



acceldata

HongKe
虹科

选择正确的数据可观察性解决方案： 一个复杂数据环境下的功能清单

爆炸性的数据量和对数据业务需求的增加，导致了现代数据环境的巨大复杂性。然而，对于那些没有成功管理其数据并将其用于实现业务成果的数据团队来说，没有任何借口可以被原谅。在今天的数字环境中，需要不断认识和了解环境中的数据和数据系统，所有这些都是为了优化数据供应链，以便在需要时提供可靠、高质量的数据。

现代数据团队是如何实现这一点的呢？答案是他们使用数据可观察性。

数据可观察性可以被定义为一个组织完全了解其数据健康状况的能力。换句话说，它是对数据复杂性问题的系统解决方案。它监测和关联跨应用、数据和基础设施层的数据工作负载事件，以解决生产分析和人工智能工作负载的问题。

实施多维数据可观察性平台，如HK-Acceldata，是数据工程师确保其数据健康和数据基础设施投资最佳运行的最有效的方法。与只监控组织应用层的传统APM工具不同，多维数据可观察性平台为你的数据、基础设施和管道提供可见性。

你应该在数据观察性平台中寻找哪些功能？

我们的清单（在本电子书的最后一页）提供了对19个重要的数据观察功能的简要介绍——所有这些都是HK-Acceldata平台的核心。

我们将这些功能分为了三组：

① 数据

② 计算性能

③ 数据管道

让我们快速浏览一下每个类别

数据

糟糕的数据会导致糟糕的决策。在一个被混乱、冗余和冰冷的数据所困扰的商业环境中，解决不良数据的问题可能是极具挑战性的。起初的原始数据湖可以迅速变成一个数据沼泽，一个很少有人敢于涉足和清理的沼泽。

实施多维数据可观察性平台可以使预测、预防和解决数据质量问题变得更加容易，特别是当它提供以下几点功能：

- 内置的数据发现
- 数据质量与核对规则
- 数据漂移和模式漂移检测

计算性能

不断添加更多的环境、更多的技术和更多的数据驱动的用例可能会导致一个不可扩展的情况，这既昂贵又容易发生。不断上升的成本与频繁的停机同时发生，造成了额外的复杂性。

多维数据可观察性平台可以帮助数据团队克服性能监控的传统障碍，它提供以下功能：

- 趋势分析
- 自动修复和警报
- 根源分析
- 配置建议
- 性能模拟
- 工作量分析
- 性能优化
- 数据温度和使用监控

数据管道

一个有问题的数据管道可以对一个组织的数据质量和效率造成巨大的破坏。不幸的是，如果对你的多云和混合云环境没有适当的透明度，识别罪魁祸首就不是一件小事。

一个可靠的数据可观察性解决方案提供了端到端的可见性，使得追踪数据流（和数据成本）在你的相互连接的系统中成为可能。HK-Acceldata平台通过提供以下功能使之成为可能：

- 性能分析
- 管道监控
- 成本效益分析
- ETL集成
- 灵活的API

向多维数据可观察性迈出下一步

多维数据可观察性提供了一种可扩展的方法来监控、检测、预防和解决你的数据处理和管道中的问题。

使用下一页的清单可以确保你对数据可观察性做出正确的决定。

数据可观察性清单

多维数据观察能力使企业能够监控、检测、预测、预防和解决其数据、处理和管道中的问题。HK-Acceldata通过提供以下功能使之成为可能：

数据

□ 数据发现

创建一个跨环境和技术的集中的数据资产清单。通过简单的分面搜索、对类似/相关数据资产的导航和RBAC，实现自助式的数据发现。

□ 数据质量检查规则

通过ML驱动的建议提高数据质量和可靠性。自动化的数据质量规则和警报使之更容易识别缺失的数据、违反数据类型、不正确的值和格式、敏感的数据，等等。

□ 数据核对规则

确保数据按预期到达并符合你的数据核对政策。电子邮件通知让你近乎实时地了解规则的失败。

□ 数据漂移检测

通过监控意外的内容变化，提高AI/ML工作负载的准确性。数据漂移规则根据关键指标的容忍阈值自动验证变化。获得过度数据漂移的通知。

□ 架构漂移检测

检测可能破坏管道或影响下游应用的模式和表的结构变化。了解列被添加、修改和删除的时间。

计算性能

□ 趋势分析

在异常情况和潜在问题影响你的操作之前，预测它们。看看哪些工作的运行时间比历史上的标准要长。

□ 自动修复和警报

利用预先建立的和可定制的运行程序，防止速度减慢和停工。自动调整、清理和配置

□ 根因分析

跨复杂环境中的事件进行关联，以快速解决事件。通过考虑资源争夺、环境健康和历史比较，找到根本原因。

□ 配置建议

将历史使用数据转换为可操作的见解，以优化资源配置、数据分配和查询性能。

□ 性能模拟

模拟预期性能并自动调整系统，以最大限度地降低资源利用率，提高性能，消除浪费，并达到服务水平协议。

□ 负载分析

分析工作负载效率，以确定瓶颈、冗余和性能改进机会。

□ 容量优化

利用资源争夺分析来优化调度，简化计划和收费，并改善你的混合云环境。

□ 数据温度和利用率监控

自动监控数据温度，检测可能表明未来问题的“热点”。

数据管道

□ 端到端的可见性

追踪数据的流动和数据的成本，一个交叉互联的系统。

□ 性能分析

识别瓶颈，分析历史记录比较，并优化数据管道性能。深入了解数据和处理问题。

□ 管道监控

监控SLA/SLOs、数据模式、分布和业务事件。跟踪数据交易、握手和转换。

□ 成本-效益分析

在做出技术决策时，评估价格和性能的权衡，以确保可扩展性和投资回报率。

□ ETL 集成

开箱即用的ETL集成降低了复杂性，节省了你的数据工程团队的时间。

□ API 集成

整合现有的基础设施，并通过灵活的API连接到现有的数据、流程、管道和应用程序。