珠联璧合：

当大数据联姻数据仓库后

Jon Deng
IBM软件集团全球信息管理InfoSphere资深产品经理

您是否被淹没在大数据集中
（从TB 级到 PB 级？）

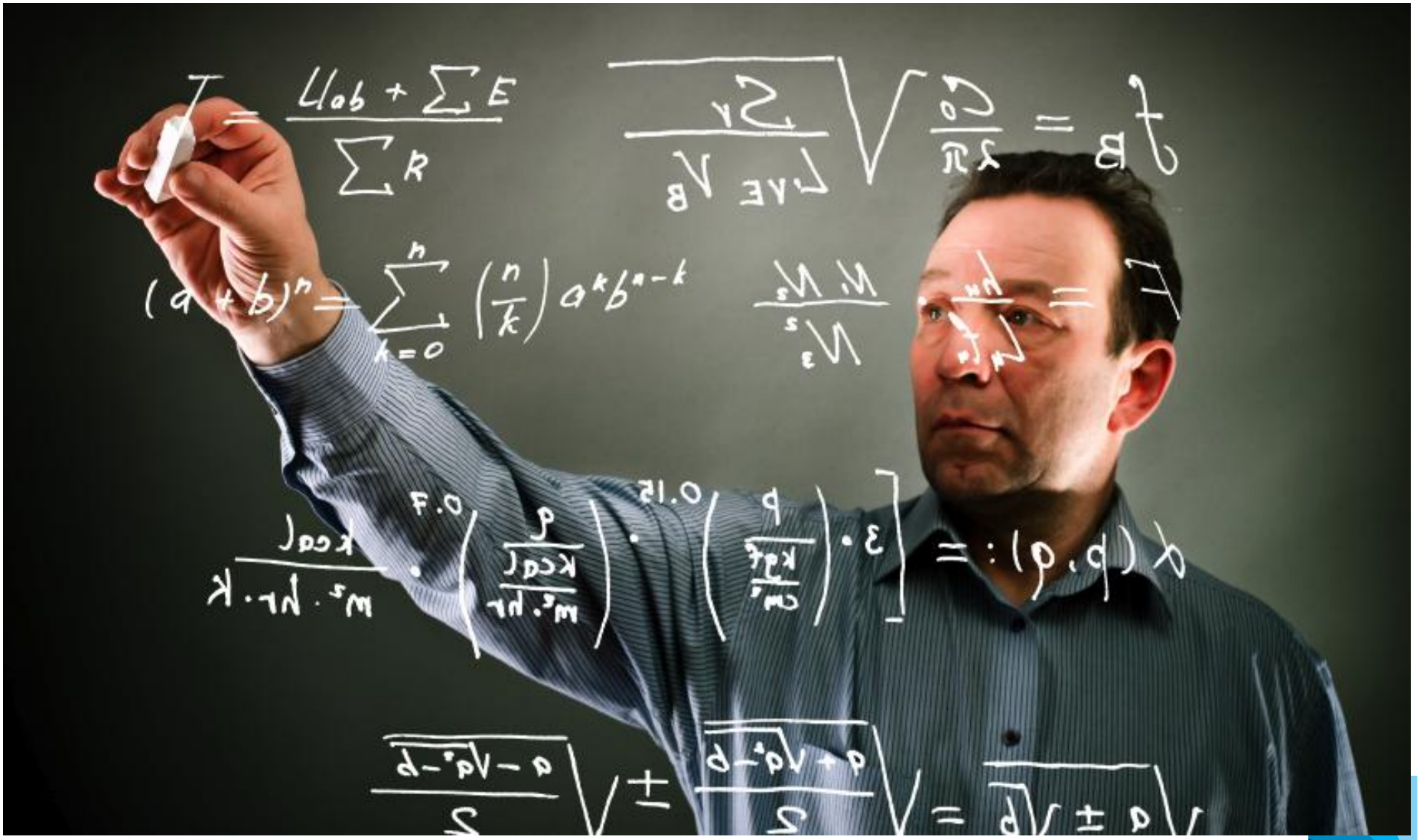
您是否用您的数据仓库环境作为
所有数据的存储库？

您是否有很多冷数据，或很少
被接触的数据？

您是否因为不能处理数据而扔掉它？

如果您能够利用自己的所有数据...

…执行复杂分析和预测分析



…方法就是利用大数据技术增强数据仓库

Hadoop
技术

实时
流

可视化和发现

优化的
设备

您的 数据仓库

目前一些组织使用这种技术的方式

从曾经无法得到利用的数据中**发现并可视化欺诈模式**、账户关闭、活动模式

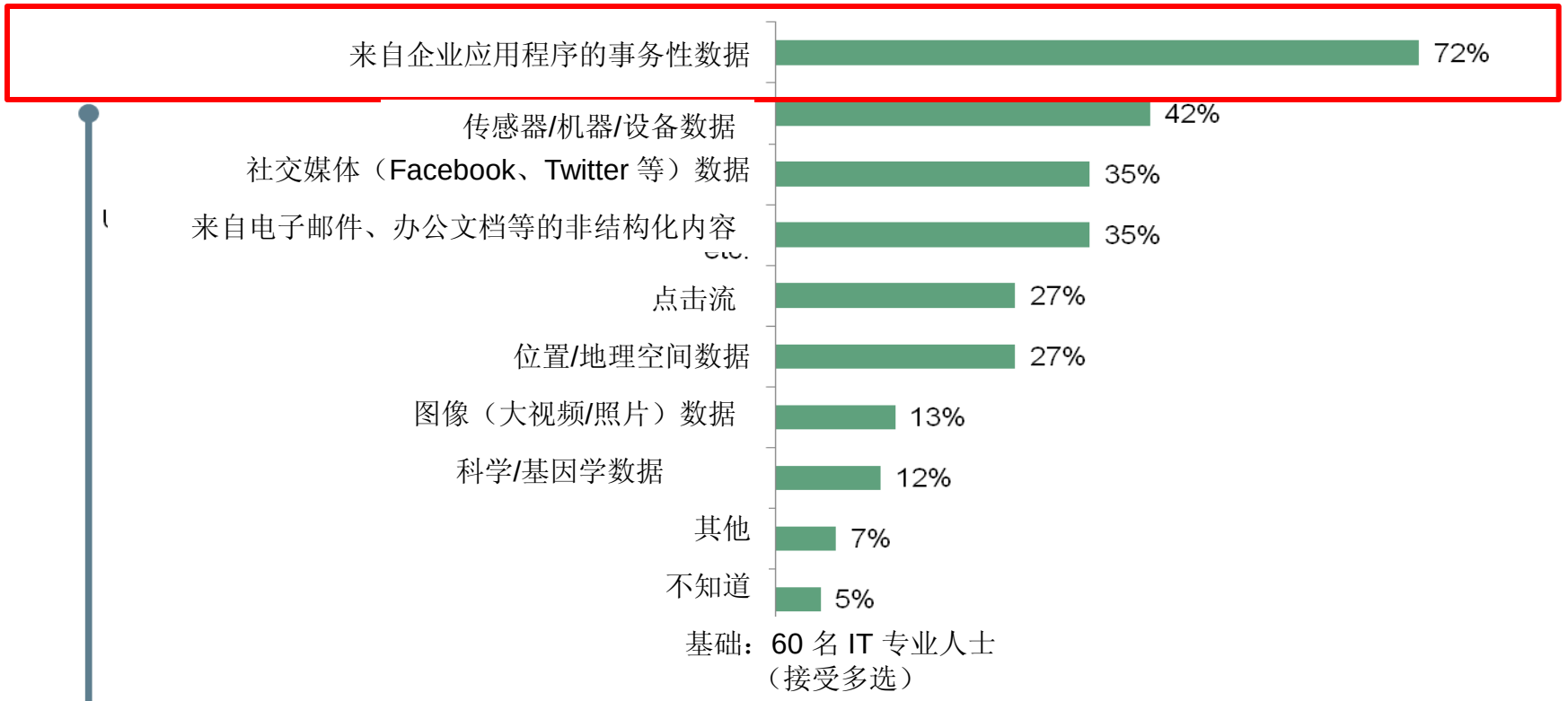
将数据分析的时间范围从 30 天增加至几年 - 支持**更准确的决策制定**

使用商用硬件将较冷的数据卸载到 Hadoop，**降低服务成本并提高服务质量**

通过分析含操作数据的社交媒体收集**有关客户个人的更多信息**

要求分析多个数据源

“您计划用大数据技术分析什么类型的数据/记录？”



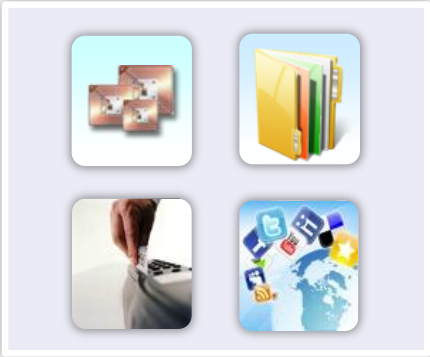
大部分大数据用例都大肆宣传其对来自社交媒体、传感器及 Web 流量的新型原始数据的分析应用，但我们发现企业很讲究实效，早期采用者使用它来分析已拥有的企业数据。

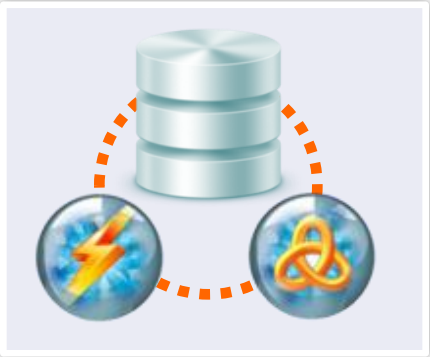
on enterprise data they already have.

资料来源：Forrester Research, 2011 年 6 月，全球大数据网络调查

概括而言，我们在考虑做什么？

集成大数据和数据仓库功能，以提高运营效率





需要充分利用各种数据

- 需要结构化、非结构化和流式传输的数据源，进行深度分析
- 低时延要求
  （几个小时，而不是几周或几个月）
- 需要对数据的查询访问权限

扩展数据仓库基础架构

- 将极少使用的数据迁移到 Hadoop，从而优化存储、维护和许可成本
- 通过对流数据进行智能处理，减少存储成本
- 通过确定馈送什么数据给仓库，提高仓库的性能

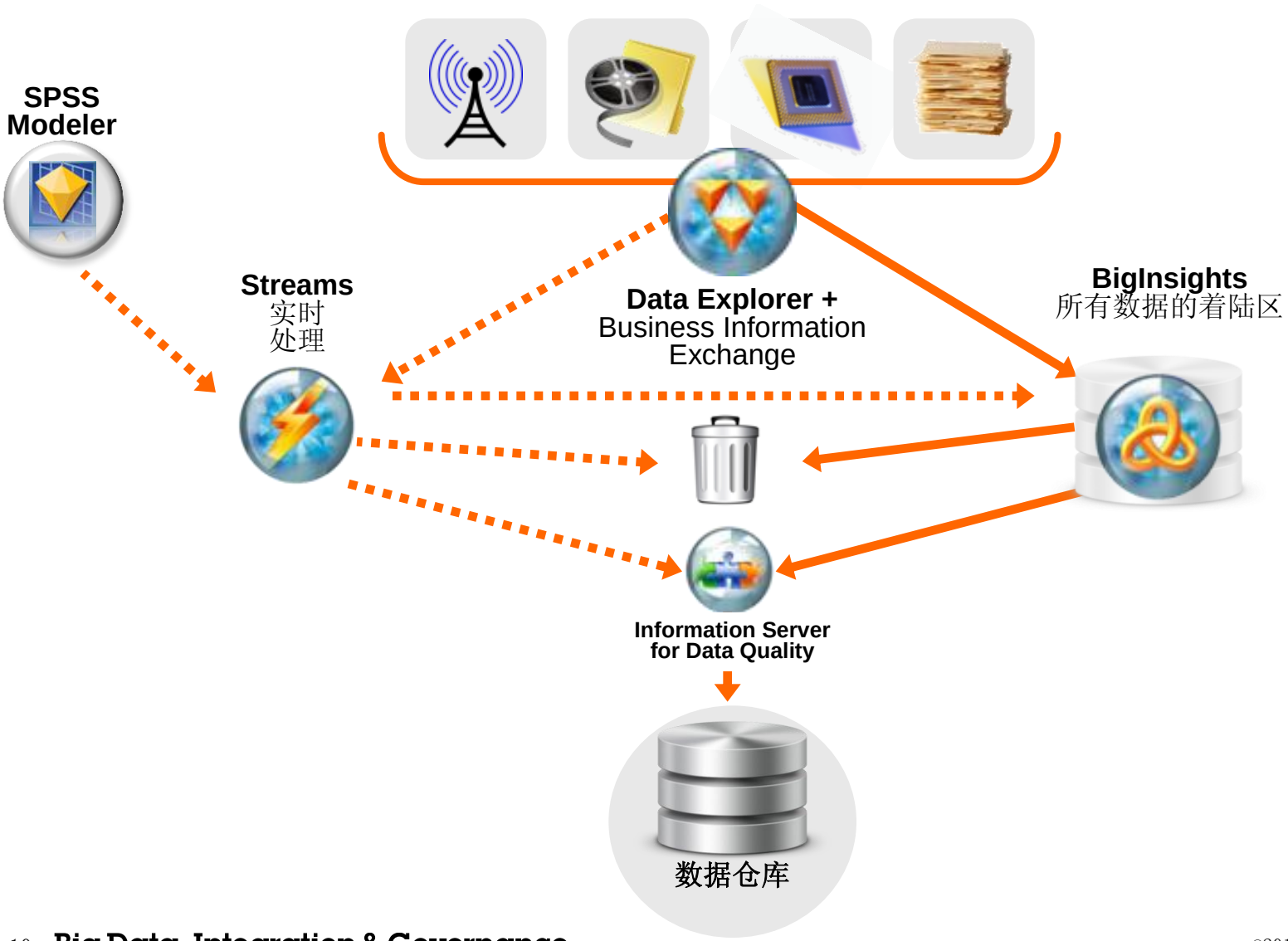
增强数据仓库的 3 种方式

❶ 预处理中心

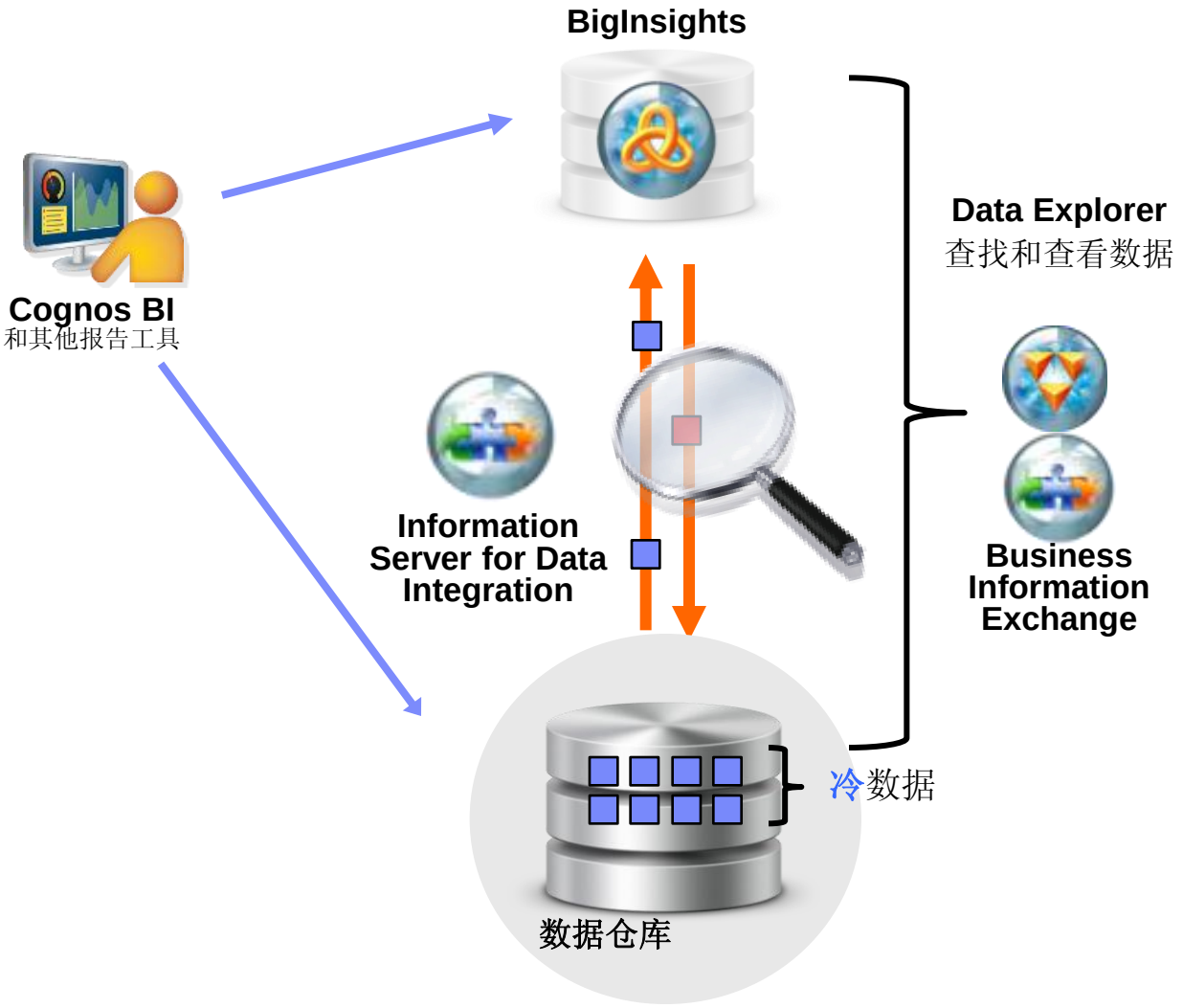
❷ 可查询的归档

❸ 探索性分析

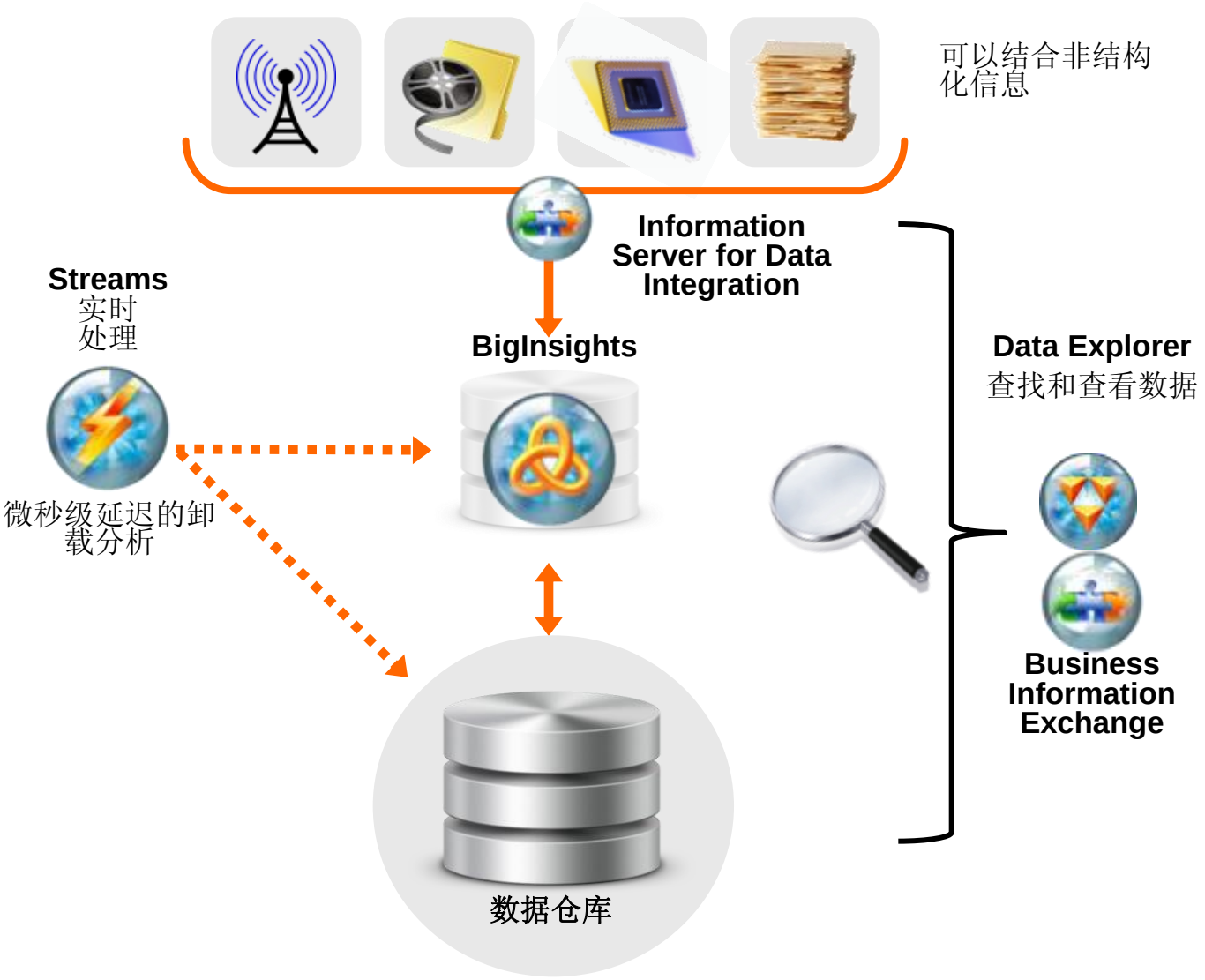
使用方法＃1 - 预处理中心



使用方法＃2 - 可查询的归档



使用方法#3 - 探索性分析



您甚至可以放弃进行“增强”，并将数据卸载到一个设备，以提高性能

针对数据服务优化：

§ 分析

§ 运营分析

专家集成：

§ 数据平台

§ 基础架构

§ 统一平台管理

§ 预置专家经验

工作负载优化的性能

数小时内完成数据加载

集成的管理

自动维护

单点支持

Constant Contact 利用 InfoSphere BigInsights 分析数十亿非结构化电子邮件

需求

客户每年发送 350 亿封电子邮件

- 分析电子邮件效果的能力对于客户成功至关重要
- 分析趋势
- 减少发送电子邮件的时间

了解什么内容在营销活动中最有效

获益

40 倍的绩效提升

- 分析时间从数小时缩短为数秒
- 直接改善客户体验
- 营销活动绩效提高 15%-25%

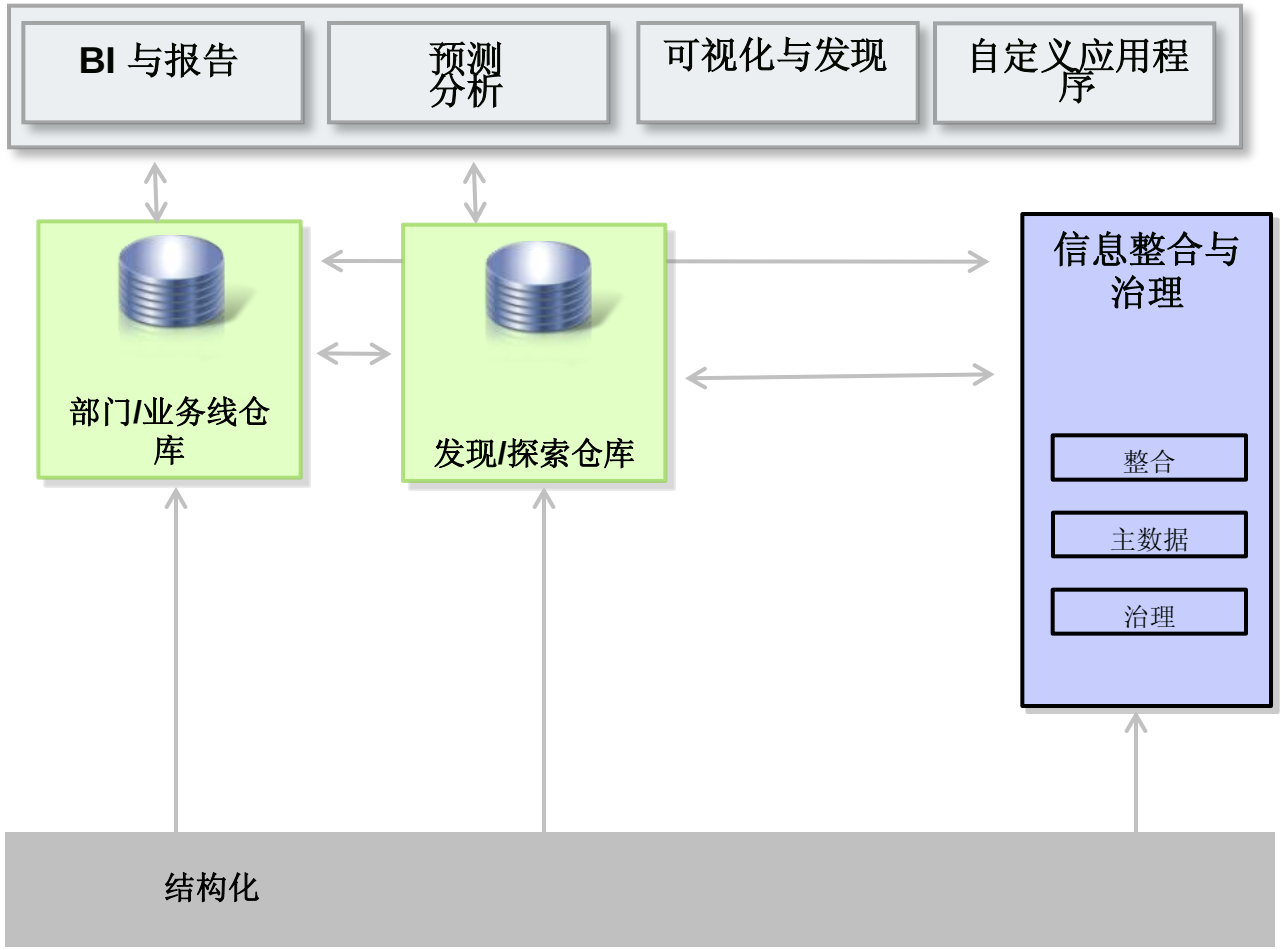
在整个组织中都易于使用

增强现有的分析和信息系统（IBM 大数据平台）

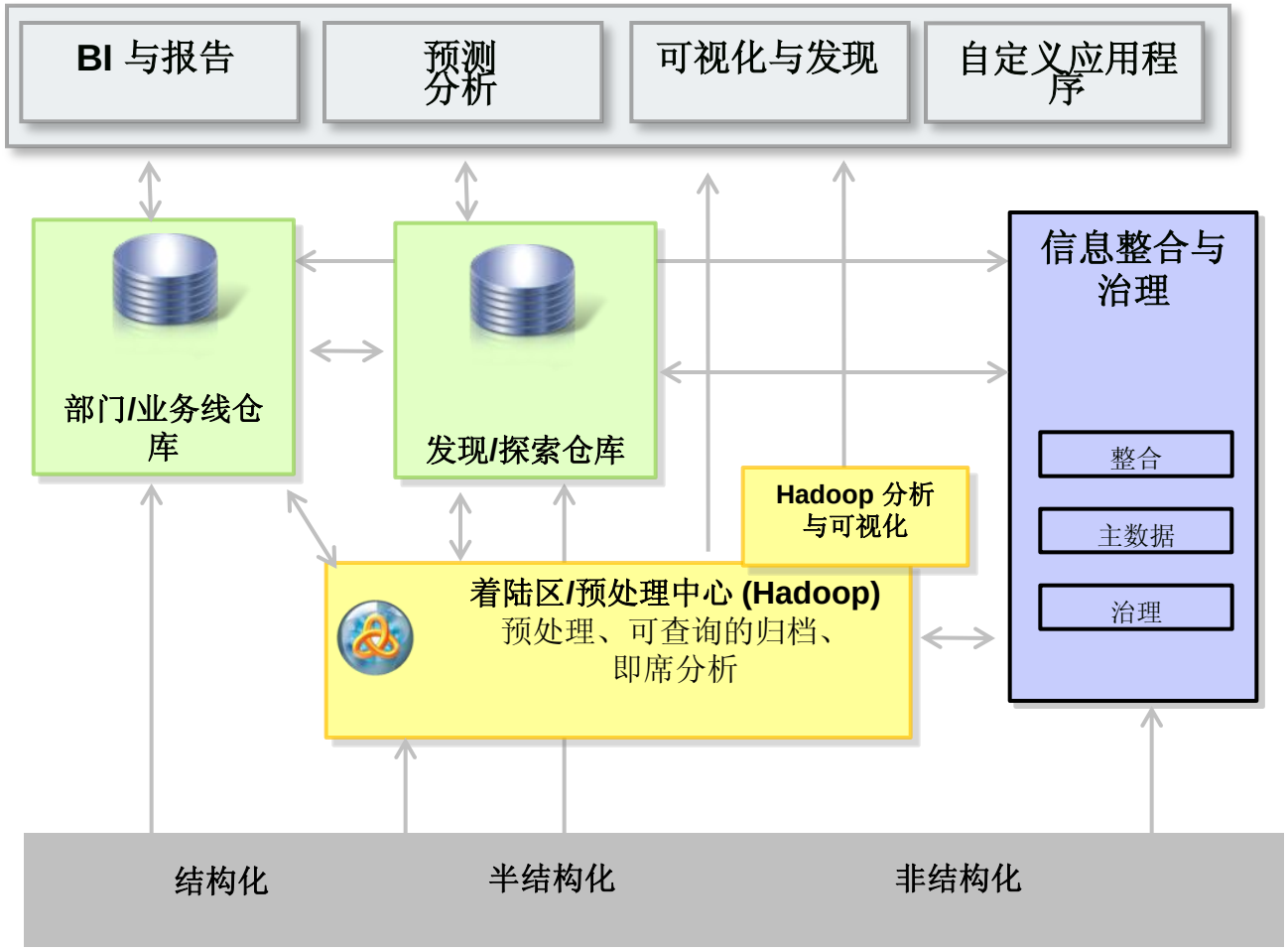
Constant Contact®

数据仓库增强实践

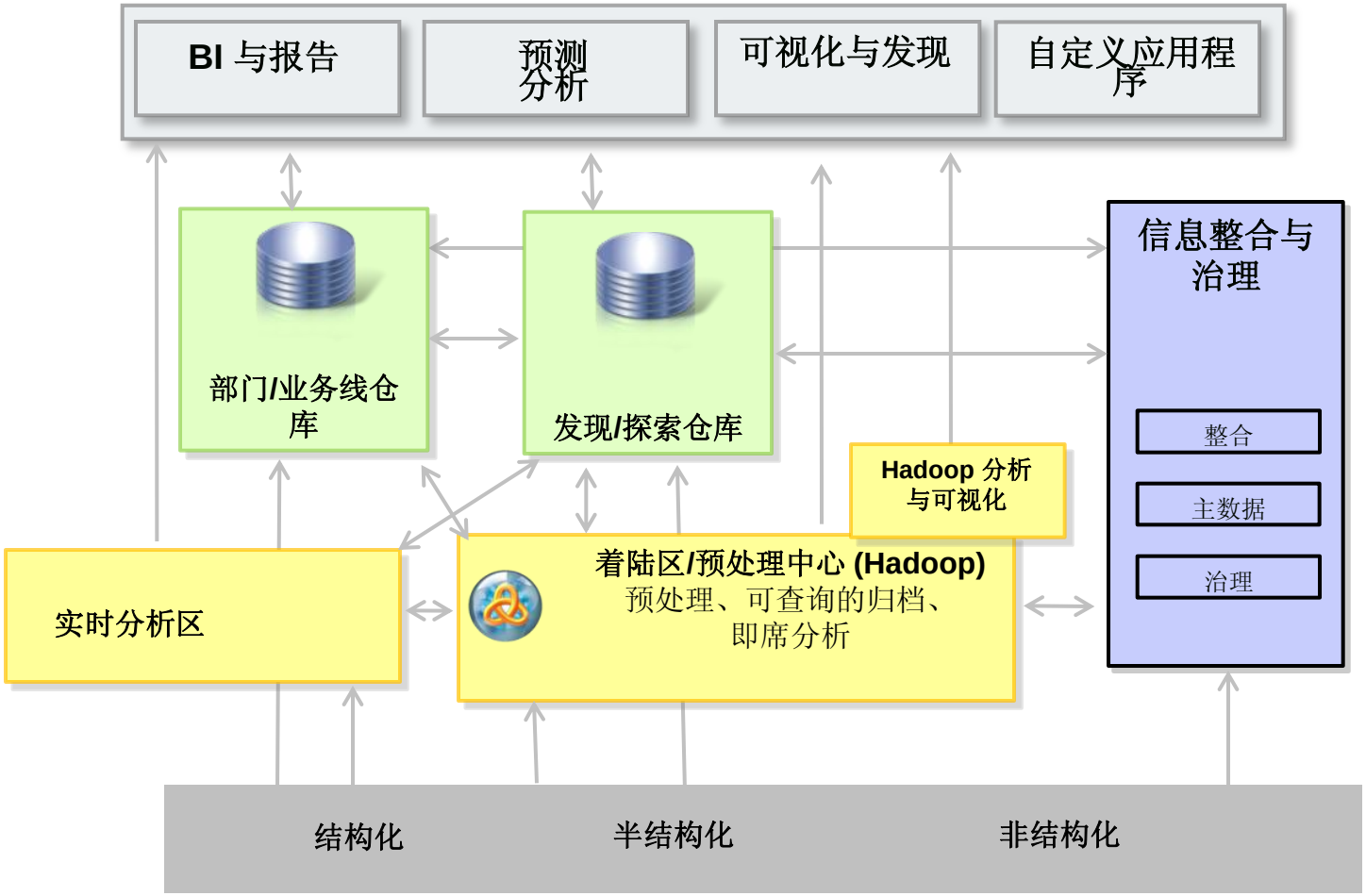
下一代的企业数据仓库架构



数据仓库增强实践
预处理中心

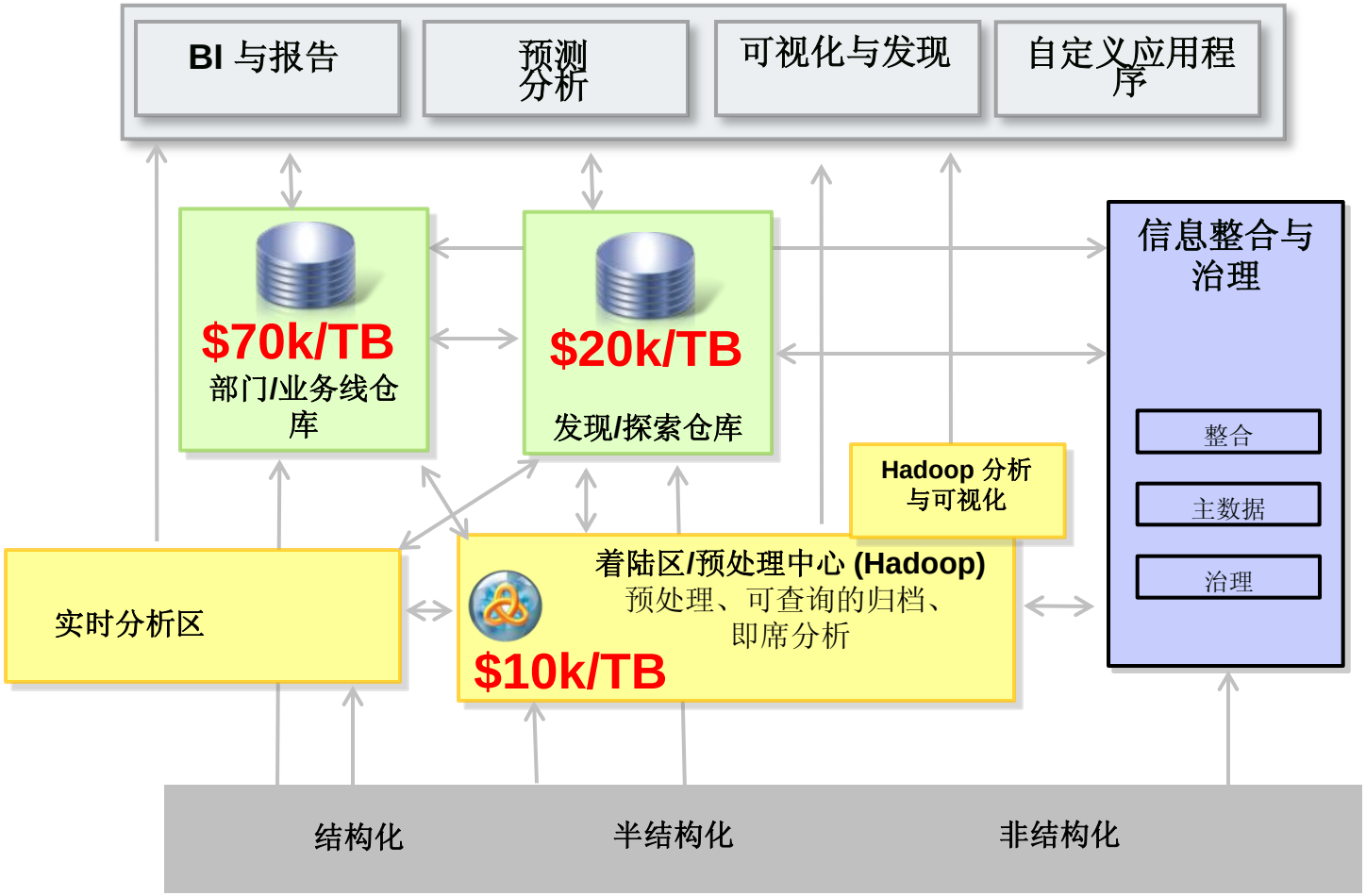


数据仓库增强实践
实时分析



数据仓库增强实践

每 TB 实售价格(*估算*)



亚洲电信公司减少开单成本并提高客户满意度

功能：

流计算
分析加速器

每天实时调解和分析 **8B CDR**

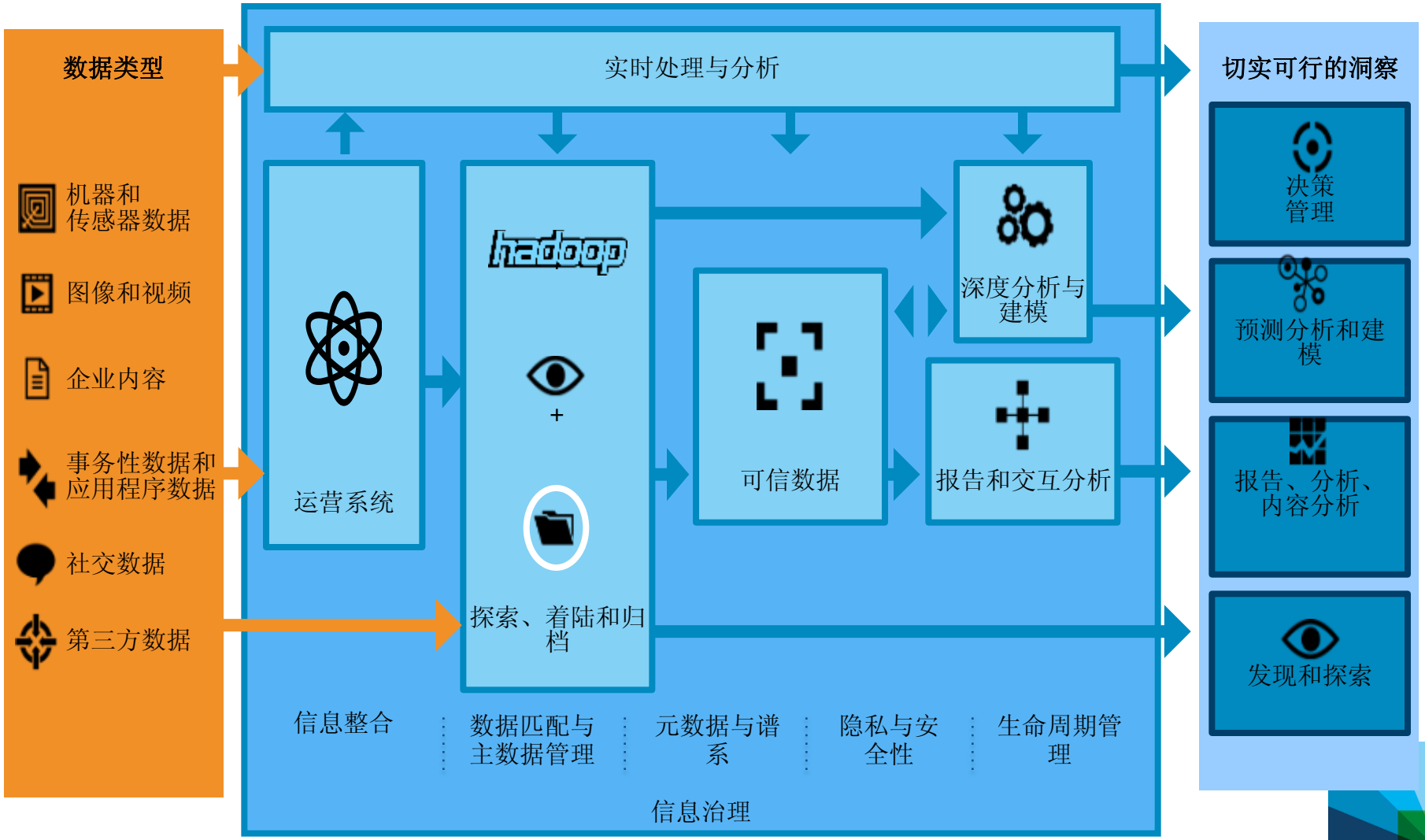
数据处理时间从
12 小时缩短为 1 分钟

硬件成本降低至 1/8

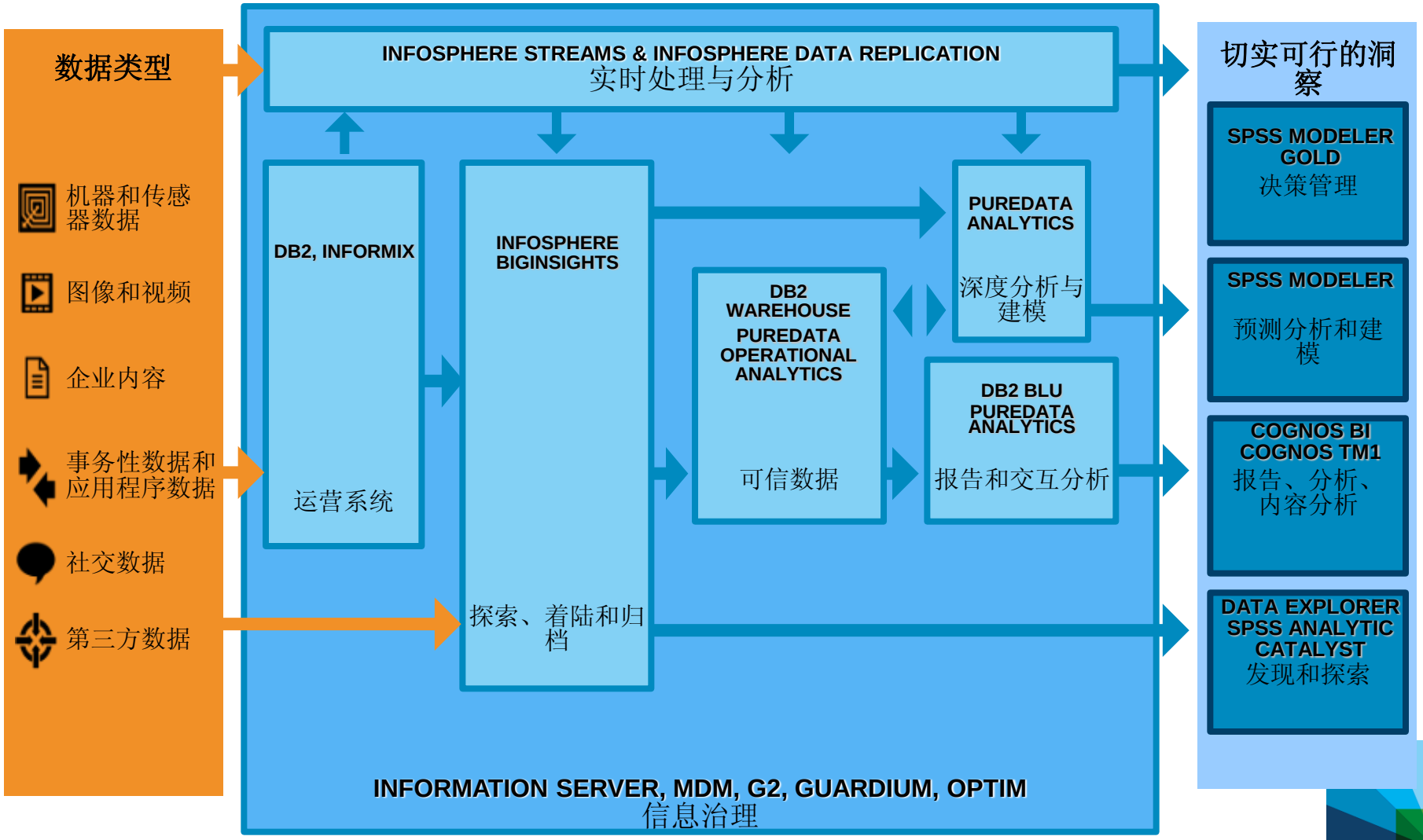
主动解决影响客户满意度的问题（如通话掉线）。

下一代企业仓库环境

将转换、匹配、安全和治理能力扩展到所有数据



IBM 大数据与分析产品



综上所述，如何开始增强数据仓库？

接受必要的培训

- 论坛内容
- IBMBigDataHub.com
- 大数据大学
- 有关大数据的 IBV 研究
- 书籍/分析师论文

预约一个大数据研讨会

- 免费
- 最佳实践
- 行业用例
- 业务使用
- 商业价值评估

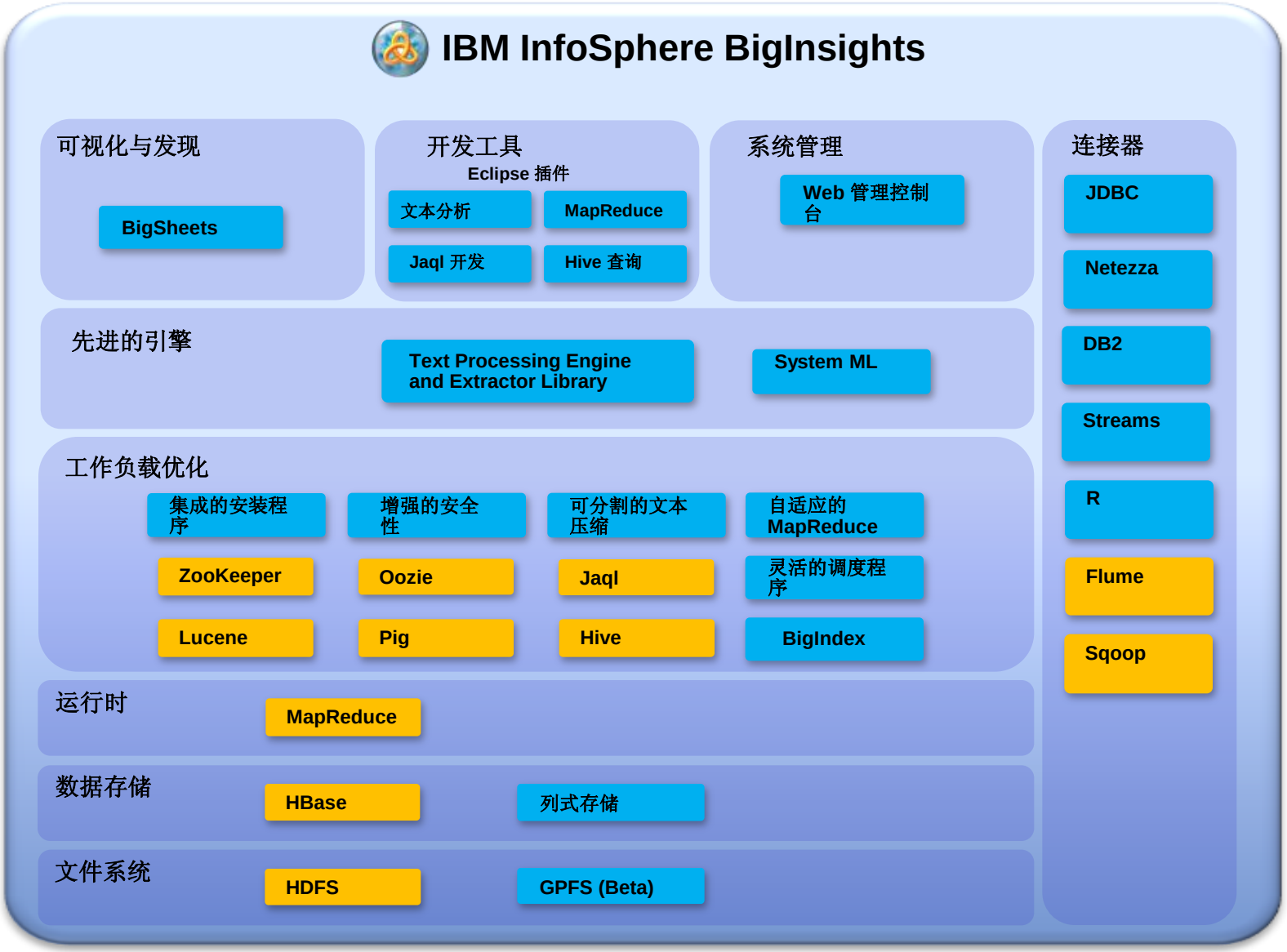

思考
BIG

BigInsights 构建于开源 Hadoop 功能之上，支持企业级部署

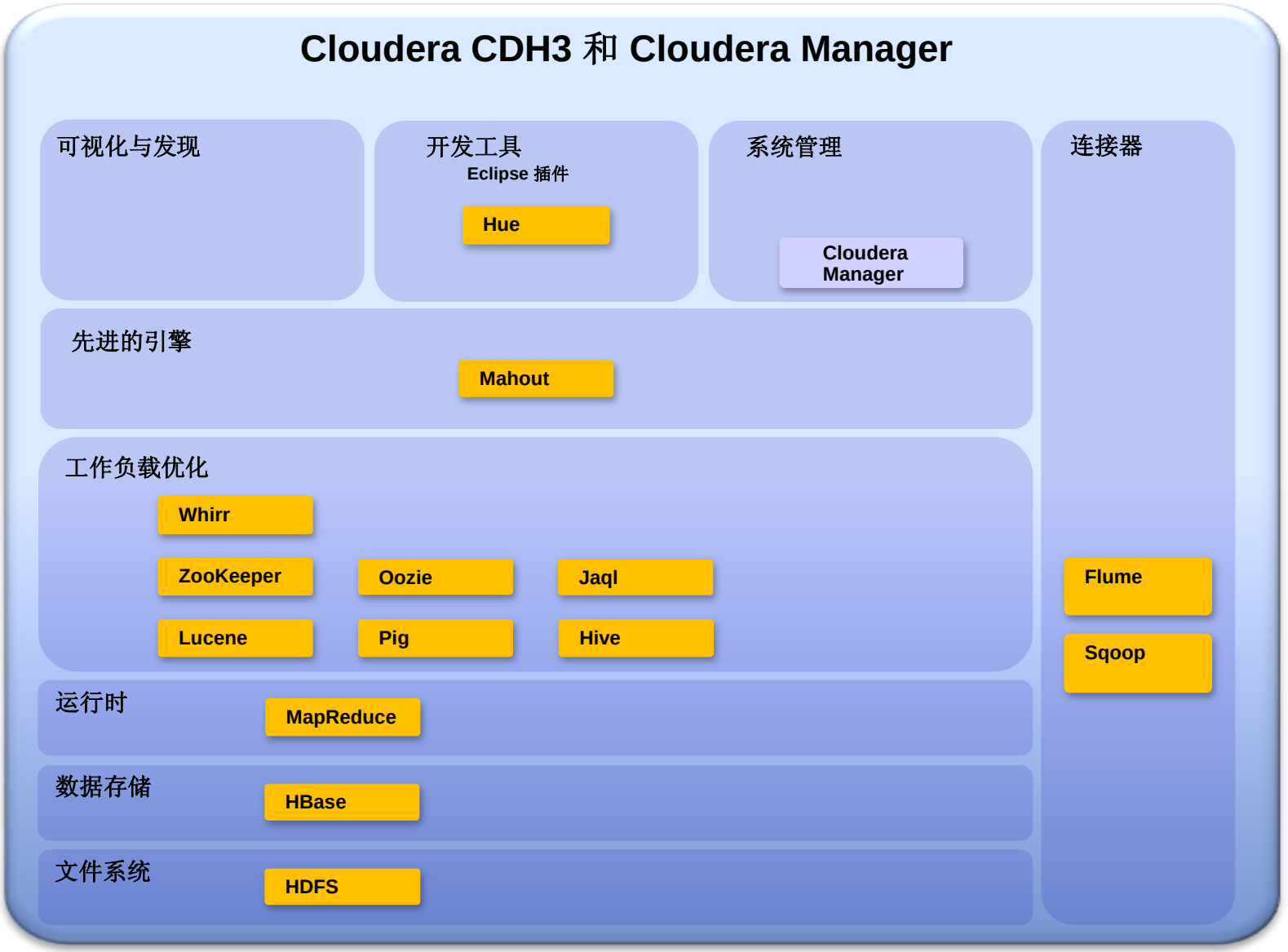
企业功能

基于开源的组件

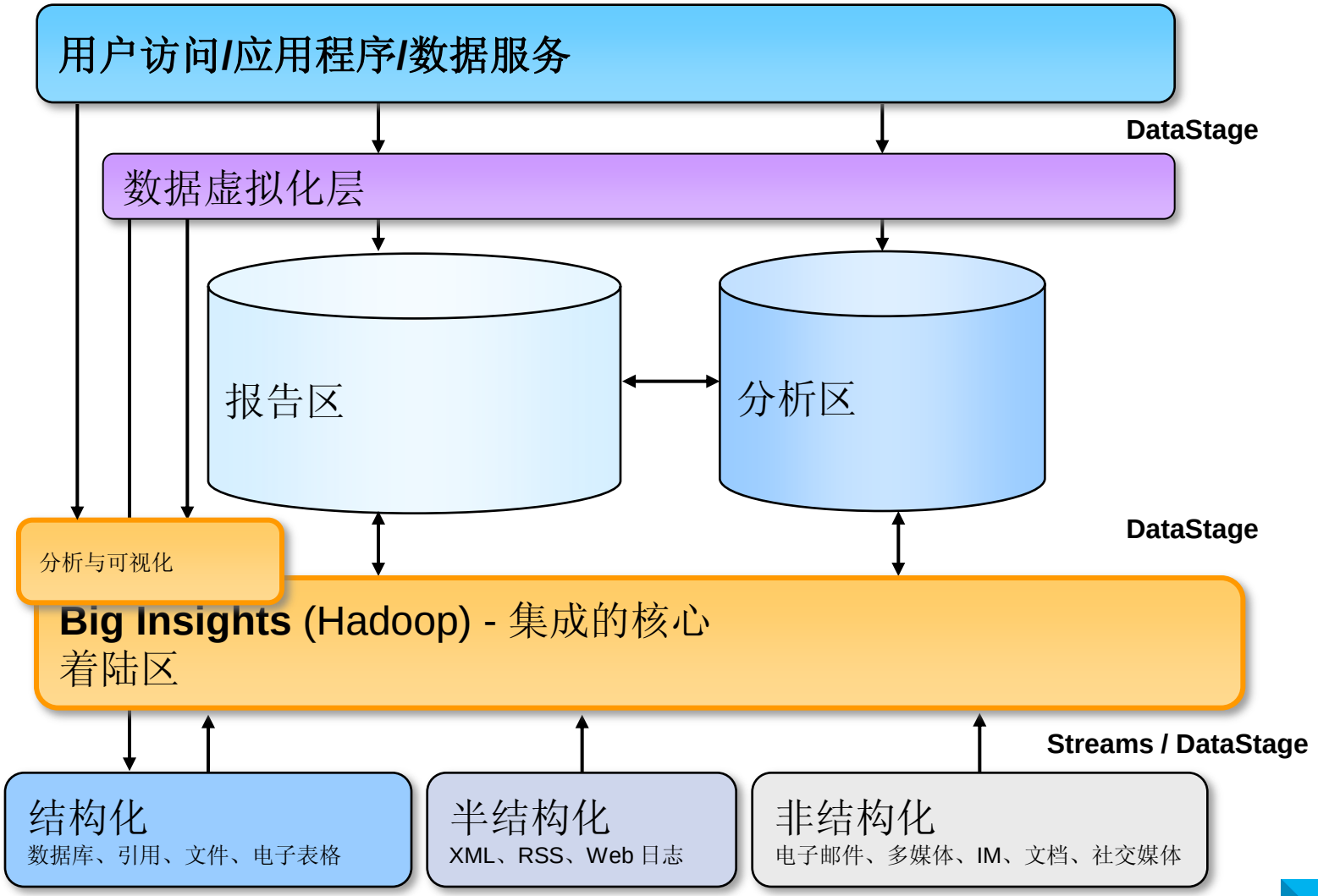
BigInsights Enterprise Edition 组件



Cloudera 组件



EDW 架构，含基于 Hadoop 的着陆区和集成的核心



IBM 主数据管理