

开放计算技术大会 2026 • RAS 工作组

构建开放的 RAS API 标准

为下一代高可用数据中心赋能

面向集群规模故障检测、分析与修复的传输无关抽象层

ByteDance

固件架构师

AMD

固件架构师

龚发强

苑方

核心挑战

日志孤岛，缺乏统一契约

01 异构芯片

CPU、GPU、加速器、Switch — 各家暴露 RAS 的方式各不相同

02 跨层流程

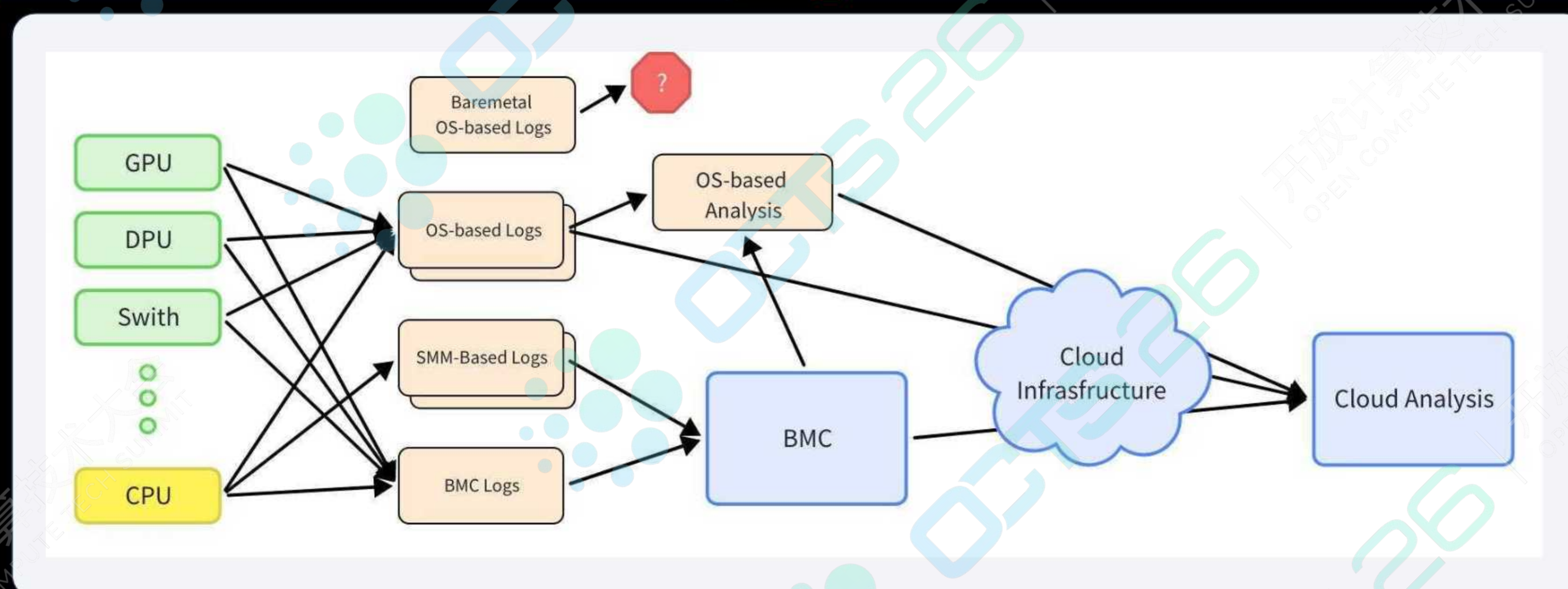
可靠性贯穿芯片、固件、操作系统与集群软件

03 接口碎片化

厂商专有的带内/外接口阻碍了可移植的、集群级自动化

04 机柜级系统

整机柜设计引入共享供电、散热与互连故障的管理需求



现状： OS、SMM、BMC 与裸金属机器日志分散流入互不相通的分析路径 —— 没有统一契约把整个集群的错误数据与纠正动作关联起来

01

设计目标

一个传输无关的 RAS 抽象层

标准化交互行为，而非底层的报文编码 —— 复用数据中心已在运行的管理协议

规范范围

RAS API 标准化了什么

范围之内 —— 已定义

- 逻辑抽象

跨所有部件的统一 RAS 交互模型

- 命令原语

发现、配置、错误读取、动作下发

- 标准载荷

错误用 CPER、动作 CPAD、结果用 Action Event

范围之外 —— 交由他方

- 不引入新的报文编码

不新增帧格式、报头、操作码或完成码

- 协议映射

Redfish / PLDM 映射由各自标准组织负责

- 强制能力

规范定义一致性，而非系统必须实现哪些功能

系统模型

发起方、端点与 RAS API 实例

RAS API 发起方 (Initiator)
请求 · 收集 · 下发动作

传输映射

Redfish · PLDM · PCIe (具体实现相关)

RAS API 端点 (EndPoint)
暴露并提供 RAS 能力

发起方 + 端点 = 一个 **RAS API 实例** —— 操作与传输无关；映射不在规范范围内

带内与带外

两个发起方可通过各自独立的队列集管理同一端点

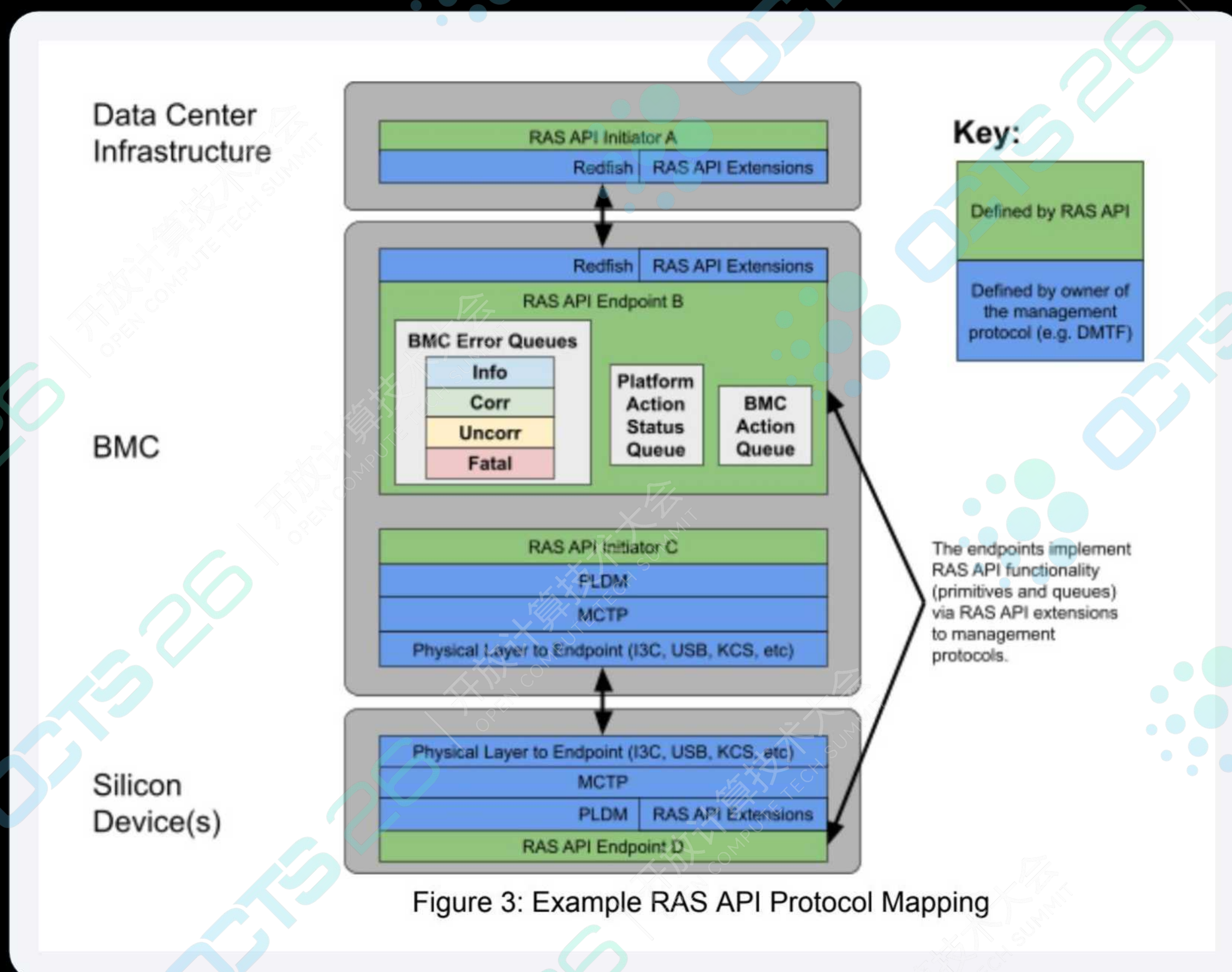
层级化拓扑

端点可向上两级串联，用于收集器与校验代理

管行为，不管字节

同一抽象操作在 Redfish 与 PLDM 上映射方式不同

RAS API 协议映射示例



RAS API 功能承载于现有管理协议之上

- **绿色 = RAS API 定义**

命令、错误队列与动作队列是被标准化的那一层

- **蓝色 = 协议组织拥有**

Redfish (DMTF)、PLDM、MCTP 与物理层仍归各自所有者

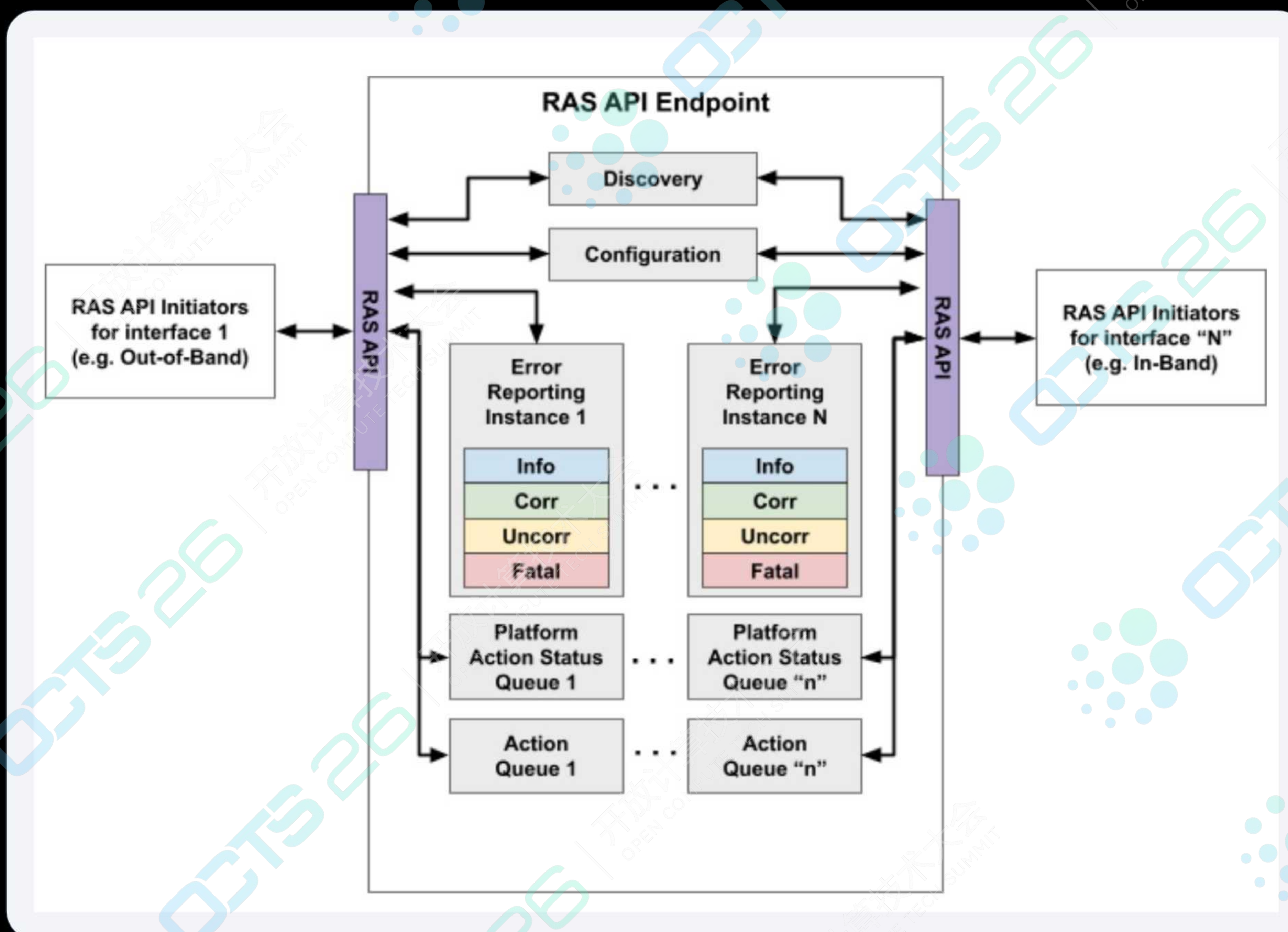
- **RAS API 扩展**

端点以管理协议扩展的形式实现 RAS API 行为

- **同一模型，任意层级**

在数据中心、BMC 与芯片设备各层级使用相同抽象

RAS API 端点的功能架构



一个端点，多个接口实例

● 多接口访问

发起方可带内与带外并行访问同一端点

● 发现与配置

控制命令用于枚举能力并设置行为

● 每实例错误队列

每个上报实例维护 Info / Corr / Uncorr / Fatal 严重级队列

● 动作与状态队列

每实例的平台动作状态队列与动作队列承载 CPAD 与结果



命令原语

RAS API 命令集 (v0.95)

规范将原语分为五类。v0.95 完整定义了三组队列命令集；发现与配置命令保留至后续版本

发现

发现端点、类型与身份；上报 RAS API 支持级别与能力

命令：后续版本

配置

读取各端点能力（配置数据）并配置端点设置。

命令：后续版本

错误上报队列

读取并清除四个严重级错误队列

Get CPER

参数：要读取的错误队列
返回 CPER + 溢出标志 + 状态

Clear CPER

参数：队列 + Record ID
清除错误队列头部

平台动作状态队列

读取并清除动作结果 CPER

Get CPER

参数：要读取的状态队列
返回 CPER + 溢出标志 + 状态

Clear CPER

参数：队列 + Record ID
清除状态队列头部

动作队列

向端点下发 CPAD (FIFO)

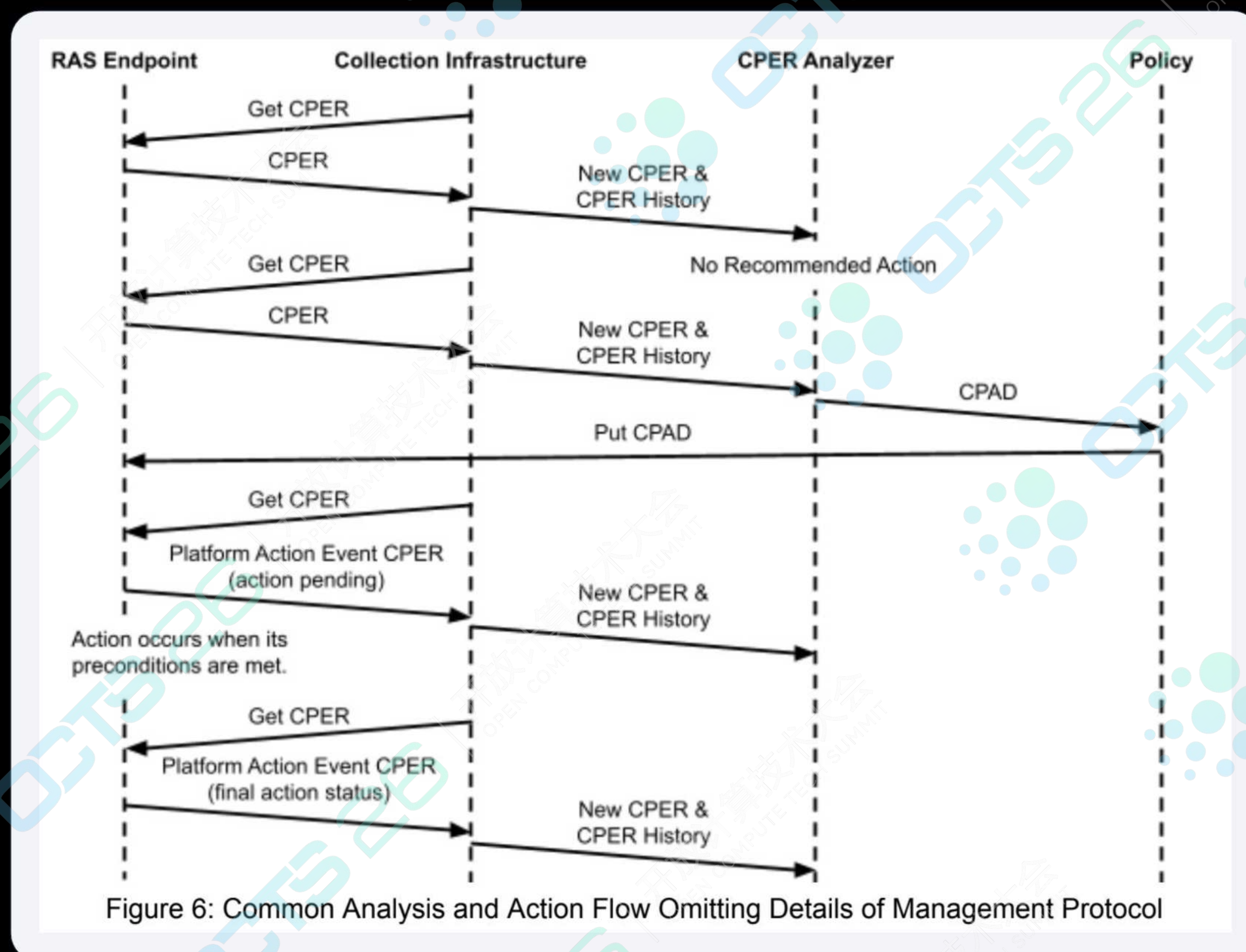
Put CPAD

参数：CPAD
状态码；队列满时报失败

(结果)

异步返回
状态队列上的 Action Event CPER

典型错误分析与 RAS 动作流程



Get CPER → 分析 → Put CPAD → 确认

● 轮询与转发

收集设施发出 Get CPER；每条 CPER 连同其历史流向分析器

● 暂不动作

首条 CPER 给出“无推荐动作”——分析器通常需要多条 CPER

● 生成 CPAD

证据充分后分析器产出 CPAD；策略授权后再 Put CPAD

● 动作待执行

端点返回平台 Action Event CPER；条件满足后执行动作

● 最终状态

第二条 Action Event CPER 将最终动作结果回传上链

02

数据结构

错误数据面与动作控制面

CPER 把错误送出，CPAD 把动作送回 —— 两者都复用 UEFI 记录模型

整体如何衔接

一个闭环，两种数据结构

错误以 **CPER** 形式离开端点；分析器把识别出的故障转化为 **CPAD**；动作执行后，结果再以 CPER 回传 —— 闭环由此闭合

CPER —— 错误与结果向外传

CPAD —— 动作向回传



↻ Action Event (本身也是一条 CPER) 反馈回"关联"环节 —— 分析器据此学习何种动作有效，避免重复下发已失败的动作

错误上报队列

四个严重级 FIFO 队列

错误落于何处。 每个端点依据记录头部的严重级字段，把每条错误记录归入四个严重级队列之一

Fatal 队列

导致无法继续运行

Recoverable 队列

需立即干预

Corrected 队列

硬件已纠正；对 PFA 有用

Informational 队列

延迟事件，如 poison

队列保证

这些队列上的每条记录都是一条 CPER →

- FIFO 顺序

- 每队列独立溢出信号

- 独立中断策略

- 非掉电重置后保持



错误记录格式

CPER —— 队列里装的是什么

记录本身。 每条入队错误都采用通用平台错误记录（CPER, UEFI 附录 N）—— 跨所有芯片共享、可解码的统一格式

CPER 记录

Header
Severit

v

Section 描述符

错误类型 · FRUID · FRUText · 每段严重级

Section 主体

厂商可解码的错误数据（如内存、PCIe）

头部严重级决定入哪个队列
· 单条记录可携带多个 section

厂商中立

复用芯片中已量产的 UEFI CPER 标准

自描述

section 类型让任意分析器都能解码陌生错误

可路由

头部 ID 把错误送至正确的供应商分析器

带严重级标记

头部严重级对记录分类并选定队列



动作控制面

用 CPAD 闭合回路

动作回传。 通用平台动作描述符 (CPAD) 是一条可路由、可策略校验的修复请求

CPAD 记录

Header

PlatformID

D · PartitionID · CreatorID · Timestamp

Section 描述符

ActionID · FRUID · Urgency

Section 主体

动作参数 (标准或私有)

每段一个动作 · 一条记录可含多个 section

可路由

头部 ID 把动作送达正确端点

可策略校验

策略引擎按 ID 允许、拒绝或替换

可扩展

标准 0x0000–0x7FFF; 私有 0x8000+

可维修

FRUID + FRUText 指导部件更换

标准动作集 Do Nothing · Platform Reset · Power Cycle · Reseat · Shuffle · Replace · Inject
Error · **0x8000+ 私有**

03

采纳与影响

从规范到集群落地

它为何重要、规范在 OCP 社区的进展，以及如何参与其中



价值回报

它为何对高可用数据中心意义重大

软件可移植

一套集群代码在不同硬件代际与供应商间持续有效。

更快修复

检测→执行→确认的闭环缩短集群级 MTTR。

降低复杂度

标准记录与动作削减逐厂商集成的工作量。

面向自治

一致的契约是自愈基础设施的基石。



标准化进展

规范在 OCP 社区的所处阶段

隶属于 OCP 硬件管理项目下、硬件故障管理子项目的一个 CLA 工作组

v0.8 2023 年 2 月

由 Intel + Google 首次提交至 OCP —— 工作组的基线版本

v0.9 2025 年 12 月

当前发布版：Discovery、四个严重级队列、CPER 与 CPAD

v0.95.rc1 进行中

工作组 wiki 上的过渡草案 —— 正收敛至 1.0

工作组负责人

- **RAS API 工作组**

Drew Walton (Microsoft) · Jose Marinho (ARM)

- **HW 故障管理子项目**

Yogesh Varma (Intel) · Rama Bhimanadhuni (Microsoft)

配套规范与节奏

- **GPU & Accelerator RAS Requirements v1.7**

2025 年 10 月 —— 标记将并入未来 RAS API 版本的 API

- **Fault Management Infrastructure v1.5**

面向服务器集群运营方的带内/带外框架

- **活跃开发中**

每周 HWFM 例会；项目例会持续至 2026 年



前进路径

采纳路线图与行动号召

阶段一

统一语义

共享词汇、CPER/CPAD 模式、
队列与命令语义

阶段二

标准化映射

经 DMTF / ACPI 组织完成
Redfish 与 PLDM 映射

阶段三

验证与采纳

跨厂商一致性、工具链与集群推广

在开放中共建 —— 向 OCP 贡献配套映射与一致性方案，让整个生态共同受益



OCP 中国

两个本地工作组推动落地

OCP 中国区域项目社区下的本地标准化工作组 —— 围绕 RAS, CloudFW构建全球协作

01 硬件故障管理工作组

中国本地小组，与全球 OCP HWFM 工作对齐

- 建设本地标准化工作组
- 建立全球协作
- 推动关键技术突破与落地 —— RAS API
 - [opencompute.org · OCP PRC / Hardware-Fault-Management](https://opencompute.org/OCP-PRC/Hardware-Fault-Management)
 - <https://github.com/OCP-China-Projects/Hardware-Fault-Management>

02 Cloud-FW —— 中国本地工作组

高度可定制、轻量级的数据中心 BIOS

- 面向 CSP 数据中心 BIOS 需求打造
- 支持深度定制的服务器硬件
- 应对多 SoC 供应链与快速迭代
 - [opencompute.org · OCP PRC / Cloud-Firmware](https://opencompute.org/OCP-PRC/Cloud-Firmware)
 - <https://github.com/OCP-China-Projects/Cloud-FW>

加入 OCP 中国硬件故障管理 —— 携手解决这些问题，把本地创新反哺到开放标准中

谢谢

开放标准，闭合可靠性缺口

OCP RAS API 工作组 · 开放计算技术大会 2026



谢谢

开放标准，闭合可靠性缺口

OCP RAS API 工作组 · 开放计算技术大会 2026



谢谢

开放标准，闭合可靠性缺口

OCP RAS API 工作组 · 开放计算技术大会 2026



谢谢

开放标准，闭合可靠性缺口

OCP RAS API 工作组 · 开放计算技术大会 2026

