



HK-Acceldata 数据可观测平台概述

HK-Acceldata 数据可观测平台

提高对现代数据平台的可见性，以优化控制和业务影响

数据对业务至关重要。它是运营、决策、预测结果和有效竞争的核心。当数据访问可靠、及时、准确时，业务才能顺利运行。

但有时，有一个障碍。随着数据在复杂的数据环境中从源移动到目的地，可能会发生影响可靠服务、及时性能和准确数据的事件。例如，服务器故障、资源饱和或计划外中断都会影响数据交付和可靠性。高用户流量会减慢查询速度并导致瓶颈。转换错误以及架构或数据漂移可能会导致数据质量不佳。

当数据管道出现问题时，影响可能是广泛且昂贵的。数据中断可能会导致创收活动停止、损害声誉并降低生产力。不可靠数据的不可靠交付可能会对决策产生负面影响，并产生与检测和解决问题相关的大量成本，更不用说对数据资产本身的潜在损害了。

越来越清楚的是，IT 部门和数据团队不能再在黑暗中运作。他们需要数据可观测性技术，以提供对数据处理、数据管道性能以及交付给应用程序和用后的数据质量的全面可见性。

什么是数据可观测性？

可观测性起源于控制理论，它是工程和数学的一个跨学科分支，处理动态系统的控制。最简单的控制理论表明，要控制一个复杂的系统，必须能够观察它以确保实现预期的结果。

数据可观测性提供对现代数据系统复杂性的全面可见性，并帮助从业者观察数据如何与技术互操作以支持业务需求和战略。真正的数据可观测性是多维的，并提供对数据、数据处理和数据管道的全面可见性，这些数据管道协调数据的移动和转换，从摄取到消费，静态数据和动态数据。但能见度只是第一步。数据可观测性技术不仅会在问题发生时通知IT和数据团队，还可以帮助他们在问题导致中断或影响业务之前预测和预防问题。而当问题出现时，数据可观测性通过在数据流经复杂数据管道时监控数据、数据流程和数据来源来发挥作用。在数据被引入系统之前，数据团队定义要在数据、管道和基础设施级别观察到的指标。随着数据流经管道，数据可观测性使用机器学习、高级分析和人工智能来监控和关联所有层的数千个事件。它将数据和数据系统的当前状态与所需状态进行比较。当未达到预期状态时，数据可观测性技术会提供反馈，以便IT和数据团队重新获得系统控制权，以确保及时交付可信赖的数据以满足业务需求。

数据可观测性可帮助团队监控、管理和改进云端或本地的数据管道。它可以快速预测、预防和解决问题，并自动执行数据工程任务，以帮助IT组织提高基础架构可靠性、管道性能和数据质量。

为什么企业需要数据可观测性解决方案？

随着现代数据架构变得更加复杂，数据可观测性应运而生。当今的现代数据格局由许多数据源和目的地组成——为关键任务应用程序提供服务的数据库、促进决策制定的数据仓库、从大数据中发现价值的数据库、用于分析动态数据的事件流处理器以及用于降低可扩展性和成本的云数据管理系统。数据管道也在发展。简单的数据管道仍然存在，其中使用单个提取-转换-加载(ETL)流程从一个或多个OLTP(在线事务处理)中提取数据。

数据库进入暂存区；将数据结构转换为关系格式以优化SQL查询，并将结果加载到OLAP(在线分析处理)数据仓库中，以方便分析和决策。然而，越来越需要更复杂的数据管道来处理新的用例、不同的数据类型和技术增强。

数据工程师现在构建的数据管道可以从不同的数据源(例如，SaaS应用程序、社交媒体、Web日志和IoT传感器)中提取不同的数据类型(即结构化、非结构化、半结构化和原始数据)。存在新的转换过程，例如提取-加载-转换(ELT)，使数据能够直接加载到弹性和可扩展的目标(例如，云数据湖、NoSQL集群和NoSQL数据库)，以仅转换支持真实所需的数据-时间分析、数据探索、机器学习模型、数据可视化和复杂查询。

在一个源系统中生成的数据可能会馈送多个数据管道，而这些管道可能会馈送依赖于其输出的其他管道或应用程序。随着数据从一个流程流向另一个流程，它不断被转换。随时评估数据质量的能力需要该过程来防止模式或数据从源转移到目的地。数据可观测性因其提供跨互连基础设施、管道和数据层的多维可见性而大放异彩。

“我们在使用Hadoop时与HK-Acceldata合作。随着我们重新平台化，HK-Acceldata为我们提供了深刻的技术见解，并让我们能够无缝地进行现代化改造。我们正在整个数据生态系统中扩展HK-Acceldata的使用，以实现GE规模和复杂性的可观测性。”

- Diwakar Goel, 首席数据官兼副总裁, 通用电气数字

谁使用 数据可观测性解决方案？

环境由组织中的许多人共享，包括但不限于：

- ◆ 数据工程师负责收集和准备用于分析的原始数据，可能包括获取数据集、维护数据管道、管理数据湖和仓库、将数据转换为有用信息以及确保符合数据治理和安全策略等任务
- ◆ IT专业人士负责管理基础设施，包括硬件可靠性、网络性能；计算处理能力和内存、存储容量、应用程序可用性等
- ◆ 系统资源工程师（SRE）负责管理云系统和流程
- ◆ 数据架构师负责调整业务和技术要求
- ◆ 数据质量经理负责管理数据确保数据准确的质量和流程
- ◆ 数据科学家负责将原始数据转化为使用机器学习算法、高级分析和可视化的有价值的信息
- ◆ 数据分析师负责交付业务通过查询数据来创建仪表板和分析报告的智能

通过利用端到端数据可观测性解决方案，数据团队可以提高数据可用性、数据管道性能和数据质量。

随着复杂的企业数据系统以前所未有的速度部署到生产环境中，数据可观测性——全面监控和了解众所周知的不透明数据系统如何在基础设施、应用程序和数据层中运行和执行的能力——对于那些寻求转型为数据驱动型企业并能在现代经济中竞争的公司来说绝对至关重要。

HK-Acceldata 数据可观测性平台简介

HK-Acceldata 数据可观测平台通过深入了解现代数据环境中的数据处理能力、数据管道性能和数据质量，帮助数据团队观察、操作和优化现代数据系统。

HK-Acceldata 数据可观测平台为所有连接的数据提供多维数据可观测性，无论其来源、技术、位置或云平台如何。随着时间的推移，它利用人工智能和机器学习来开发上下文、学习和优化可见性和预测能力。该平台不仅可以监控单个系统组件，还可以关联信号

跨多维层了解数据、用户和应用程序之间的交互。这使IT和数据团队能够确保数据以及底层流程和系统针对性能、可扩展性和成本进行了优化。

与竞争对手不同，HK-Acceldata 数据可观测平台提供了多种工具，它们作为一个集成的数据可观测性解决方案运行，如图1所示。该平台为所有产品提供了一个通用用户界面，以提高生产力。IT组织可以从一个工具开始，快速启动并运行以解决最紧迫的业务问题。或者他们可以同时安装和使用所有使用工具来降低复杂性、有效扩展并进一步改善业务成果。

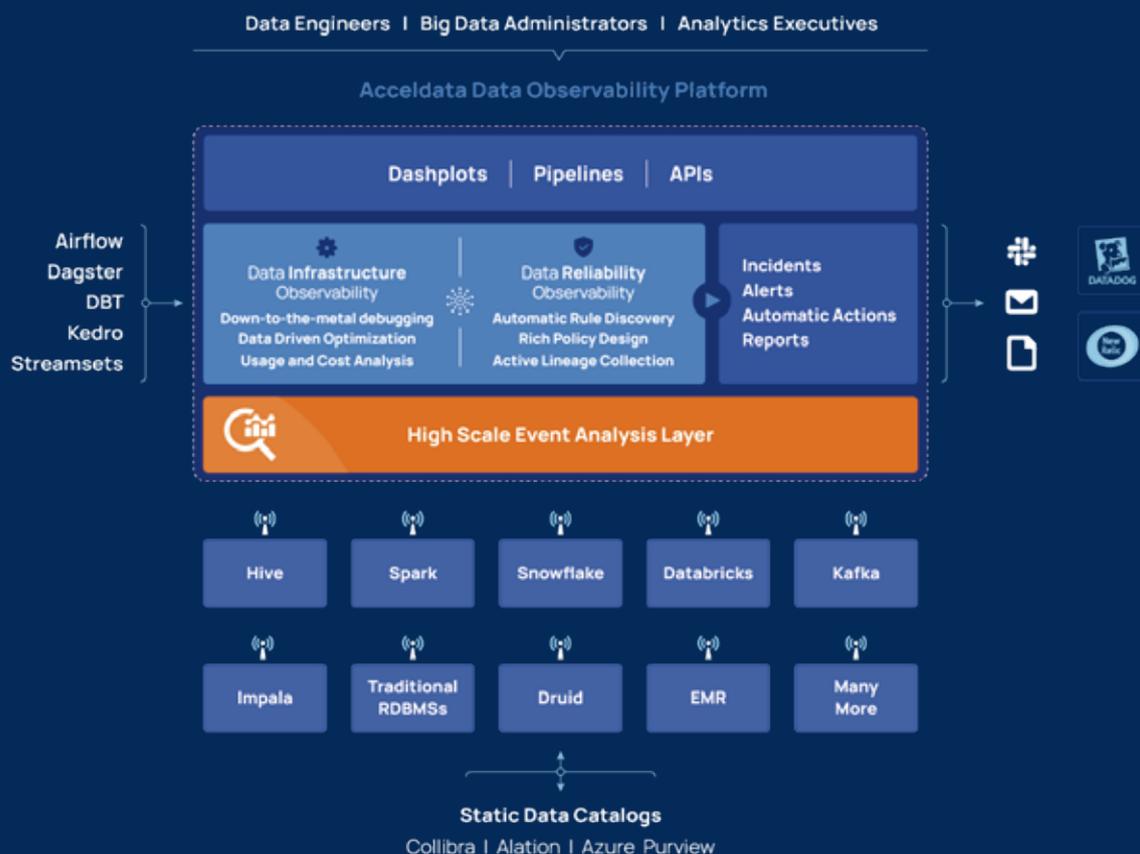


图1: HK-Acceldata 数据可观测平台

HK-Acceldata

数据可观测平台包括

HK-Acceldata Pulse:

根据关键性能指标观察混合数据环境中的数据流，以防止数据停机、有效扩展以支持增长和创新，并优化资源以降低成本。

HK-Acceldata Torch:

在数据通过复杂的数据管道时根据策略观察数据可靠性，以提高数据的可发现性和使用率，增加对数据的信任并提高数据可靠性。

HK-Acceldata 数据可观测平台与现有系统、数据源和工具配合良好，因为它与云提供商和基础架构无关。这有助于 IT 组织利用现有投资，同时了解其企业数据的全部潜力。

HK-Acceldata Pulse: 观察混合数据环境以优化可靠性、性能和成本

HK-Acceldata Pulse 是一种数据可观测性工具，可全面了解混合数据环境的基础架构和计算能力。借助 Pulse，数据团队可以定义指标来衡量分析和 AI 应用程序的服务器和节点可靠性、网络服务带宽、分布式计算处理能力、查询性能、资源利用率和交付服务性能。数据团队使用 HK-Acceldata Pulse 来监控系统运行状况，包括稳定性、可扩展性、负载、响应能力和成本。

当发生异常或超过阈值时，Pulse 会提醒数据团队并推荐快速解决问题的方法、预测问题的趋势分析、查明根本原因的事件关联以及防止再次发生的自动解决方案。工作负载分析、成本规划和模拟建模等功能辅助。数据团队有效扩展、优化资源和控制成本。

有关详细信息，请参阅 HK-Acceldata Pulse 解决方案详细信息。

HK-Acceldata Torch: 通过数据管道观察数据流以优化数据可靠性、质量和使用

HK-Acceldata Torch 是一种数据可观测性工具，可提供对流经复杂数据管道的数据的可用性、可靠性和质量的可见性。Torch 根据为数据质量、协调、数据漂移和架构漂移定义的用户定义策略监控数据，从而帮助增加对数据的信任。如果规则被破坏，Torch 会自动发出警报并提供基于 AI 的建议，以帮助数据团队减少数据损坏和流程故障。

Torch 还监控数据管道本身的可靠性和效率。软件开发工具包 (SDK) 可帮助程序员将 Torch 与现有工作流解决方案集成，以编程方式监控数据管道、收集有关事件的信息并向其他团队成员提供见解。而且由于 Torch 直接与 HK-Acceldata Pulse 集成，它使数据管道了解支持它们的底层数据和数据源。当性能下降时，数据团队可以使用 Pulse 来查找处理能力、资源利用率或容量问题。当数据质量受到影响时，数据团队可以使用 Torch 跟踪数据沿袭，并将错误追溯到数据分析或机器学习过程中的根本原因。

此外，Torch 帮助数据团队发现位于大型数据湖、NoSQL 存储或数据仓库中的资产。搜索、过滤和注释等功能允许数据分析师、数据工程师和数据科学家标记和探索数据，以用于实时分析、复杂查询或机器学习工作负载。

有关详细信息，请参阅 HK Acceldata Torch 解决方案简介。

把它们放在一起

HK-Acceldata 数据可观测平台提供的功能可帮助数据团队提高现代数据平台的可靠性、规模和效率。

IT 部门和数据团队使用该平台来稳定和扩展现代数据环境，以确保可靠、按时交付为现代业务流程提供支持的高质量数据。通过调整数据计划和业务战略，HK-Acceldata 数据可观测性解决方案帮助企业实现卓越运营、创新敏捷性和更高的数据投资回报。

HK-Acceldata 性能优势

- ◆ 提高整个分布式数据平台的可见性
- ◆ 提高数据可靠性并消除数据停机时间
- ◆ 优化数据驱动的业务运营绩效
- ◆ 有效扩展以支持增长和创新
- ◆ 优化资源以增加回报投资
- ◆ 在满足 SLA 的同时降低 IT 成本获取

获取演示

点击这里获取个性化演示

<https://hongcloudtech.com/contact/>

虹科 Pulse

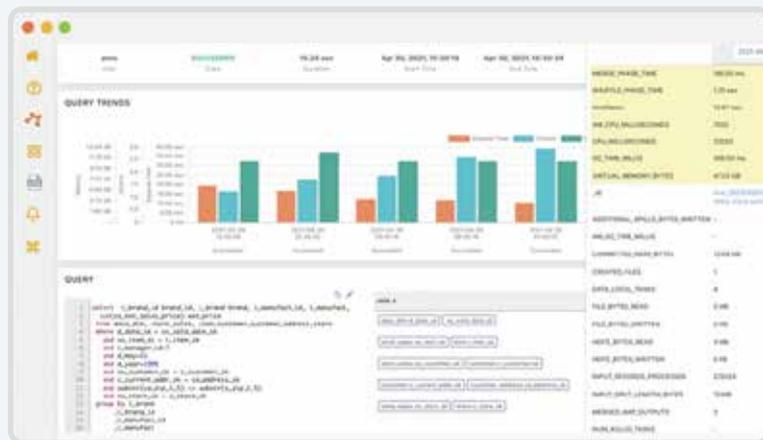
数据可观察性和计算性能监控

HK-Pulse 提供了跨混合数据湖和仓库的企业分析和人工智能系统的横截面可见性。

Pulse 将跨基础架构、应用程序和数据层的事件关联起来，以提供对单个组件、数据管道和系统性能的全面了解，所有这些事情都在一个单一的 web 控制台中可以看到。

典型的客户挑战

- ◆ 计划外数据系统的中断和减速
- ◆ 由于操作问题，导致开发人员工作效率低
- ◆ 基础设施无法扩展，无法满足不断扩大的业务需求
- ◆ 爆炸式的数据系统成本
- ◆ 有限的人才和技术资源来处理快速变化的技术以及不断增加的复杂性



关键功能

预测和预防操作事故:

检测和预测异常，自动化预防性维护，以及关联相关事件以加速根原因分析。

加速数据消耗:

优化查询和算法性能，识别瓶颈和多余的开销，并利用定制的指导策略来改进部署配置、数据分布、代码和查询执行。

优化数据操作、容量和数据工程:

将部署配置和资源与业务需求对齐，监控和预测共享资源的成本，用对数据使用和热点的深度可见性去管理管道数据流。

与关键的数据系统集成:

可以轻松地连接到数据库、Spark、Kafka、Hadoop，以及其他流行的开源发行版、数据仓库、查询引擎和云平台。

PhonePe (沃尔玛公司) 节省500万美元， 实现大规模的可靠性

主要结果

- ◆ 数据基础设施规模扩展到 700 个节点，增加10 倍
- ◆ 0 次计划外停机（12 个月并还在继续）
- ◆ 每年节省 500 万美元的软件许可费

客户业务价值

消除计划外停机：

从事件解决与预防：许多 Pulse 客户已经消除了 12 个月以上的计划外停机。

一键缩放：

自动分析工作负载，以优化配置和设置。加速实验、测试和实施，以便提高新技术的采用率。

降低成本：

Pulse 客户通常会节省每年数百万美元的成本，通过卸载过度配置的软件和平衡计算成本与业务需求。

适用范围

电信、金融服务、能源、物联网、电子商务行业的公司，以及任何依赖大量静态数据和动态数据以满足运营和分析需求的企业。

拥有 30 TB 以上数据和 500 个内核的公司

正在寻求采用和扩展高级开源数据技术

拥有 Amazon EMR, Amazon Redshift, Apache HBase, Apache Hive, Apache Spark, Cloudera, Databricks, Flink, Google Cloud Platform, Google Cloud Storage, Hortonworks Data Platform, Impala, Kafka, Pulsar, Snowflake 的组织

想了解更多 HK-Pulse，请访问：hongcloudtech.com/pulse/

“Acceldata 支持我们的高速增长，并帮助我们管理全球最大的即时支付系统之一。如果没有 Acceldata，PhonePe 有史以来最大的数据基础设施计划是不可能实现的”

- Burzin 工程师 PhonePe
创始人兼首席可靠性

虹科 Torch

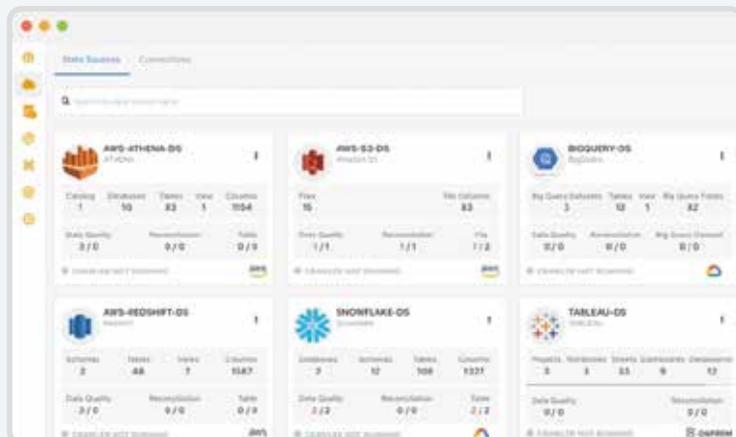
数据驱动型企业的数据可靠性

数据质量并不是一个一次性的活动。这是一个连续的过程，HK-Torch帮助确保过程在整个数据系统中顺利和高效地运行。

HK-Torch在规模上实现了数据质量和可靠性的自动化规模。企业数据团队使用Torch来确保所有的数据在整个数据管道中都是准确、完整和及时的，从摄取到消耗的整个数据管道。

典型的客户挑战

- ◆ 数据质量问题阻碍了云迁移，特别是在互联数据系统之间的握手期间
- ◆ 由数据质量、管理和可靠性问题导致的数据停机时间
- ◆ 为类似的数据资产大规模地管理数据质量
- ◆ 与不透明和不必要的数据集相关联的过度成本
- ◆ 难以理解和增加对大规模分布式数据集的访问



关键功能

超越数据监控:

通过整个数据管道，确保跨算法、模型、特性和数据源的数据可靠性和质量。

自动化数据分类:

甚至可以理解大量的非结构化数据集，包括敏感的、类似的和相关的数据资产，并在几分钟内进行自动分析、协调和分类。

检测模式漂移和数据漂移:

提高动态数据处理，维护ML和AI的准确性。

增加数据消耗:

发现数据，探索数据配置文件，并通过自助服务访问快速查看元数据。

可以轻松地与企业数据源快速集成:

Torch可以连接到任何流行的数据源，无论是在云中还是在本地。

客户业务价值

消除数据停机时间：

通过混合数据湖和数据仓库以深度和精度监控数据，以确保数据质量、一致性、可靠性。

规模数据使用：

快速分析大规模的 分布式数据集——结构化或非结构化——以识别有价值的信息，并改善对未充分利用的数据资产的访问。

自动化数据验证：

保留动态客户中的数据分类、分类和自动化业务规则。

想了解更多 HK-Torch，请访问：hongcloudtech.com/torch/

适用范围

电信、金融服务、能源、物联网、电子商务行业的公司，以及任何依赖大量静态数据和动态数据以满足运营和分析需求的企业。

依赖大量静止数据和运行中的数据来满足操作和分析需求的公司。

拥有 Amazon EMR, Amazon Glue, Amazon Redshift, Apache HBase, Azure SQL, BigQuery, Databricks, Google Cloud Storage, Hive, Impala, Kafka, MySQL, PostgreSQL, Snowflake 的组织。

虹科Flow

数据管道可观察性

数据管道就像现代的数字信息供应链一样。当他们被破坏时，业务就会停止运转。

通过确保在混合数据湖和数据仓库环境中进行可靠的数据事务、握手和转换，HK-Flow 消除了数据管道停机时间。

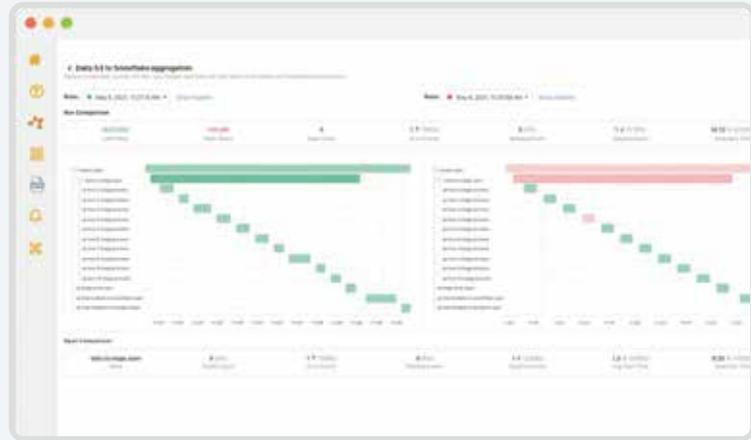
典型的客户挑战

- ◆ 不可靠的数据管道和爆炸式的成本
- ◆ 依赖于耗时的、手动的数据操作和故障排除过程
- ◆ 数据和业务策略错位
- ◆ 数据投资的回报较差

客户业务价值

- ◆ 根据资源、成本和时间优化数据管道性能
- ◆ 自动化操作，以扩展到数百个，甚至数千个数
- ◆ 据管道调整业务和数据策略，使数据投资回报率最大化
- ◆ 加快由数据驱动的转变

想了解更多 HK-Flow，请访问: hongcloudtech.com/flow/



关键功能

消除数据管道停机时间:

在问题影响系统性能或导致计划外中断之前，预测和预防这些问题。

优化数据工程工作流程和生产率:

自动化管道检测，并通过摄入、转换、数据移动和复制生成高保真遥测数据。

简化数据管道操作:

使用单一控制台进行监控、调查和补救。

与流行的工具和平台集成:

轻松连接与常用的 ETL 和 ELT 工具，定制管道建立在 Databricks, Snowpipe, Hive, Spark, 还有编排引擎，如 Airflow 和 Oozie。

适用范围

电信、金融服务、能源、物联网、电子商务行业的公司，以及任何依赖大量静态数据和动态数据以满足运营和分析需求的企业。

拥有 Airflow, Amazon Kinesis, Amazon Redshift, Amazon Web Services-Kafka, Apache HBase, Apache Spark, Confluent, Databricks, Google Cloud Storage, Hive, Impala, Kafka, MySQL, PostgreSQL, Samza, Nifi, Snowflake 的组织

HK-Flow 目前可供早期访问客户和设计合作伙伴使用。

深入了解和控制您的Snowflake环境

通过深入了解成本、数据可靠性和最佳实践，最大限度地提高Snowflake的投资回报

Acceldata 针对 Snowflake 的数据可观察性 结合了跨多个数据操作维度的监控、分析和自动化



控制

确保资源得到有效利用，
并有护栏以使成本与价值
保持一致



相信

每次都按时交付
高质量数据



最佳实践

充分利用 Snowflake
的性能和配置

想了解更多 Acceldata 信息，请访问: acceldata.io

支出智能

提高资源效率并使成本与价值保持一致

利用率洞察和护栏:

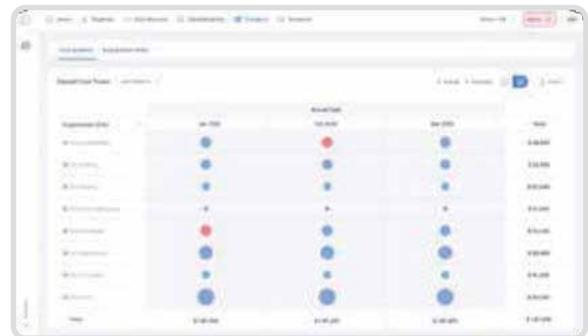
成本智能仪表盘、异常检测（峰值、长时间运行的查询等）和使用护栏

消除浪费:

检测过度配置、未使用的资源和性能低效

支出预测和 FinOps:

合同计划、当前和预计支出分析以及部门级跟踪、
预算和退款



数据可靠性

提高数据信任度

数据管道可观察性:

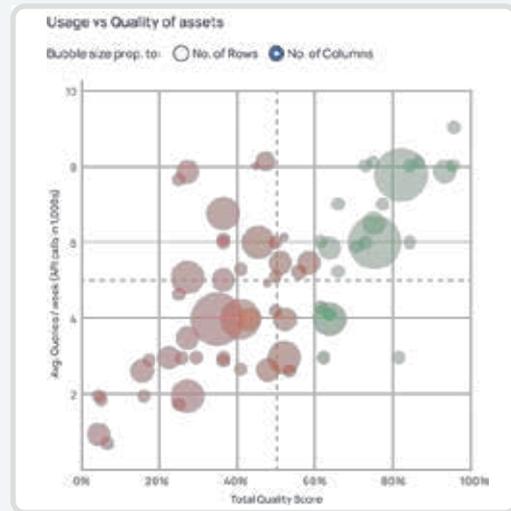
跟踪 Snowflake 内部和外部的端到端管道性能和质量

数据可靠性:

灵活、自定义和自动化的 DQ 监控、数据协调、架构漂移、数据漂移和异常检测

数据发现和分析:

内置可搜索数据目录



最佳实践

监控和分析性能和配置

监控:

当出现违反最佳实践或其他问题时，接收通知和建议或触发外部补救流程

性能分析和优化:

分析瓶颈、高溢出、编译时间、大量扫描、数据量、资源分配、文件大小、数据布局、集群键、微分区等

账户健康监控:

确保在安全和访问、配置等方面实施 Snowflake 最佳实践



像专家一样使用 Databricks 的 全部功能

通过深入了解数据可靠性、性能、成本等，最大限度地提高 Databricks 投资回报。

为什么将 Acceldata 与 Databricks 一起使用？



信任：

每次都按时交付高质量数据



规模：

提高数据工程生产力并加速采用



控制：

确保资源得到有效利用，并设置 guardrails 使成本与价值保持一致

想了解更多 Acceldata 信息，请访问: acceldata.io

数据可靠性 提高数据信任度



数据管道：

追踪 Databricks 内部和外部的端到端管道性能和质量

数据可靠性：

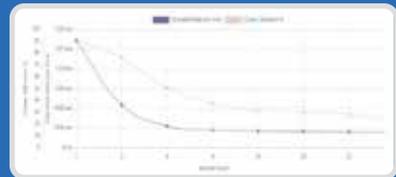
灵活、定制和自动的 DQ 监控、数据调节、模式漂移、数据漂移和异常检测

数据发现和分析：

内置可搜索的数据目录，用于 Databricks 及其他。

性能优化

消除瓶颈& 预防事故



预测和预防事件：

利用自动稳定性跟踪预测未来的事件。创建细粒度的警报、通知和触发器给外部程序，以加速补救。

监控和故障排除：

追踪 Databricks 环境的健康状况，从多集群的角度直至单个作业的执行。简化根本原因分析，并通过事件关联、历史比较、瓶颈分析等加快解决问题的平均时间。

优化：

通过自动配置建议、模拟、数据布局分析等，最大限度地提高效率 and 合适的价格/性能。

支出智能

提高资源效率， 使成本与价值相一致



使用洞察力和护栏：

成本智能仪表板、异常检测（峰值、长期运行的查询等）、使用护栏、访问日志、配置等。

消除浪费：

检测过度配置、未使用的资源和性能低下的情况

支出预测和财务运营：

合同计划、当前和预计支出分析以及部门级跟踪、预算编制和收费。

HongKe



虹科电子科技有限公司

www.hongcloudtech.com
hongcloudtech@hkaco.com

广州市黄埔区神舟路18号润慧科技园C栋6层

T (+86)400-999-3848
M (+86)155 2866 3362

各分部: 广州 | 成都 | 上海 | 苏州 |
西安 | 北京 | 台湾 | 香港 | 美国硅谷



联系我们



行业交流群



获取更多资料



hongcloudtech.com