

## 全面彻底的运营分析

### 面向各种规模企业分析的集成平台

2013 年5 月

白皮书

作者：Barry Devlin 博士，9sight Consulting  
[barry@9sight.com](mailto:barry@9sight.com)

运营分析已经成为了现代企业必须具备的能力。随着大数据从炒作转为价值，随着市场要求以更快的速度、更高的准确度把握机遇和应对威胁，运营分析正在成为近实时决策和行动的基础技术。

本白皮书针对精通业务的 IT 专家和掌握计算机相关知识的业务经理概述运营分析的业务价值和技术基础。概括介绍“弥补业务需求与技术能力之间的差距”这一观念的演变过程。描述及时、详细决策和行动的业务驱动因素；通过样品实施展现广义价值。

我们将探讨市场对利用完美集成、高度连贯和易于扩展的平台实现所需性能级别的需求。该平台是一系列经过妥善试验和测试的数据仓库技术的演变成果，同时采用了更多先进的开发成果。

最后，我们将讨论分析 IBM 的运营分析产品，重点介绍 IBM InfoSphere Warehouse 以及相关组件作为软件库满足全面实施的各种需求的功能。此外，我们将简单介绍一些部署方案，例如 IBM PureData Systems for Operational Analytics，它可以为所有行业中各种规模的企业创建运营分析环境。

简而言之，即全面彻底的运营分析。

赞助商：

International Business Machines

[www.ibm.com](http://www.ibm.com)



## 目录

### 2 从数据仓库到运营分析

### 4 运营分析——详细、及时的高价值

### 6 集成的技术平台

### 7 闭路运营分析

### 11 部署选项

### 12 结论

一家亚洲公共事业公司跟踪 200 万用户的实时燃气使用情况，预测需求模式，快速确定泄漏故障和盗窃行为。一家汽车零部件零售商分析客户网站行为和销售数据，密切管理库存，根据市场机遇即时调整营销工作。一家英国移动通信公司仔细检查掉线率和数据传输量，根据实时使用形式划分客户群，预测客户流失，检测欺诈行为。

在不同的行业和业务部门之间存在共同的思路。大量的新数据来源与传统数据库相结合并接受深入分析。结果近实时地应用于呼叫中心和网站，影响客户行为，采取不同措施，或者通常改变运营事务的参数和环境。捕捉详细结果并反馈到分析当中。信息运营环路正在快速闭合。

自二十世纪九十年代以来，Stephan Haeckel 的感知和响应<sup>1</sup>已演变为倾听和预测<sup>2</sup>。倾听客户和预测客户需求的企业一直保持着更高的收入增长，比未这样做的企业高出多达 60%。实际上，感知和倾听在本质上是相同的：尽可能广泛、快速地捕获数据，通过全面、管理妥善的流程将数据集成一个一致的环境中。然而，预测远远超出了单纯响应的范围。预测要求能够根据潜在成果创造触发事件，预测当前行动的影响，模拟最佳成果。正如我们将要在下文看到的那样，这些都是理解运营分析并从中获益的关键。

业务影响十分显著，所需的行为变化相当彻底。麻省理工学院斯隆管理学院与 IBM 商业价值研究院最近联合发布的一份报告<sup>3</sup>指出：“利用新业务模式和新数据，意想不到的竞争对手正在涌现，熟悉的客户要求前所未有的关注。诸如此类的突发状况为各个级别的竞争者创造了越来越多的机会。利用综合分析战略制约下的全面分析能力，企业能够更好地拉大或缩小自己与竞争对手之间的差距，实现自身优势最大化。”

运营分析适用于整个供应链上所有行业中所有规模的组织。无论客户群有多大，了解客户行为与优化客户体验和盈利能力同样重要。零售商、电信运营商和公共事业公司都能通过提升客户保留率获益。此外，运营分析还有利于改善企业与供应商和合作伙伴的互动关系。下面，我们将探讨这个全新世界是如何从以前的商业智能事业中浮现出来的。

运营分析能够产生即时业务影响，但要求用户和 IT 行为发生重大变化。

## 从数据仓库到运营分析

万变不离其宗<sup>4</sup>

**万变** 不离其宗。原来被称为商业智能 (BI) 的领域在过去 20 年里一直深受各种术语的困扰——很多术语尝试暗示实际上几乎不存在的新奇性。二十世纪八十年代的数据仓库 (DW) 在二十世纪九十年代变成了商业智能 (BI)，后来又被人们转变成了业务分析 (BA)。二十世纪九十年代的运营数据存储 (ODS) 改容换面，到本世纪头十年成为了运营商业智能 (BI)，最近则演变成了运营分析。现在，很难找到一个从未放到“分析”前面并且声明最新的决策支持必备要素的形容词。将无处不在的大数据添加到组合中，您会完全混淆。然而，万变不离其宗……

事实上，业务决策制定的技术支持已经发展了多年。重点已经从收集和清理数据转变为向业务部门交付各种各样的信息，从回顾过去的业绩转变为跟踪当前的业务运营状况，从简单的事实报告转变为基于预测情况的高度自动化决策。图 1<sup>5</sup> 显示了两个方面的演进：推动业务用户的行为和需要 IT 提供商做出的响应。

随着这种演进飞速发展，业务要求增加数据利用的灵活性以及数据供应的及时性。在第一阶段，每周利用数据生成一次基本固定的报告，IT 认为这是适当的，并且会提前提供数据。此类报告完全是对过去的回顾，描述业绩的某些方面。如今，由于有限的投资回报，很少有企业会停留在这一阶段；大部分已经进入到第二阶段。

第二阶段的特点是越来越关注更广泛的业务范围内的信息广度，并且能够利用信息预测未来成果。在这一阶段，IT 高度专注于从公司里的多个运营数据来源集成数据。这是传统的商业智能环境，由一个数据仓库基础设施提供支持，详情请见公司白皮书《Integrating Analytics into the Operational Fabric of Your Business》（将分析整合到贵公司的运营架构中）<sup>6</sup>第 5-6 页。如今，大多数企业正处于提供全面决策支持的第二阶段，领导者开始向第三阶段过渡。

对大多数企业而言，第三阶段标志着在决策支持方面做出重大变革。在业务方面，需求从主要的临时、被动和管理驱动型向流程视图转变，允许预测分析结果直接（通常实时地）应用于业务运营。这是运营分析行为的本质。在这一阶段，IT 必须具备高度适应能力，以预测业务对信息的新需求。此类变化要求在思想上做出转变，从独立运营和信息系统转变为综合运营信息环境。这是今天的行动所在。这是领先企业发现投资回报的地方。以下两个部分将分别定义运营分析和介绍支持运营分析所需的架构平台。

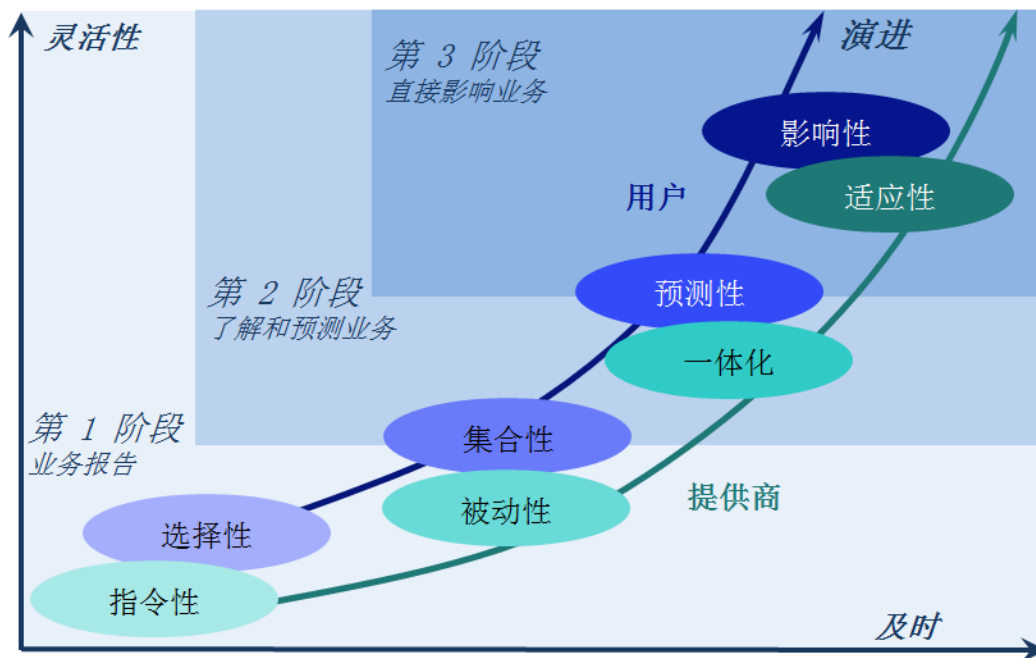


图1：决策支持的演进

从“简单地”预测客户行为向真正影响客户行为的转变，要求高度适应但以流程为导向的 IT 环境

## 运营分析——详细、及时的高价值

“下列三项依然是竞争基础：高效执行、明智决策和从业务流程挖掘价值的 ability——所有这些都可以通过高效利用分析来获得。”<sup>7</sup>

### 上文

提到的麻省理工学院斯隆管理学院与 IBM 商业价值研究院联合报告将分析定义为“利用数据以及通过应用分析学科（例如，统计、情境、定量、预测、认知和其他模型）分析得出的洞察推动基于事实的规划、决策、执行、管理、评估和学习”。运营这个词指的是经营企业所需的日常行动——联机事务处理 (OLTP) 系统，记录和管理企业、客户、供应商等各方之间详细的实时活动。这与分析数据和报告分析结果的信息系统形成了鲜明对比。

运营行动需要实时决策。有时，答案一目了然，以至于我们甚至都没有看到问题。一家在线零售商收到了登录客户的库存夹克订单；毫无疑问，订单被接受了。但是，如果商品缺货，如果我们有一件利润更高并且客户可能会喜欢的夹克，或者当我们知道客户最近在搜索卡其裤时我们正好有一条与夹克配套的裤子，那么隐性问题——我们应当如何处理这份订单？——会更加明显。这是一个微妙的渐进过程，从响应到假设再到根据最佳现有信息进行预测。决策可能非常明显，但有时候值得提出这样的问题：如果我们探

每项运营事务都有一个相关联的决策；每次行动之前都要制定一个决策。

索不同的据测并采取不同的行动，结果是否会更好？

运营分析是 *根据通过运用统计模型和分析现有及/或模拟未来数据得出的洞察，为实时运营决策提供最佳或切合实际的建议，并且将这些建议用于实时互动的过程。*

## 如今，为什么运营分析至关重要

*“分析本身不是战略，但利用分析优化特征鲜明的业务能力肯定是一项战略。”<sup>8</sup>*

以上两个引述摘自 Kaplan 和 Harris 在 2007 年出版的书籍《*Competing on Analytics: The New Science of Winning*》（*分析法层面之竞争：获胜的新科学*）。无论过去还是现在，它们都是真实的，但运营分析是目前的话题，这是因为：

1. 业务运营通过适应性网站以及越来越自动化和数字化的流程传感器数据收集为统计分析提供了更多的数据。
2. 社交媒体正在为分析提供数量越来越多、种类越来越繁杂的信息
3. 随着业务变革持续加速，业务竞争将日趋激烈
4. 数据存储和处理功能日益强大，成本却逐渐下降，使运营分析在经济上成为了小型企业的理想选择
5. 制定许多更加明智的小规模、低价值决策通常比制定少量的高价值决策带来更多的净收益；故障风险的覆盖面越来越广

企业决策管理专家 James Taylor 指出<sup>9</sup>，运营数据量大得足以提供具有统计意义的结果，决策实施成果可以在相对较短的时间内看到和跟踪。因此，运营分析可以提供一个理想平台，开始将技术进步应用于预测分析并测试其有效性。

### 业务使用案例——简要概述

领先企业正在将运营分析应用于整个价值链上的所有业务流程。《*Integrating Analytics into the Operational Fabric of Your Business*》（*将分析整合到贵公司的运营架构中*）<sup>6</sup> 提供了一个以功能为导向的视图，强调营销、客户保留和防欺诈。在这里，我们采用以行业为导向的视图。

金融服务业是最早采用数据仓库解决方案，跨多个多样化系统实现单一客户视图的行业。通过向近实时事务和客户交互数据访问转变，基本集成基础设施获得了延伸，将欺诈检测提升到了全新级别。北美一家大型金融机构一直利用运营分析，跳出发现个人可疑交易的限制，识别可能影响分散地区的多个客户的欺诈行为模式，快速做出响应，防止损失。此外，此类使用模式跟踪可以进一步扩展，预测客户需求，改善客户服务。

零售商是最早采用运营分析的主体之一，当时运营分析叫作数据挖掘。由于收银机具备了强大功能，足以存储购买记录，购物篮分析——尤其是在结合忠诚度卡使用时——自二十世纪九十年代早期就已成为规范。但是分析销售模式只是部分情况；许多实际价值来自将这种知识与供应链相集成——从店面回溯至仓库乃至供应商——确保货架上有充足的最有价格产品，尤其是当客户需要时。需要一个与运营查询和深入分析同样易于使用的系统，像在韩国领先零售连锁店 GS Retail 看到的那样，将运营存储管理功能和营销分析融合在单一运营分析环境中。

随着电话连接到个人（一个更大、更关注成本的客户群）而非位置，以及广泛的市场自由化促使供应商之间展开激烈竞争，手机/移动电话在二十世纪九十年代中期使电信业发生了翻天覆地的变化。智能手机延续了这种趋势，2012 年移动用户超过 60 亿<sup>10</sup>。在这个行业，了解和预测客户流失和采取行动管理客户流失是生存问题。运营分析在营销、服务中心和运营中占据核心地位，提供客户行为相关信息，以便在呼叫中心管理客户流失，

*在一个业务领域部署的运营分析系统可以扩展至或在其他领域重复使用。*

开展吸引或留住客户的营销活动。品牌名为 3 的英国领先移动提供商 Hutchison 3G 利用行为细分——基于电话和移动宽带使用情况分析——而非传统的人口特征细分，在用户离开之前确定用户行为趋势和模式，进而确定可能流失的客户并进行主动干预。该系统也用于检测欺诈性使用模式。这只是一次尝试，看看公司正在做什么。在成本节省和创收方面取得了界定明确的实质性成效。正如我们看到的，安装后，这些系统通常能够在初期目标领域以外的其他业务领域带来同样显著的收益。

## 传统

商业智能（在图 1 中描述为第一阶段和第二阶段）很久以来都是通过一个严格分层的数据架构交付。此架构以一个集成、一致、可靠的历史信息存储库——企业数据仓库 (EDW)——为中心，有专门用于决策支持的数据集市<sup>12</sup>。运营分析和“大数据”的出现要求翻新旧架构，引入集成信息平台，如图 2 显示，详细描述请见《The Big Data Zoo—Taming the Beasts》（大数据动物园——驯服野兽）<sup>13</sup>。该平台适用于所有信息类型，由大量的数据库和分析技术构成，每项技术都针对特定的处理和访问类型进行了优化，称作支柱，以它们支持的业务角色命名。

第一个处于中心位置的**核心业务数据**支柱是在 EDW 和 MDM 系统中找到的一致且有质量保障的数据。IBM DB2 等传统关系数据库用作基础技术。明确排除目前通常存储在 EDW 中的应用特定报告和决策支持数据。第二个支柱**核心报告和分析数据**涵盖这些被排除在外的数据类型。就技术而言，在理想情况下，该支柱也是关系数据库。数据仓库平台，例如 IBM InfoSphere Warehouse、IBM Smart Analytics System 和新的 IBM PureData System for Operational Analytics 在这里发挥着重要作用。要求更高查询性能的业务需求可能要求分析数据库系统构建于大规模并行处理 (MPP)、纵列数据库或其他专业技术之上，例如全新的 IBM PureData System for Analytics（基于 Netezza 技术）。

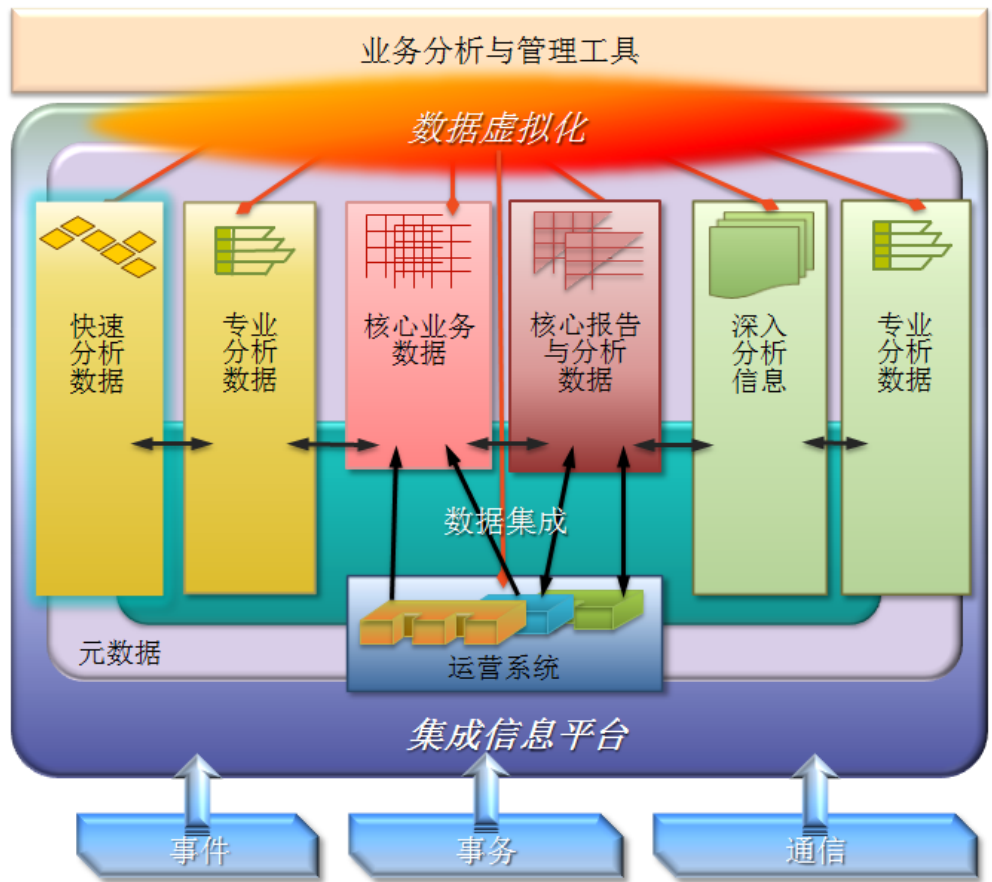


图 2：  
集成信息  
平台

**深入分析信息**要求高度灵活的大规模处理，例如通常在 Hadoop 环境中执行的统计分析和文本挖掘。**快速分析数据**要求高速分析处理，它必须在活动数据上完成，例如利用 IBM InfoSphere Streams。这部分数据通常来自多个数据源，需要为实时分析和决策制定进行近乎零延迟的连续分析和聚合。综合速度和灵活性，我们有**专业分析数据**，使用 NoSQL、XML、图形和其他数据库和数据存储的特殊处理。它在平台中出现两次，因为它应用于由左边事件生成的机器生成数据和右边通过沟通产生的人力资源信息。

因此，我们可以看到，平台的中心支柱与经典商业智能和传统数据仓库架构紧密相连，不同于数据集市的数据用于报告和分析。数据可以——而且经常应该——来自运营系统，以减少旧的三层架构固有的延迟。运营分析也集中于平台的同一部分，

集成平台包含企业生成和使用的**所有信息**。

但是清楚地了解按需涉及的其他支柱和跨支柱提供的环境集成，是确保所用数据的一致性——语义和时间——和维护运营分析所需的闭路流程的关键因素。

## 闭路运营分析

“我们知道……过去不会为将来提供有用的指导，而且需要我们处理的突发事件永无止境。”<sup>14</sup>

**在**理论上，传统商业智能应该是一个闭路流程——数据从运营环境复制到并在商业智能中分析；这可以推动决策制定，影响运营环境，并且可以对比实际结果与预期结果。实际上，在战略和战术商业智能中，前三个步骤进展地非常缓慢，最后一个闭路步骤很难执行，因为采用的决策范围偏大。过去实施运营商业智能的企业出于技术方面的考虑，也曾想方设法关闭环路。运营分析要求关闭环路。今天需要的实际环路在某种程度上也更加复杂。幸运的是，近期的技术进步可以满足这种要求。

图 3 详细介绍了在集成信息平台上的运营分析闭路流程（经过简化，以便清楚展示）。

1. 收集信息用于分析：从各种来源收集信息——现有的仓库数据、运营系统、社交媒体等。
2. 深入分析：对收集的数据执行统计分析，发现行为和行动之间的关联，构建模型。根据涉及的数据类型和数据量，此项任务可以在关系数据库和 Hadoop 组合中执行。
3. 模型应用：将一个或多个模型应用到运营环境。在某些情况下，它们可以应用到所有互动中；在其他情况下，实时分析可以确定是否激活以及何时激活模型。
4. 实时分析：实时地分析行为和行动（例如，网页点击量），在发生运营活动时，酌情实时应用模型。
5. 深入分析反馈：记录结果，在深入分析阶段提供反馈，以进行进一步建模。

从 IT 角度来讲，步骤（1）（2）和（5）与步骤（3）（4）在处理特征方面存在显著差异，尤其是在拥有大量数据和快速决策需求的组织中。前者需要阅读和大规模计算潜在的海量数据，用时限限制相对来说不那么苛刻。后者则恰恰相反——要求快速响应少量数据的读写操作。现在，我们使用 IBM InforSphere Warehouse 和相关的 IBM 产品探讨一下这些特征。

### (1) 收集信息用于分析

长期以来，数据整合一直是商业智能的重心。过去，焦点一直集中于协调和整合来自分散运营系统的数据——经典 ETL（提取、转换和加载）。尽管这一方面依然重要，但运营分析要求它必须由

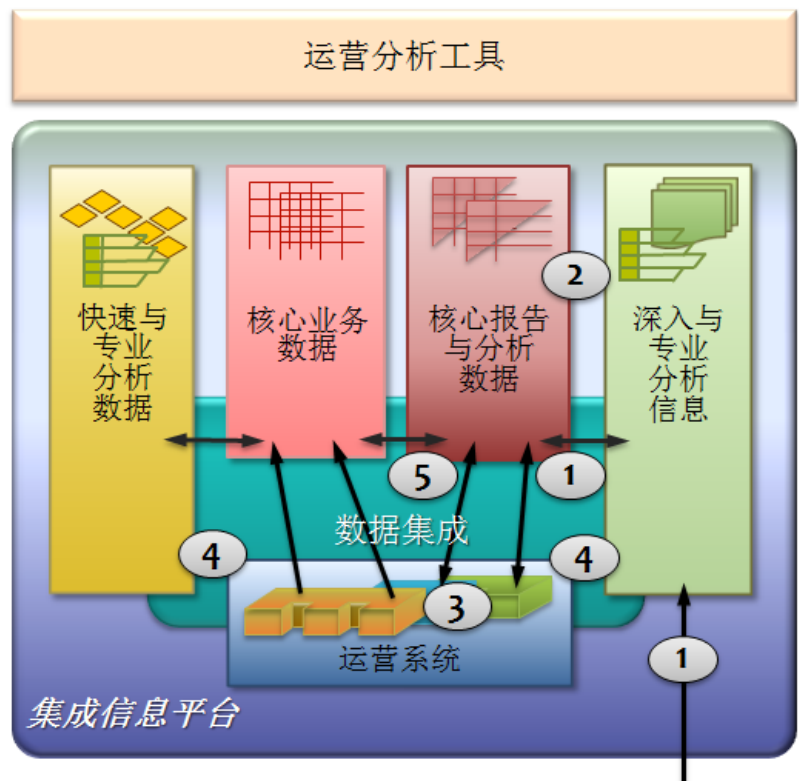


图3：  
闭路运营  
分析

运营分析是一个要求将运营与信息处理特征结合在一起的流程。

另外两个特性做补充：(1) 数据采集的速度和时间，(2) 传统运营系统以外的来源，包括“大数据”和仓库内部的来源。

**IBM InfoSphere Warehouse V10** 利用新的持续数据获取功能满足第一个要求，允许从管道或文件加载 **Linux、UNIX 和 Windows (LUW) 版 DB2** 表格，避免升级表格，允许其它应用在加载表格时使用数据。通过消除或减少传统脚本或批量 ETL 方法带来的停机故障，所需数据可以更早地提供给用户或自动化流程，支持更快的分析和战术决策。

相比传统商业智能，运营分析要求从广泛的来源收集信息，在更多的结构中处理信息。**InfoSphere Information Server** 和 **DataStage** 可以提供全套信息收集和供应功能，从所有常见的关系和非关系数据来源以批量模式和实时模式（在许多情况下）收集和供应信息。此外，内建的 **InfoSphere Warehouse Replication** 功能可以简化在 DB2 环境中在核心业务与报告/分析数据区域之间迁移和复制数据的流程。

总之，这是一个完整的功能集，适用于信息收集的所有方面。

## (2) 深入分析

在运营分析的这一阶段，主要强调广泛的统计调查，以及以更传统的方式查询大量的相关数据。目标是发现成百上千个行为属性、特征和行动之间的意外关联。鉴于有如此多的属性和百万份报告，不应期望这一阶段是实时的。然而，及时性仍然是一项要求，尤其是在初步清理、准备和探索阶段之后。过去运行夜间（或更长时间）数据挖掘和不同平台之间的频繁导出/导入数据的经验不再适用于持续分析。

平衡数据范围需求和及时性是设计运营分析时的关键考虑因素。

规模和及时性可以通过许多不同的方式获得，具体情况取决于所用数据的特征以及分析需求的紧迫性。初步准备和探索通常可以从使用基于 **Hadoop** 的平台中获益，在图 3 中显示为深入和特殊分析。在这一方面，**InfoSphere BigInsights** 可以提供一个管理和使用大量多结构化数据的平台。当充分了解、优化数据并准备将其用于生产时，通常需要将部分数据迁移到关系环境中，在这里，**IBM** 可以提供三个基础平台：

1. **IBM DB2 Advanced Enterprise Edition**，一款适用于 Linux、Unix 和 Windows 的产品，是本白皮书重点介绍的产品之一
2. **IBM DB2 Analytics Accelerator**，一款运行在 System z 上的硬件/软件产品，使用 **Netezza** 技术，与 z/OS 上的 DB2 实现了紧密集成，并且通过 DB2 进行透明访问，详细描述请见其他地方<sup>6</sup>
3. **IBM PureData System for Analytics**，一个完全基于 **Netezza** 技术的系统，针对大规模的集中战略性和预测性分析进行了优化

在运营分析领域，需要一个支持 (i) 深入分析阶段——这些需求不那么极端——和 (ii) 后续实时分析阶段以及 (iii) 为核心业务数据提供平台的关系平台。

LUW 版 DB2 通过其数据分区功能可以满足这三点需求，基于大规模无共享架构，针对关系以及本地 AML 数据提供高性能查询处理。**Version 10** 提供适应性压缩，自动在基于行和页的方法之间确定最佳压缩，有可能降低存储成本，提升性能，尤其适用于大型 I/O 绑定仓库应用和查询工作负载。多温数据管理功能使 DBA 能够按温度划分数据类别，进一步降低存储成本。典型类别包括热（频繁访问、响应时间短）、暖（不太频繁的访问、响应时间更长）、冷（访问不频繁、从不更新）和休眠（已存档），存储在不同速度/成本的硬件上，涵盖固态硬盘、传统旋转磁盘乃至磁带（通过 **Optim** 管理）。这种分类法也可以让 **DB2 Optimizer** 按优先顺序排列热数据访问。

**InfoSphere Warehouse** 可以提供嵌入式数据挖掘、建模和评分功能，包括标准数据挖掘模型算法，例如集群、关联、分类和预测（在行业标准的 **Predictive Model Markup Language** 格式中可以导入其他算法），为这一运营分析阶段的深入、预测性分析特征类型提供支持。此外，文本分析功能允许直接使用自由格式信息，例如呼叫中心记录、客户反馈等作为总体分析的一部分。高性能特征意味着这种深入分析可以与后续实时分析阶段在相同的平台和数据上执行。



### (3) 模型应用

模型应用——快速更改企业及其客户之间的互动以提升业绩的流程——显然是一个运营流程。在传统的商业智能方法中，我们会说它发生在商业智能系统以外，并且进展迅速。归根结底，运营系统不在商业智能范围之内。然而，仅运营分析这一术语就暗示我们必须处理运营和分析两个方面的任务。复杂性分为四个级别：

运营和分析之间的集成程度取决于企业需要的自动化水平。

- 1. 影响人工流程：**在呼叫中心，代表按照以前分析客户的总体生命周期价值、类似标准和基于人口特征的响应或产品预测时开发的模型与客户进行互动。呼叫中心应用通过预加载模型数据（无实时数据）松弛地连接到分析系统。代理拥有很大的权力来指导互动。
- 2. 指导人工流程：**与上一个情况类似，但呼叫中心和分析应用之间的连接接近实时。必须激活运营分析的第四阶段（实时分析），使代理有实时数据和基于实时数据的建议；代理在互动方面的自由更受约束。
- 3. 自动化供应流程：**零售网站的特征，客户看到的商品都会自动根据过去的互动以及当前点击流从模型中生成。分析环境与运营系统紧密相连，实时分析按闭路方式执行。
- 4. 集成的模型与供应流程：**在一个集成系统中开展实时分析和运营活动，允许多个模型针对不同的客户同时运行，并根据结果进行自我调整。几乎无法分辨实时分析和运营。

随着我们逐渐完成这些阶段，从单独但相连的运营和信息环境到完全集成的实时自动化运营信息环境，对数据库系统的需求大幅增加。因此，高效查询处理和优化变得至关重要。通过利用 DB2 中的特定优化技术，例如星型结构优化，数据库可以为分析工作负载交付高性能以及高性能点查询访问。Workload Manager 允许按优先顺序精细地排列特定类型的查询和应用。改进的索引管理，包括跳跃扫描和预先获取，Predicate Evaluation Avoidance 也有利于性能改进。

### (4) 实时分析

与上一阶段相同，性能是实时分析的重中之重。我们已经详细讨论了 DB2 在这一方面的表现。另一个特别适用于实时分析并且在任何商业智能相关函数中都非常有用（以及很久以来在许多 EDW 开发中进行手动编码的函数）的功能是 Time Travel Queries，也叫作 Bi-temporal Tables。借助这一功能，DB2 可以接管表格中指定时间戳列的管理工作，实现更快速、更轻松的历史和趋势分析查询。随着将新数据不断从实时活动添加到分析环境，此功能允许更轻松地构建和维护历史表格，将分析集中在特定时间段内进行。

在这一阶段以及深入分析阶段，包含的 **IBM Cognos Business Intelligence** 功能都可以为业务用户提供丰富的商业智能能力，从整个运营分析环境轻松访问数据；而且报告和分析功能允许它们按需要的方式、在需要的时间和地点交付相关信息。此外，客户能够利用 **IBM Cognos Dynamic Cubing** 功能部署高性能 OLAP 报告集群。**IBM Mashup Center** 也包括在内，是满足小型组织和部门使用的分析和可视化需求的基础。

### (5) 深入分析反馈

运营分析的第五个阶段在技术上相对简单：只需更新在深入分析中使用的数据以及实时发现的其他数据和新结果。然而，不应低估它的重要性；这一步骤最终会关闭运营活动与信息活动之间的环路。长期以来，它一直是商业智能系统中的绊脚石，因为两个环境在设计和所有权方面不匹配。它之所以在这里更容易，是因为这两个环境在运营分析中实现了概念统一。但是，在实践层面，单一数据库环境的广泛使用是主要推动因素。我们讨论的 InfoSphere Warehouse 特性和功能——尤其是 Replication——使其成为可能。Advanced Security 特性（DB2 透明地管理表格中行和

集成软件平台是运营分析的关键，它可以为广泛的运营和信息需求提供支持。

列的数据访问) 也有助于在这种混合运营/信息环境中使用不同类型的数据。

---

## 部署方案

**IBM** 是最早认识到决策支持的重要性以及关系数据库技术支持决策的程度的厂商之一。通过 DB2 和 SQL/DS 上的实施, 他们在 1988 年<sup>15</sup>发布了第一份数据仓库架构描述。从那时起, IBM 创建和/收购了大量的软件和解决方案组合, 涵盖整个决策支持——从战略商业智能到业务分析, 从预测性分析到运营分析, 从中小企业到大型跨国公司。这一部分概括介绍适用于前面描述的广泛功能的主要部署方案。

如我们前面看到的, 任何运营分析的核心都是数据仓库环境, 但大部分核心业务数据源自与运营环境的某些重要方面紧密集成的数据仓库环境。因此, 从软件角度来讲, IBM 的运营分析环境围绕 DB2 (在任何平台上) 以及 InfoSphere Warehouse 及其支持和补充工具构建。在需要更深入和/或更专业的分析的情况下, 可以添加基于 Netezza 或 Hadoop 技术。因此, 任何在 IBM 基础上构建预订解决方案的公司都会从这些软件组件开始。

然而, IBM 也提供大量的预构建和集成解决方案。

2012 年 10 月发布的 IBM PureData System for Operational Analytics 是 IBM PureSystems 专业集成系统家族的最新成员。它可以提供预集成硬件、软件和专业平台, 此类平台针对运营分析进行了优化, 以 IBM AIX 平台上的 LUW 和 InfoSphere Warehouse 版 DB2 为构建基础。它最大可以扩展至 96 个 Power 7 核心, 可以存储超过 1PB 的未压缩原始数据。它设计用于处理 1000 个以上的并发运营查询。

PureData System for Operational Analytics 以 IBM PureData System for Analytics (基于 Netezza 技术) 为补充, 针对预测性分析进行了优化; IBM PureData System for Transactions 基于 LUW 版 DB2, 针对单纯的运营工作负载进行了优化。它们都是 PureSystems 专业集成系统的一部分。PureData System for Operational Analytics 在技术层面从 **IBM Smart Analytics System** 演变而来, 包含 System x 和 System z 平台上的最新产品。前者由面向中小企业的 IBM Smart Analytics System 5600 和 5710 构成。后者由 IBM zEnterprise Analytics System 9700 和 9710 以及 **IBM DB2 Analytics Accelerator** 构成, 详细描述请见《Integrating Analytics into the Operational Fabric of Your Business》(将分析整合到贵公司的运营架构中)<sup>6</sup>。

尽管命名在不断进化, 但所有这些解决方案都源自一个想法: 提供逐渐针对现代企业环境中的常见数据使用类型进行优化的预配置、易实施系统。

---

## 结论

“我们的未来定下了我们今天的规矩。”<sup>16</sup>

**运营** 分析是实时、即时业务的基础, 在一个快速发展的市场中有助于提升业务价值和客户满意度。它是美丽新世界的预言者。在新世界里, 运营行动和信息决策之间的旧有障碍完全消失。调查和评估实际变化——是机遇还是威胁, 决定响应, 采取行动, 在一个连续的闭路中评估影响。环路的关闭速度或频率仅取决于业务需求。

选择集成的全功能运营分析平台是立即实现业务价值的关键一步。迈向全新的运营/信息一体化世界。选择所基于的三个基本因素:

1. 核心数据库环境, 支持深入分析和快速响应运营需求, 足以处理大多数日常工作

2. 支持和鼓励使用专业化存储和工具处理非常规工作的环境

3. 完全实现的数据集成和管理环境，支持最广泛的数据迁移、转换和访问需求

为了使用一流组件设计、构建和部署这样一个复杂环境座位预订开发，有可能超出了除大型企业以外其他所有企业中大多数 IT 部门的能力。幸运的是， InfoSphere Warehouse 的 IBM 软件库以及相关产品提供了一个功能齐全且高度连贯的解决方案。利用 IBM PureData System for Operational Analytics 举例证明的专业集成系统类型，所有准备实现飞跃的公司都可以轻松部署。

Barry Devlin 是业务洞察方面首屈一指的权威人士之一，是数据仓库创始人之一，在 1988 年发表了第一篇有关这一话题的架构论文。拥有 30 多年的 IT 经验，其中包括 20 作为 IBM 杰出工程师的经验，他是公认的分析师、顾问、讲师，是重要著作《Data Warehouse—From Architecture to Implementation》（数据仓库——从架构到实施）以及大量白皮书的作者。



Barry 是 gsight Consulting 的创始人兼负责人。他专门研究融合了运营、信息和协作环境的深入业务洞察解决方案对人力资源、组织和 IT 产生的影响。Barry 定居于南非开普敦，工作地点遍布世界各地，是 [BeyeNETWORK](#)、[Focus](#)、[SmartDataCollective](#) 和 [TDWI](#) 的定期撰稿人。

本白皮书提到的品牌和产品名称均为 IBM 的商标或注册商标。

---

<sup>1</sup> Haeckel, S.H., 《The Adaptive Enterprise; Creating and Leading Sense-and-Respond Organizations》（适应性企业：创建和领导感知响应型组织），哈佛商业出版社（1999）

<sup>2</sup> 《Outperforming in a data-rich, hyper-connected world》（在一个数据丰富的超连接世界里脱颖而出），IBM Center for Applied Insights, YTE03002-USEN-01, （2012 年 3 月），<http://ibm.co/QZJ8XL>

<sup>3</sup> Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., Haydock, M., 《Analytics: The Widening Divide》（分析：日益扩大的鸿沟），（2011 年），<http://mitsmr.com/DataReport1>

<sup>4</sup> Jean-Baptiste Alphonse Karr, 著于其杂志《Les Guêpes》（《The Wasps》）1849 年 1 月刊中。

<sup>5</sup> 基于 2007 年与来自 IBM 的 Mike Kehoe 的私人谈话

<sup>6</sup> Devlin, B., 《Integrating Analytics into the Operational Fabric of Your Business》（将分析整合到贵公司的运营架构中），<http://bit.ly/IDAA-WP>

<sup>7</sup> Davenport T. H. 和 Harris J. G., 《Competing on Analytics: The New Science of Winning》（分析法层面之竞争：获胜的新科学），哈佛大商学院出版社，（2007 年）

<sup>8</sup> 同上。

<sup>9</sup> Taylor, J., 《Where to Begin with Predictive Analytics》（预测分析从哪里开始），<http://bit.ly/yr333L>, 2011 年 9 月 1 日，摘自 2012 年 2 月 8 日

<sup>10</sup> 《The State of Broadband 2012》（2012 年宽带发展现状），联合国宽带数字发展委员会，（2012 年），<http://www.broadbandcommission.org/Documents/bb-annualreport2012.pdf>

<sup>11</sup> Kelly, K., 《How technology evolves》（技术如何演进），TED Talk（2005 年 2 月）

<sup>12</sup> Devlin, B., 《Data warehouse—From Architecture to Implementation》（数据仓库——从架构到实施），Addison-Wesley, （1997 年）

<sup>13</sup> Devlin, B., 《The Big Data Zoo—Taming the Beasts: The need for an integrated platform for enterprise information》（大数据动物园——驯服野兽：对企业信息集成平台的需要），2012 年 10 月，[http://bit.ly/Big\\_Data\\_Zoo](http://bit.ly/Big_Data_Zoo)

<sup>14</sup> Handy, C., 在《Learning Organizations》（学习组织）中，作者：ed. Chawla, S. 和 Renesch, J., Productivity Press, （1995 年）

<sup>15</sup> Devlin, B. A. 和 Murphy, P. T., 《An architecture for a business and information system》（业务与信息系统架构），IBM Systems Journal, 第 27 卷第 1 册，第 60 页（1988 年）<http://bit.ly/EBIS1988>

<sup>16</sup> Nietzsche, F., 《Human, All Too Human》（颇具人性化），剑桥大学出版社，（1878 年）