

面向高级驾驶辅助系统及无人驾驶的 IBM Storage 解决方案

借助可交付一流可扩展性及高性能的 IBM Storage，构建端到端数据管道

汽车正在快速从独立的大型机械设备变为地球上最具技术先进性的互联平台之一。从娱乐和导航到驾驶辅助与防撞功能，当今的汽车已经与数年之前的汽车大不相同。在分享互联汽车（尤其是无人驾驶汽车）的未来方面所潜藏的巨大机会已经引起了技术企业与汽车制造公司的兴趣。

这些互联汽车计划主要可以分为两大类：

- 高级驾驶辅助系统 (ADAS)，此类系统包括导航、远程诊断和碰撞预警等服务
- 无人驾驶 (AD)，此类系统包括自动泊车、自动操舵等服务

这些计划有一个共性，即数据 - 海量的数据。数据来源包括传感器数据、天气数据、卫星数据、行为数据及其他个人数据、诊断数据等。每台互联汽车都会生成数据，每天生成的数据量从数 MB 到数 GB 不等。对于用于训练 AD 模型的测试车辆而言，每台汽车每天产生的数据量可达到数 TB。这些互联汽车计划会生成约 200 EB/天的存储需求。

借助 IBM Storage 构建端到端数据管道

按照数据管道的阶段，存储可以分为以下类别：边缘存储；瞬态存储；全球摄入层；快速摄入/实时分析所用存储；提取、转换、加载 (ETL) 或数据处理所用存储；Hadoop 或 Spark 数据湖所用存储；机器学习/深度学习所用存储；归档存储。在数据流经数据管道时，每个阶段的存储需求各不相同，每个阶段都可能在为其提供支持的存储的可扩展性、性能和位置方面有着独特的需求。

亮点

- ☒ 支持端到端数据管道的设计和部署：
 - ☒ IBM Spectrum Scale 文件存储可提供可扩展高性能
 - ☒ IBM Cloud Object Storage 可确保全球规模的效率和耐久性
 - ☒ IBM Spectrum Discover 可提供现代化元数据管理
 - ☒ 通过容量和性能层的集成，实现轻松管理
 - ☒ 最大程度地减少数据孤岛和副本，以改善数据治理
 - ☒ 能够以服务、仅软件或集成的形式提供软件定义存储
 - ☒ 提供永久性、按月定价和公用定价模式，以满足不同的业务需求
-

下图所示为 IBM Storage 的产品如何帮助您构建端到端的数据管道，以满足数据管道中每个阶段的存储需求。

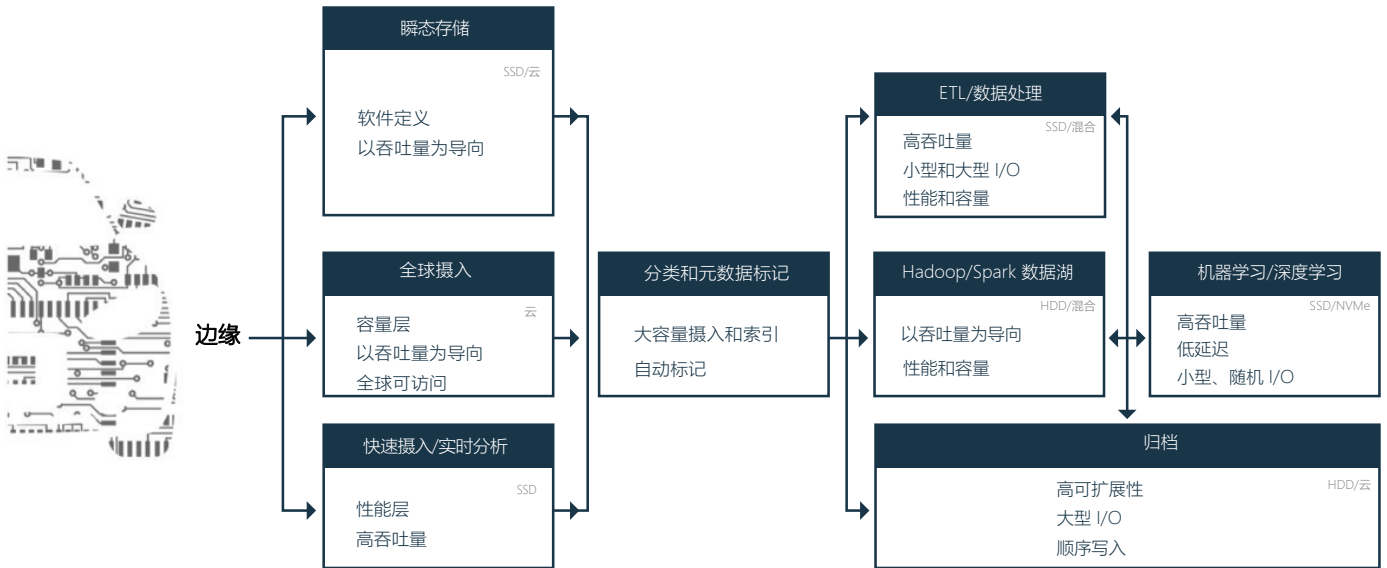


图 1: 面向 ADAS 和 AD 的数据管道。

IBM Spectrum Scale 文件存储

对于 AD 系统的开发和测试而言，需要存储和分析比以往任何时候都要多的数据。唯有能够更快地交付洞察力同时高效管理基础架构快速增长的客户，才有望成为行业领导者。若要交付此类洞察力，底层存储必须要能同时支持新的大数据应用和传统应用，而且具备一流的安全性、可靠性和高性能。若要处理海量非结构化数据的增长，相应的解决方案必须能够无缝扩展，并使数据价值与不同存储层次和功能成本相匹配。IBM Spectrum Scale 作为 IBM Spectrum Storage 系列产品的一部分，能够帮助您应对这些挑战及其他挑战。

IBM Spectrum Scale 是一款高性能的文件存储解决方案，它能够高效地管理数据，能够以独特的方式执行原地数据分析，同时综合支持 POSIX、NFS、SMB、HDFS 和 S3/Object 等数据访问协议。它可为所有这些数据提供一个命名空间，并通过直观的图形用户界面提供单点管理。IBM Spectrum Scale 可提供系统可扩展性、高可用性和可靠性，确保大规模存储基础架构中不会出现单点故障。

如图 2 所示，IBM Spectrum Scale 文件系统适用于面向原地分析工作负载的文件（POSIX、NFS、SMB）、对象（S3、SWIFT）和/或 Hadoop 分布式文件系统（HDFS）。

久经验证的高性能和一流的可扩展性

IBM Spectrum Scale 是一款成熟的文件存储系统，可在现场实现大规模部署，存储数十亿个文件和数 PB 的数据。

它的高性能特性也非常突出。在可交付顶级 SPEC SFS 基准结果的所有文件存储系统中，有 50% 的系统运行在 IBM Spectrum Scale 软件之上。美国最新、速度最快的超级计算机 SUMMIT 也是运行在 IBM Spectrum Scale 之上。

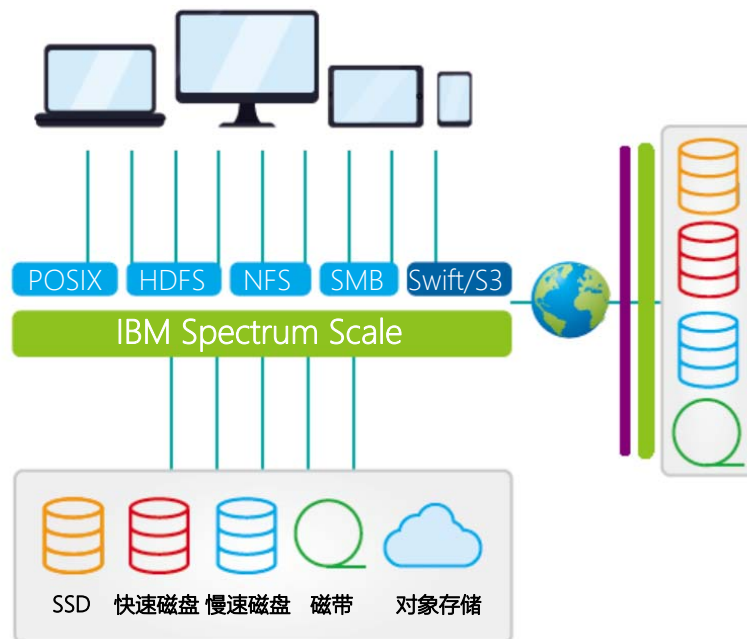


图 2: IBM Spectrum Scale 提供了一个通用名称空间，该空间具有智能存储分层功能，可帮助您构建可扩展的数据湖。

高级数据管理

ADAS 的开发需要从合约和监管方面做好测试数据的保留工作。在开发流程的模拟和验证阶段，需要维持数十甚至是数百 PB 的高性能数据存储。这些数据必须保留数十年的时间，而且服务合约通常会针对恢复和再次模拟的时间（按天数计算）给出硬性规定。

IBM Spectrum Scale 有助于提升绩效、降低成本、增加弹性并简化协作，同时可提供基于算法的策略驱动型数据移动功能，包括拷贝、缓存等。IBM Spectrum Scale 能够对多个存储池（包括云端存储）中的数据进行分类。它能够跟踪使用情况、存储延迟及各种广泛的标准元数据和自定义元数据，您可以借助这些元数据构建数据移动策略。

IBM Spectrum Scale 能够了解数据使用情况及其底层存储，进而高效管理磁带、云等多种存储层中的数据。强大的数据感知智能引擎能够基于性能、位置或成本对设备进行分组（闪存、固态硬盘 (SSD)、磁盘或磁带），进而构建优化的分层存储池。通过迁移策略可透明地将数据从一个存储池移动至另一个，同时确保无需更改文件在目录结构中的位置。借助针对数据使用模式的自动分析，有助于根据需要将数据提升到性能更高的存储层。IBM Spectrum Scale 中内置的信息生命周期管理工具能够针对数据放置提供额外的控制，进而简化数据管理。这些工具包括存储池和一个高性能、基于规则的策略引擎。

消除 HiL/SiL/MiL 测试中的数据相关瓶颈

如果存储速度过慢，将会对应用造成不利影响，导致任务延误并浪费昂贵的基础架构资源。IBM Spectrum Scale 可提供硬件回路 (HiL)、软件回路 (SiL) 和机器回路 (MiL) 测试所需的并行数据访问功能，进而帮助您缩短获取结果所需的时间并实现存储利用率的最大化。此外，借助共享式磁盘和富存储服务器，有助于改善高性能工作负载的可扩展性。IBM Spectrum Scale 基于并行文件系统而构建，可在客户端提供智能功能，能够在所有存储集群节点（包括单个文件）中分布工作负载。在传统的向外扩展网络附加存储中，每个客户端每次只能通过单个节点访问文件。由于采用了这种并行文件系统架构，使得 IBM Spectrum Scale 能够无缝地处理数十乃至数千个客户端、数十亿个文件及数 YB 的数据。

促进全球协作

借助活动文件管理 (AFM) 分布式路由和高级缓存技术，IBM Spectrum Scale 可实现对全球任何位置的数据的低延迟读写访问。AFM 可跨越地理距离对 IBM Spectrum Scale 全局名称空间进行扩展，并可提供快速读写性能以及自动命名空间管理。当数据在某个位置进行编写或修改时，所有其他位置均能够以最小的延迟得到相同的数据。AFM 可利用 IBM Spectrum Scale 固有的可扩展性，提供一款与位置无关的高性能解决方案，该款解决方案可屏蔽网络故障并隐藏广域延迟和中断。这些颠覆性的功能可加速遍布全球各地的团队的项目进度并提高效率。

跨存储层分层 - 规模化地简化数据管理

IBM Spectrum Scale 包含了各种集成式管理工具及直观的图形用户界面，可帮助用户实现规模化数据管理。该文件系统能够跨越多个存储环境及遍布全球的数据中心，以消除数据孤岛和“文件管理器蔓延”。IBM Spectrum Scale 能够智能地在多个异构存储设备之间扩展数据，进而优化可用存储容量利用率、减少管理负担并在需要时交付高性能。该软件提供有多个部署和配置选项，这些选项可通过通用访问功能满足全局名称空间中当前 NFS 文件管理器、块存储和富存储服务器的需求。

与 Hadoop 和 Spark 工作负载相集成

IBM Spectrum Scale 支持 Hortonworks Hadoop 工作负载及 HDFS，同时无需对应用进行任何更改。借助 IBM Spectrum Scale Hadoop 连接器，能够将多个 IBM Spectrum Scale 集群或其他 HDFS 存储库联合到单个 HDFS 实例中。IBM Spectrum Scale 可帮助您减少数据移动需求，进而简化 Hadoop、Apache Spark 及相关软件包的部署和工作流。

IBM Elastic Storage Server

IBM Elastic Storage Server 是一个基于 IBM Spectrum Scale 软件的预集成系统。该系统包含有 IBM Power 服务器、闪存和硬盘驱动器存储，以及 IBM Spectrum Scale 软件。该系统还配有一个高效的纠删码实施，相比传统的硬件 RAID 实施，这种实施可提供一流的数据保护和可靠性。

IBM Cloud Object Storage

数据存储是 ADAS 和 AD 开发中非常关键的一部分。当今的汽车制造商需要能够处理“智能”汽车生成的海量数据的系统。车载电脑及 AD 所需的各种传感器及其他技术会生成数 PB 的非结构化数据。

因此，数据存储需要高度可靠，而且不能影响数据处理的速度。在实际运用中，必须快速处理数据并将其传输至车辆或驾驶员，以免出现严重或是致命的交通事故。

借助 IBM Cloud Object Storage 的离散存储技术，ADAS 和 AD 的开发基础架构可以快速、安全地存储并高效地处理海量数据。通过 IBM Cloud Object Storage 在无人驾驶汽车中管理高效的离散存储可实现大量的优势，包括挽救生命、降低保费、减轻拥堵、减少污染并最终提升效率等等。

什么是离散存储？

IBM Cloud Object Storage 采用了一种创新的方法，能够以成本高效的方式存储大量非结构化数据，同时确保安全性、可用性和可靠性。IBM Cloud Object Storage 利用信息离散算法 (IDA) 将数据分离成不可识别的“切片”，这些“切片”通过网络连接分布到本地或全球存储节点。这种分布式存储设备集合即构成了 IBM Cloud Object Storage 系统。借助 IBM Cloud Object Storage 的离散存储技术，能够从根本上确保数据传输与存储的私密性和安全性。没有完整的数据副本存在于任何一个存储节点中，只要节点的子集可用，便可完全检索网络数据。

何谓信息离散？

IBM Cloud Object Storage 的核心技术是“信息离散”技术，该技术采用纠删码构建数据传输与存储所需的冗余性。

您的数据在线且可访问

在以下所示配置中，最多允许有 6 个系统停机，包括一个完整的站点。在此情况下，数据仍旧会保持在线且可用。



图 3：借助数据离散技术，一旦出现系统故障，系统能够重新装配之前分散的数据切片，进而实现恢复。

纠错码是一种前向纠错 (FEC) 编码，它将一条 k 符号消息转换为更长的 n 符号消息，这样原始消息便可从 n 符号的子集 (k 符号) 中恢复。

IDA 可由任何前向纠错编码制成。在 IDA 算法中，已编码的数据会被额外拆分为多个分段，随后这些分段被存储于不同设备或介质上，以获得高度的故障独立性。举例来说，如果某个计算机的硬盘驱动器发生故障，单独针对该计算机上的文件使用 FEC 并无太大效果，但如果使用 IDA 将数据分散到各个机器中，这些机器便可容忍多个故障，同时依然能够重新装配这些数据。

IBM Cloud Object Storage 内置的单站点或多站点容错

借助内置的离散存储技术，仅需要一个切片子集便可检索数据。如此一来，IBM Cloud Object Storage 便可同时容忍单个站点和多个站点内的设备故障。

IBM Cloud Object Storage 可为 ADAS 和 AD 开发基础架构带来的优势

ADAS 和 AD 的开发需要确保能够防止 PB 规模及以上的数字信息出现数据丢失，而若要做到这一点，依靠 RAID 等传统存储保护技术已然不够。传统的存储架构在设计上无法扩展到 PB 范围，因为它们的可靠性和安全性均有所欠缺。

IBM Cloud Object Storage 非常适于“温数据”，即必须保持活动且可供使用的数据，比如以下位置的数据：

- 备份和归档系统中的数据，此类系统将数据保存在高效且受保护的存储库中，以确保轻松、安全且快速的访问
- 具有在线可扩展容量、供分析使用的大型大数据存储库
- 可移植的云原生应用，此应用经过开发可支持各种云规模及任意位置的云访问

ADAS 和 AD 与 IBM Cloud Object Storage

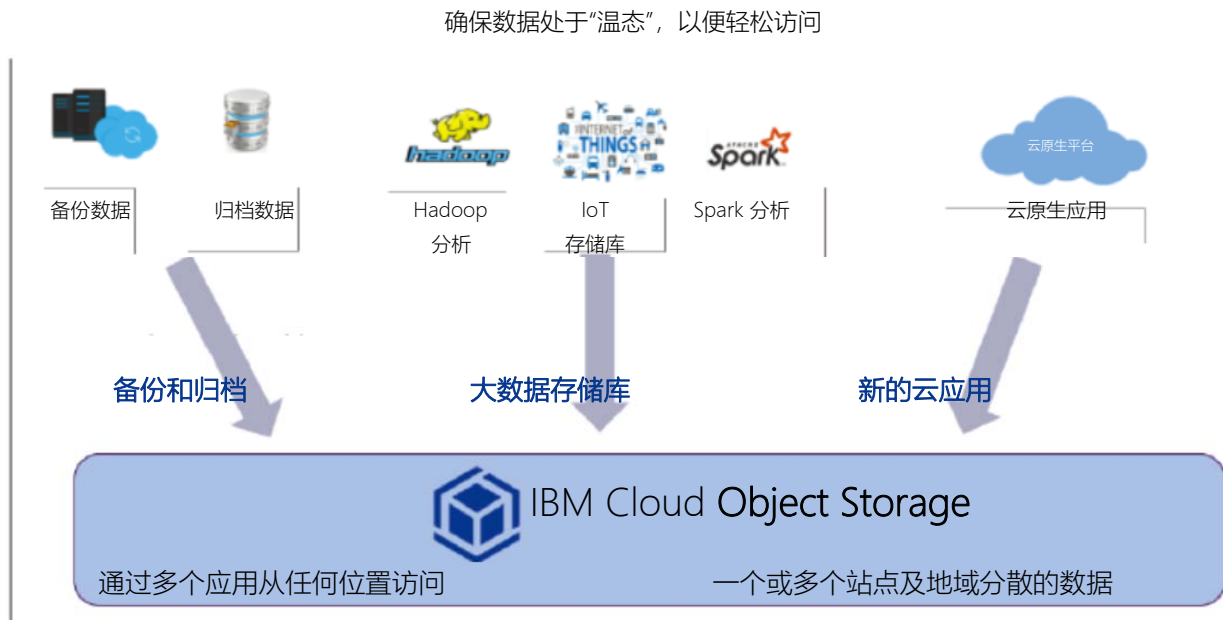


图 4: IBM Cloud Object Storage 支持便捷的温数据访问。

IBM Cloud Object Storage 可提供的关键优势主要体现在以下五个方面：

- **可用性**：无论是计划停机还是非计划停机，数据始终保持可用。
- **可扩展性**：系统能够从 TB 级扩展到 PB 级，甚至是 EB 级。
- **安全**：即使多个驱动器、服务器、容器或站点被攻击，也可确保数据的机密性。
- **成本高效性**：无需昂贵的复制，能够显著降低 PB 及以上级别存储系统的总体拥有成本。
- **效率**：每个管理员都能够管理数 PB 的存储。

IBM Spectrum Archive 可为 ADAS 和 AD 开发基础架构提供成本高效的磁带存储

IBM Spectrum Archive 是 IBM Spectrum Storage 软件组合中的产品之一，旨在通过智能软件层改善存储的经济性，进而解决数据存储方面的低效问题。

IBM Spectrum Archive 是一款成本高效的解决方案，可用于保留智能汽车生成的海量数据。对于无人驾驶汽车开发人员而言，它是一款理想的基础架构解决方案，因为它提供了一种轻松的方法，便于开发人员将测试车辆的数据移动到分层式存储基础架构中成本高效的 IBM 磁带驱动器和库中。在该解决方案中，第 2 层和第 3 层长期存储使用的是磁带库（而非磁盘），如此一来，ADAS 和 AD 开发基础架构便可提升效率、降低与存储不断增长的数据相关的成本。

IBM Spectrum Archive 能够与 IBM Spectrum Scale 的可扩展性、可管理性和高性能无缝集成，进而提供一个理想的 ADAS 和 AD 开发基础架构存储解决方案。

IBM Spectrum Archive 采用 IBM Linear Tape File System (LTFS) 格式，能够以直接、直观且图形化的方式访问存储在使用 Linear Tape-Open (LTO) Ultrium 第 7、6、5 代磁带盒的 IBM 磁带驱动器和磁带库中的数据，以及存储在使用 IBM TS1150 和 IBM TS1140 磁带驱动器的磁带库中的 IBM 3592 磁带盒中的数据。如此一来，无需额外的磁带管理和软件即可实现对数据的访问。

借助 IBM Spectrum Archive，能够以透明的方式访问存储在 IBM 磁带盒中（而非磁盘上）的数据。IBM Spectrum Scale 能够为磁带上的数据保留一个指针，一旦收到请求，便可检索数据，无需用户或操作人员的参与。借助 IBM Spectrum Scale 的功能，IBM Spectrum Archive 能够在整个存储基础架构中实现高性能的可靠数据访问。反过来，IBM Spectrum Archive 能够以更少的媒介、占地面积和电源成本为 IBM Spectrum Scale 的安装增加广泛的容量。通过基于策略的迁移，可以在不影响数据可用性的情况下扩展归档容量。

借助 IBM Spectrum Archive，组织可以通过以下方式改善云及分析基础架构的成本高效性：

- 运营存储层采用磁带，而不是将冷数据存储在本成本高昂的磁盘存储上。
- 长期存储数字资产，如此便可在未来多年确保资产的可参考性和货币化。
- 在运营存储上构建数据副本，进而改善分层式存储基础架构的效率和成本高效性。

借助 IBM Spectrum Discover 从 ADAS 和 AD 系统生成的 EB 规模的数据中发掘一流的数据洞察力

ADAS 和 AD 系统由大量物理传感器构成，包括雷达、LIDAR、超声传感器、光混合器设备 (PMD)、摄像头和夜视设备等。这些设备的作用是监控每个方向上的近场区和远场区，在车辆移动过程中，这些设备会生成海量的非结构化数据。正确地处理这些数据有助于基于交通、天气、不可预测的危险驾驶条件等因素确保车辆、驾驶员、乘客及行人的安全。不过，若要从如此海量的数据中发掘有用的洞察力，就需要能够识别这些非结构化数据并对其进行分类和管理，以供分析使用，同时确保规模化的性能和效率。

作为 IBM Spectrum Storage 软件组合中的产品之一，IBM Spectrum Discover 是一款现代化的元数据管理软件，旨在从 EB 规模的非结构化数据中发掘数据洞察力，因此非常适于 ADAS 和 AD 系统生成的海量数据。此外，IBM Spectrum Discover 非常灵活，可支持多个用例，包括大规模分析、数据治理和存储优化等。

IBM Spectrum Discover 如何规模化地提供一流的数据洞察力和灵活性？

- **自动元数据摄入：**IBM Spectrum Discover 能够扫描现有的存储数据并在添加或修改文件和对象时响应流事件，以此方式自动摄入来自 IBM Spectrum Scale 和 IBM Cloud Object Storage 的元数据并将其输入索引。
- **超出系统级别的元数据：**IBM Spectrum Discover 支持用户创建和管理自定义元数据标记，这些元数据标记可基于用户定义的策略运用到数据上。通过这些自定义标记，数据管理人员可以使用域特定的标记对数据进行分类；举例来说，这些域特定标记能够将数据关联到负责收集数据的特定传感器，或关联到特定的事件或条件。通过在摄入时或在之后尽快地使用此类重要的语义标签扩充数据，用户便可轻松地管理和准备特定的、具有高度针对性的数据，以便进行后续操作，比如分析、运营模型训练等，进而缩短实现结果所需时间并提升准确性。
- **可扩展的平台：**IBM Spectrum Discover 提供了一个可扩展的平台，可通过其操作代理应用编程接口 (API) 实现与 IBM 解决方案和开源软件的集成。通过结合使用 API、自定义标记和策略引擎，可支持内容检查等额外用例，或者实现数据移动及更复杂工作流的统筹安排。

为什么选择 IBM Spectrum Discover?

IBM Spectrum Discover 可为面向 ADAS 和 AD 基础架构的 IBM Spectrum Scale 和 IBM Cloud Object Storage 实施交付巨大的增值。它有助于组织以下列多种方式更好地管理他们的数据和存储环境:

- 更高效地分析数据, 识别上游分析处理所需的相关数据, 进而发现潜在价值, 以便数据科学家和研究人员更快速地生成结果并提升准确性。
- 改善治理水平, 以减缓风险, 同时识别按照内部策略要求需要予以保留的数据, 或者识别会给组织带来风险因而需要从存储环境中移除的数据, 以此方式改善数据质量。它能够针对大规模环境的快速搜索功能, 有助于提高对法律发现或监管审计请求的响应速度。
- 识别可移动到低成本存储的数据、组织不再需要保留的数据或冗余数据, 以此方式优化存储基础架构并改善利用率。如此一来, 便可为新创建的数据回收容量, 进而降低成本。IT 部门可以根据数据本身的价值或重要性更高效地分配存储资源。

IBM Spectrum Discover 在 ADAS 和 AD 工作流中的示例用例

- 创建自定义标记。例如“车辆型号”、“交通条件”和“地理位置”等。
- 添加一个自定义操作代理, 用于检查专用文件格式、提取与 ADAS 和 AD 分析相关的信息。
- 创建策略。例如“车队信息”。将该策略关联到之前创建的自定义标记和操作代理, 并将执行时间设置为“现在执行”。
- 捕获输入到 IBM Cloud Object Storage 中的车队数据, 一旦有数据进入, 该软件会向 IBM Spectrum Discover 发送通知。
- 在 IBM Spectrum Discover 中, 针对这些通知运行“Fleet-Information (车队信息)”策略。
- 在 IBM Spectrum Discover 中, 使用操作代理从入站文件中提取信息并将其用作入站数据的自定义标记。
- 在 IBM Spectrum Discover 中, 使用自定义元数据标记识别相关数据。
- 将此类数据拉取到高性能层, 以执行大规模分析, 然后在 IBM Spectrum Discover 中, 使用从分析结果中衍生出的新的自定义标记对数据进行标记, 以进一步扩充数据。

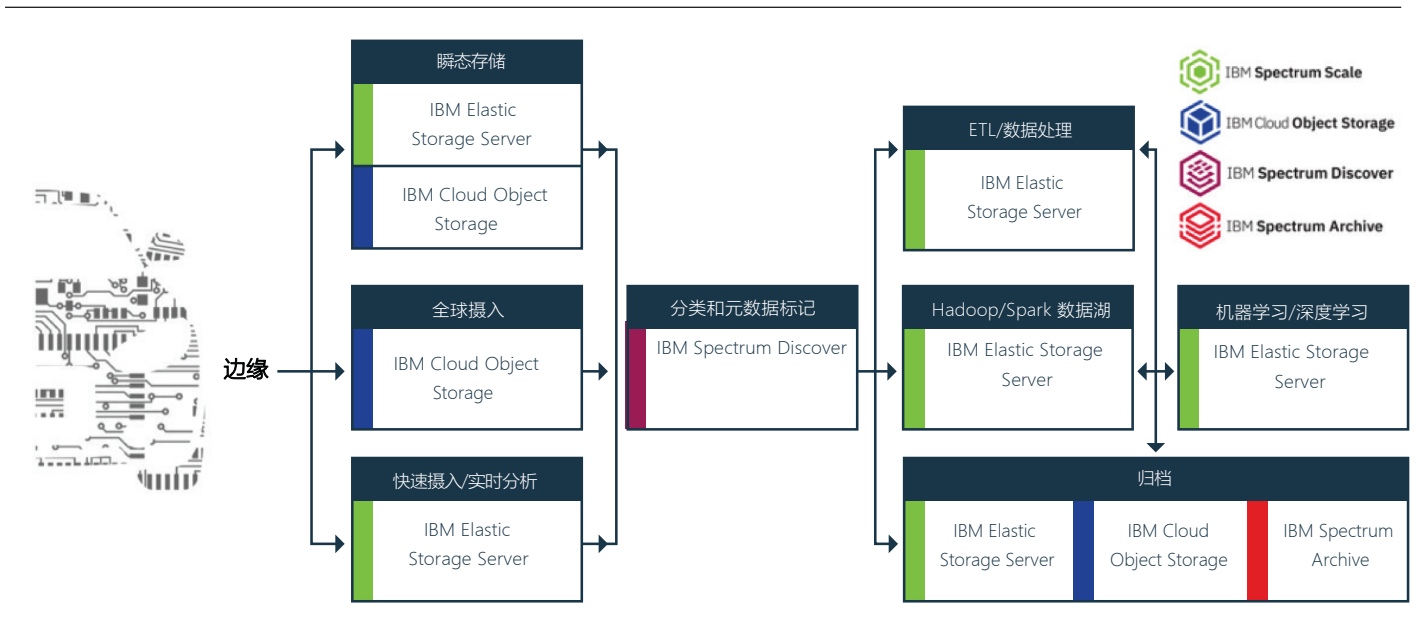


图 5: IBM Spectrum Storage 智能存储软件产品组合及 IBM Elastic Storage Server 可满足数据管道中各个阶段的需求。

为什么选择 IBM?

IBM 长期以来一直致力于存储产品和数据管理方面的创新，积累了大量久经验证的经验，可帮助企业设计并部署面向未来的综合性存储基础架构。IBM 的存储解决方案不仅仅为数据提供一个可信赖的存储位置，其在用户界面和管理方面的创新还有助于简化配置并通过持续的元数据分析和增强扩展数据的可访问性。IBM Spectrum Storage 产品组合还有助于确保数据的无缝移动，进而实现适用于业务且成本高效的存储，这种存储能够兼顾各个存储层的速度、成本和可靠性。正因为这些优点，IBM 的存储创新非常适于当前要求苛刻的海量数据登台、归档和分析等工作负载。

有关更多信息

如欲了解有关 IBM Spectrum Storage 的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或访问以下网站：<https://www.ibm.com/cn-zh/storage>

此外，IBM 全球融资部可提供各种支付选项，进而帮助您获取开发业务所需的技术。我们可提供 IT 产品和服务的全生命周期管理（从收购到处置）。有关更多信息，敬请访问：<https://www.ibm.com/financing>

拨打免费咨询热线

手机请拨打：400-810-1818 转 5122

座机请拨打：800-810-1818 转 5122



© Copyright IBM Corporation 2018.

IBM、IBM 徽标及 www.ibm.com 是 International Business Machines Corporation 在世界各地司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。Web 站点 <https://www.ibm.com/legal/us/en/copytrade.shtml>，包含了 IBM 商标的最新列表；Web 站点 https://www.ibm.com/legal/us/en/copytrade.shtml#section_4 包含了可能在本文中提及的所选第三方商标列表。

本文档中包含了与以下 IBM 产品（IBM Corporation 的注册商标）相关的信息：

IBM Spectrum™、IBM Spectrum Archive™、IBM Spectrum Scale™、IBM Spectrum Storage™、IBM Elastic Storage™

有关 IBM 未来发展方向及意图的声明如有变更或撤销，恕不另行通知，且仅用于说明目标之用。
