

大数据是任务关键组成部分，您的大数据是否保护到位？



全新存储解决方案专为当下的基础架构量身定制

短短几年，大数据就从零零散散的实验性项目演变成数字化企业的顶梁柱，并且其重要性还在不断攀升。IDC 分析指出，到 2020 年，如果企业有能力分析相关数据并提供可操作信息，其生产效益将比不太注重分析的同行多出 4300 亿美元。¹ 诸多大型企业也将大数据分析从过去的偶尔运行调整为每日运行，例如 Amazon、Walmart 和 UPS。

大数据备份挑战

大数据已攀上信息阶梯之顶，这意味着我们必须谨慎对待大数据。备份一度是大数据的亡羊补牢之举，而现在已变成了基本任务，同时还要兼顾速度和性能。IT 领导者们也开始重视备份。在 ESG 最近一份关于 2018 年数据中心现代化优先项 Top 5 报告中，31% 的受访对象表示，改进数据备份和恢复是重中之重。备份和恢复资源会逐渐集中到大数据这一领域。IDC 预计，到 2020 年，大数据市场价值将达到 2100 亿美元，而 2017 年这一数字为 1508 亿美元。²

为了提高备份数据的效率，很多企业都不得不从运行层面做出巨大变革。旧版体系架构能够应付过去一代的工作负载，但现在必须将其转换为现代化架构，保护运行大数据工作负载的环境，例如 Hadoop。例如：

- ▶ **很多企业实施了客户端-服务器备份架构**，该架构在传统工作负载游刃有余，但向来无力应对大量大数据或非结构化数据的持续分析处理流程。这种数据，包括响应的时间点历史数据，与相对传统的 CRM 或 ERP 系统中的大量事务性数据同等重要。
- ▶ **备份相同数据集的多个备份会严重浪费存储资源**。面向大数据环境的有效备份架构应消除毫无必要的相同数据集，节省存储资源，实现数据的快速处理并降低相关成本。
- ▶ **对于企业而言，数据完整性无比重要**，因为企业制定任何业务决策都依赖于关键任务数据分析。如果数据因人为错误而遭致损坏，那么一般的 Hadoop 大数据环境中也会复制不准确的信息。
- ▶ **鉴于每日分析流程成为部分企业的日常需求，因此服务级别协议 (SLA) 对备份任务提出了更高的执行速度要求**。此外，监管规定通常也要求企业能够快速检索数据。
- ▶ **过时的备份架构会阻碍大数据流量**，例如，来自多个数据节点的复制数据块流入 Hadoop NameNode 时会产生瓶颈，导致大数据流运行缓慢。

1 IDC FutureScape: 《全球大数据、业务分析和认知软件预测》(Worldwide Big Data, Business Analytics, and Cognitive Software Predictions), International Data Corporation, 2016 年。

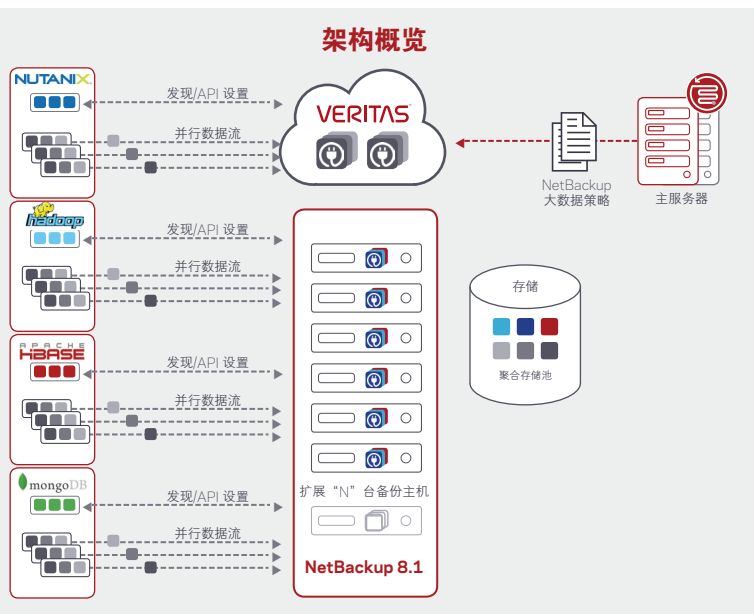
2 《全球半年度大数据和分析支出指南》(Worldwide Semiannual Big Data and Analytics Spending Guide), IDC, 2017 年。

VERITAS[™]

The truth in information.

IDG
STRATEGIC
MARKETING SERVICES

► **备份 Hadoop 环境** 迫使大多数企业不得不创建耗时繁冗的解决办法，除了要部署次要 Hadoop 群集，还要扩大存储空间。这些操作的复杂性导致每一步都可能面临数据丢失风险，并且很可能错过还原时间目标 (RTO)。



Veritas 解决方案: NetBackup 8.1

Veritas NetBackup 8.1 旨在满足大数据备份的要求。它包含 Veritas NetBackup 并行流框架，专为 Hadoop 环境中大型横向扩展备份和多节点的群集工作负载设计。NetBackup 8.1 无须代理即可本机备份 Hadoop 分布式文件系统 (HDFS)，因此根本无需企业采用复杂的解决办法。无代理体系架构是现代环境的重要概念之一。因为群集节点中不存在代理，因此您无须担心要管理各个代理，从而节省了大量时间和金钱，而且您也不必担心 Hadoop 的升级问题。

该产品的运行方式如下：NetBackup 8.1 的可下载插件会扫描 Hadoop 群集以查找元数据，然后通过整合 NetBackup 介质服务器或客户端中资源的方式进行分区并在备份主机中分发工作负载。管理作业资源，从而在 Hadoop 节点发送数据至备份主机时最大限度优化并行流处理的备份和恢复性能。大数据策略完全集成到 NetBackup 用户界面。运用插件之后，您无需等待 Hadoop 的下一个更新版本。

Veritas 解决方案与 Hadoop 大数据架构不同，大数据架构会通过 NameNode 从每个节点复制三份数据，然后通过备份主机将全部数据发送至存储设备，从而造成存储空间的不必要浪费，也导致存储成本的上涨。相比之下，NetBackup 8.1 仅捕获和存储不重复的数据，从而将存储空间节省近三分之二，相

应地节省了 67% 的次要存储成本。不仅如此，备份速度也提升了三倍。

尽管 Hadoop 执行数据复制，但并不提供时间点数据保护，因此无法保证不会出现人为错误和数据损坏等局面。NetBackup 8.1 弥补了这一缺陷，通过执行时间点备份来保留重要数据并轻松满足 SLA 和合规要求。

由于 Veritas 解决方案运用无代理的 NetBackup 插件（而非位于 Hadoop 群集上的代理），因此群集的分析性能不会受到备份和恢复流程的影响。此外，代理相关的管理难题，例如安装、管理和升级也全部迎刃而解。

可扩展性对于大数据环境而言必不可少，而 Veritas 解决方案还在不断改进和增强以满足不断扩展的 Hadoop 数据量的需求。Hadoop 群集中添加节点后，NetBackup 节点也可以相应地增加。

除了 Hadoop，NetBackup 8.1 还支持多个环境，例如 Nutanix 超融合基础架构、Apache HBase 和 MongoDB。NetBackup 8.1 通过集成式云接口支持 40 种云环境。

总结

大数据分析迈出测试实验室，走进企业环境并成为企业的重要支撑之一，在诸多领先企业中奠定了战略性价值地位。在业务成败未卜之际，企业必须争分夺秒，快速实时分析大量数据。

支持大数据备份和恢复的架构必须适应这种任务关键角色。Veritas NetBackup 并行流框架解决了大数据环境下大数据备份的固有难题。Veritas NetBackup 8.1 提升了性能，降低了风险并减少存储资源消耗，开拓了数字化企业的新视界，让企业运用大数据分析奠定自己的战略性前沿地位。

如需了解更多信息，请单击 [此处](#)