



星环助力中泰证券大数据创新应用



中泰证券是经中国证监会批准设立的大型综合类创新试点证券公司，集证券、基金、期货、直投为一体的综合性证券控股集团。多年来，公司积极致力于为广大投资者提供证券代理买卖、投资咨询、财务顾问、证券发行与承销、收购兼并、资产重组、资产管理、融资融券、证券投资基金代销、股指期货中间介绍、向保险机构投资者提供综合服务 etc 全方位的专业化证券投、融资服务。



凭借良好的专业能力和业绩表现，中泰证券得到了社会各界的广泛认可。2007年以来，公司多次被山东省政府授予“山东省金融创新奖”、“山东省金融发展贡献奖”；公司先后荣获“金钥匙奖”、“中国最具成长性证券经纪商”、“中国最佳证券经纪商”、“最具成长性投行”、“最佳套利服务券商”、“最具创新力证券公司”、“优秀保荐机构”、“金牛投行进步奖”等称号；在2011年中国证监会证券行业分类评价中，公司成为16家A类AA级券商之一。

而近年来，随着业务的发展与规模的扩张，中泰证券数据呈指数级增长，各类IT系统数据量已经达到20多TB的规模。每日还有大量新增的日志数据、交易数据需要存储和处理。中泰证券的IT系统也面临着一些问题的困扰。

首先是数据存储量庞大。现有系统存储了5年的数据，总共20多TB，日志数据超过一半，使用分区存储方式，历史数据采用离线存储方式，存储资源紧缺，存储扩展花费非常高。

其次是现有系统计算负载高、延迟长。现有系统在运行中跨历史范围查询延迟长，一次计算的数据量大，计算和存储资源都存在瓶颈；大范围查询时，对生产业务影响较大，例如：持仓分析、对账流水情况、区间查询等应用，严重影响日常业务的正常运行。

同时，历史数据服务请求带来额外工作负担。历史数据查询时需要额外将离线的历史数据导入，再等到系统资源空闲时进行查询，不仅效率低，而且工作负担繁重，也极易出错。

最后，现有系统的资源已经严重紧缺，CPU负载高、存储空间不足，已经影响到业务的正常发展。

由此背景，本着不断发展、不断进步的原则，中泰证券力求解决现有IT系统的这四大问题，并据此做了一些解决方案的调研。希望能够建设具有高效数据处理能力，同时兼具性价比的新平台。

调研结果发现，在解决分布式存储、计算问题上，Hadoop技术近年来得到了广泛的应用。Hadoop是一个由Apache基金会所开发的分布式系统基础架构。它是针对大规模分布式数据而开发的软件框架，目前已经成为企业管理大数据的基础支撑技术，是解决企业数据中心大数据存储、大规模数据计算、快速数据分析的优秀基础数据平台。

现在物联网、多媒体、移动互联时代，Hadoop最适合用于解决由于数据爆发式增长所带来的传统架构下存储和计算瓶颈的问题。

Hadoop技术发展迅速，同时也有众多国内、外厂商为企业客户提供了的成熟度一站式的大数据平台产品，个别技术领先的厂商也已经将Hadoop平台用于了金融客户的数据仓库系统，经调研，恒丰银行、民生银行、平安银行等金融机构，就已经在用Hadoop技术解决现有数据仓库存储、计算、分析等问题。

从行业内成功经验及案例来看，Hadoop平台已成为现有数据仓库平台的必要补充，主要体现在以下五个方面。

首先是数据类型支持，Hadoop平台扩展了企业数据平台对数据处理类型的支持。传统的数据仓库仅仅能够处理结构化数据，而对半结构化、非结构化数据的处理，只能依赖于Hadoop平台，例如：来自新闻的个股/主题新闻聚合、智能研报，情感指数，热度统计、事件研究、主题跟踪和发现等等。

其次是数据处理能力方面，Hadoop平台使得企业数据处理平台的处理能力变得更强大，它能够处理从中小数据量到大数据量的数据。能够作为传统数据仓库之前的数据统一存储和计算平台，将大规模的数据先进行清洗、计算、建模、汇总等，并将最终的精细化数据传递给数据仓库。例如恒丰银行就已经将数据仓库中复杂的、高负载的拉链表程序迁移至Hadoop平台。

同时，在业务应用的支持方面，相比于传统的数据仓库平台，大数据平台可以完成更多的体现企业特质的有价值的应用，例如：通过数据挖掘、机器学习算法和模型，结合相关数据，完成财经新闻搜索、新闻类型分析、新闻聚



类、情感分析、知识图谱等等。

技术架构上，基于x86服务器集群的Hadoop平台，通过横向扩展的方式，线性扩展存储和计算资源，避免基于传统IOE架构的下只能进行基于硬件资源的纵向扩展。从而避免计算资源的瓶颈和IO资源的瓶颈。

最后，仅花费40%左右价格，便可以达到甚至超越传统IOE架构的性能。在性价比方面也极具可行性。

经过对国内、内外各个厂商的技术调研，中泰证券选取了星环科技的TDH (Transwarp Data Hub) 大数据平台。星环科技是目前国内极少数掌握企业级大数据基础软件技术的数据库厂商。在全球大数据去IOE的大背景下，Hadoop已成为公认的传统数据库的替代品。星环产品Transwarp Data Hub (TDH) 功能比肩硅谷同行，五大组件Hyperbase, Stream, Inceptor, discovery,和TOS的性能和SQL兼容性在业界处于领先水平。

测试结果显示，针对海量数据的统计型业务需求，星环科技的产品TDH Inceptor能够提供快速的查询支持，查询时间在秒级或分钟级，大大提高了查询效率。

针对海量数据的多维度查询业务需求，TDH Hyperbase能提供极快的查询支持，测试数据显示，查询结果基本在10s以内，极大的提高相应业务场景的查询效率。而星环科技的TDH平台 能够提供很好的存储过程支持，以满足中泰证券的业务需求。

经过对现有成熟Hadoop产品、技术，以及金融行业相关应用的调研，星环科技的TDH平台技术领先、性能优越、功能全面、运行稳定，能够很好的解决现有平台存储、计算、分析等问题，同时能够满足中泰证券未来业务发展的需要。利用星环大数据平台，将为打造中泰证券新一代数据平台，为上层应用提供更高效、可扩展的数据服务。利用整体规划、分布实施的策略，依托星环大数据平台改造和提升现有数据仓库系统以及历史数据查询系统的性能和稳定性，同时建设财经新闻搜索、新闻类型分析、新闻聚类、情感分析、知识图谱等基于大数据的分析、挖掘应用，让数据真正发挥价值。

星环信息科技（上海）有限公司

🏠 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233 ☎ 电话：4008 079 976

🌐 网址：www.transwarp.io

TRANSWARP
DATA HUB