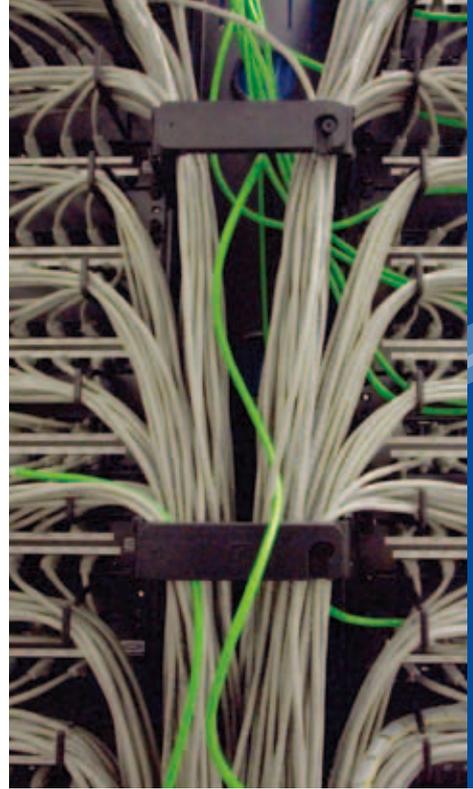


IBM BIG DATA PLATFORM SOLUTION

IBM 大数据平台解决方案



行业需求概述

大数据时代已来临。随着移动互联网的发展、智能手机的普及，全球有接近1亿的3G用户。电信运营商面临着从传统语音业务向数据业务转变、存储的数据数量急剧增长挑战。这些数据来自于网络上的大量实时信令和CDRs数据、互联网的海量内容数据和多媒体数据，传统数据库技术在处理海量数据——尤其是非结构内容数据时存在着瓶颈，在扩展性、运营和维护成本方面均面临着挑战。

解决方案概述

IBM一直是大数据技术的领导者和实践者，结合全球各大实验室处理大数据的实际经验，IBM推出了具有企业级高性能海量数据存储、强大文本分析与挖掘、丰富集成能力、易于使用与维护的大

数据平台——IBM Big Data Platform，它是一个面向云计算环境将分布式海量数据存储分析的流计算实时处理技术与传统数据仓库技术完美结合的、实时处理海量结构化与非结构化数据的平台。



- 高效处理各种海量静态、动态数据，包括：结构化、半结构化、非结构化数据
- 满足不同需求的专业处理引擎
- 实时数据分析
- 数据治理
- 企业数据集成

解决方案价值主张

IBM大数据平台Big Data Platform中的Infosphere Big Insights可提供Internet级别的海量存储和企业级的高可用高可靠性，是一个被证实的高性能、可扩展的海量数据分布存储的方案，解决了海量数据(半结构化和非结构化)的存储和访问问题，Big

Insights底层存储GPFS-SNC基于GPFS发展而来，其借鉴了HDFS的一些设计理念，与HDFS相比，GPFS-SNC在性能、可靠性方面具有巨大优势，消除了HDFS单点故障问题。



	BigInsights GPFS-SNC	开源HDFS或其他方案
健壮性	无单点故障99.99%	NameNode存在单点故障
数据一致性	高	数据可能会丢失
可扩展性	数千节点, 实测4000+	数千节点
POSIX兼容	完全兼容	有限
数据管理能力	安全、备份、快照、缓存、广域网复制	有限
传统应用能力	好, 兼顾读写性能	随机读写性能差
安全性	支持ACL, 容量限制, 安全认证	不支持

Big Insights提供开放性的接口和集成能力, Big Insights可以与IBM数据分析软件深度集成, 可以提供更为强大的分析能力, 不存在技术壁垒, 方便用户自定义开发和客户化, 该平台的高度的集成性还体现在可与数据仓库中运行的第三方分析模型、分析应用和分析模型进行端对端集成, 避免海量数据的加载等。这些开放的能力可以继续利用客户现有的分析平台的投资, 降低整体的拥有成本。包括企业级别的数据仓库集成能力(Netezza, DB2, InfoSphere Warehouse), Big Insights具有企业级管理、易用性和强大的数据分析能力, 诸如: System T文本分析、具有强大的扩展能力的JAQL语言、统计分析平台project R以及机器学习systemML、可视化工具Big Sheet等。

- IBM大数据平台Big Data Platform中的InfoSphere Streams可基于PC服务器构建流计算处理集群。



IBM BIG DATA PLATFORM SOLUTION

IBM 大数据平台解决方案

InfoSphere Streams支持一种新的信息处理模式，提供了许多独特的功能，包括：

- 以亚毫秒级的响应时间持续分析流数据
- 扩展现有的应用，而无需重新启动运行时或应用
- 将现有的预测模型标记语言(PMML)模型重用到流数据上
- 支持InfiniBand
- 与Internet Protocol version 6 (IPV6)网络通信
- 强有力地支持开发人员和管理员生产力
- 包含关系、实用程序和适配器操作符的标准工具包
- 拥有持久化和扩充信息流的数据库工具包
- 拥有内置函数、输入和分析以及示例应用的财务工具包

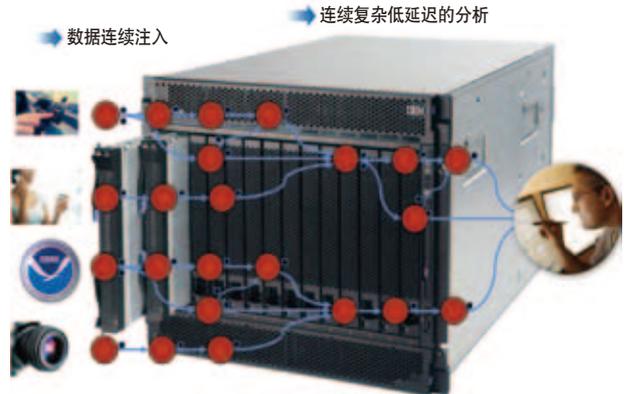
IBM软件产品

- IBM InfoSphere BigInsights Enterprise Edition
- IBM InfoSphere Streams Enterprise Edition

典型业务场景

大数据存在于电信运营的各个环节，例如：通过打通传统BSS域与OSS域的壁垒，采用流计算技术对O域中大量网络信令进行实时处理和分析的，将分析的结果与B域中的客户信息相关联，当网络出现故障或拥塞时，快速定位受影响的高端客户群体，进

流计算演示



而通过呼叫中心的外呼或短信平台及时与客户沟通交流，提升客户满意度、防止客户流失。

运营商都在关注智能管道的建设和流量经营，在实际的运营中，数据业务中大量非价值业务占据了60%以上的流量总带宽，通过IBM大数据平台，可以帮助运营商在存储的海量数据中，通过分布式数据处理和流计算技术快速实现“用户可识别”和“业务可区分”，做到根据用户使用业务的种类制定资费套餐，进而，通过对用户行为的实时跟踪与分析，培养客户的使用习惯、增强客户粘性。

此外，互联网用户行为分析、基于内容的舆情分析也都是大数据的典型应用场景，以下是通过IBM大数据平台所提供的能力，体现的对运营商大数据的应用场景的描述：

运营商大数据应用场景



成功案例

IBM大数据平台帮助亚洲电信运营商减少计费成本，提高客户满意度。



亚洲电信公司减少计费成本并提高客户满意度

利用的功能

- 流计算
- 分析加速器

成果

- 实时调解和分析每天50亿元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
- 主动解决可影响顾客满意度的问题（如掉话）

IBM大数据平台帮助美国大型无线电信公司增加收入，提高客户满意度。



美国大型无线电信公司增加收入并提高客户满意度

利用的功能

- Hadoop 系统
- 数据仓库分析设备
- Streams
- IBM Netezza

成果

- 合并/加载通话记录数据的时间减少90%以上
- 存储减少90%以上
- 提高了网络质量，提高了客户满意度，降低了用户流失率

“合并/加载时间和存储需求减少了90+”

© 版权所有IBM Corporation 2013

IBM、IBM徽标、ibm.com是国际商业机器公司在美国和/或其他国家或地区的商标或注册商标。如果上述和其他IBM商标在本文中初次出现时带有商标符号(®或™)，则表示在此信息发布时，这些商标是IBM拥有的、在美国的注册商标或普通法商标。此类商标在其他国家/地区也可能是注册商标或普通法规定的商标。可在网络上获取IBM商标的最新列表，请查看ibm.com/legal/copytrade.shtml的“Copyright and trademark information”部分。未经IBM公司书面许可，不得以任何方式复制或传播本文档的任何部分。

到发布之日止，产品数据都进行了准确性审核。产品数据可能随时更改，恕不通知。关于IBM未来方向或打算的声明仅代表IBM的发展目标，如有变更，恕不另行通知。IBM“按原样”提供本出版物，不进行任何明示或暗示的保证，包括推销期间或出于某种目的而做出的任何暗示的保证。一些法律法规不允许在不预先通知的情况下在某些交易中表达或暗示质量免责声明。

本文中针对IBM和非IBM产品及服务的性能数据是在特定的操作和环境条件下得出的。由任何该产品或服务的执行方获得的实际成果取决于大量特定于该方操作环境的因素并可能有很大差异。IBM不保证此类产品或服务的任何实现能够获得或包含此类成果。本文档中包含的有关第三方的任何材料基于从该方获得的信息，并没有独立验证信息的精确性。本文档不等于来自IBM对任何第三方产品或服务的明示或暗示的建议或认可。

客户应自行保证遵守法律法规要求。获取有能力的法律顾问关于确定和解释任何可能影响客户的业务的相关法律和法规要求，以及读者为遵守法律可能必须采取的任何措施的建议是客户自己的责任。IBM不提供法律建议，也不表示或保证其服务或产品将确保客户遵从任何法律或规定。

