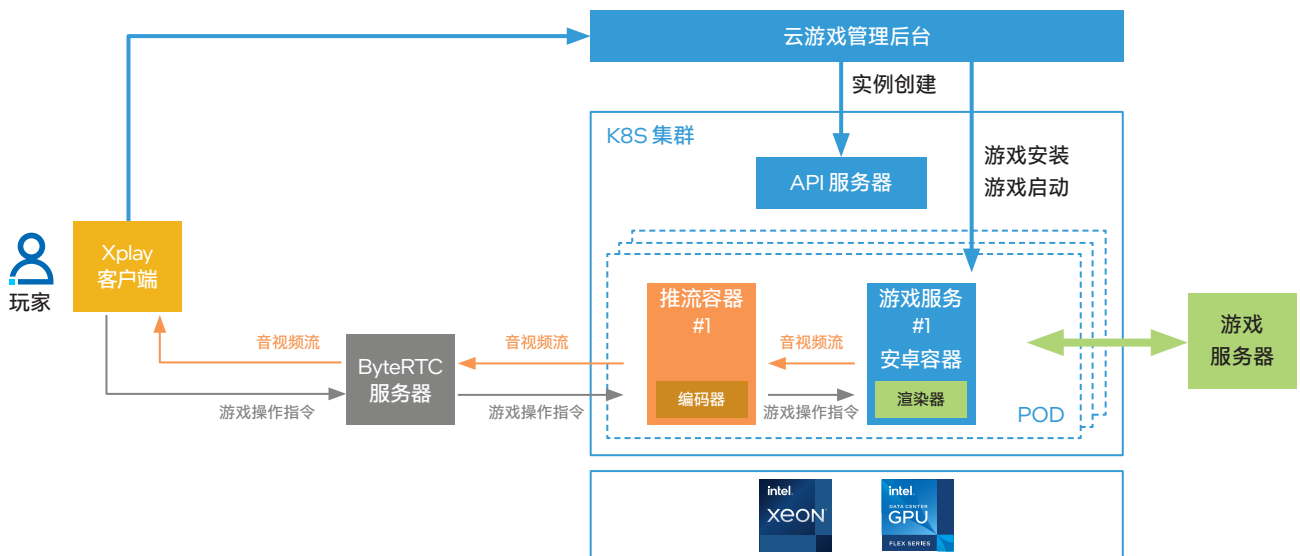
图二 火山引擎云游戏解决方案<sup>3</sup>

## 解决方案：引入英特尔® 数据中心 GPU Flex 系列，打造高品质云游戏方案

瞄准以上目标，火山引擎携手英特尔，以自研 RTC 产品 ByteRTC、智能调度技术等为基础，引入英特尔新一代数据中心 GPU，即英特尔® 数据中心 GPU Flex 系列，作为云端平台渲染和编码工作的核心引擎，释放其自身强大的云服务、

边缘云技术以及对自研软件的优化能力，打造高效能、兼容多体系架构和多形态硬件的火山引擎云游戏解决方案。

新方案整体架构如图三所示，当玩家想要开启游戏之旅时，无须下载游戏本体，借助 Xplay 客户端即可向云游戏管理后台（部署在 PaaS (Platform as a Service, 平台即服务) 层）发起请求。后台收到请求后，会通过 API 服务器在云端的 K8S



图三 火山引擎云游戏解决方案核心架构

容器集群中创建一个对应的云游戏实例, 并完成用户身份鉴权、用户订购信息确认等流程。

云游戏实例会在容器中激活运行游戏所需的一系列工作组件, 包括渲染、编码等, 然后云游戏管理后台, 就可在实例中安装并启动游戏。此时, 玩家的操作指令可通过 ByteRTC 服务器上传至游戏实例, 并最终在相应的游戏服务器中完成场景、人物等数据的交互。而更新后的游戏数据, 则会借助实例渲染为新的音视频流, 并经过编码压缩后通过 ByteRTC 服务器推送至 Xplay 客户端, 形成闭环。

为了让玩家体验到可媲美本地游戏的性能, 火山引擎首先借助其在边缘计算领域的经验积累和服务能力, 在新方案中将云游戏实例广泛地部署到全国乃至全球范围内的边缘节点, 并通过火山引擎自研的智能调度技术来为玩家分配适宜的云游戏实例, 就近接入、就近渲染和编码, 为玩家构建低时延的交互体验。

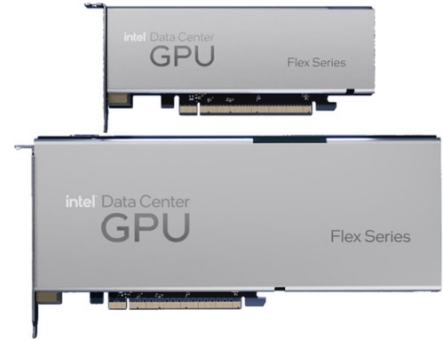
其次, 新方案也引入了经过亿级 DAU (Daily Active User, 日活跃用户量) 产品验证打磨过的自研 RTC 产品 ByteRTC, 这一产品内置了实时音视频推拉流和弱网对抗策略等技术, 可针对恶劣网络环境预测、优化和动态调整码率, 从而大幅优化全网络链路的延迟状况, 并支持最高 4K/60FPS 的视频流输出, 打造更佳画质。

同时在降低单路成本方面, 火山引擎引入英特尔® 数据中心 GPU Flex 系列, 与英特尔® 至强® 可扩展平台一起作为新方案的核心处理引擎, 在提供强大渲染能力和硬件编码能力的同时, 为新方案再添两大助力:

- 提升云游戏实例中渲染和编码工作负载的效能;
- 推动 H.265 编码格式在云游戏场景中的落地, 降低对网络带宽的需求。

作为英特尔新一代数据中心 GPU 产品, 也是英特尔在数据中心领域首款支持 AV1 编解码的独立 GPU 产品, 英特尔® 数据中心 GPU Flex 系列拥有大量 X<sup>e</sup> 内核和光追单元, 以及 X<sup>e</sup>SS 超级采样技术等。其中, X<sup>e</sup>SS 超级采样技术能通过 AI (Artificial Intelligence, 人工智能) 超分技术, 在画面质量相

近的前提下有效缩短渲染时间, 提升云游戏方案运行效率; 而利用 AV1 编解码器, 新方案还可在同样画质分辨率下, 大幅降低所需带宽。



图四 英特尔® 数据中心 GPU Flex 系列

面向不同的云游戏解决方案, 英特尔提供了 Flex 140 和 Flex 170 两种型号。其中, 单张 Flex 170 GPU 卡可支持多达 68 路的 720p30 (720P 分辨率、30FPS, 以下同)《王者荣耀》游戏的渲染和推流负载, 而单张 Flex 140 GPU 卡则可支持多达 40 路的 720p30《王者荣耀》游戏的渲染和推流负载, 或 40 路的 720p30 激流快艇游戏的渲染和推流负载<sup>4</sup>。

[如欲了解更多英特尔® 数据中心 GPU Flex 系列详细信息, 请访问: https://www.intel.cn/content/www/cn/zh/products/docs/discrete-gpus/data-center-gpu/flex-series/overview.html](https://www.intel.cn/content/www/cn/zh/products/docs/discrete-gpus/data-center-gpu/flex-series/overview.html)

同时, 英特尔也为 Flex 系列 GPU 产品提供了丰富的软件栈, 来提升其渲染与编码性能以及可运行的游戏种类, 例如其中的英特尔® ICR (Intel® Cloud Rendering, 英特尔® 云渲染) 提供了高品质的云渲染技术, 其可利用经英特尔优化的 Mesa 3D 图形库, 来提升 GPU 产品的利用率。

而得益于对硬件 SR-IOV (Single Root I/O Virtualization, 单根 I/O 虚拟化) 技术的支持, Flex 系列 GPU 所具有的良好可扩展性, 也让云游戏方案能通过多张 Flex 系列 GPU 卡的组合, 在云游戏服务器集群中实现高密度的渲染和编码能力。例如在使用 6 张 Flex 140 GPU 卡进行扩展时, 配合 SR-IOV 技术可实现多达 228 路的 720p30 激流快艇游戏的渲染和推流负载<sup>5</sup>。

与此同时, 借助 Flex 系列 GPU 强大的硬件解码能力, 新方案也获得了强大硬件编码能力, 突破了云游戏一直存在的硬件编码能力瓶颈, 为云游戏画质从 720P 进一步向 1080P、2K 的演进提供了高性价比的选择和可信赖的方案。



## 收益与展望

现在, 这一全新的云游戏解决方案已在抖音云游戏平台等互联网产品中获得了成熟化运营, 并面向玩家推出了《航海王热血航线》等一系列精品云游戏大作<sup>6</sup>。这些游戏画质精美且含有大量精彩的格斗特效和对战要素, 改变了过去游戏要保证高画质下的流畅性, 对终端性能配置就会有较高的要求, 玩家使用低配终端可能会出现掉帧、卡顿等严重影响体验的窘境。



图五 新方案已成功应用于抖音云游戏平台

与用户体验一致, 经测试显示, 新的云游戏解决方案性能数据同样亮眼。在英特尔® Flex 140 GPU 卡的加持下, 游戏在

720p60 帧场景中可获得 60 路的编码能力和 20 路的渲染能力, 远高于同等能耗的其它 GPU 产品。而在 1080p60 帧场景中, 编码能力和渲染能力则分别达到 28 路和 10 路<sup>7</sup>。

同时在英特尔® Flex 140 GPU 卡提供的硬件编码加持下, 新方案使用 H.265 编码格式相比于传统的 H.264 编码格式, 能使云游戏在同等画质下减少约 40% 的网络带宽<sup>8</sup>, 无疑可有效帮助客户压降成本。

《航海王热血航线》	分辨率	GPU 编码	GPU 渲染	单 GPU 卡并发数
FLEX 140	720P 60FPS	15 * 4	10 * 2	20 (路)
	1080P 60FPS	7 * 4	5 * 2	10 (路)
FLEX 170	720P 60FPS	15 * 2	30	30 (路)
	1080P 60FPS	7 * 2	15	14 (路)

表一 新方案为《航海王热血航线》带来的性能提升<sup>9</sup>

面向未来, 双方还计划通过更深入的技术协作, 完善和优化英特尔® 数据中心 GPU Flex 系列、英特尔® 至强® 可扩展处理器等在云游戏解决方案中的运用, 进一步满足云游戏服务在处理性能、端到端时延、部署成本以及服务稳定性等方面的高品质要求, 进而在推动云游戏体验持续提升的同时, 为云游戏产业乃至数字经济的发展提供更强动力。



<sup>1</sup>数据援引自公开媒体报道: 艾瑞咨询《2022 年中国云游戏行业研究报告》, <https://www.iresearch.com.cn/Detail/report?id=4093&isfree=0>

<sup>2</sup>数据援引自火山引擎内部测试数据, 如欲了解更多详情, 请联系火山引擎 <https://www.volcengine.com/>。

<sup>3</sup>图片来源于火山引擎官网: <https://www.volcengine.com/docs/6512/75569#%E4%BA%A7%E5%93%81%E6%9E%B6%E6%9E%84>

<sup>4, 5</sup>数据援引自: <https://edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-data-center-gpu-flex-series/>

<sup>6</sup>相关游戏应用援引自公开媒体报道, 如欲了解更多详情请访问: [https://www.sohu.com/a/621176624\\_114760](https://www.sohu.com/a/621176624_114760)

<sup>7, 8, 9</sup>数据援引自火山引擎内部测试数据, 如欲了解更多详情, 请联系火山引擎, <https://www.volcengine.com/>。处理器: 英特尔® 至强® 铂金 8336C 处理器, 128 超线程; Linux 内核版本: 5.4.143 英特尔并不控制或审计第三方数据。请您审查该内容, 咨询其他来源, 并确认提及数据是否准确。

### 法律声明

英特尔并不控制或审计第三方数据。请您审查该内容, 咨询其他来源, 并确认提及数据是否准确。

英特尔技术特性和优势取决于系统配置, 并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。没有任何产品或组件是绝对安全的。更多信息请从原始设备制造商或零售商处获得, 或请见 [intel.com](https://www.intel.com)。

描述的成本降低情景均旨在在特定情况和配置中举例说明特定英特尔产品如何影响未来成本并提供成本节约。情况均不同。英特尔不保证任何成本或成本降低。

英特尔、英特尔标识以及其他英特尔商标是英特尔公司或其子公司在美国和 / 或其他国家的商标。

© 英特尔公司版权所有