



ALEIYE

让 | 大 | 数 | 据 | 更 | 简 | 单

IDC 大数据解决方案

目 录

1. IDC 行业现状与挑战	2
2. 传统解决方案的局限性.....	3
3. ALEIYE 大数据平台解决方案	4

1. IDC 行业现状与挑战

在未来几年中, IT 部门面临的挑战将是存储、治理和管理信息,以推动获得洞察力和实现创新,同时控制数据中心的支出。这些挑战是由数据增长的四个维度所致。

挑战一: 数据量

数据中心的数据量最低为 PB 级别, 并且数据量在不断上涨,而这需要更多的存储空间、更长的保留时间以及更多的管理资源。但是, IT 部门是否准备好应对随之产生的更高的运营开支? 这些开支会降低数据中心的价值回报,并会为可扩展性设立障碍。

而传统工具越来越难以进行数据预处理、备份以及 ETL 过程。为支持激增的数据量,使用传统增量方法构建基础架构的手段已不再有效。按需增添组件来支持不断增长的数据量无法提供必要的可扩展性,从而无法应对现在面临的海量数据的挑战。

而大部分数据量来自于非结构化数据。企业需要通过全新的工具和响应能力,以文件或对象的形式搜索和访问数据。服务水平协议 (SLA) 的要求和业务灵敏性取决于全天候数据可用性。要达到最高可用性,数据量需要能够不停机地扩展。

挑战二: 数据处理速度

目前数据中心面临的数据处理速度挑战是数据量激增以及实时访问数据分析。随着业务发展,数据中心的服务器、新型的数据类型越来越多,数据处理速度成为一种特别具有挑战性的管理难题。

数据中心能否从数十亿对象或文件中快速查找数据可以体现出较高的处理速度。要应对数据处理速度挑战并实现企业所需的数据处理速度,必须将数据存储于统一数据平台中。而且,这种数据平台能够按需进行纵向扩展和横向扩展。

挑战三: 数据准确性

由于数据中心存在的数据量、每天的增量很大,常用的数据分析处理软件很难保证数据的准确性。如果将原始数据完整的输入、输出出来,成为数据中心选择核心软件应用的重要指标。

2. 传统解决方案的局限性

传统的业务解决方案一般包括，IT 运维、网络安全、数据审计业务数据分析等环节。这里对每个环节进行分析。

- IT 运维

- 1) 传统的网管软件有较好的采集能力，可采集粒度最小 5 分钟，实时性不强。对各设备告警信息。但误报率高；
- 2) 设备类型多，监管系统同样也多，管理不便；
- 3) 出现设备故障现象，需逐一排查，耗时费力。

- 网络安全

- 1) 针对不同设备会有不同安全设备，如防火墙 IPS、IDS、WAF 等，安全事件独立存放。

- 数据审计

- 1) 人工将各业务系统、网络日志数据导出，手动统一格式，提取关键业务数据信息；
- 2) 耗时费力：需要花费几天，甚至两、三周时间，动用几乎员工全体。

- 业务数据分析

- 1) 呼叫中心、网络交易系统等各系统都有独立的客户管理；
- 2) 当想查看客户全面画像时，需要将所有系统结合展示。

可见，传统的解决方案已经无法满足银行业务的实际需求。为了更好的满足不同岗位的业务分析要求，为了提高 IT 综合运维能力，银行急需一套具备强大数据处理能力和针对性的高效、完整、可行解决方案。

3. Aleiye 大数据平台解决方案

(一) 需要可扩展性开放架构做支撑

大数据量必然要求金融企业 IT 基础设施更易于数据的整合与集中、扩展与伸缩, 以及管理与维护, 同时还必须具备良好的可靠性、可控性、安全性。同时, 在稳定性、可用性及服务性也足以胜任海量数据对基础架构能力的要求, 因此, 具备高扩展性的开放架构正逐步成为数据中心行业应对大数据的优选方案。

- 1) 海量数据存储: Aleiye 大数据平台采用了分布式文件系统、分布式存储、高性能的索引机制, 保障银行数据存储性能以及可靠性的要求。
- 2) 可弹性扩展系统: 面对银行数据规模和复杂度持续增加, Aleiye 实现了存储系统的高可扩展性, 支持随时硬件扩展。
- 3) 分布式数据索引: 将全部索引数据水平切分后存储到多个节点上。这样可解决单个节点无法存储庞大的索引数据和单个节点构建索引的效率瓶颈。

(二) 通过整合进行数据统一

对持续增长的海量数据进行高效管理, 需要数据中心将所有数据并作为一个整体进行管理。数据通过 Aleiye 分布式数据采集, 可更好地共享和使用数据, 数据中心可以管理、分析更多关系复杂的数据。他们将能实时提供具有完全可扩展性的数据。通过数据整合, 将能以更低的设备开支和运营开支实现更高的投资回报率。

- 1) 通过数据采集器、协议采集等方式完成多数据源、多类型的数据实时汇总, 解决不同系统下的数据相互关联的前提。
- 2) Aleiye 预处理功能可以解决数据重复、无标签等问题, 可以按用户要求在数据存储前完成提前处理使数据在 Aleiye 处理效率大大提高。

(三) 具有大数据分析平台性能优势

通过 Aleiye 大数据平台中良好的调度以及消息队列机制充分利用硬件配置的 CPU、内存、IO 资源，并且对任务自动容错，可完成数据自行恢复。保障的数据完整性以及准确性。根据性能测试，平台的完整性、准确性皆可达到 99%。

(四) 通过分布式数据计算提供复杂数据分析

对数据中心运维人员而言，快速、灵活地将不同维度的数据进行统计分析是提供工作效率的方法。利用 Aleiye 分布式数据引擎不仅将数据中心的历史数据高效的提供分析报表、如将如 PV(网站访问量)、UV(有效用户访问量)、频道流量情况分析展示出来，还可以根据相对应的算法模型为数据未来的情况预测出来。

- 1) 分析网站运营情况，按用户访问情况、页面被访问情况等指标对用户、站点、计费等我维度进行分析。
- 2) 精准定位，更了解用户。精准定位访问地址及路径，通过用户的行为特征，帮客户更加了解用户。