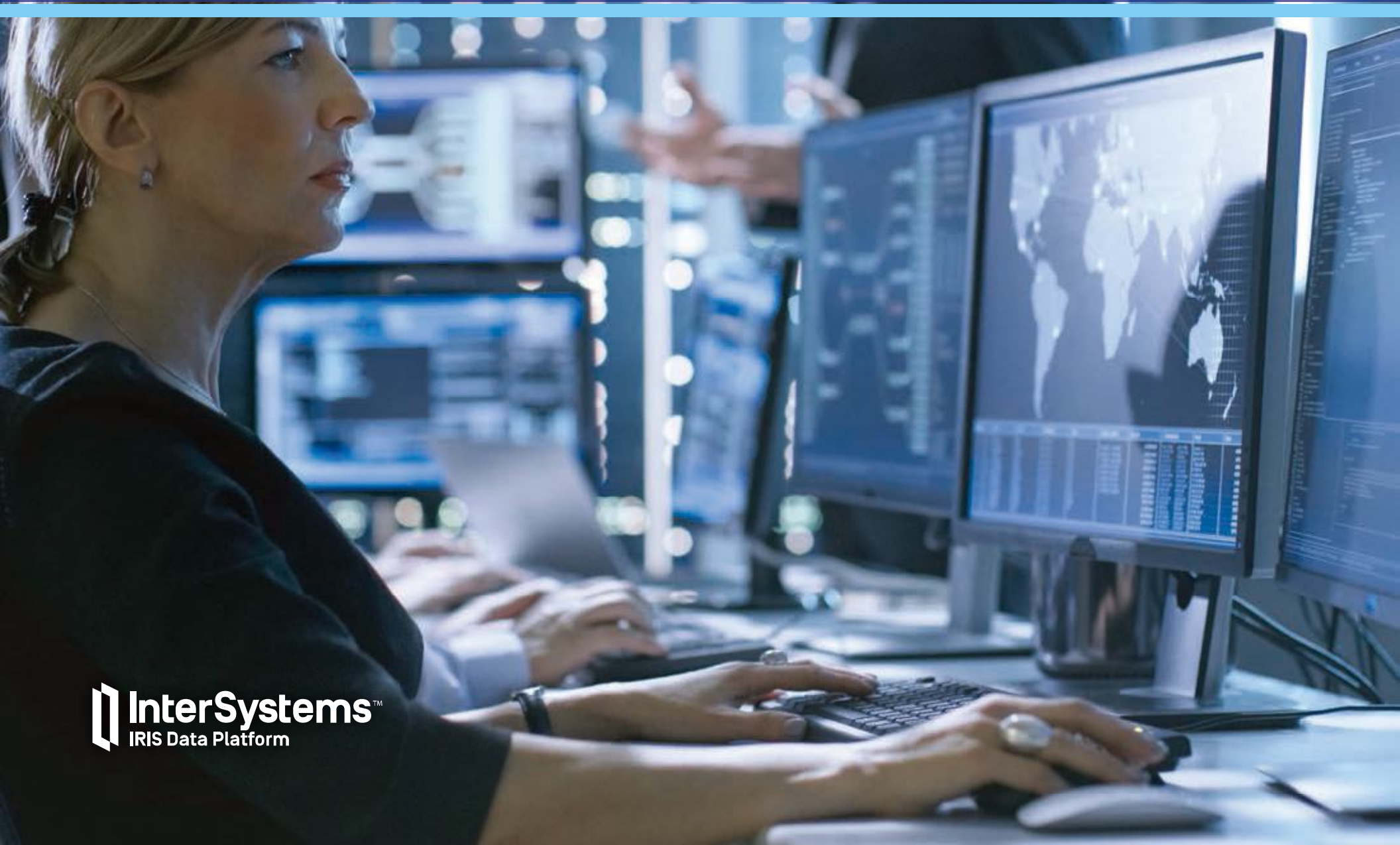


InterSystems IRIS 数据平台： 助力数据密集型实时应用程序的统一平台



InterSystems IRIS 数据平台： 助力数据密集型实时应用程序的统一平台

概要

各行业的企业都希望通过缩短和消除事件、洞察及行动之间的时间延迟，以获得策略及运营方面的优势。企业们也都在力图将数据驱动的智能嵌入到他们的实时业务流程中。

一旦成功实现这些目标，企业将能受益匪浅，包括：

- 提供新颖和创新的业务服务
- 提高收入
- 改善客户体验
- 简化操作
- 识别并降低风险
- 遵循不断改变的最新行业法规
- 节约成本

这一份白皮书描述了与缩短和消除这些延迟所带来的机会与挑战，并介绍了一种能面向各行业，简化数据密集型实时解决方案的开发、部署和维护的新技术。

引言

企业拥有的数据比以往任何时候都要丰富。但是，对很多企业而言，如何从这些数据中获得洞察力并及时采取行动以获得竞争优势仍是一种挑战。企业期望通过构建数据密集型实时应用程序来更好地把握商业机会，这类应用程序所需技术必须能够：

- 无延时的分析实时事件和事务型数据----以及历史性数据和参考数据构成的大型数据集
- 支持多种数据模型和表示，包括关系型数据、文档数据、键值数据、对象数据和非结构化文本数据。
- 创建能无缝整合到不同应用程序和数据源的实时复合流程
- 可以随工作量、数据规模和用户数量增多而实现扩展

- 将SQL查询、机器学习、预测分析和自然语言处理（NLP）嵌入到数据驱动型应用程序中
- 可使用支持本地、云端和混合部署的灵活选项，同时支持持续交付和DevOps
- 以具有良好性价比的方式提供上述功能，企业无需额外招募精通多种专业的专家

实现企业实时运营

行业技术分析公司 IDC 近期采访了全球各领域超过 500 多家企业。超过 75% 受访者认为，不能分析当前实时数据已经削弱了他们利用新商业机会的能力。同时，超过一半受访者称，这还限制了企业的运营效率¹。

这一项研究发现，在使用 ETL（抽取、转换、加载）技术将数据从业务系统移动到数据仓库的公司中，有 64% 的公司事务数据分析延迟时间达到 5 天或更多。

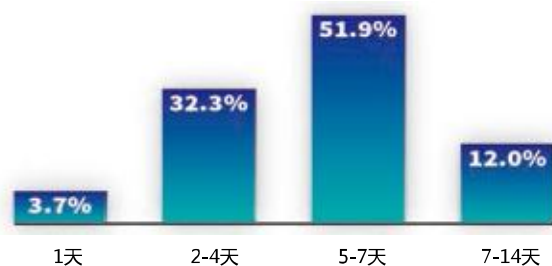


图1: 使用 ETL 将事务数据移动到分析数据库所需平均时间

来源: 第三代信息管理平台需求调查, IDC, 2016年10月; n=502

对于依赖于变更数据获取 (CDC) 过程的实时应用程序, 受访企业称, 超过 95% 的 CDC 过程耗时超过 1 分钟, 而 65% 的 CDC 过程耗时超过 10 分钟。在每一毫秒都有意义的关键实时用例中, 这种速度过于缓慢。

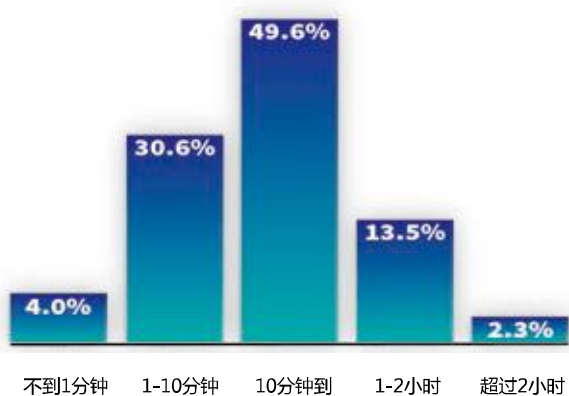


图2: 完成 CDC 过程所需平均时间

来源: 第三代信息管理平台需求调查, IDC, 2016年10月; n=502

几乎每一种行业都在使用对多种数据源实时数据进行实时分析的应用程序:

- 金融服务, 用于遵循强制性标准和联邦法规、欺诈检测和风险管理计划
- 离散制造/原始设备制造, 用于预见性维护
- 航运与物流, 用于集装箱和货物实时追踪
- 零售业, 用于客户和访客的定位及个性化服务
- 公共安全, 用于一线应急人员态势感知
- 医疗保健, 用于现场及时提供个体化和前瞻性治疗

这类应用程序需要一种能消除延迟和复杂性的数据平台, 实现在同一个引擎中同时支持事务型工作和分析型工作, 而无需移动、映射或转换数据。

InterSystems IRIS 数据平台™能满足上述需求。它能合并多个独立且异构的数据源, 支持嵌入式实时分析, 能方便地随着数据量和用户数量的增多而扩展, 能与其他系统无缝集成; 能提供灵活、敏捷, 与 DevOps 兼容的部署功能。

¹ “选择一种数据库管理系统来迎接第三代平台的挑战” (IDC, 2017)



InterSystems IRIS 数据平台

InterSystems IRIS 数据平台是一个完整的统一平台，能简化实时、数据丰富的解决方案的开发、部署和维护。它提供并行事务处理和分析处理的能力，支持多种完全同步的数据模型（关系型、层次、对象和文档数据模型）；是一种能整合不同数据孤岛和应用程序的具有完整互操作性的平台，能够支持批处理和实时处理复杂结构化及非结构化分析功能的场景。

平台同时提供一种开放式分析环境，用于将同类最佳的分析方法整合到 InterSystems IRIS 解决方案中，它也提供灵活的部署功能、用于支持任意组合的云端和本地部署。

InterSystems IRIS 是一款独立产品，它采用单一架构构建，能支持广泛的应用程序和应用场景。

InterSystems IRIS 数据平台提供下列关键特性：

- 用于支持实时应用程序的混合事务/分析处理功能
- 多种数据模型
- 嵌入式分析和开放式分析功能
- 集成 Apache Spark
- 商业智能（BI）
- 将高级分析功能整合到实时流程的能力
- 自然语言处理（NLP）
- 互操作性
- 统一的部署环境
- 灵活的部署选项

用于支持实时应用程序的混合事务/分析处理

InterSystems IRIS 数据平台的核心是一个已经过验证的企业级分布式混合事务/分析处理（HTAP）数据库。它可以以极快的速度接收和存储事务数据，同时处理基于实时数据（包括符合ACID原则的事务型数据）和非实时数据的大量分析工作。这种体系架构消除了将实时数据移动到不同环境以进行分析所导致的延迟。

InterSystems IRIS 通过一系列技术创新获得了高效执行大规模HTAP的能力。

更好的 Sharding

InterSystems IRIS 提供一种强有力且高效的方法对大型数据集执行查询任务。InterSystems IRIS 中的 Sharding 集群能在一层应用服务器中水平分配工作任务和数据集，从而将特定大型表格中的数据划分到多个数据节点中（这被称为数据 Shards²）。

Sharding 有益于多种应用程序，但涉及下列一种或多种任务的用例能从分区中获得最大收益：

- 需要扫描超大型数据集的查询
- 基于大型数据集的复杂查询
- 高数据摄取率和/或容量

当 InterSystems IRIS Sharded 集群接收到应用程序查询任务时，shard master 会将分解后的查询任务推送到数据 Shard，并行执行任务、汇总单个 Shard 返回的结果并将最终结果返回给应用程序。如果某一 Shard 需要来自另一 Shard 的数据来完成工作任务，那么它可以直接连通另一个 Shard 访问所需数据，而无需动用 shard master。

因此，InterSystems IRIS 能够实现一致的高性能、高效和可靠性，即使对于涉及多个表格的复杂查询也是如此。与之相反，其他很多数据库平台需要依靠传送整个表格来实现 Shard 架构，这会导致性能损失和超时。

由于 Sharding 创建了不相交的 Shard 数据，每一个数据服务器的缓存都完全独立，线性添加数据服务器则会增加 Sharding 集群的总内存。因此，通过适当的大小调整，InterSystems IRIS 就能实现内存数据库的性能优势，而无需将所有数据存储在内存中。

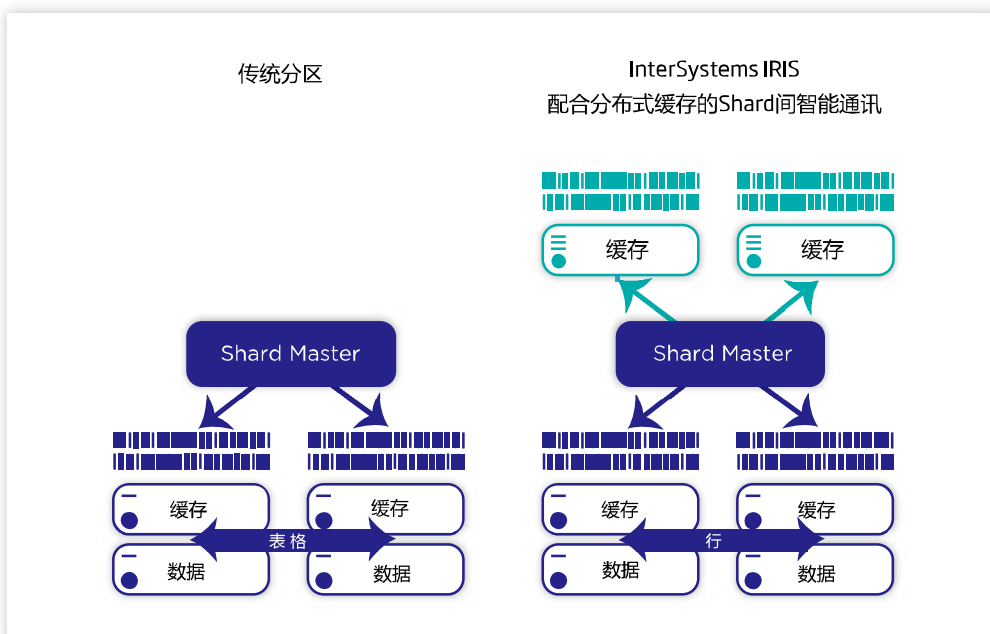


图3: 为分析大型分布式数据集而进行的Shard间智能通讯

² 一个数据Shard是指， 储存Sharding集群shard master定义的每一个Shard表格中一个横向分区的一个InterSystems IRIS实例。储存这一实例的节点被称为一个Shard数据服务器。

InterSystems IRIS Sharding 集群还能提供其他性能优势：

- InterSystems IRIS Java数据库连接 (JDBC) 驱动程序具有透明并联负载能力，可以支持用户在不同Shard中使用基于Java的工具进行非常快速的并行数据提取。
- 大型多用户查询工作任务可能在 shard master 中遇到瓶颈，InterSystems IRIS 可以在 shard master 前面添加一层应用程序服务器，通过分布式应用程序逻辑和缓存进行针对用户数量的扩展。

Sharding 对应用程序透明可见，因此用户不需要更改或只需要更改极少量应用程序代码。Shard 表格和 Non-Shard 表格之间的区别也对应用程序完全透明；这些都是应用程序时要考虑的时间因素。

InterSystems IRIS 的体系架构使复杂的多表连接能够识别分布式 Sharding 数据集中的模式和关系，无需进行 cosharding³、无需复制数据、也无需在网络中广播整个表格。

³ cosharded的数据指的是被分割到使用同一个公共密钥的分布式数据。

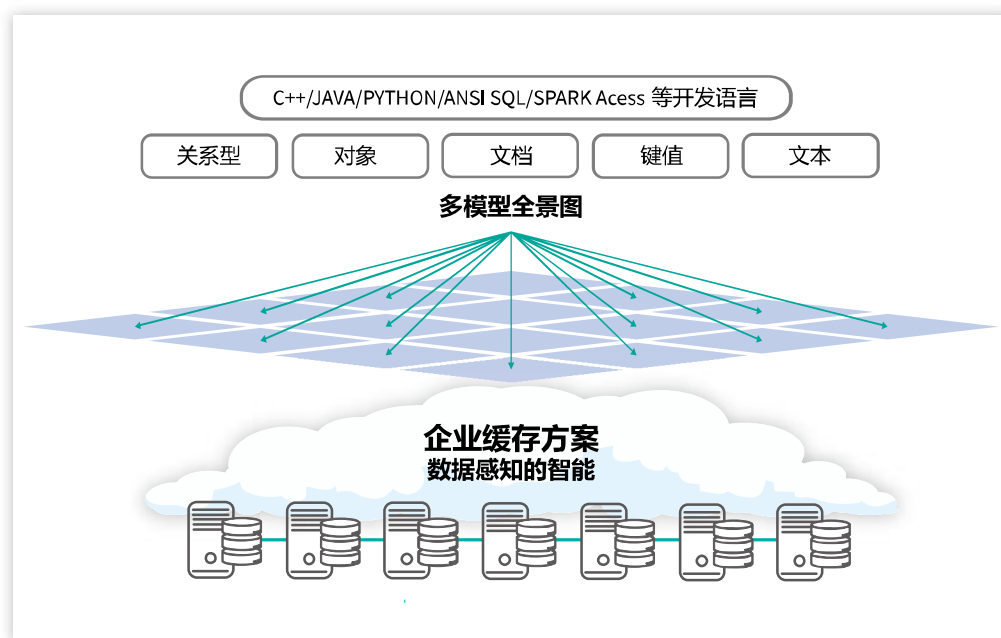


图4: InterSystems IRIS对多模型分布式数据的统一访问

更高性能，更低成本

除了高效执行分析处理，InterSystems IRIS 还能高效和规模化地并行处理事务型工作和分析型工作，而无需将事务型数据移动到不同环境中进行分析。InterSystems IRIS 能处理事务、使数据长期保存于持久性储存中，并可以使用商用硬件在数十纳秒内使事务型数据可用于分析查询。

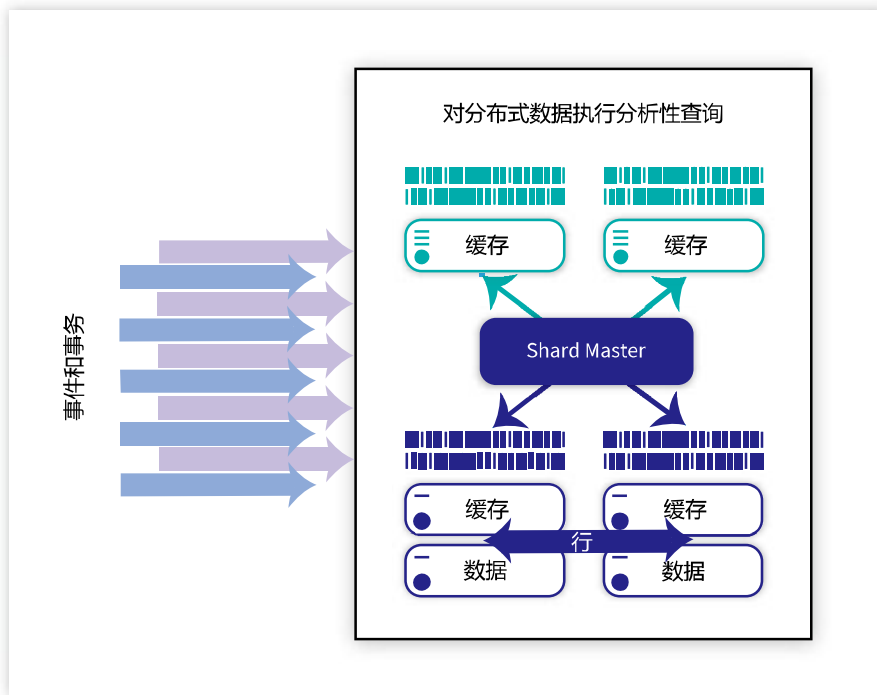


图5: 横向分布的HTAP

InterSystems IRIS 能同时支持直接的共享内存写入和用户 / 服务器分布式 SQL 处理，从而支持高性能的并行事务型 / 分析型用例。因此，InterSystems IRIS 能可靠地处理和分析实时数据以及存储在分布式数据集和分区数据集内的数据，耗时更短且运行成本更低。

为了保证非分区表格和分区表格的高可用性，所有存储数据的节点都可以建立镜像。用户可以轻易添加和移除计算节点，以支持用户工作量波动。InterSystems IRIS 提供强有力的企业级安全；集成 Kerberos、LDAP (轻量级目录访问协议) 和 KMIP (密钥管理互操作性协议)；基于角色的权限控制；以及对中转数据和静置数据进行加密。

多种数据模型

InterSystems IRIS 的构建基础是真正的多模型数据库。这意味着数据一旦被储存，就能被多种数据模型获取使用，包括总是同步的关系型模型和对象模型。这避免了复制数据或在不同的数据表现形式之间提供映射（例如对象 - 关系映射）的麻烦。原生支持多种数据类型的能力使得企业能以最为适合的格式和表现形式来建模、储存和使用数据，从而灵活地开发解决方案、实现更高性能并降低复杂性。

新的数据类型有多重要？

(评分量表：1=不是非常重要，5=非常重要)

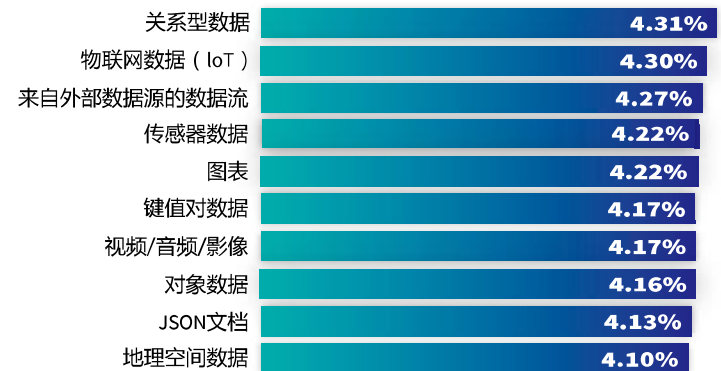


图6: 同一数据平台支持多种数据类型的重要性

来源: 第三代信息管理平台需求调查, IDC, 2016年10月; n=502

嵌入式分析和开放式分析

InterSystems IRIS 支持多种分析方法，以满足当代数据密集型的实时应用程序的多样需求。InterSystems IRIS 提供嵌入式的先进分析技术，可用于进行分布式 SQL、商业智能和自然语言处理，同时按照用户需求可纳入各种各样的第三方分析包和开源分析包。

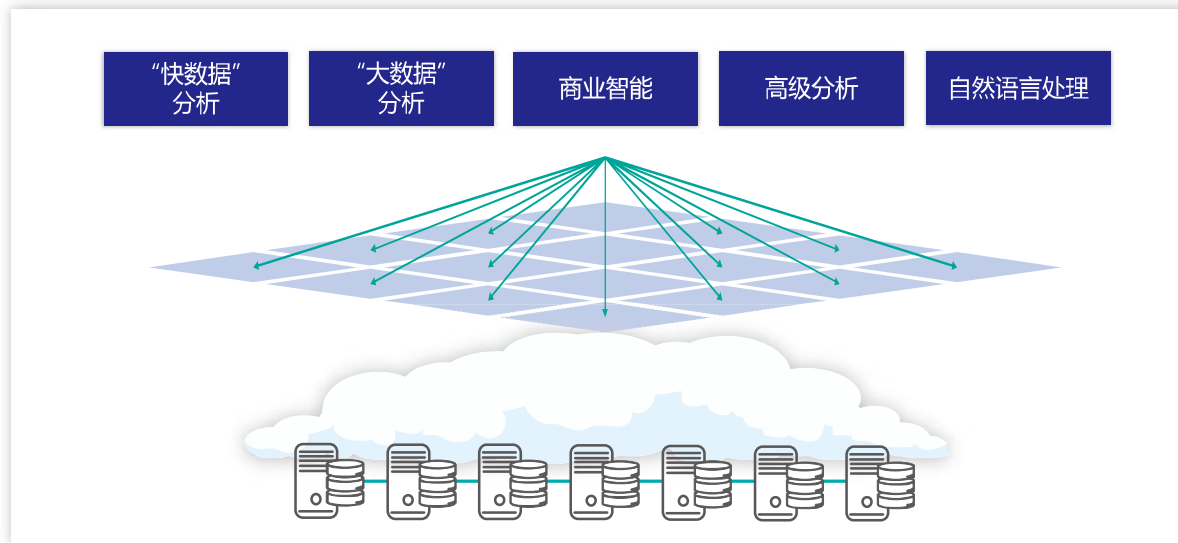


图7: InterSystems IRIS的嵌入式和开放式分析能力

根据研究公司Gartner在2017年进行的一项大型企业调查，在1931名受访者中，有45%称他们计划使用数据挖掘和预测分析，39%计划使用Apache Hadoop或 Spark，25%计划使用由Apache Hadoop或 Spark提供的高级分析功能⁵。

用户们正迅速采用高级分析技术。这些方法和技术包括机器学习、预测分析、人工智能和Apache Spark等实时大数据处理框架。

除了实时（HTAP）和大数据（分布式）处理能力，InterSystems IRIS还提供下列分析功能和功能集成。

集成Apache Spark

Apache Spark 是一种高性能、开源的集群运算框架，通常被用于有高性能需求的大型分布式数据集分析。Apache Spark 比 Apache Hadoop(MapReduce) 运算速度快 100 倍，多种常用机器学习和统计算法可在 Spark 上应用。

InterSystems IRIS 通过一个能识别分区原生 Spark 连接器直接集成 Apache Spark，因此 InterSystems IRIS 应用程序能纳入 Spark 处理过程，而 Spark 应用程序也能通过 InterSystems IRIS 获取分布式数据。Apache Spark 连接器将 InterSystems IRIS sharded 集群中的数据 shards 作为本地分区来呈现，从而实现最高性能。连接器能识别 InterSystems IRIS 数据库的分区本质，允许 Apache Spark 工作节点自动直接连接至 shards，对不相交的数据块执行并行操作。这种并行的直接连接也能允许更高的吞吐量（因为需要通过每一个连接的数据更少）和支持 sharded 集群进行高速数据摄取。

⁵Rita L. Sallam, 等。“调研分析：BI和分析功能相关支出意向，2017”（Gartner,2017）

商业智能

InterSystems IRIS 为商业智能 (BI) 建模、分析和终端用户面板提供完全内建的支持。商业智能模型能呈现对业务有价值的维度,包括总体概念(例如产品线、销售区域、细分市场等等)以及量化指标(例如收入、支出、同比增长、不良率等等)。InterSystems IRIS 可以直接基于事务型数据和其他可能需要的数据来建立 BI 模型。完全自动化的同步选项能避免 ETL 处理。通过拖拽操作即可执行数据分析的能力使得非技术用户也能轻松地检阅所有层级的数据和执行复杂检索。InterSystems IRIS 面板能展示在线的业务指标并为最终用户提供一定的分析能力。

将高级分析融入实时处理的能力

基于 InterSystems IRIS 对预测模型标记语言 (PMML) 的内嵌支持,用户可以使用外部工具和应用程序来将数据挖掘和机器学习创建的预测模型纳入到实时处理中。PMML 是一种完整定义了预测模型中所有参数的 XML 标准,可通过外部分析应用程序或框架构建。当 PMML 模型被加载到 InterSystems IRIS 平台中,平台将生成本地编码,使得模型能被实时执行,而无需任何外部工具或任何可能降低性能的系统间数据传输。这种集成使得数据科学家和其他专业人士创建的预测模型能无缝融入 InterSystems IRIS 内的数据处理过程和业务流程中。

⁶Rita L. Sallam, 等。“调研分析:BI和分析功能相关支出意向,2017”(Gartner,2017)

自然语言处理

InterSystems IRIS 提供自然语言处理 (NLP) 功能,可以从自然语言文本中推断出含义及段落。InterSystems IRIS 能够自动识别文本中的概念和关系,无需前期工作或专业领域知识。这些高级 NLP 功能内嵌于 InterSystems IRIS,可以被纳入业务流程,使得企业能从注释字段、社交媒体和其他富含数据的应用程序中获取信息。

目前存在很多不同的特定 NLP 工具,每一种工具具有特定类型的功能或适用于特定专业领域,一部分应用程序也许需要逐一使用这类工具。InterSystems IRIS 支持 Apache 非结构化信息管理架构 (UIMA) 标准,这使得用户能定义和执行基于该标准的可插拔 NLP 过程。Apache UIMA 为 InterSystems IRIS 内置的 NLP 功能带来了开放式互操作性。

46%的大型企业计划在2017年让他们的应用程序具有非结构化文本内容情感分析功能⁶

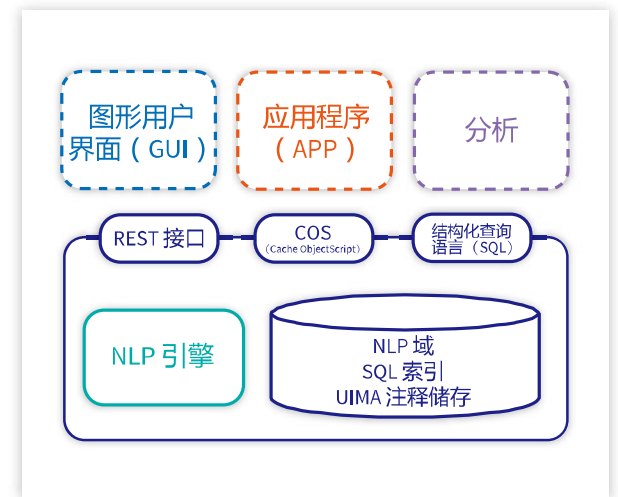


图8: InterSystems IRIS自然语言处理功能

互操作性

InterSystems IRIS 提供一整套内嵌的集成和互操作特性。它为多种的应用程序包、数据库、行业标准、协议和技术提供即时可用的连接和数据转换。灵活的数据转换功能使得 InterSystems IRIS 能解决应用程序或服务之间存在的语义和数据模式差异。

应用程序开发者能够创建无缝业务流程，连通内部和外部数据源、应用程序和服务。InterSystems IRIS 提供图形工具以可视化地呈现流程、规则以及工作流，尽可能降低开发者关于应用程序界面、适配器或中间件机制的顾虑，从而集中精力于系统之间的逻辑交互。图形模型允许业务人员

和 IT 人员进行合作，从而更快速地开发出能满足业务需求的解决方案、更容易地进行修正和扩展现有流程。平台内置了基于角色的工作流引擎，可支持业务流程中的人工交互，从而在用户中自动化分发任务和合并他们的决策及行动。

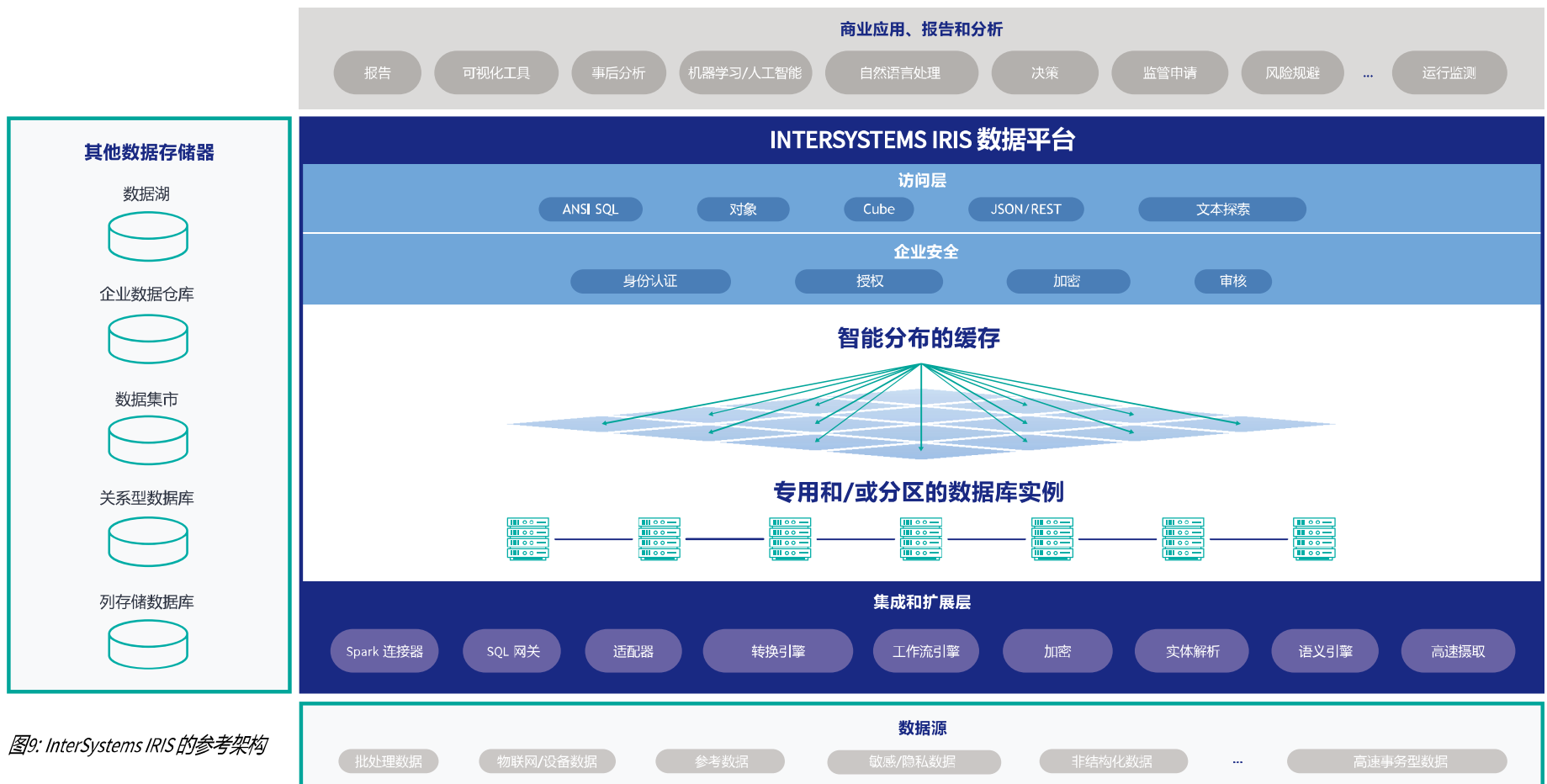


图9: InterSystems IRIS 的参考架构

由于 InterSystems IRIS 包括内嵌的数据库和分析功能，同时利用存量数据和实时数据的复杂分析可以无缝地融入业务流程。所有数据，包括中转数据或与长期异步事务相关的数据，都能自动储存在数据库内，可被用于制定报表和分析。

平台支持医疗、金融服务、零售业和电讯业等各种行业使用的多种标准，包括REST架构和Web Service（例如JSON、XML、XPath、XSLT、SOAP和DTDs）。

统一的开发环境

InterSystems IRIS统一的开发环境能够基于代码工作或基于图形工作，能够对各种编程模型、编程界面和数据格式提供一致的表达，从而为所有功能提供统一的开发环境。

灵活的部署选项

InterSystems IRIS 提供一种简单直观的方式在云端和本地基础架构上预览和部署服务。通过 InterSystems 基于 IRIS 的应用程序，IRIS 提供代码、不可变基础架构和容器化部署等等能力。它能避免在新技术和相关培训上进行大量投资，以及避免进行试错性的系统配置和管理。

InterSystems IRIS 允许企业利用云端计算和容器化软件具有的高效、敏捷和可重复优势，而无需大量开发或再加工。它能在现有虚拟或实体集群提供和部署 InterSystems IRIS 配置，也支持在包括已有基础架构和商业云平台的企业级别操作系统平台上部署容器化软件。

结论

InterSystems IRIS 是完整、统一的数据平台，能简化富数据实时解决方案的开发、部署和维护。InterSystems IRIS 提供并行事务处理和分析处理功能；支持多种完全同步的数据模型（包括关系型、层次、对象和文档数据模型）；是一种能集成数据孤岛和分离应用的具有完整互操作性的平台；具有成熟的结构化和非结构化分析功能，能用于批量和实时的数据处理场景。IRIS 平台提供一种开放式分析环境，可将同类最佳分析方法纳入 InterSystems IRIS 解决方案，也提供灵活的部署功能可支持云端部署、本地部署或任何形式的混合部署模式。

很多行业用户正在使用 InterSystems IRIS，通过利用更多数据和消除事件、洞见及行动之间的延时来提供一系列战略上和运营上的收益。

我们特别为您提供 InterSystems IRIS 体验项目，您能自主探索、亲身感受 InterSystems IRIS 的强大。如需体验请访问

[InterSystems.com/Experience](https://www.intersystems.com/Experience)

InterSystems 公司
全球总部
One Memorial Drive
Cambridge, MA 01142-1356
电话: +1.617.621.0600
InterSystems.com

InterSystems 中国
系联软件 (北京) 有限公司
北京市朝阳区建国门外大街乙12号
双子座大厦 (东塔) 1902室
电话: +86 10 8524 9700
InterSystems.com/cn



InterSystems.com/cn