



星环大数据助力平安银行数据仓库全面升级

去年九月，银监会发布了39号文，中心思想要求银行信息技术“安全可控”。它要求2015年起，各银行业金融机构对安全可控信息技术的应用以不低于15%的比例逐年增加，直至2019年达到不低于75%的总体占比。虽然没有明确指出，但是“安全可控”的要求显然对国产IT产品更有利。事实上，39号文提出的指导方向，包括“优先选择愿意在核心知识和关键技术领域进行合作的机构，避免对单一产品或技术的依赖，”“有序推进整体架构自主设计、核心应用自主研发、核心知识自主掌握、关键技术自主应用”也证实了银监会的态度——鼓励银行使用国产技术，国外厂商要在银行有一席之地则必须要愿意分享核心技术。

现阶段，我国银行的IT系统高度依赖“IOE”（以IBM为代表的小型机，以Oracle为代表的数据库软件、以EMC为代表的高端存储）等国外厂商。39号文的发布让大家认为银行业的信息技术立刻就要掀起国产化的大浪，但是今年四月，银监会宣布了暂缓39号文新规的决定。这个决定并不出乎意料。首先，国外IT厂商不会对39号文新规无动于衷。事实上，包括美国商会在内的多个国外商业团体向我国政府提出了抗议，声称新规违反了WTO承诺，干涉了商业机构的采购活动。然而，对39号文新规的阻力还来自于我国银行业本身。众所周知，银行业对IT系统的可靠性和稳定性要求极高，系统故障常常意味着直接的金钱损失。所以，银行对IT产品的故障抱着“零容忍”的态度。我国的信息技术因为起步晚，确实和国外还存在着差距，尤其在操作系统和数据库这样的基础软件领域。同时，我国的银行业正在高速发展，不断地推出新产品新服务，还要应对互联网金融等各种变化，更需要稳定可靠的IT系统作为业务的保障。银行愿意使用经过长期验证的“IOE”配备情有可原。在政策上要求银行冒业务风险选用在先进性、可靠性、稳定性以及成熟度方面还和国外有差距的自主产品是不现实的，选用什么样的产品还需要银行依据业务情况自己来决定。



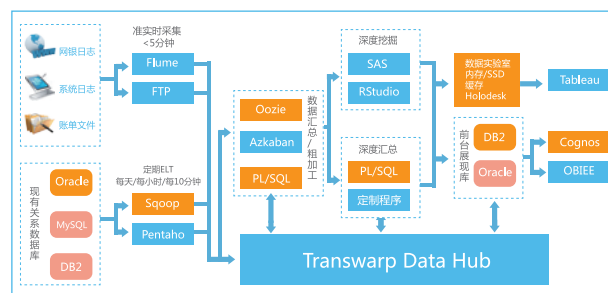
如此看来，我国的IT厂商似乎短期内无法得到银行业的青睐。然而，大数据时代的到来改变了这一格局。随着网上银行、手机银行的发展和银行新业务、新产品、新用户的增多，银行的数据量出现了爆发。数据分析的规模和复杂度都超过了传统数据库的处理能力，本应该提供坚实后盾的“IOE”配备在海量数据面前渐渐显得力不从心。一天一次的常规“跑批”（批量数据计算）正在花费越来越多的时间，面临着无法在当天完成的可能。为了解决这样的问题，银行需要对数据处理系统进行升级。以“IOE”为代表的架构采用纵向扩展（提升单机性能）的策略来升级，单机性能不可能无限提高，这样的策略有着显而易见的瓶颈。而银行的数据量每天都在增长，对系统不断的升级是必然的。可以预见的是，在高性能端，数据处理系统的升级将越来越困难，同等资金换来的性能将越来越少。在这样的情形下，银行需要采用完全不同的数据处理技术。近年来得到长足发展的大数据技术就是为了体量巨大且不断增长的数据而生的，而大数据技术就是国内IT厂商的机会。计算机技术发展时，我国远远落在国外之后，这个差距在互联网发展时已经得到了大大缩短。今天，在大数据时代的开端，我国已经有了不输国外的经济土壤、政策环境和人才储备，国外的技术不再有不可望其项背的优势。事实上，我国已经有了性能比肩甚至超越硅谷同行的大数据产品。

日前，平安银行便面临着将数据仓库业务从传统数据库迁移到大数据平台的任务。在大数据平台的选择上，平安银行进行了多方考量。平安银行的首要目标是解决现有数据库处理能力不足的问题。所以，采购的大数据平台必须有极强的数据处理能力，它需要能够轻松应对涉及超宽表（宽度在几万字节）、多张表（多达几十张）的关联和聚合，需要能够快速地完成银行每天TB级别数据量的分析。第二，平安银行原先使用Oracle数据库，所以业务逻辑由SQL和Oracle的过程语言PL/SQL写就，这也是银行的分析人员最熟悉的数据分析语言。所以，大数据平台需要提供SQL和PL/SQL的支持，以便银行将分析业务直接迁移到新的平台运行，否则，银行将需要对业务进行大量改

写，银行员工也需要重新学习和适应，迁移成本过高。第三，大数据平台需要良好的扩展性。银行的数据还在快速增长中，可预见的是，银行需要对数据处理系统进行进一步的扩容和处理能力的升级。银行要求采购的大数据平台的存储和计算能力能够方便地扩展，为未来更大的数据量做好准备。平安银行对市场上常见的大数据数据仓库产品的可用性、架构优势和平台依赖性三个方面就银行的数据仓库场景——数据量在TB级别的分析场景——进行了调研。可用性的内容包括对SQL和PL/SQL的支持以及是否有界面化的运维工具。架构优势的内容包括产品的扩展性、性能和容错性。平台依赖性描述该产品是否必须在特殊的硬件设备上运行。

调研中，平安银行发现，主流大数据数据仓库各有其优势和劣势，不依赖平台（可以部署在通用商业服务器上）的产品可用性和架构优势往往较低。而可用性高而架构优势明显的产品往往对平台依赖较高，必须使用自有的小型机或者一体机，直接导致硬件成本高昂。经过多方调研和全面测试，平安银行选择了具有综合优势的Transwarp Inceptor，它是上海星环科技开发的Transwarp Data Hub（TDH）分布式一站式大数据处理平台下的交互式内存分析引擎，具有高可用性和明显的架构优势，同时不依赖特殊平台，可以直接部署在商用服务器上。产品虽然年轻，但是在交通、物流、电信、能源等行业已经有很多落地案例，经过了大量的验证，加上在测试中稳定而出色的表现，打消了该银行对国产IT产品可靠性、稳定性方面的担忧。

星环科技为平安银行提供的解决方案如下图所示：





- 使用TDH下的交互式SQL分析引擎Transwarp Inceptor支持平安银行的数据仓库业务。Transwarp Inceptor全面支持SQL2003，且支持80%PL/SQL语法，经测试，Transwarp Inceptor 100%支持平安银行数据仓库业务，银行的数据分析任务不需要经过改写便可轻松迁移至TDH平台。
- Transwarp Inceptor采用分布式内存计算方式，将涉及海量数据的复杂任务分割成小任务交给多台机器同时处理，加快计算速度。同时，Transwarp Inceptor将计算中间结果放在内存中，利用内存的高速随机读写进一步提升计算速度。TDH平台处理银行的数据仓库业务相对于平安银行原系统速度有了十几倍的提高，使平安银行可以轻松完成每天的分析任务。
- 不同于传统数据库的纵向扩展策略，TDH采用向计算集群添加服务器的方式来横向扩展处理能力，这样的策略保证投入和处理能力的提升是线性关系——多少投入就带来多少提升。横向扩展的策略让TDH的扩展性极佳，可以无限提升存储和计算能力。

锦上添花的是TDH计算集群无须超高性能的服务器，经济的通用服务器便可以用于搭建TDH集群，使得TDH解决方案具有极高的性价比。而且TDH配备的JDBC接口完美对接上层报表系统，平安银行的数据分析人员只需按照以往的工作习惯继续使用上层应用，无需重新适应新的系统。下面是一张贷款销售分析截图：

销售日报

所属省份	城市	名单总量	首播名单量	拨打完成名单量	接触名单量	拨打完成接触名单量	剩余跟踪名单量
广东省	佛山市	131	131	21	93	16	110
广东省	佛山市	15	13	0	10	0	13
广东省	佛山市	13	13	0	70	23	56
广东省	佛山市	21	10	0	19	0	21
广东省	佛山市	10	10	0	8	0	10
广东省	佛山市	38	38	1	34	1	37
广东省	佛山市	41	41	0	37	0	41
广东省	佛山市	24	24	0	23	0	24
广东省	佛山市	57	57	0	52	0	57
广东省	佛山市	16	16	0	9	0	16
广东省	佛山市	54	54	23	43	18	31
广东省	佛山市	116	116	1	89	1	115
广东省	佛山市	20	20	0	14	0	20
广东省	佛山市	50	50	50	37	37	0
广东省	佛山市	9	9	0	5	0	9
广东省	佛山市	3	3	0	3	0	3
广东省	佛山市	6	6	6	4	4	0
广东省	佛山市	60	60	0	45	0	60
广东省	佛山市	6	76	28	76	28	48
广东省	潮州市	123	123	0	106	0	123

平安银行在业务中对一家国内厂商产品的选择无疑是对国产信息技术的鼓励，更值得一提的是这个选择并不是政策要求的，而是基于业务的需求和产品的先进性做出的。几年前，国产的个人IT产品还仅仅是“廉价低质量”的代名词。现在，联想、小米、华为等品牌不仅被国人青睐，还在世界范围内广受欢迎。这些品牌的成功不在于政策的扶持，而是真正做出了客户需要的产品。在企业级的IT产品上，国产厂商和国外厂商虽然还有一定差距，但是在技术不断革新的今天，这个差距在不断缩短。相信我国会有越来越多的企业级IT厂商做出客户需要的优秀产品，跻身世界一流的行列。

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