

THE FORELAND OF
TRADING TECHNOLOGY

内部资料 免费交流
《准印证》编号沪(K)0671

交易技术前沿

2022年 第二期 总第49期

本期主题

金融科技

No.2



内部资料 2022 年第二期（总第 49 期）

准印证号：沪（K）0671

NO.2

主管：上海证券交易所

主办：上海证券交易所

总编：邱勇、蔡建春

副总编：王泊

执行总编：唐忆

责任编辑：徐广斌、徐丹、陆伟、王昕

上海市杨高南路 388 号

邮编：200127

电话：021-68607129, 021-68607131

传真：021-68813188

投稿邮箱：ftt.editor@sse.com.cn

篇首语

近年来,金融科技蓬勃发展,新兴数字技术和传统业务相融合,在提高效率的同时降低成本和门槛,重塑现有的金融体系,提高服务能力和水平,增强综合竞争力,赋能行业数字化升级。本期《交易技术前沿》以“金融科技”为主题,收录行业金融科技及前沿技术探索的优秀文章,其中。

《关于交易所运用金融文本分析的建议》对证券监管行业的金融文本分析应用现状进行深入调研,以数据驱动业务转型的视角重新梳理了业务数字化转型中遇到的问题,提出了完善数据治理及培养数据分析专业人才的建议。

《汇聚知识,赋能行业——上交所探索构建资本市场知识库》旨在搭建统一的知识运营管理平台,融合专家构建、众包共建和自动化抽取的知识获取模式,以提升知识库内容获取的效率和准确性,形成庞大的、高质量的证券行业知识库。

《企业级证券业务中台探索与实践》通过对用户业务旅程进行分析,遵循业务引领、领域划分、统一视角及能力复用的建设原则,形成“薄应用 厚中台 稳后台”的企业架构全景,助力企业数字化转型。

《数据治理体系建设的研究与实践》介绍了郑商所在数据治理体系建设方面的研究与实践,逐步探索数据治理从管控式理念模式向服务式理念模式的转换。

《数据中台数据标准化设计与实践——上交所数据主题域顶层设计》结合上交所数据中台的实际情况,以最基础的主题数据层为例,探讨了数据模型标准化的设计思路。

《交易技术前沿》编辑部

2022年9月1日

目录 Contents

本期热点

- | | |
|---|----|
| 1 关于交易所运用金融文本分析的建议 / 顾渊智、周彬、吴贤 | 4 |
| 2 汇聚知识，赋能行业——上交所探索构建资本市场知识库 / 何浩，苑阳，马信达 | 8 |
| 3 企业级证券业务中台探索与实践 / 樊建、严伟富 | 12 |

实践探索

- | | |
|--|----|
| 4 数据治理体系建设的研究与实践 / 张如意、宫朝辉、李斌、雷晓全、徐祯、于三禄 | 27 |
| 5 数据中台数据标准化设计与实践——上交所数据主题域顶层设计 / 王芳、韩凤宁 | 37 |
| 6 核心业务系统连续性保障的探索与实践 / 张如意，刘永亮，孙维，王建超，刘相富，雷晓全，于三禄 | 45 |
| 7 构建邮件纵深安全体系，提升互联网重要入口安全运营水平——海通证券安全运营实践分享 / 沈俞超、徐文韬、秦军、郑熊、马冰、王东 | 54 |

前沿技术应用

- | | |
|---------------------------------|----|
| 8 基于 FPGA 异构加速平台的高维美式期权定价 / 喻伟 | 67 |
| 9 基于图技术的基金抱团识别方法的研究 / 武凯，刘宏，王洪涛 | 77 |

信息资讯采撷

- | | |
|----------|----|
| 监管科技全球追踪 | 86 |
|----------|----|

AF03457-DJ-JK

本期热点

- 1 关于交易所运用金融文本分析的建议
- 2 汇聚知识，赋能行业——上交所探索构建资本市场知识库
- 3 企业级证券业务中台探索与实践

关于交易所运用金融文本分析的建议

顾渊智¹、周彬²、吴贤³/¹ 上海证券交易所 信息公司 ² 上海证券交易所 办公室 ³ 上海证券交易所 数据管理部
邮箱 : yzgu@sse.com.cn, binzhou@sse.com.cn, xwu@sse.com.cn



为适应金融数字化转型和科技监管需求，在证券监管中运用金融文本分析手段提高监管水平，提升监管效率成为监管层主动转型的选择。报告对证券监管行业的金融文本分析应用现状进行深入调研，以数据驱动业务转型的视角重新梳理了业务数字化转型中遇到的问题，提出了完善数据治理及培养数据分析专业人才的建议，以期为证券行业数字化转型提供参考。

自 2016 年以来，上交所在上市审核、公司监管、债券发行、基金要素审核、舆情监控等领域探索运用金融文本分析，取得积极成效。但由于场景应用复杂、需求迭代频繁、数据质量参差及技术边界等原因，上交所运用金融文本分析支持业务监管的潜能尚未得到充分发挥。为落实相关监管指示要求，切实加强科技监管能力建设，推进科技与监管深度融合，近期上交所组织力量就核心业务运用金融文本分析现状和建设规划挑

战开展了情况调研，并结合当前交易所数字化转型趋势与要求，提出了运用金融文本分析支持业务数字化转型的相关工作建议。

一、上交所业务运用金融文本分析现状

在证监会大力倡导“提升监管数字化智能化水平”“数据让监管更加智慧”的时代背景下，

近年来交易所积极探索运用金融文本分析技术支持业务监管实践，多措并举推动科创板试点和全市场注册制改革，取得积极成效。

分业务来看，发行审核方面，为配合科创板和试点注册制改革，上交所于2020年启动了科创板智能辅助审核平台的建设，目前已完成前期建设并全面投入使用，初步实现了项目审核、公司画像系统和知识库等功能开发。该平台从审核人员的需求出发，通过模拟线下审核的业务场景，为审核人员提供企业概要信息、关系图谱、舆情搜索、参考规则等辅助工具，运用金融文本分析实现了申报审核在线检索、比对，自动化生成问询函等功能。

上市公司监管方面，上交所于2019年上线了“公司画像系统”并向相关监管机构开放使用。该平台借助金融文本分析技术，对上市公司公告、财务报表附注、招股说明书等信披申请文件中的数据信息进行提取采集，并整合包括工商、权益、舆情等外部数据信息，实现勾画企业全方位特征的多维度监管视角。“公司画像系统”旨在帮助监管人员快速全面掌握公司情况，及时发现上市公司潜在风险，辅助监管人员定期报告审核，成为科技赋能监管的有益实践。此外，上交所还采用机器采集、人工复核模式对临时公告文本进行了智能采集，首批业绩预告公告采集程序已完成开发部署，后期将扩展至其他更多主板及科创板高频披露重要监管公告。

债券监管方面，上交所运用金融文本分析结合业务实践起步较晚，目前在债券业务管理、债券招标发行等系统实现利率债公告数据自动化提取，后续将继续探索开发关键要素分析功能，进一步提升债券审核效率。

基金与衍生品监管方面，上交所已运用金融文本分析实现部分产品方案要素审核、外部数据填报的自动化处理功能，后续计划逐步推进金融文本分析在基金产品及衍生品日常统计与信息披露工作的融合应用。

二、交易所业务运用金融文本分析面临的主要问题

金融文本分析作为监管智能化的核心技术，借助人机结合的自动化处理技术在提高工作效率和防范金融风险上发挥了重要作用，但目前的文本技术水平离监管智能化目标仍有一定差距，特别是在数据质量、深度语义识别和涉及主观判断的业务场景中均存在进一步优化空间，预计短期内无法取代人力，甚至在线初期可能“加大”原有工作量，造成实际使用者额外的工作负担，以致于出现推广受阻现象。目前看，主要存在以下三方面问题：

一是场景应用复杂，业务理解专业性要求较高。监管机构的业务场景目前主要涉及招股说明书、上市公司公告等金融文档，文档理解专业性较强，且不同的审核或监管业务需求也存在较大差异。但由于监管业务人员需求不明确，加之双方沟通存在壁垒，最终产品往往不能满足业务需求，业务与技术预期成效不一致现象依然存在。

二是数据质量参差，源数据填报准确度待提升。一方面，目前我国证券期货行业文本书写缺乏明确规范，各类金融文本中往往会出现诸如财务科目名称不统一、语义含糊、名词创新、数据单位缺失等问题，文本识别难度较大；另一方面，由于在规则与技术层面未强制性约束上市公司数据录入规范，上市公司使用XBRL等信息报送工具时，往往存在不报、少报、漏报、错报或未更新修正等情形，使得数据从源头已出现差错，进一步降低了金融文本分析效果。此外，业务审核需结合大量外部数据信息进行综合判断，目前实际业务运用数据过程中仍存在数据源分散和关键信息缺失的问题。

三是技术存在边界，无法根本替代人力操作。识别准确率不高仍是当前金融文本分析存在的主要瓶颈，由于文本格式不统一、用词不规范、重命名等复杂因素，现有技术往往难以精准识别文

本含义。此外，受限于用户习惯、软件版本差异、平台兼容性问题，目前无法完全独立地依靠金融文本分析技术实现全自动化处理，仍需借助人工复审纠偏确保准确率，某种程度会削弱使用的积极性。

三、业务监管实践中运用金融文本分析的主要需求

交易所交易所各业务条线工作规划和推进方式虽各有侧重，但一线业务监管人员对于金融文本分析的应用需求整体具有高度相似性，主要集中在优化数据质量、替代重复劳动、监管线索识别等核心功能。

一是丰富数据资源，加强数据规范。金融文本分析应用的前提基础，必须具备完整的结构化数据库。当前交易所主要标准数据录入系统（XBRL系统）已实现对包括年报、季报及部分公告等信息录入，但存在数据缺失、准确率不高等问题。业务监管人员希望数据标准化采集范围扩大至临时公告、招股说明书、债券募集说明书等金融文本资料，实现数据精准结构化入库，同时对于行业参与者数据质量缺乏标准提出了关切。

二是实现自动作业，提高工作效率。面对日益增长的监管人员队伍规模，迫切需要利用金融文本分析等技术实现从纯人工操作过渡至人机结合操作，最终实现全自动化流程作业，通过机器自动化处理取代大量人工重复劳动，降低人为差错、执行统一标准、提升工作效率。应用场景包括制作日结清单、统计披露报表、复核校验数据、基金产品要素比对等。

三是开发智能模型，实施科技监管。业务监管人员还提出，希望结合现有业务监管经验，加强金融文本分析，发现规律并构建智能模型，实现科技辅助人工审核监管、实施风险预警。适用场景包括智能舆情监控、知识图谱、重大事件分析、上市发行审核、财务舞弊识别、交易风险核

查等。

四、交易所运用金融文本分析工作建议

从助力提升科技监管及提高业务数字化建设水平，推广金融文本分析应用角度出发，提出以下建议：

一是统筹系统建设，探索整合全所金融文本分析应用需求。金融文本分析应用是交易所数字化转型升级的重要内容，建议决策层自上而下组织业务部门、技术部门共同参与全所金融文本分析应用需求规划，加强对具有通用性、普遍性的需求进行整合，争取实现数字化转型一盘棋，提升整体建设效率；集中优势力量配合关键业务重大项目建设，以公司画像系统二期建设为契机，推动融合各业务部门多样化需求，“小步紧走”促进技术滚动迭代，逐步实现从点到面、由易及难的过渡优化，持续推进金融文本技术对交易所业务的全方位覆盖；协调技术部门设置专岗维护各部门文本分析需求，确保系统建设延续性与稳定性。

二是完善数据治理，推动数据标准在全所各业务系统落地。建议将数据治理纳入业务部门职责范畴，推动业务部门积极参与并牵头全所数据治理组织、制度、标准与安全相关工作，尤其是加强上市公司、证券基金公司等市场参与者定期报告等信息报送的流程规范，推动将完善规范指引和上市公司数据披露年度考核评价，促进上市公司形成数据报送第一责任人意识，重视XBRL数据填报的准确性，从源头把控数据质量管理，为建设满足金融文本分析应用标准的系统奠定坚实基础；总结标准化监管关键信息，推动制定行业信息书写与录入规范；积极收集业务金融文本数据需求，配合拓展数据资源内容，完善数据资产管理；鼓励业务部门梳理数据逻辑，优化业务流程，从执行层面落实数据标准，推动业务场景

标准化。

三是加强人才培养，构建数字化思维促进业务技术协同。通过建立数字化能力培训课程体系、组织开展各类共创会及分享会、加快开发便捷系统工具激发业务人员数字化能力建设的广度与深度，促进数字化专业人才积累，构建由数据专家、数据分析师、基础数据分析人员构成的多层次人才梯队；组织数据分析人员到系统内外金融文本技术应用成熟的单位部门交流学习，总结应用场景及成果经验；推动建立业务与技术人员常态化的双向交流机制，促进业务技术协同融合与交叉

赋能；推动数据分析人员参与重大业务项目建设，以岗代训，加快数字化人才与业务融合的进程，在工作中发现和锻炼人才，构建适合交易所业务发展需求的数字化人才梯队。

四是跟踪科技前沿，挖掘金融文本分析应用开发潜力。加强与其他金融监管机构和行业金融文本分析领先机构沟通交流，探索金融文本分析技术与应用创新模式，深化产学研合作与行业联合攻关，争取在舆情智能监控、知识图谱等方面取得进展，推动业务场景产品化，合力共建金融科技监管发展互助生态。

汇聚知识，赋能行业

——上交所探索构建资本市场知识库

何浩¹，苑阳¹，马信达² / ¹ 上交所技术有限责任公司 ² 恒生电子股份有限公司



资本市场知识库旨在搭建统一的知识运营管理平台，融合专家构建、众包共建和自动化抽取的知识获取模式，以提升知识库内容获取的效率和准确性，形成庞大的、高质量的证券行业知识库。此外，利用知识图谱技术中的图存储和图计算，结合行业应用场景，梳理领域知识体系，支撑智能化工程建设，助力于科技监管和智能审核。

近年来，上交所着力推动全行业数字化转型和信息技术应用创新，积极承接证券行业重大课题研究，旨在加强创新合作，推动金融科技取得重大发展。2021年，上交所联合恒生电子承接了证监会科技监管研究课题——资本市场知识库，经过双方共同的研究开发，形成了一系列重要研究和工程实践成果，助力于科技监管与智能化辅助审核，并积极探讨知识库在行业内的应用落地，推动构建共享、共建、智能的知识应用生态。

课题价值：支撑监管科技的重要基础

随着存储技术和图数据库技术的不断发展，

知识库的规模呈现出指数级增长，并逐步展现出极大的应用潜力，形成了行业或者领域级别的知识图谱，已成为大数据应用的重要基础。在此背景下，本课题拟通过搭建统一的资本市场知识库运营管理平台，编写《资本市场知识库工程方案》，以高效聚合证券行业知识内容，挖掘潜在关联关系，实现知识赋能，提升智能化服务水平。知识库既可支持资本市场主体信息库的建设，也可提供智能查询、智能问答、智能推荐、智能决策等智能化分析功能，并能够基于已有知识进行新知识的推理发现，实现自我学习、自我创新，为实现智能化辅助监管和审核提供强有力的支持。

在传统的知识库建设方案中，一般分为专家

构建方式、自动化构建方式和众包构建方式。由于资本市场自身的复杂性和监管需求的特殊性，前述三种建设方案并不能完全满足资本市场知识库的建设需求，挑战主要表现在以下几个方面：一是资本市场知识库本体设计困难。资本市场变化快速且复杂，对资本市场的认识也很难一步到位，难以自顶向下设计出全面清晰的本体；二是资本市场知识库缺乏相关的标注数据，大量的知识存在于非结构的文本信息中，很难直接采用自动化构建的方式从非结构化文本中提取有效知识；三是知识的高效获取和准确性之间存在矛盾，从获取知识的效率来看，自动化抽取模式的效率最高，但知识的准确性和真实性存疑，相较而言，众包构建和专家构建的准确性更高。

基于以上问题，本课题提出了以众包构建方式为主，同时结合专家构建与自动化构建方式来建设资本市场知识库。首先通过众包的方式构建本体，辅以专家审核来确定本体，可以降低资本市场本体设计难度。随着大量专业人士的参与，知识库的内容不断得以丰富，被标注的信息不断增加，便于结构化知识的获取，同时也为自动化抽取模型提供了大量的训练样本，进一步提升了智能化抽取的准确性和效率。因此将知识众包模

式与专家和自动化构建方式融合，可以在前期数据较少时快速积累有用数据，快速增加知识库的服务受众和贡献用户，从而进一步助推了知识的量级扩展，不断优化自动化抽取模型，形成效率提升与准确度提高的正向反馈机制。本次课题研究项目，就是对该建设方式的工程实践探索，并希望基于知识库建设工程支撑相关业务创新和发展，提升智能化服务水平，助力于科技监管和智能化辅助审核。

课题成果：构建统一的、应用级别知识库

上交所结合自身业务发展和行业实践需要，深入分析课题的实现需求和价值内核，确定构建统一的、应用级别的证券行业知识库，汇聚行业知识，推动行业智能化发展。其实现的需求来源包括：1、课题实现要求，如：搭建协同工作系统（知识众包）、实现自动化导入模块（数据汇聚）、构建统一且便捷运营的知识库系统、利用图存储和计算技术形成多个领域图谱、拓展知识库应用等；2、上交所业务发展需要，如：科创板审核、公司监管及其他业务活动中的知识沉淀、管理、查

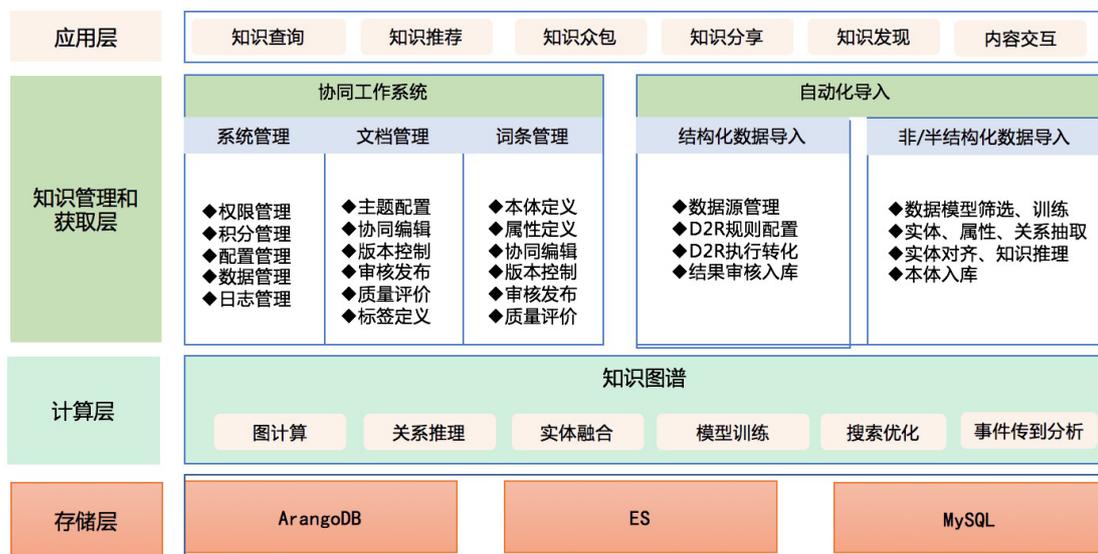


图 1：资本市场知识库整体架构图



图 2：资本市场知识库工作界面

询、分享以及智能化探索;3、行业赋能所需,如:行业知识的集中存放、管理、查询与分享,基于知识库的智能化应用探索(如智能查询、关系推理等)等。本课题的设计思路为:构建统一的、应用级别知识库,实现各类业务知识的便捷管理与获取、高效检索与分享,动态更新和维护,包括基于主题的文档管理、基于本体论的层次化知识管理以及基于知识图谱的网状知识管理。

依据前述三方面的需求,我们搭建统一的知识库运营管理平台,融合了协同工作系统(知识众包)和自动化导入模块(智能化抽取),实现了基于主题的文档管理、基于本体论的层次化知识管理及基于知识图谱的网状知识管理的协同统一,通过图计算和图存储,实现各类知识的有效转换和便捷获取,并形成了多个领域知识图谱。此外,为进一步挖掘领域图谱的价值,开展了相关的应用实践,实现了基于图计算和NLP的智能查询,基于关系链的新知识发现。以及结合舆情系统,构建了重点事件的传播影响分析等。

上交所在推进资本市场知识库的课题研究过程中,同时联合恒生电子深入探索人工智能底层基础技术,结合大量的行业应用场景,训练、优化智能抽取模型,并将人工智能领域先进的工程技术应用于丰富的富文本等非机构化知识的抽

取,如:将NLP、NER、OCR、CV等技术应用于在大量文档内容的自动化抽取,以此快速积累丰富的知识内容,支撑资本市场主体信息库的建设,为后续提供智能查询、自助问答、关联推荐、辅助决策等智能化分析功能夯实基础。

随着知识库本体工程的不断完善,本体特征不断明确和丰富,再结合人工智能技术的高效率和自动化,可快速提升知识库有效信息的抽取能力,使得知识库的内容获取更加智能化,更为高效和准确。庞大的高质量知识库数据平台,结合知识图谱技术,再融入特定领域的应用场景,可实现对离散、孤立的数据进行快速连接、组合,构筑有效实用的知识体系,输出强大的知识应用服务,例如:基于知识库构筑的多个领域知识图谱,可实现快速的关系推理和新知识发现,在关联关系挖掘、相关案例推荐、审核质量控制等应用场景中发挥着重要作用,可提供更加高效、准确的知识服务能力。

课题拓展：助力科创板智能审核

随着设立科创板并试点注册制的实施,上交所积极落实国家重大改革任务,坚持服务国家发展大局,推进科创板各项改革任务行稳致远。上

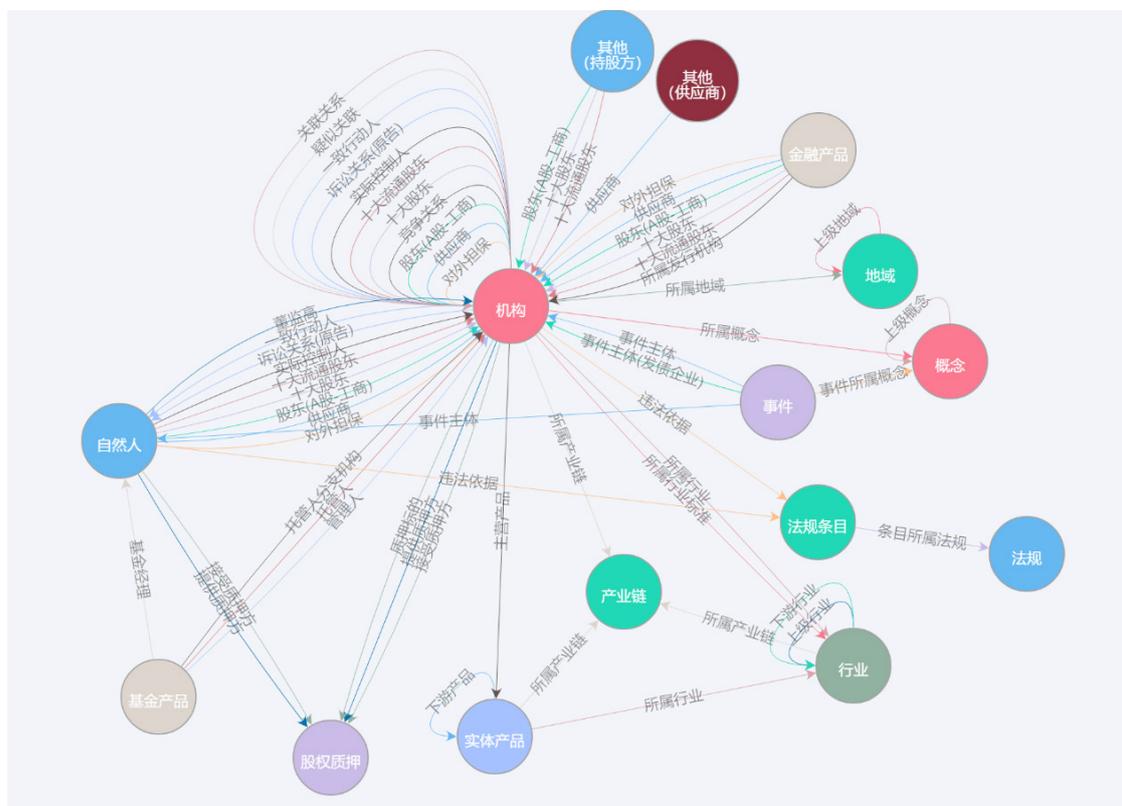


图3：资本市场知识库内的部分领域知识图谱

交所也在同步探索金融科技在审核工作中的落地实践，提升审核工作的智能化水平。在资本市场知识库研究中，上交所深入分析科创板审核的应用场景，为其设置专属功能模块，通过将拟上市公司招股说明书中的章节内容与知识库的丰富内容（如法律法规、违规案例、审核问询样例等）进行智能比对、快速匹配、直观展示，从而减轻审核人员的查阅工作量，提高审核效率。

科创板拟上市公司的审核涉及的法规内容较多，查阅工作量较大，为便于审核人员核对，知识库通过运用NLP技术对法律法规进行了条目自动抽取，并根据关键词自动标注，按照条目进行存储。当审核人员需要检索招股书某章节相关的法规条例时，可依据向量计算的结果，匹配关联度较高的法规条例，并支持审核人员进行关联标记，进一步提高匹配的可信度。此外，在审核过程中，涉及大量的案例研究和关联案例分析，历史案例及审核意见均为非结构数据，知识库通

过融入NLP、OCR等技术，将大量的案例数据进行了结构化抽取和存储，并对其进行词向量处理，可快速实现关联案例的推荐和查询，提升审核人员的工作效率。

更为关键的，对于拟上市公司的审核，还需要挖掘出隐含在招股说明书内的关键信息，如：拟上市公司的股权关系、所属行业的景气度和发展前景、供应链的完整性等。资本市场知识库聚合了丰富的内外部数据，并进行了大量的图谱化处理，构建了相关领域的知识图谱，如：股权关系链、产业链、供应链等，以此便于快速对公司进行画像，给审核工作提供了重要的参考数据。

未来，上交所将积极携手行业用户，共同探索知识库在行业内的应用拓展，推动构建统一的基础数据平台，发展知识共享、共创的内容生态，助力于行业的数字化转型和智能化发展，不断提高信息技术安全自主可控能力，促进科技与业务深度融合，为中国证券行业的发展保驾护航。

企业级证券业务中台探索与实践

樊建、严伟富 / 东方证券股份有限公司系统研发总部 邮箱: fanjian@orientsec.com.cn



东方证券以金融科技赋能业务发展和管理提升，在行业内率先对企业中台做出先行探索与实践，通过对用户业务旅程进行分析，遵循业务引领、领域划分、统一视角及能力复用的建设原则，将企业业务中台主要划分为账户、产品、财富、资产、行情、资讯、交易、认证、权益、投研等能力中心，形成“薄应用 厚中台 稳后台”的企业架构全景，并基于API化的开放式模块化架构核心思想，将核心业务知识进行沉淀，以模块化、服务化、共享化的形式建设企业级能力，从而快速响应市场变化和客户需求，大幅提升业务交付效率，金融科技核心竞争力取得重要突破，助力企业数字化转型。

1 背景

近年来，随着移动互联网、大数据和人工智能等新技术的全面深入应用，以及新商业模式的涌现和财富管理转型，对证券公司的IT能力提出了新的挑战。如何快速响应业务的需求、提供差异化的服务、满足投资者个性化的需求，成为摆在证券公司IT建设面前的巨大挑战。总体来说，

券商在系统建设过程中仍存在不少的痛点，主要如下：

►“烟囱式”建设，与大多数公司一样，东方证券系统以往多采用“烟囱式”发展方式，系统各自为政，功能大量重复，例如证券公司面向零售客户的网上交易系统，一般都有通达信/同花顺等，面向机构客户的PB产品端往往有恒生/迅投/根网等系统，架构层面也缺乏统一规划和

管控，技术成果更加无法共享；

- ▶ 单体架构，牵一发而动全身，且系统相互之间耦合度高，相互影响，无法保证 7*24 小时业务开展；

- ▶ 技术架构异构化，各个系统的功能调用方式、支持的开发语言、调用入口等等也不尽相同，形成了系统间的技术壁垒。在此基础上再进行系统开发以及进一步的迭代，其技术难度和风险也非常大；

- ▶ 交付速度慢，核心系统建设多以购买为主，需求响应缓慢，受制于人；

2 建设方案

2.1 中台定义

从 2015 年开始，以阿里巴巴为代表的各互联网巨头，陆续开启中台化进程，随后，“中台化”的理念与实践开始快速向各行业渗透和发展；对于金融行业，打造中台能力，无论是银行、证券或是保险等细分行业，均已是高度共识的战略举措之一，与此同时，证券行业财富管理转型、客户需求日新月异、IT 改造难度加大等现实状况也日益严峻，为此，为了促进公司数字化转型，基于企业现状及未来展望，东方证券也提出了“薄应用 厚中台 稳后台”的企业中台战略规划。

中台的定义各有不同，如阿里 [1] 官方定义“业务中台就是将企业的核心能力随着业务不断发展以数字化形式沉淀到平台，形成以服务为中心，由业务中台和数据中台构建起数据闭环运转的运营体系，供企业更高效的进行业务探索和创新，实现以数字化资产的形态构建企业核心差异化竞争力。”Gartner[2] 将中台定义为企业应用系统的 SOD 层，是灵活响应的前台与稳定可靠的后台之间的“变速齿轮”，ThoughtWorks[3] 定义为企业级能力复用平台，是面向用户与创新的新型平台型企业组织。

东方证券企业级业务中台的核心建设目标是基于 API 化的开放式模块化架构 [4] 核心思想，将核心业务知识进行沉淀，以模块化、服务化、共享化的形式建设企业级业务能力，从而快速响应市场变化和客户需求，提升业务交付效率，破除证券前台业务快（敏捷响应）和后台稳（坚实支撑）之间的“发展速率脱节和失配”的突出矛盾。

2.2 建设原则

企业业务中台建设对于东方证券来说也是一个新生事物，行业内亦无可参考的经验，为此在实际建设过程中，遵循如下建设原则：

▶ 业务引领

中台的建设是以提高业务响应为目标，所以需从自身商业模式和市场需要出发，围绕业务目标，按照中台的理念推进业务 / 技术架构变革，而不是简单跟风和模仿，为中台而中台。

▶ 领域划分

按照领域驱动的原则，在战略阶段划分问题域，确定核心领域，将系统划分为多个业务能力中心，当系统划分为多个业务能力中心后，中台建设就进入战术阶段。在战术阶段，针对已确定的各业务能力中心，结合业务需求进行具体的领域设计。

▶ 统一视角

企业业务中台面对众多业务线，需要站在业务整体视角，如 PC 网上交易、APP、临柜等，梳理业务流程，统筹考虑建设，要将后台资源抽象、沉淀和整合，包装成便于前台使用的可复用、可共享的核心能力，实现后台资源到前台易用能力的简化。

▶ 能力复用

中台是针对“商业模式”和“业务模式”的抽象与复用，沉淀共享能力，以可重用和可复制方式输出给各渠道产品线，以组件和能力编排实现业务场景化应用，并以服务化的形式输出能力。

2.3 用户业务旅程

区别于互联网行业及其他传统行业，证券行业也有其自身行业特点，我们站在用户的视角，具体分析下券商用户的核心业务旅程，以券商用户目前应用最为广泛的 APP 应用为例，其典型具体业务旅程如下：

- ▶ 打开 APP，进入首页，查看各类资讯；
- ▶ 进行手机号注册，注册后点击登录进行认证登录，进行积分等权益登记；
- ▶ 提示开立资金账号，点击开立后，上传各类资料，进行风险评测及双向视频验证；
- ▶ 资金账户开立，适当性验证，后续可进行各类业务办理；
- ▶ 点击银证转账，设置银行账号，转入 / 转出资金；
- ▶ 点击理财页面，查看各类理财产品信息；
- ▶ 点击产品购买，进行适当性判断，进入理财产品购买流程；
- ▶ 点击基金投顾页面，配置基金投顾服务；
- ▶ 点击行情页面，查看市场行情；
- ▶ 点击交易页面，点击买入 / 卖出，进行场内交易；
- ▶ 点击资产页面，查看资产详情；

▶ 点击商城页面，查看 LEVEL2 等非金融产品，可使用积分 / 第三方 / 保证金等方式购买非金融产品（见图 1）。

2.4 总体架构

通过对用户旅程进行分析，可以得知，券商客户的核心业务领域主要涉及行情、资讯、账户、资金、认证、产品、基金销售 / 投顾、场内交易、资产及商城等核心场景，因此，根据业务单一原则，如图 2 所示，东方证券将企业业务中台主要划分为账户、产品、财富、资产、行情、资讯、交易、认证、权益、投研等能力中心，形成“薄应用 厚中台 稳后台”的企业架构全景（见图 2）。

2.5 业务中台组成

企业业务中台主要划分为账户、产品、财富、资产、行情、资讯、交易、认证、权益、投研等能力中心，各个能力中心定位如下：

- ▶ 账户中心，建设权威、完整、标准的账户主数据中心，进行账户类业务受理及办理，提供各类账户全生命周期及适当性管理，并为各业务渠道提供数据服务（见图 3）。
- ▶ 财富中心，对接场外交易系统原子能力，

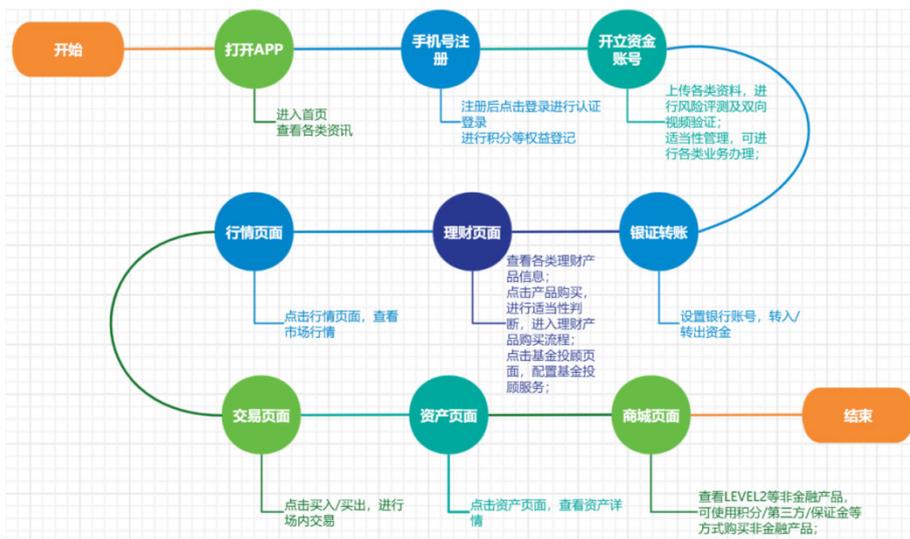


图 1：用户业务旅程图



图 2：东方证券企业架构全景图

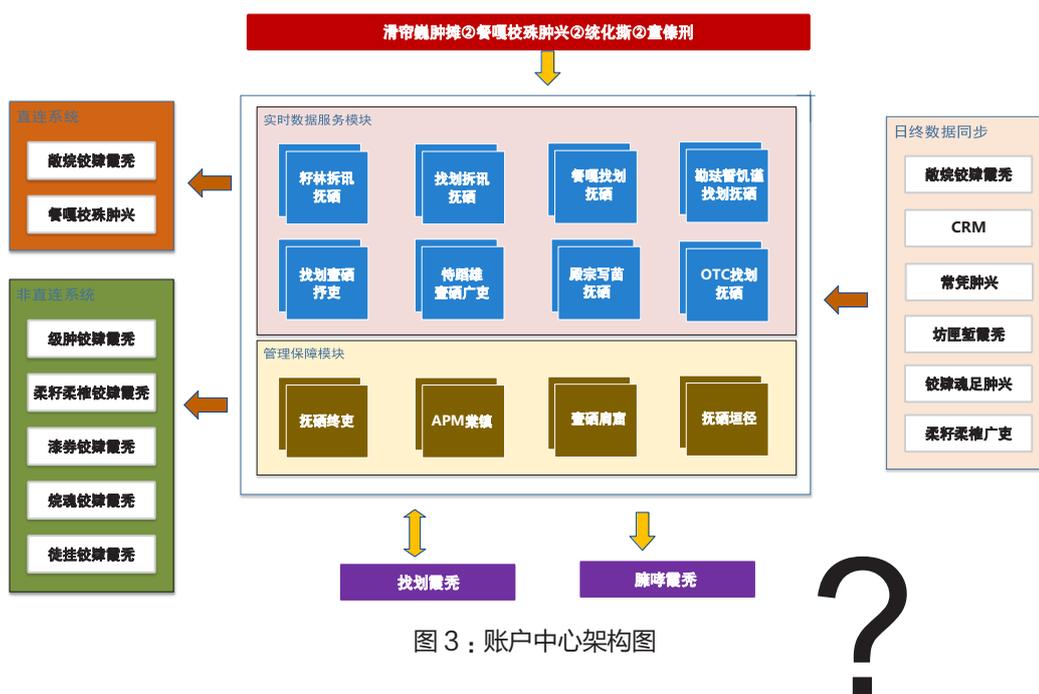


图 3：账户中心架构图

进行金融产品销售业务流程封装，提供场外交易统一接入服务能力（见图 4）。

► 产品中心，建设公司级金融产品仓库，覆盖公司全业务各类型的金融产品及产品化业务，从产品引入、产品上架、产品库管理、销售支持、营运管理到售后的分析报表与绩效考核，实现金融产品全生命周期管理，成为公司金融产品标准和权威的来源（见图 5）。

► 资产中心，整合各个相关业务系统的底层数据，汇总交易和回报的实时数据，承接交易清算数据，进行交易明细数据、资产查询及各类衍

生指标的计算和服务，为客户提供更加深度的资产交易查询分析服务等功能（7*24 小时服务），统一提供用户资产整体解决方案（见图 6）。

► 交易中心，统一制定场内交易协议，一方面屏蔽各柜台接口差异性（各交易中心须按交易接入中心协议进行对接），进行交易系统路由，对交易业务进行管控，另一方面对接场内竞价原子能力，对竞价业务流程进行组合包装，统一提供场内竞价业务能力（见图 7）。

► 资讯中心，统筹管理公司内外各类资讯数据源，对资讯数据进行提取、清洗、加工、存储

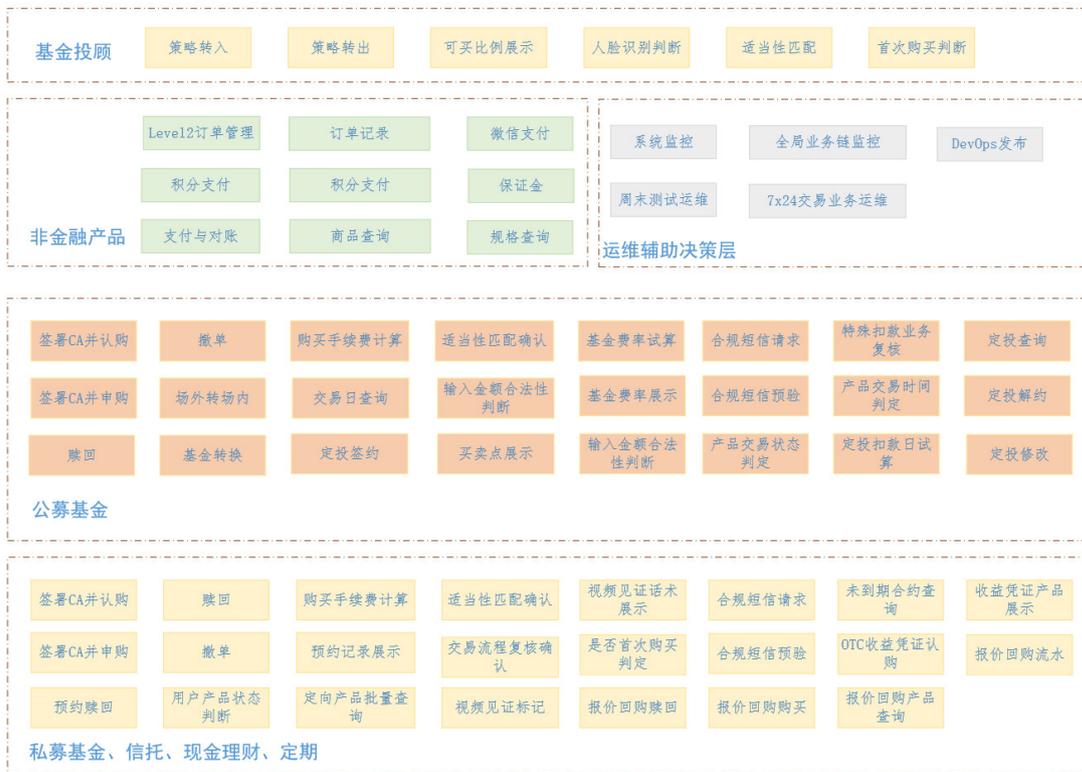


图4：财富中心架构图

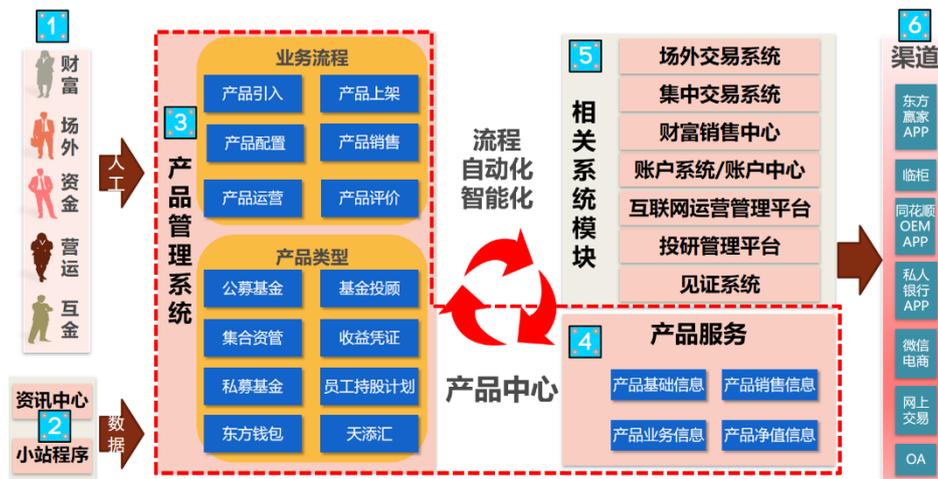


图5：产品中心架构图

等操作，形成资讯数据标准，对资讯数据进行全生命周期数据管理，并对外整体提供资讯类服务（见图8）。

▶ 行情中心，整合接入了国内外主要金融市场的交易行情，提供了行情接入与推送、存储、回放、计算及分析等领域的一体化解决方案（见图9）。

▶ 权益中心，旨在为公司客户的数字化商品权益和卡券类等虚拟资产提供管理功能，为相关的运营体系提供基础服务，系统提供的主要服务包括权益管理、卡券管理和内控管理（见图10）。

▶ 投研中心，贯穿投前、投中、投后整个投资研究过程，进行分析、决策、投资的整体投研流程生命周期管理，提供资产配置，数据服务，



图 6：资产中心整体架构图



图 7：交易中心架构图

算法服务，因子计算、策略回测与分析等投研相关服务（见图 11）。

► 认证中心，以用户身份管理为核心，加强管理 B/S、C/S、移动 APP 等结构的多应用的安全访问机制，集身份管理、身份认证、授权管理、应用资源访问控制及其安全审计于一体，构建多信息资源的应用整合、集约管理和安全防护的安

全基础服务平台（见图 12）。

3 建设成果

3.1 行业探索

中台这个概念脱胎于互联网行业，对于如何快速地完成业务创新并占领市场，带来了一个全

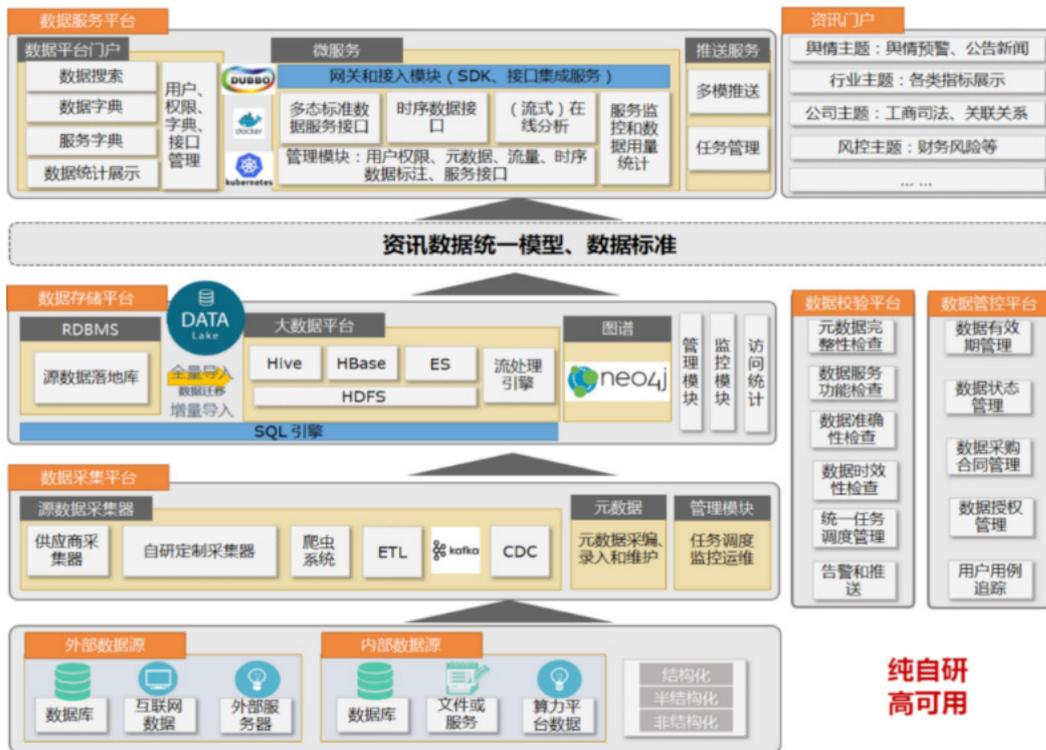


图 8：资讯中心架构图

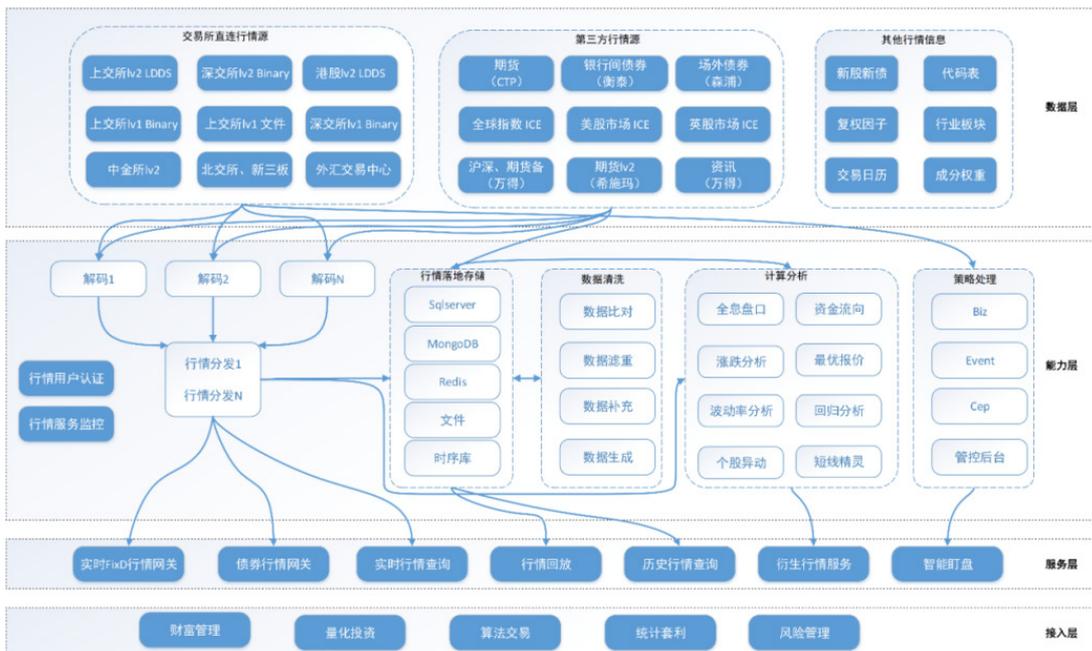


图 9：行情中心架构图

新的发展思路，其中对于业务中台、技术中台、数据中台都有相应的成功案例，而对于证券行业，其仍是新鲜事物，没有太多可借鉴经验，从 2019

年开始，东方证券开始在中台战略上进行探索和创新，遵循业务引领、领域划分、统一视角及能力复用的建设原则，实现数据 / 流程的共享，业

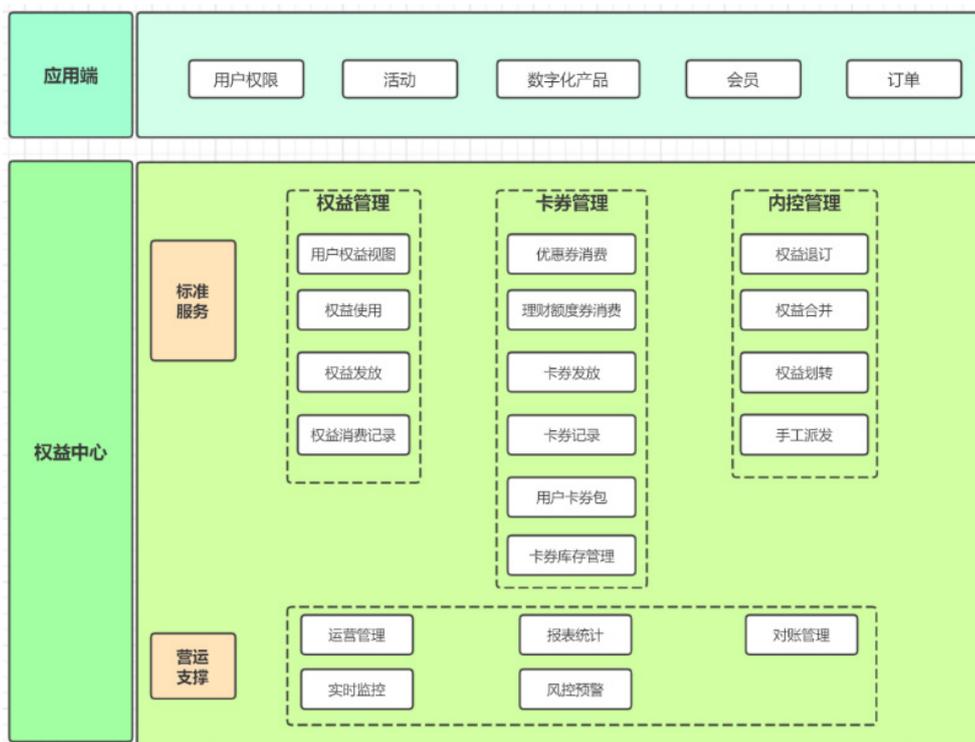


图 10：权益中心架构图

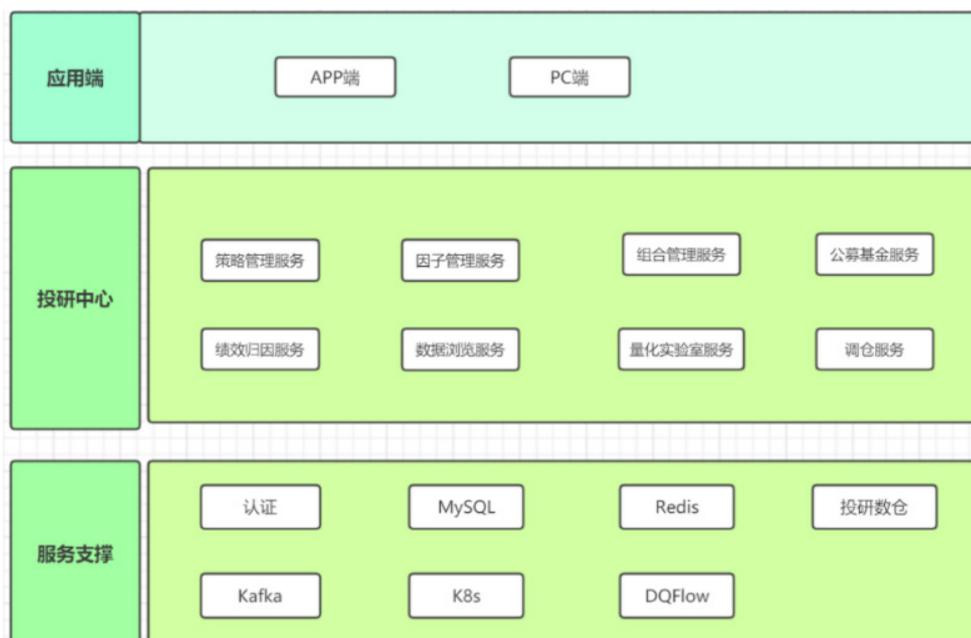


图 11：投研中心架构图

务能力的复用，形成以服务为中心，构建起数据闭环运转的运营体系，供企业更高效的进行业务探索和创新，实现以数字化资产的形态构建企业

核心差异化竞争力。

► 业务领域，通过对用户旅程分析，将业务域划分为账户、产品、财富、资产、行情、资讯、

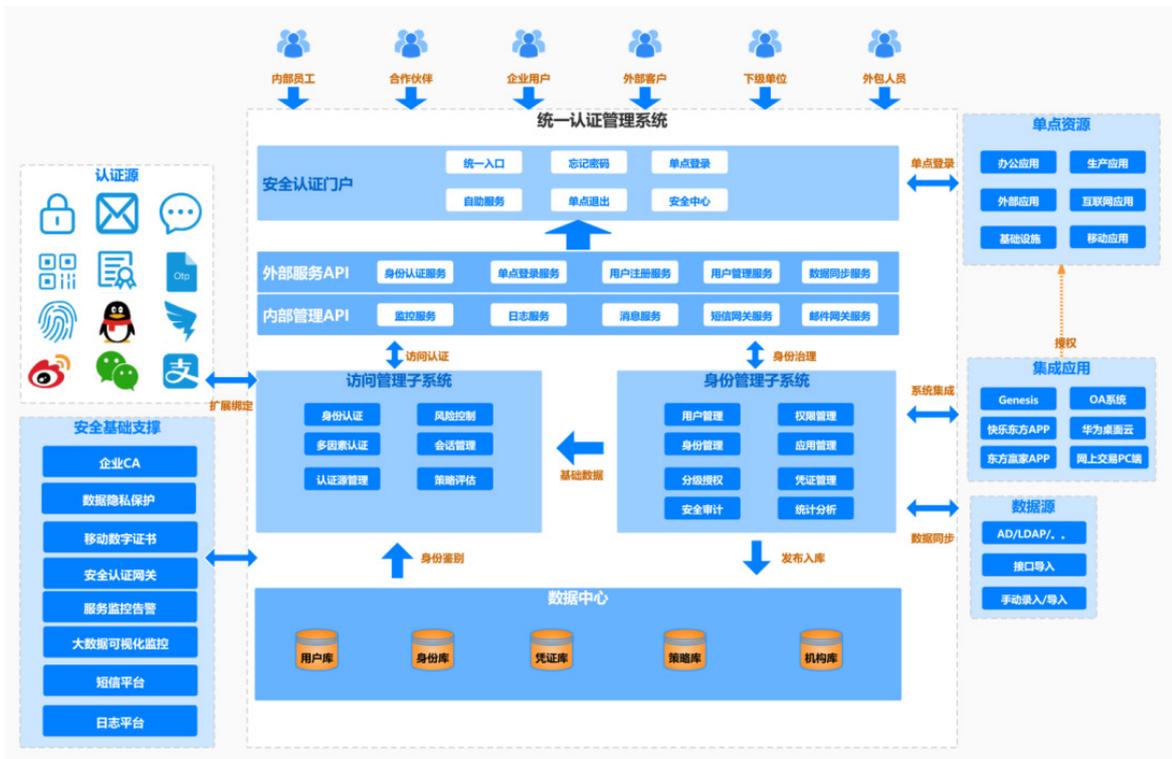


图 12：认证中心整体架构图

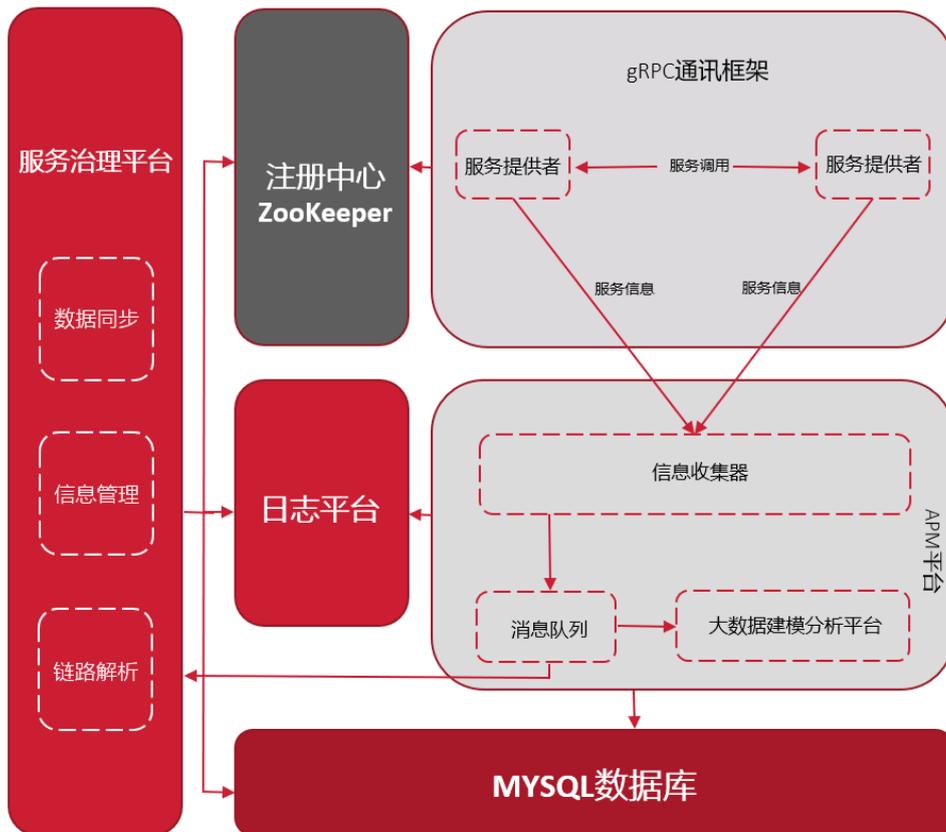


图 13：东方证券服务治理总体架构

东方证券 IT 架构决策

| | | | |
|--------|--|------|------------------|
| 架构决策 | 服务治理平台接入规范 | 编号 | TA. 2019. 0001 |
| 提出方 | 胡长春 | 提出日期 | 2019/5/30 |
| 架构领域 | TA: 技术架构 | 主题 | 各系统须按照规范接入服务治理平台 |
| 决策人员 | 架构委员会 | 决策时间 | 2019. 6. 30 |
| 问题 | 大量的业务及支撑系统开始上线运营并提供服务, 目前公司内部系统间接口调用无统一标准, 异构化非常严重, 各厂商都有各自私有协议, 且存在有 SPX、T2、Web Service、REST、TCP 等各类型异构接口, 无法以全局视角对内部服务进行统一治理, 进一步增加了系统开发运维优化的难度。 | | |
| 假设及约束 | 1、所有新建系统需接入服务治理平台, 旧有系统逐步改造; 2、应用通过服务治理框架完成服务注册和发现, 通过服务治理平台进行服务治理治理, 使用 APM 探针或 SDK 完成数据采集。 | | |
| 动机 | 采用统一的接口调用方式, 支持东方证券 IT 技术架构从传统架构向微服务为核心的现代面向服务架构全面转型。 | | |
| 可选决策意见 | 1、各系统各自选择接口标准 2、建立统一的接口标准各系统使用 3、须支持跨语言相互调用 | | |
| 决策 | 制定统一的系统间调用标准, 所有系统采用统一接口标准接入服务治理平台。 | | |
| 理由 | 制定统一的系统间调用标准, 解决跨语言难题, 所有系统采用统一接口标准接入服务治理平台, 从而达到服务治理目标。 | | |
| 影响分析 | 应用接入服务治理平台, 需要部署 APM 探针或使用 APM 探针 SDK 进行埋点, 对应用系统性能会有 3%-5% 的性能影响。 | | |

图 14 : 架构标准示意

交易、认证、权益、投研等能力中心, 制定了各领域架构全景并进行实践探索, 《交易技术前沿》发表论文两篇, 获 2021 年度证券期货业金融科技研究发展中心 (深圳) 课题三等奖;

► 技术架构领域, 自研 gRPC-Nebula 成为行业内首个开源社区 (<https://github.com/grpc-nebula>), 并在《交易技术前沿》《金融电子化》《InfoQ》发表论文三篇, 获 2020 年度证券期货业金融科技研究发展中心 (深圳) 课题二等奖 (见图 13)。

► 架构管理领域, 形成了数十项架构决策, 使整体研发活动具备了遵循架构标准进行开发的能力 (见图 14)。

► 数据治理领域, 以业务中台为基准形成公司级业务术语、标准词根及公共代码, 并将数据

治理流程嵌入整体研发流程 (见图 15)。

► 数字化管理领域, 探索版本火车模式, 打通整体的研发运行一体化流程, IT 自身交付效率、质量上了一个新台阶 (见图 16、图 17)。

整个证券行业目前仍处在中台战略的摸索期, 东方证券提出的业务中台解决方案, 为整个行业提供了参考, 希望能达成行业共识, 形成统一技术标准, 大幅降低今后行业各系统间对接成本。

3.2 架构演进

常见的应用多活架构分为同城多活、异地多活、混合云多活, 和传统容灾相比, 应用多活具备 RTO 低、资源充分利用、切换成功率高、流量精准控制等优势。如图 19 所示, 我们主要在同城多活做了以下工作:

| 接口名称 | | | | | | |
|-----------------|--------------------------------------|------------|--------|--------|----|---------------------|
| 功能描述 | | | | | | |
| 基金信息查询 | | | | | | |
| 输入参数 | 字段名 | 域名名称 | 标识 | 类型及长度 | 必要 | 描述 |
| | req_address | 请求地址 | 第1域 | String | Y | 请求来源的网卡地址，或TCP/IP地址 |
| | branch_code | 营业部代码 | 第4域 | String | Y | 营业部代码 |
| | fund_acc | 资金账户 | 第7域 | String | Y | |
| | ta_code | 基金公司代码 | 第8域 | String | N | (空表示查询所有) |
| | prod_code | 产品代码 | 第9域 | String | N | (空格表示查询所有) |
| | entrust_method | 委托方式 | 第10域 | String | Y | 见数据字典 |
| | req_num | 请求个数 | 第12域 | String | N | 默认100 |
| | position_str | 定位串 | | String | N | |
| | channel_type | 渠道类型 | | String | Y | 见数据字典 |
| 输出参数 | 字段名 | 域名名称 | 标识 | 类型及长度 | 必要 | 描述 |
| | error_code | 错误代码 | | String | | 成功返回0 |
| | error_msg | 错误消息 | | String | | |
| list | ta_name | 基金公司名称 | | String | Y | |
| | issue_begin_date | 发行起始日期 | | String | Y | |
| | issue_end_date | 发行结束日期 | | String | Y | |
| | purchase_begin_date | 申购起始日期 | | String | Y | |
| | ta_code | 基金公司代码 | | String | Y | |
| | prod_name | 产品名称 | | String | Y | |
| | prod_code | 产品代码 | | String | Y | |
| | indv_single_purchase_highest_amount | 个人单笔申购最高金额 | | String | Y | |
| | indv_purchase_lowest_amount | 个人申购最低金额 | | String | Y | |
| | indv_first_purchase_lowest_amount | 个人首次申购最低金额 | | String | Y | |
| | indv_single_subscribe_highest_amount | 个人单笔认购最高金额 | | String | Y | |
| | indv_subscribe_lowest_amount | 个人认购最低金额 | | String | Y | |
| | indv_first_subscribe_lowest_amount | 个人首次认购最低金额 | | String | Y | |
| | redeem_lowest_share | 赎回最低份额 | | String | Y | |
| | redeem_highest_share | 赎回最高份额 | | String | Y | |
| | convert_status | 可转换状态 | | String | Y | |
| | avl_fixed_invest_flag | 可定投标志 | | String | Y | |
| | fixed_invest_lowest_amount | 定投最低金额 | | String | Y | |
| | prod_code_type | 产品代码类型 | | String | Y | |
| | indv_holding_total_uplmt | 个人持仓累计上限 | | String | Y | |
| prod_risk_level | 产品风险等级 | | String | Y | | |

图 15 : 数据治理接口评审示意图

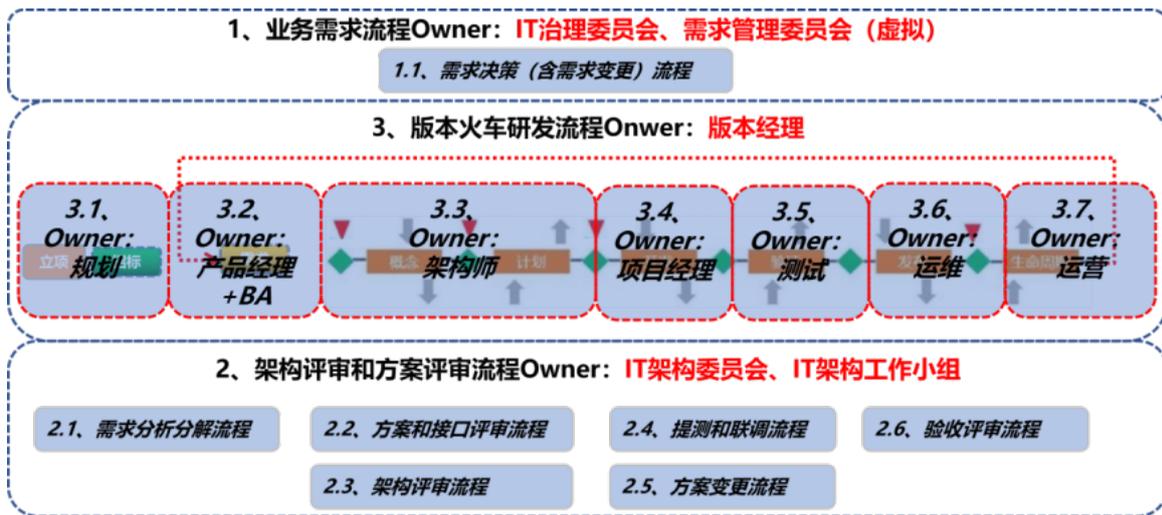


图 16 : 研发版本火车发布模式

首先，将各业务应用中经常使用的 PaaS 中间件 (Zookeeper/kafka/redis) 进行多活机房部署；

其次，由于 gRPC-Nebula 框架所依赖的 PaaS 中间件 Zookeeper 已实现多机房部署，并配合框架本身的多机房分组功能，实现 gRPC-Nebula 的多活架构；

最终，依赖 gRPC-Nebula 开发框架与 PaaS

中间件的结合，从而实现各中台应用的多活架构，实现企业架构整体应用多活，促进了整体架构演进升级（见图 18）。

3.3 应用成果

从 2019 年初开始，东方证券确定了企业大中台战略，围绕业务价值与 IT 用户体验，大力

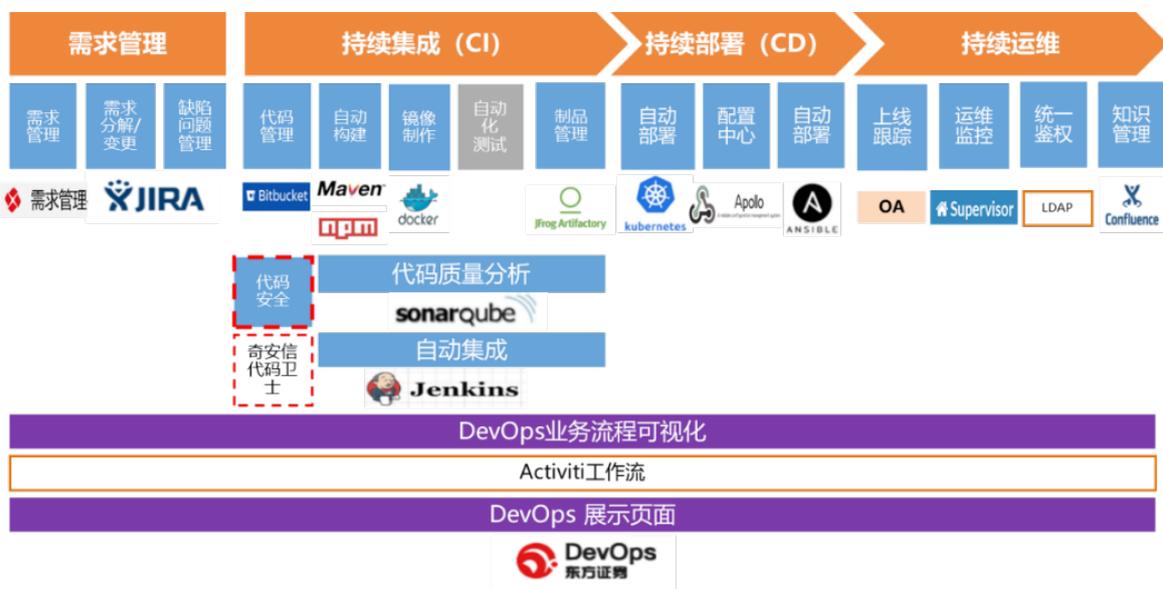


图 17 : 研发运行一体化平台架构图

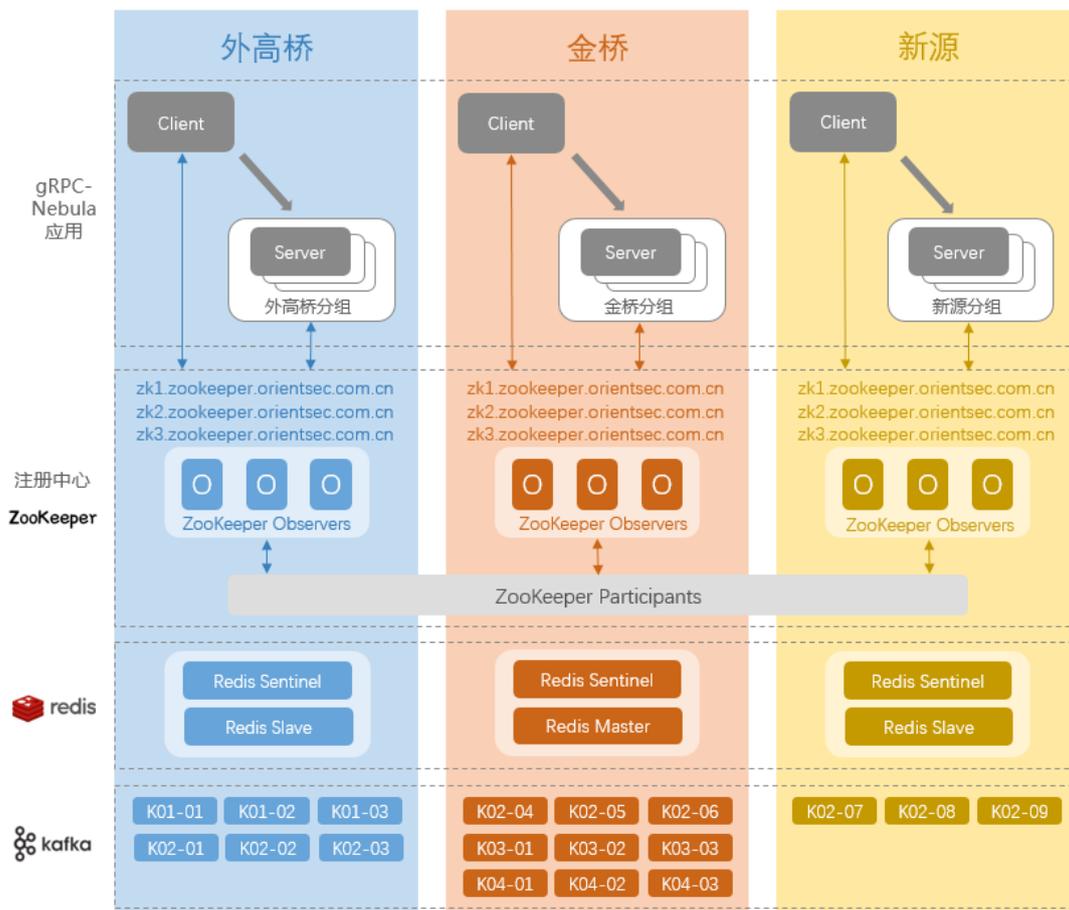


图 18 : 中台多活应用架构

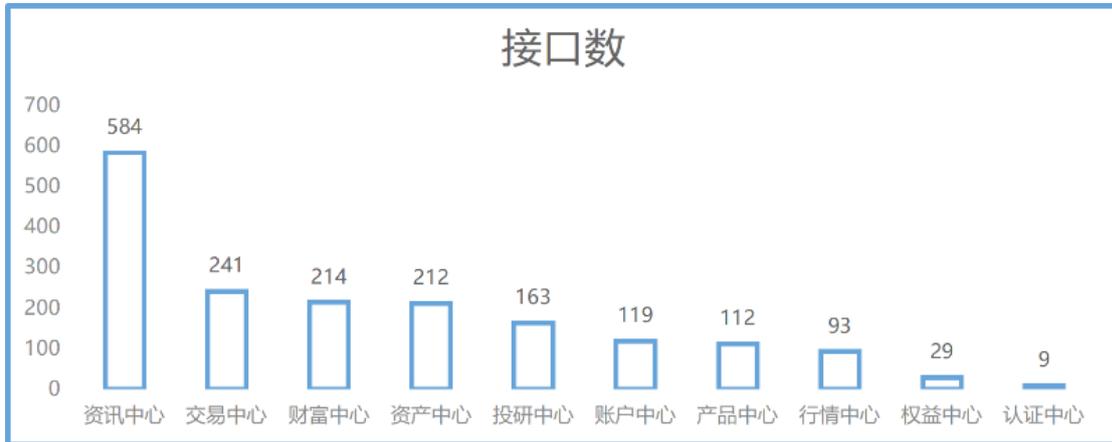


图 19：业务中台对外接口数



图 20：业务中台接口峰值访问量

推进“薄应用，厚中台，稳后台”的架构转型，并通过制定架构标准推动相应领域建设，经过三年多的建设，截止到 2022 年 8 月底，形成了体系化的技术架构和数字化管理能力，各业务中台上线对外服务数 195 个，对外接口总数达 1822 个，日峰值请求量 4600 万+次，对接各类业务需求 450+（不完全统计），形成数据治理词根 2436 个，字段 1554 个，公共代码 13 个，已深入经纪、财富管理、期货、资管、PB、自营等业务领域，初步形成了集团协同能力，为东方赢家 APP、网上交易 PC 端、机构交易客户端、机构理财客户端、

CRM 等业务线提供共享能力服务，并成功实现了财富管理业务领域需求全覆盖，助力东方证券以客户数排名 38 名，经纪业务排名 20 名，取得了公募基金保有量在行业排名第 7 位 [12] 的成绩，整体上线业务需求达 90%，通过技术共享、服务共享、数据共享、研发规范，有力的提升了开发效率，降低了对开发商的依赖和系统研发成本，避免了系统的重复建设和异构化，提升了业务需求的交付速度，沉淀了证券核心领域业务知识库和领域，对于增强公司科技创新应用能力，激发行业技术创新动力，发挥了重要作用。

4 总结

东方证券经过三年多的建设，并基于 API 化的开放式模块化架构核心思想，形成了“薄应用厚中台 稳后台”的企业架构全景，并将核心业务知识进行沉淀，以模块化、服务化、共享化的形式建设企业级业务能力，从而快速响应市场变化和客户需求，大幅提升业务交付效率，金融科技核心竞争力取得重要突破。

未来，东方证券将继续以开放共享、合作共赢为原则，以金融科技规划为牵引，持续推进中台战略转型，并以新一代交易核心平台建设为契机，与交易系统配合形成更为全面的领域服务划分，合理清晰的原子及业务流程服务，精细化各系统内部及之间细粒度的服务、接口及数据库设计，持续推动金融科技研究应用，以金融科技赋能业务，引领创新，不断助力公司数字化转型与行业创新发展。



实践探索

- 4 数据治理体系建设的研究与实践
- 5 数据中台数据标准化设计与实践——上交所数据主题域顶层设计
- 6 核心业务系统连续性保障的探索与实践
- 7 构建邮件纵深安全体系，提升互联网重要入口安全运营水平——海通证券安全运营实践分享

数据治理体系建设的 研究与实践

张如意¹、宫朝辉²、李斌³、雷晓全¹、徐祯¹、于三禄^{1/1} 郑州商品交易所 信息科技一部 郑州 450018² 郑州易盛信息技术有限公司 风控系统部 郑州 450018³ 郑州易盛信息技术有限公司 金融创新实验室 郑州 450018

邮箱: ryzhang@czce.com.cn



本文主要介绍了郑商所在数据治理体系建设方面的研究与实践，包括数据治理体系框架和运行机制，数据标准的统一、数据质量的稽核、数据安全的提升、多层级的数据使用以及数据建模方法等实践路径，逐步探索数据治理从管控式理念模式向服务式理念模式的转换。

1 数据治理背景概述

十九届四中全会首次将数据增列为生产要素，《关于构建更加完善的要素市场化配置体制机制的意见》中明确要求加快培育数据要素市场，全面提升数据要素价值。在 ABCD（人工智能、区块链、云计算、大数据）技术的创新驱动下，新兴金融科技逐渐应用于监管科技、风险防范、运营决策、运维自动化等方向，取得了较好的应用成效。在数字化转型深入推进的大背景下，数据治理体系建设是数字化转型依托的基础、精细化管理不可或缺的部分，“是一切信息化工作

的重中之重”。郑商所作为重要的金融基础设施，聚焦“数据让监管更加智慧”的愿景，持续加强数据治理能力建设，不断完善数据治理体系，加快推进科技和业务的深度融合，充分挖掘数据潜能，释放数据要素价值。

2 数据治理体系探索实践

数据治理主要围绕“建制度、盘资产、落标准、核质量、保安全、搭应用”六个方面开展工作。一是构建起一套符合组织的数据治理体系框架、与业务需求高度契合的数据管理制度体系，

将数据治理的理念与工作内容融入现有的管理体系，保障数据治理工作的长效运转。二是持续盘点内外部数据资产，为信息数据的“目录化、集成化、可享化”打下坚实基础。三是落地基础数据标准和指标数据标准，统一技术语言和业务语言，实现组织内数据“书同文、车同轨”。四是核验数据质量，编制并应用数据质量审核规则，持续提升数据质量。五是确保数据安全，通过采用访问终端白名单制、智能数据脱敏、数字水印、日志埋点和审计提升技防水平，数据分类分级、访问权限控制、数据服务流程约束等夯实人防基础，求取数据安全和数据使用的“最大公约数”。六是搭建大数据平台，构建“湖仓一体、冷热数据分离”的分布式底座，融合数据中台，打造敏捷、共享、统一的数据服务，满足日益增长的数据需求。

2.1 建立有效的数据治理体系框架

数据治理体系是指通过搭建多层次、专业化的数据治理组织架构，从数据标准、数据质量、数据安全、数据使用等不同维度，树立“一盘棋”的数据治理理念，加强顶层设计，从上至下指导和从下而上推进的多层次、多维度、多视角体系框架。

建立组织机构、明确职责分工。郑商所数据治理工作在金融科技委员会（以下简称委员会）领导下，由数据管理方、数据归口方、数据使用方和数据服务方分工负责。委员会统筹数据治理工作。数据管理方牵头制定数据治理相关制度，构建以数据管理办法为纲领的“1+N”制度体系，建立多层次、相互衔接的运行机制，协调落实运行、组织推动数据治理工作，建立、完善、维护数据架构，搭建专业的数据管理平台与工具。数据归口方负责数据标准的业务解释和数据质量的业务优化，盘点识别本业务域的数据资产。数据使用方根据业务需要和数据安全的要求，提出数据需求，获取相关数据，开展数据应用。数据服

务方承担数据治理的技术支撑和信息安全职责，提供数据标准的技术定义、数据使用的服务提供、数据质量的技术支持、数据安全的技术保障。

建立工作运行机制、确保高效运行。一是建立数据治理服务顾问机制，搭好业务、科技、数据的桥梁，确保业务需求及时有效转化。二是成立跨部门的数据治理工作组，推动数据治理工作有效开展，形成治理合力。三是制定金融科技及数据治理参与度绩效考核机制，通过合理设立扣分、加分项，明确数据使用需求中的数据内容、使用频次与时效要求、指标统计口径和数据级别，着力解决数据服务需求排队问题、统计指标与业务发展不匹配问题，提升业务需求质量和业务人员参与测试的积极性，鼓励业务人员盘点业务域数据资产、打造标签体系、编制数据质量业务检核规则。四是定期组织数据治理、数字化转型、数据分析相关培训，培养业务人员数字化思维，打造数据分析师团队和多层级数据服务体系，鼓励业务人员通过可视化自助分析系统开展数据分析、研究、探索工作。

2.2 形成统一数据标准

数据标准是指保障数据的内外部使用和交换的一致性和准确性的规范性约束。数据标准管理的目标是通过统一的数据标准制定和发布，结合制度约束、系统控制等手段，实现大数据平台数据的完整性、有效性、一致性、规范性、开放性和共享性管理，为数据资产管理活动提供规范依据。一是按照交易、结算、交割、监查、会员管理等业务条线和业务职责范围，制定业务数据归口表确定各类数据归口方，数据归口方负责归口数据标准的业务解释和命名规范。二是结合数据资产盘点，对数据的标识、格式及定义的一致约定，确保组织内数据业务属性、技术属性和管理属性的统一定义。其中指标数据标准参考《证券期货业统计指标标准指引》。三是开展标准体系建设、标准评审发布、标准落地执行、标准运营

维护等工作，确保数据标准贯穿业务数据化全过程。基础数据、指标数据标准示例如图 1、图 2 所示。

2.3 提升数据质量

数据质量指数数据反映业务真实情况，满足

数据使用要求的程度，即“描述、翻译业务的语法标准”的指标。数据质量衡量的六个维度 [2]：规范性，即数据的描述是否满足业务要求；完整性，即数据是否缺失；准确性，即数据是否准确表达业务信息；一致性，即数据的取值是否存在信息意义上的冲突；时效性，即数据的获取和更

数据标准检索

新增数据标准 下载数据标准

数据标准名称/编号 数据敏感度

全部(627) 资产管理(183) 主体管理(122) 交易管理(103) 交割管理(88) 结算管理(70) 公共信息(28) 产品管理(26) 风控管理(7)

| 数据标准名称 | 数据标准英文名称 | 数据敏感度 | 数据类型 | 业务管理部门 | 操作 |
|-----------|--------------------------|-------|------|--------|---|
| 交割预报申请编号 | PREDICT_APPLY_ID | / | 编码类 | 交割部 | 查看详情 删除 |
| 交割预报方式 | PREDICT_MODE | / | 代码类 | 交割部 | 查看详情 删除 |
| 入库数量 | IN_WH_QTY | / | 数值类 | 交割部 | 查看详情 删除 |
| 交割预报申请状态 | PREDICT_APPLY_STATUS | / | 代码类 | 交割部 | 查看详情 删除 |
| 预报数量 | PREDICT_QTY | / | 数值类 | 交割部 | 查看详情 删除 |
| 交割预报总撤销数量 | PREDICT_TOTAL_REMOVE_QTY | / | 数值类 | 交割部 | 查看详情 删除 |
| 定金清退状态 | DEPOSIT_RETURN_STATUS | / | 代码类 | 交割部 | 查看详情 删除 |
| 报检号 | INSP_DECLARE_NO | / | 编码类 | 交割部 | 查看详情 删除 |
| 样品单号 | SAMPLE_ID | / | 编码类 | 交割部 | 查看详情 删除 |

图 1：基础数据标准示例图

指标标准检索

新增指标 下载指标

指标名称/编号 数据敏感度

全部(757) 交易(321) 资金结算(173) 会员客户(76) 交割(68) 做市商(55) 仓单(47) 风险(17)

| 指标名称 | 业务域 | 一级分类 | 二级分类 | 指标归口业务部门 | 指标使用部门 | 业务定义 | 指标别名 | 适用公共统计规则 | 度量单位 | 数据类别 | 数据值 | 操作 |
|----------|------|------|------|----------|--------|-------------|------|----------|------|------|-----|---|
| 会员数 | 会员客户 | 会员统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，经郑州... | / | 本年 | 个 | 数值类 | | 查看 编辑 |
| 营业部数量 | 会员客户 | 会员统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，经中国... | / | 本年 | 家 | 数值类 | | 查看 编辑 |
| 客户数排名 | 会员客户 | 会员统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，每个会... | / | 本年 | 无 | 排名类 | | 查看 编辑 |
| 自然人客户数排名 | 会员客户 | 会员统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，每个会... | / | 本年 | 无 | 排名类 | | 查看 编辑 |
| 法人客户数排名 | 会员客户 | 会员统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，每个会... | / | 本年 | 无 | 排名类 | | 查看 编辑 |
| 账户数 | 会员客户 | 账户统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，在郑州... | / | 当日 | 个 | 数值类 | | 查看 编辑 |
| 自然人账户数 | 会员客户 | 账户统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，客户类... | / | 当日 | 个 | 数值类 | | 查看 编辑 |
| 法人账户数 | 会员客户 | 账户统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，客户类... | / | / | 个 | 数值类 | | 查看 编辑 |
| 一般法人账户数 | 会员客户 | 账户统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，客户类... | / | 当日 | 个 | 数值类 | | 查看 编辑 |
| 特殊法人账户数 | 会员客户 | 账户统计 | 基本 | 会员管理部 | 会员管理部 | 统计期末，客户类... | / | 当日 | 个 | 数值类 | | 查看 编辑 |

图 2：指标数据标准示例图

新是否及时；可访问性，即数据能被访问的程度。为确保数据质量，一是在系统需求、设计、开发、测试、运行等阶段，结合数据标准，定义数据质量检核维度，设计数据质量检核规则，加强数据质量管控。二是借助工作流引擎工具，对交易、结算、业务参数录入等关键业务环节中，执行数据集中抽取入库过程时，对抽取进接口库的数据进行质量核验和业务逻辑校验，通过系列跑批任务正确执行与有效监控，确保当日入库数据质量达标。在业务空闲期，对全库数据进行数据质量规则校验扫描。每日跑批任务执行情况如图 3 所示。三是建立数据质量检查和监控系统，开展数据质量日常监控，对数据的规范性、完整性、准确性、一致性、时效性、可访问性进行监测，定期或手动触发生成数据质量报告，如图 4 所示。四是不断总结解决数据质量问题的的工作经验，并对数据质量问题的发现、发生频率、解决周期进

行评价，与金融科技及数据治理参与度绩效考核机制挂钩，主动预防、严格控制、及时处理数据质量事件。

2.4 严控数据安全

数据安全指通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力 [3]。数据安全遵循“谁管理谁负责、谁使用谁负责”原则。依据证监会数据管理办法、《证券期货业网络安全等级保护基本要求》《证券期货业数据分类分级指引》等制度要求和标准规范，制定数据分类分级细则，明确数据分类分级的工作方法与过程，区分数据的重要性和敏感度差异，制定业务数据分类分级表，明确高级别敏感数据的权限范围，确保数据分级使用权责一致。其中，四级数据，特定范围内使用，主要包括：个人开户信息、资金类数据、



图 3：每日跑批任务执行情况示例图



图 4 : 数据质量巡检报告示例图

监查类数据，逐笔及客户粒度的交易、结算、仓单、交割数据。三级数据，指定业务领域范围内使用，主要包括：会员管理信息、统计维度的交易、结算、交割数据。二级数据，组织内共享使用。主要包括会员基本信息、交易基本信息、结算基本信息、外部数据等。一级数据，公开数据。主要包括行情、官网公布数据等。制定数据安全管理制度，加强数据的采集、传输、存储、使用、删除、销毁等全生命周期安全管理，健全完善应对数据泄露、数据篡改、数据损毁的应急预案，细化应急处置措施，夯实“制防”基础。另一方面，采用分部门分岗位两级权限管理、数字水印、终端访问白名单机制、日志审计溯源、安全巡检等技防措施，做到数据安全“事前可管、事中可控、

事后可查”。

2.5 方便数据使用

数据使用管理遵循分级使用、权责一致的原则。数据安全和数据使用是“车之两轮、鸟之两翼”，制定数据使用管理细则，明确数据使用流程、权限管理和使用安全保障，通过明确各级数据的使用范围和访问流程，定期梳理高敏感度的四级数据的权限分配和使用现状，确保敏感数据的知悉范围，并根据使用情况调整数据级别，保障数据安全的基础上方便数据使用。业务人员通过临时数据查询、需求开发、已有功能权限变更等三种方式获取数据服务。其中临时数据查询，用于特殊情况下，临时查看四级数据。需求开发用于

业务人员根据需求模板书，提交数据开发需求，经权限和级别评审，开发上线后使用。需求开发的需求评审阶段，“前置争议、确保安全”，对于数据级别不清晰的，由数据管理方、使用方、服务方组成评审小组，通过数据脱敏、统计维度、明确时间范围等降级方式，在安全审慎的前提下，满足数据使用需求。另一方面，通过需求评审，不断补充完善数据分类分级表，形成需求评审与数据分级的闭环管理。已有功能权限变更，主要是避免功能重叠、重复开发，做到“一次开发、共享使用”，已开发的功能报表变更部门级使用权限时，由数据管理方、数据归口方审核后，由数据管理部门进行功能权限变更。同时，金融科技及数据治理参与度绩效考核机制设置加分项，鼓励业务数据归口方，根据业务领域特点，研究并制定本业务领域功能报表，不断拓宽拓深各业务领域数据的使用范围。

在数据服务的提供形式方面，一是提供了统计分析、可视化自助分析、数据沙箱等多层次的数据服务方式，满足各业务场景个性化数据服务需求，极大便利业务人员的日常业务处理以及研究分析工作，赋能“研究型交易所”建设。其中，统计分析指通过固定报表的方式，快速响应日常业务中标准业务指标查询。可视化自助分析提供主题和维度两种数据模型，方便业务人员通过图形化拖拽方式，进行数据加工、数据研究等探索式分析。数据沙箱主要响应交易、结算、监查等交易后数据分析“大户”的研究型需求，采用脱敏数据子集，通过可视化算法建模、自助分析、快速编程和 AI 组件，形成模型探索、模型分享、模型固化的模式，实现业务经验、专家智慧和技术的深度融合。二是构建数据分析师团队。数据归口部门和数据服务部门利用 BI (Business Intelligence) 分析工具提供的可视化自助分析功

- | | | |
|---|--|--|
| <p>会员当日资金查询 需求部门: 结算部 归口部门: 结算部 安全等级: 4 上线时间: 2021-09-28 13:04:42</p> | <p>自然人客户月度交易 需求部门: 市场监督部, 商品一部, 商品二部, 商品三部 归口部门: 交易部, 市场监督部 安全等级: 3 上线时间: 2021-09-28 13:18:04</p> | <p>特殊客户组合持仓明细 需求部门: 结算部 归口部门: 结算部 安全等级: 4 上线时间: 2021-09-28 13:12:22</p> |
| <p>会员客户仓单注册注销 需求部门: 交割部 归口部门: 交割部 安全等级: 4 上线时间: 2021-08-18 20:24:43</p> | <p>会员资管账户成交盈亏 需求部门: 市场监督部 归口部门: 结算部, 市场监督部 安全等级: 4 上线时间: 2021-08-20 09:04:24</p> | <p>客户数结存统计 需求部门: 会员管理部, 商品一部 归口部门: 会员管理部 安全等级: 3 上线时间: 2021-08-20 09:07:07</p> |
| <p>分品种产业客户统计 需求部门: 商品一部, 商品二部, 商品三部 归口部门: 会员管理部 安全等级: 3 上线时间: 2021-08-20 09:34:04</p> | <p>私募基金公司成交持仓 需求部门: 会员管理部 归口部门: 交易部, 结算部 安全等级: 4 上线时间: 2021-08-20 09:37:09</p> | <p>私募基金客户成交持仓 需求部门: 会员管理部 归口部门: 交易部, 结算部 安全等级: 4 上线时间: 2021-08-20 09:39:29</p> |
| <p>法人客户月度交易情况 需求部门: 市场监督部, 商品一部, 商品二部, 商品三部 归口部门: 交易部, 市场监督部 安全等级: 3 上线时间: 2021-08-20 09:41:26</p> | <p>投资者结构数据分析 需求部门: 商品一部, 商品二部, 商品三部 归口部门: 交易部, 结算部 安全等级: 3 上线时间: 2021-08-20 09:46:31</p> | <p>分客户归属地会员账户 需求部门: 会员管理部, 商品一部 归口部门: 会员管理部 安全等级: 3 上线时间: 2021-08-20 08:59:20</p> |
| <p>账户数结存统计 需求部门: 会员管理部, 商品一部, 商品二部, 商品三部 归口部门: 会员管理部 安全等级: 3 上线时间: 2021-08-20 10:00:13</p> | <p>境外客户持仓 需求部门: 结算部 归口部门: 结算部 安全等级: 4 上线时间: 2021-08-20 09:50:36</p> | <p>一般法人账户信息变化 需求部门: 会员管理部 归口部门: 会员管理部 安全等级: 4 上线时间: 2021-08-20 10:04:48</p> |

图 5 : 数据地图功能示例图

能，加工数据以满足数据使用部门的使用场景，一键分享给数据使用部门，数据使用部门“收藏后反复使用”，提升数据服务质效。三是提供数据地图导航功能。一站式展示所有已上线的报表

功能，包括报表名称、需求部门、归口部门、数据级别等信息，支持下钻查看报表说明、数据指标计算公式和示例图表等信息，如图5、图6所示，便于使用人员全局了解数据资产。四是提供

报表名称: 公共行情主题

报表类型: 多维分析

报表目录: 多维分析/公共主题

上线时间: 2021-11-05 10:17:04

需求部门: 结算部

归口部门: -

安全等级: -

报表说明: 自助分析行情中结算价、最高价、最低价等指标在合约类、品种类维度下的情况。

数据来源: BI_BI_PUB_QUOT_SUBJ

关键词: 业务日期,品种代码,品种简称,合约代码,合约月份,前收盘,今开盘,今收盘,最高价,最低价,结算价,昨结算,涨跌,持仓量,昨持仓量,持仓变化,交易量,交易额,交割结算价,隐含波动率

备注: -

报表示例:

| 合约月份 | 交割月 | 品种代码 | 品种简称 | 合约代码 | 合约月份 | 前收盘 | 今开盘 | 今收盘 | 最高价 | 最低价 | 结算价 | 昨结算 | 涨跌 | 持仓量 | 昨持仓量 | 持仓变化 |
|--------|------|------|------|--------|------|--------|--------|--------|--------|--------|--------|--------|----|---------|------|---------|
| AP2109 | 2019 | AP | 平糖 | AP2109 | 2019 | 6.185 | 6.199 | 6.500 | 6.698 | 6.177 | 6.210 | 6.210 | 0 | 1,308 | 0 | 1,308 |
| AP2011 | 2011 | AP | 平糖 | AP2011 | 2011 | 6.403 | 6.615 | 6.615 | 6.405 | 6.518 | 6.442 | 6.442 | 0 | 24,534 | 0 | 24,534 |
| AP2102 | 2012 | AP | 平糖 | AP2102 | 2012 | 6.850 | 7.005 | 7.014 | 6.840 | 6.917 | 6.893 | 6.893 | 0 | 18,468 | 0 | 18,468 |
| AP2101 | 2101 | AP | 平糖 | AP2101 | 2101 | 7.340 | 7.407 | 7.509 | 7.340 | 7.403 | 7.354 | 7.354 | 0 | 225,744 | 0 | 225,744 |
| AP2108 | 2108 | AP | 平糖 | AP2108 | 2108 | 7.875 | 7.909 | 7.919 | 7.875 | 7.449 | 7.874 | 7.874 | 0 | 11,840 | 0 | 11,840 |
| AP2105 | 2105 | AP | 平糖 | AP2105 | 2105 | 7.796 | 7.881 | 7.899 | 7.782 | 7.851 | 7.788 | 7.788 | 0 | 22,200 | 0 | 22,200 |
| CR2011 | 2011 | CR | 郑棉 | CR2011 | 2011 | 12.385 | 12.445 | 12.540 | 12.385 | 12.405 | 12.415 | 12.415 | 0 | 18,000 | 0 | 18,000 |
| CR2101 | 2101 | CR | 郑棉 | CR2101 | 2101 | 12.740 | 12.755 | 12.850 | 12.715 | 12.800 | 12.750 | 12.750 | 0 | 272,790 | 0 | 272,790 |
| CR2108 | 2108 | CR | 郑棉 | CR2108 | 2108 | 12.805 | 12.900 | 13.040 | 12.805 | 12.870 | 12.915 | 12.915 | 0 | 29,224 | 0 | 29,224 |
| CR2105 | 2105 | CR | 郑棉 | CR2105 | 2105 | 13.000 | 13.100 | 13.200 | 13.040 | 13.115 | 13.075 | 13.075 | 0 | 125,024 | 0 | 125,024 |

图 6 : 数据报表明细描述示例图

数据智能搜索

搜索

功能(17) | 数据资产(79) | 分析指标(10)

- 会员客户仓单注册注销 功能来源: 统计分析
报表说明: 统计时间段范围内会员名下仓单注册注销数量及客户数量。
- 客户仓库仓单情况 功能来源: 统计分析
报表说明: 统计时间段范围内客户在不同交割地点仓单注册注销明细。
- 分品牌仓单注册注销 功能来源: 统计分析
报表说明: 查看分品牌的不同品种仓单信息, 包括等级、生产厂家等。
- 仓库仓单注册注销 功能来源: 统计分析
报表说明: 统计不同交割地点仓库注册注销情况及客户明细。
- 仓库仓单查询 功能来源: 统计分析
报表说明: 查看查询时段内分品种或分仓库每天的仓单量。
- 仓库仓单平均在库情况 功能来源: 统计分析
报表说明: 统计查询时段内每个品种在不同仓库仓单平均在库时间, 仓单费、仓单注册量及注册量情况。
- 不同交割方式交割情况 功能来源: 统计分析
报表说明: 在选定的时间范围内, 逐月查询每月挂牌合约的期转现量、总交割量、分交割类型(集中/滚动) 分别的交割量、分交割方式(仓单/车船板) 的交割量, 以及分交割类型、交割方式的占比情况。

大家都在搜

- 1 仓单
- 2 交割率
- 3 客户
- 4 定单
- 5 持仓
- 6 交割
- 7 交易
- 8 单边市
- 9 会员数
- 10 期权

共 17 条记录 第 1/3 页
< 1 2 3 > 跳至 页

图 7 : 数据搜索功能示例图



图 8：智能搜索功能示例图

数据智能搜索功能。基于 NLP（Natural Language Processing，自然语言处理）的数据搜索，便于用户快速定位功能报表、数据资产和分析指标，变树形查找为符合用户习惯的一键搜索，如图 7 所示。此外，正在研究数据智能搜索引擎技术，利用 NL2SQL（Natural Language to SQL）技术，实现期货业务语义的 NLP 识别，支持输入自然语言后输出结果图表、数据智能问答、数据需求及时收集，如图 8 所示。

2.6 构建数据模型

数据模型指数据组织和存储方法，从业务使用角度，分类组织和存储数据，实现对数据内容以及数据实体和属性之间关系的一个可视化的表示。数据模型设计的六大原则：高内聚和低耦合、公共处理逻辑下沉及单一、数据可回滚、核心模型与扩展模型分离、成本与性能平衡、规范性。目前主流的建模方法论主要有 Bill Inmon 提出的范式建模和 Kimball 提出的维度建模。两者主要

表 1：主流建模方式的对比

| 特点 | 范式建模 | 维度建模 |
|--------|----------|---------|
| 数据一致性 | 高 | 低 |
| 数据冗余度 | 低 | 高 |
| 项目投产周期 | 长 | 短 |
| 建模团队要求 | 高 | 中 |
| 后期运维难度 | 低 | 高 |
| 模型敏捷度 | 低 | 高 |
| 组成要素 | 主题、实体和关系 | 维度表、事实表 |
| 适应业务场景 | 稳定 | 迭代快 |

特点对比如表 1 所示。其中，维度建模更适合敏捷开发的数仓、数据集市领域。

维度建模实施过程主要由以下六步组成：第一步数据调研，进行充分的业务调研和需求分析；第二步面向业务分析，将业务过程进行抽象，划分并确立主题域（详见表 2. 典型主题域及其内容）；第三步构建业务过程矩阵，确立事实表（委托、成交、持仓等度量信息）；抽象出维度信息，

和维度表（会员、客户、品种、合约、日期、地区等描述信息），并确认事实表、维度表的关系；第四步指标规范化定义，对业务需求进行抽象，整理出相关指标体系；第五步设计模型，依据事实表、维度表，构建公共汇总模型（公共维度层）和应用汇总模型（应用集市层），详见图 9 数据模型架构）；第六步，最后进行 ETL 代码开发。数据模型架构示例如图 9 所示。后续将参考证券

表 2：典型主题域示例

| 主题域 | 内容 |
|------|-------------------------|
| 会客 | 会员及客户类信息 |
| 资产 | 仓单、债券等生命周期管理 |
| 交易 | 委托、成交、行情、行权等 |
| 交割 | 交割环节内容 |
| 结算 | 持仓、手续费、盈亏、保证金等 |
| 风控 | 套保额度、监控事件、大户报告、实控组信息等 |
| 财务 | 记录凭证、科目、总账等 |
| 外部数据 | 现货档案、工商信息、宏观经济等 |
| 公共维度 | 编码、参数、品种及合约、各种参与实体等维度信息 |



图 9：数据模型架构示例图

期货业数据模型等相关标准规范，不断优化完善数据模型设计。

3 总结思考

“数据治理没有完成时”。以体系建设、规范制度为基石，定期开展“体检”工作，不断优化

数据治理工作的技术实践与业务运行治理，逐步建立科学、有效、安全、稳定的数据治理体系，探索并固化“最优流程解”，形成可持续的长效机制，确保数据治理体系运行得益。持续将数字元素注入金融科技服务全流程，将数字思维贯穿业务运营全链条，切实发挥数据的核心要素作用，推动交易所数字化转型。

数据中台数据标准化设计与实践

——上交所数据主题域顶层设计

王芳、韩凤宁 / 上交所技术有限责任公司 技术开发总部 上海 200120 邮箱: fangwang@sse.com.cn



近年来，随着企业数字化转型工作在各行各业的深入开展，数据中台已逐步在金融、互联网等行业建成落地。数据中台的建设理念相比于传统的数据仓库和大数据平台有着较大的不同，数据中台更加重视打破数据孤岛及解决烟囱式开发问题，注重数据的体系化管理、标准化建模和产品化展示，强调数据的复用能力和共享能力。其中，数据的标准化建模是构建数据中台模型体系的核心环节，本文结合上交所数据中台的实际情况，以最基础的主题数据层为例，探讨了数据模型标准化的设计思路。

1 背景介绍

上交所（以下简称“本所”）肩负着组织证券市场的有序运转以及对市场参与主体的监管重任。随着业务的不断发展，现有的大数据平台已表现出诸多不足，主要问题有：一是数据源缺失，存在数据孤岛，无法支撑监管业务的全链路融合分析；二是缺乏标准化的数据模型，数据摆

放散乱，指标口径多样化，不便于取用；三是烟囱式开发模式使得大数据平台与各应用系统耦合性高，增加了系统维护成本。为了解决以上问题，2022年伊始，本所启动了数据中台建设项目，项目内容包括丰富数据源、实现数据模型标准化、建设统一的指标库标签库等，以沉淀业务数据、充分发挥数据价值为目标，打造出一个以数据赋能新业务新应用的支撑性平台。

在本所实际的业务场景中，各业务部门常常要结合业务场景做一些数据的统计分析，通过输出报表等方式服务于部门的运营与决策。如果把业务需要的数据比作果实，那么数据中台就是果园，数据模型就是果树，要想收获不同品种的优良果实，就必须要把果园里的果树修剪嫁接好。同理，数据中台数据标准化的过程就是修葺数据模型的过程，数据模型构建的好坏直接影响到数据中台的数据质量和服务能力。所以，在数据标准化的过程中，一方面要有科学的方法论作指导，另一方面也要充分借鉴行业案例，以理论联系实际的方式构建出最佳实践结果。要把纷繁复杂的数据源进行标准化，首先要建立起清晰的数据层次，然后逐层制定出数据体系并打磨和完善，本文将围绕着数据中台最核心的主题数据层阐述数据标准化的建设思路。

2 数据中台模型体系

本所数据中台的模型体系整体可以归纳为原始数据层、主题数据层和汇总数据层三个层次。对数据进行分层设计的目的主要有：一是将复杂问题简单化，把数据的处理逻辑拆分成多层来完成，明确每一层的功能定位，降低任务之间的耦合度；二是减少重复开发，规范数据分层，通过主题层能将明细数据标准化，通过汇总层能减少指标重复计算，做到一处计算多处使用，提高数据的复用性；三是隔离原始数据，从数据的敏感性等方面考虑，将原始数据与统计数据隔离开，做好数据安全的分类分级管理。整体上做到结构清晰、层次分明，提升了开发效率，降低了维护成本。

原始数据层，又称为 ODS 层（Operation Data Store），主要用来存放从上游业务系统采集的或外部采购的原始业务数据，定位于保持数据原貌，并且需要保存全部历史数据。

主题数据层，又称为 DWD 层（Data Ware-

house Detail），主要是对原始数据层进行标准化建模，基于每个具体的业务特点，构建最细粒度的明细层事实表。主题数据层需要对原始数据层进行清洗（去除脏数据）、融合（按数据内容合并或拆分），将原始数据层数据按业务主题归类整理，并明确每个主题之间关系。

汇总数据层，又称为 DWS 层（Data Warehouse Service），是在主题层基础上，结合业务分析视角，将数据按维度和指标进行汇总，基于上层的应用需求构建通用的指标事实表。

从上述三层体系的描述中可以看出，主题数据层是数据建模的核心区域，这一层要对数以万计的上游接口进行清洗、归纳和融合，是整个数据中台最基础的一层。

3 主题层建模原则

为保证模型的稳定性和对业务支持的灵活性，本所数据中台在设计主题层模型时一般遵循以下原则：

1) 业务继承性原则：数据中台在进行模型设计时，对于业务实体的相关概念，原则上以继承业务系统定义为主，不提出新的概念。

2) 稳定性原则：在设计模型时，需要根据数据内容的特性选择稳定的模型结构，以屏蔽上游数据接口变化对数据中台带来的影响。

3) 前瞻性原则：在设计模型时，基于当前已有的数据做适当超前的规划，以适应业务的未来的发展，尽量保持模型的稳定性。

4) 兼顾实际原则：尽量兼顾上下游系统实际情况及数据中台本身的技术特性、性能特性，以建模理论为基础，结合实际，在运用建模理论的同时做一定的变通。

5) 实用性原则：在进行主题层设计时，兼顾模型结构一致、数据内容清晰、数据无交叉和数据不冗余等方面因素，以方便使用为原则进行整合，形式上可以是多张表合并成一张表，也可

以是一张表拆分成多张表。

6) 数据源可靠度：分析本所内部数据与外部数据可靠度和数据质量，尽量保证所内数据相互融合，所外数据单独融合；对于同一数据存在多头数据源的，优先使用所内数据，其次使用所外数据。

7) 按需整合原则：原始数据层整合到主题数据层时需要着眼于数据内容，优先整合有明确分析价值的的数据，对于一些过程数据、备份数据、零散数据等价值尚不明确的数据，暂时放在在原始数据层，待后续明确后再提到主题数据层。

4 主题层模型构建思路

4.1 主题域划分 (Level 0)

在构建主题数据层模型时，首先要划分数据的业务主题。本所数据中台在划分业务域主题时，是以证监会 SDOM 主题划分标准为指导，同时参考了 FS LDM 金融行业主题划分标准，结合交易所业务全景视图，将数据范围扩大至金融行业、资本市场和宏观经济等层面，对所涉及的全部业务数据进行抽象与归纳，最终形成本所业务数据主题的顶层视图 (Level 0)，如图 1 所示：

顶层视图中描述了八大主题以及主题之间的关系，下面详细介绍每个主题包括的具体内容：

1) 参与人主题

参与人 (Participant, 简称 PAR) 是指证券市场活动中的所有参与主体。包括发行人、中介机构、服务机构、从业人员、高管、自律性组织和证券监管机构的基本信息及以参与人为主体的附加属性，例如评价、知识产权、董监高、主要科研人员、交易权限，以及其之间的关系等。参与主体由名称、证件类型和证件号码、统一社会信用代码唯一标识。

2) 账户主题

账户 (Account, 简称 ACC) 是指参与人持有资产及参与各类证券行为 (如委托、交易、结算、过户、分红等) 的载体。包括证券账户、资金账户、期货账户、资管账户、银行账户、账户权限等。

3) 资产主题

资产 (Asset, 简称 AST) 是指参与人进行证券产品投资所拥有的资产信息，主要包括证券、期货、基金持仓及其变动、资金及其变动等。

4) 产品主题

产品 (Security, 简称 SEC) 是指在交易所交易或被监管的各类投资品种，包括证券市场上

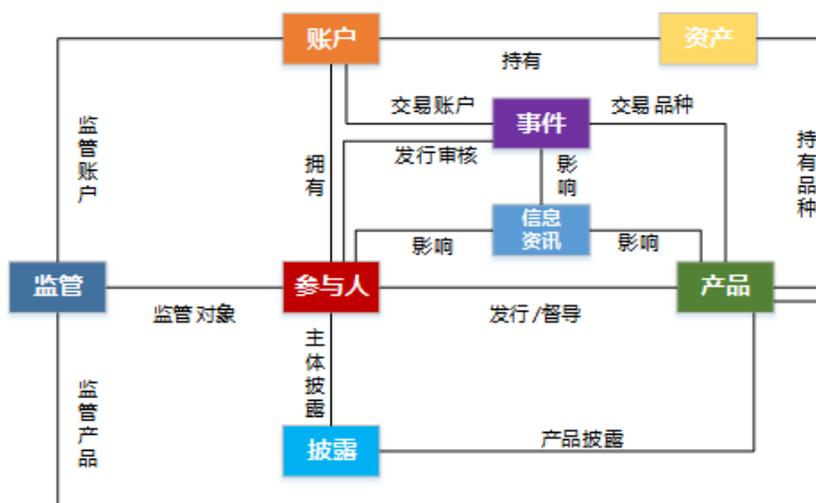


图 1：数据中台主题域划分 Level 0

的股票、债券、基金、衍生品等，期货市场上的期货合约及期权，场外市场的产品（收益凭证、资管计划、私募基金等）及衍生品等，以及由上述品种编制所得的指数。除记录其基本信息外，还记录其行情、业务控制信息等。

5) 事件主题

事件（Event，简称 EVT）是指投资者参与证券期货市场的所有行为，以及该行为触发的证券期货市场其他参与者的所有行为。如上市公司的审核、发行、回购行为，投资者新股申购、竞价委托及交易行为，结算公司为上市公司及投资者的权益登记行为等。

6) 监管主题

监管（Regulation，简称 REG）是指依照法律法规和上级单位授权，监督管理证券市场秩序，保障其合法运行的行为。监管主题的数据内容包括司法处罚、行政处罚、行业监管、内部监管等。

7) 披露主题

披露（Disclosure，简称 DSC）是指参与主体依照法律规定将其自身的财务变化、经营状况等信息和资料向监管部门和证券交易所报告，并向社会公开或公告的数据。

8) 信息资讯主题

信息资讯（Information，简称 INF）是指第三方机构或自然人对证券市场的评价，以及其他与证券市场有关的数据。包括宏观数据、市场舆情、法律法规等。

4.2 主题一级分解（Level1）

根据自顶向下、逐层分解的模型设计方法，在明确划分了数据模型业务主题后，需要进一步对各个主题进行分解，将其分解到子主题的层次。对数据主题进一步分解后形成的一级视图（Level 1）如图 2 所示。



图 2：数据中台主题模型一级视图 Level1

参与人的子主题包括：发行人（Issuer，简称ISR）、机构（Institution，INS）、参与人关系（Relationship，简称RLT）、业务控制（Control，简称CTL）。

账户的子主题包括：账户信息（Basic Information，简称BSC）、账户分类（Classification，简称CLS）、业务控制（Control，简称CTL）。

资产的主题有：金融产品持仓（Financial Product，简称FIN）、资金持仓（Capital，简称CAP）、商品持仓（Commodity，简称CMM）、业务控制（Control，简称CTL）。

产品的子主题有：产品信息（Basic Information，简称BSC）、产品行情（Quotes，简称QTS）、业务控制（Control，简称CTL）。

事件的子主题包括：发行（Issue，简称ISS）、交易（Transaction，简称TXN）、权益变动（Equity Change，简称CHG）。

监管的子主题包括：司法监管（Judicial，简称JUD）、行政监管（Administrative，简称ADM）、行业监管（Industrial，简称IND）、内部监管（Internal，简称ITR）。

披露的子主题包括：招股公告（IPO，简称IPO）、定期公告（Periodical，简称PER）、临时

公告（Temporary，简称TMP）。

信息资讯的子主题包括：研究报告（Research，简称RSH）、舆情信息（News，简称NWS）、宏观经济（Macroeconomics，简称MCR）、法律法规（Laws，简称LAW）。

以上主题的简称都将运用在后续的物理建模过程中，作为数据元标准贯彻于整个数据中台的模型定义。

4.3 主题二级分解（Level2）

对一级视图进行进一步分解，可以得到各个一级主题包含的主要业务实体。关于这些实体的概念和实体间的关系将在后续文章中对每个主题的概念模型进行详细描述。下面先以图的形式展示各主题所包含的一级主题及主要业务实体。

- 1) 参与人主题
- 2) 账户主题
- 3) 资产主题
- 4) 产品主题
- 5) 事件主题
- 6) 监管主题
- 7) 披露主题
- 8) 信息资讯主题



图 3-1：数据中台主题模型 Level2- 参与人主题



图 3-2：数据中台主题模型 Level2- 账户主题



图 3-3：数据中台主题模型 Level2- 资产主题



图 3-4：数据中台主题模型 Level2- 产品主题



图 3-5：数据中台主题模型 Level2- 事件主题



图 3-6：数据中台主题模型 Level2- 监管主题



图 3-7：数据中台主题模型 Level2- 披露主题



图 3-8 : 数据中台主题模型 Level2- 信息资讯主题

5 总结

综上所述，本文围绕着上交所数据中台的主题数据层探讨了数据标准化的设计思路，在介绍了数据中台模型体系后，重点讨论了主题数据层中八大主题的划分，并对每个主题进行了逐层分

解，最终得到各个主题包括的主要业务实体。后续将会随着业务的拓展，持续对标准数据模型进行推敲打磨和优化完善。下一篇，我们将会重点讨论各主题的概念模型，从业务视角描述每个主题包含的核心业务实体以及实体之间的主要业务关系。

核心业务系统连续性保障的探索与实践

张如意¹, 刘永亮², 孙维², 王建超³, 刘相富², 雷晓全¹, 于三禄¹ / ¹ 郑州商品交易所 信息科技一部 郑州 450018 ² 郑州商品交易所 信息科技二部 郑州 450018 ³ 郑州易盛信息技术有限公司 清算系统部 郑州 450018
邮箱: ryzhang@czce.com.cn



本文主要介绍了交易、结算等核心业务系统连续性保障方面的一些探索与实践，重点讲述了核心交易系统主备中心异构部署方案，交易网关和结算系统异构并行的实践，通过软件异构的方式丰富风险防范处置工具箱，提升核心业务系统连续性保障能力。

1 背景概述

近两年国外交易所安全运行事件多发，安全运行和应急响应能力建设受到了越来越多的关注。期货市场的杠杆交易、保证金制度、“T+0”双向买卖、持仓限额、夜盘连续交易、每日无负债结算等机制，增加了核心业务系统安全运行的复杂性和艰巨性。交易所作为中央对手方，交易、结算等核心业务系统的连续性保障，事关国家金融安全、社会秩序和投资者合法权益。如何强化交易、结算等核心业务系统的连续性保障，通过人防技防相结合，不断提高系统容错能力和应急处置能力，筑牢安全生产防线，是行业技术人员

面临的“终身课题”。

2 连续性保障面临的风险与挑战

信息系统天生就是一个存在很多不确定、不稳定因素的复杂脆弱系统，各类硬件设施的故障、软件功能设计的缺陷、恶意的网络攻击、日常运维的疏忽，以及暴雨、地震等自然灾害，重大公共卫生事件、暴恐事件、火灾等外部因素，都有可能直接或间接导致信息系统故障，引发网络安全事件。《SRE Google 运维解密》一书中，将信息系统比喻成养育孩子，生育（开发）过程很痛苦，但是养育孩子（运维）更耗费精力。生产系统运

行保障，经历了手工操作、脚本运行到自动化工具运维等阶段。近年来随着大数据、人工智能等前沿技术的广泛应用，基于运维大数据、机器学习、预测分析、日志聚类、态势感知等技术的数据化、智能化运维逐渐成为新的焦点。运维体系和方法论方面，ITIL、双态运维，SRE、DevOps等百花齐放，核心是持续优化完善监控系统，高度重视异常事件、变更管理，精确定位并消除安全生产中存在的风险隐患，制定应对各种突发情况的应急预案和常态化的应急演练，畅通开发、测试、运维流程环节，敏捷迭代快速修复、回退，强化应急处置能力建设等。新技术的快速迭代，使我们拥有了更丰富的运维手段，但是不断深化业务理解、持续优化故障分析与决策、丰富完善风险防控工具箱和方法论，仍然是做好交易、结算等核心业务系统安全运行的圭臬。

通过对历年来信息系统运维类事件的梳理分析，常见的故障事件类型、故障点、解决方案和预防措施如下表 1 所示。

其中，机房的电力保障、精密空调、制冷系统、硬件服务器、存储设备、网络设备及专线等，在设计及建设阶段，通过多路冗余部署，避免单点故障，确保系统的高可用性。单点或多点故障发生时可通过自动切换、主备切换、温备替换、主备中心切换等方式避免异常中断。软件类故障

事件，一是每周举行变更评审例会评估重大变更，结合业务流程、突发情况，推演各种灾难场景，制定并测试回退方案和应急预案；二是深入剖析已发生的事件，举一反三预防其他系统发生类似事件；三是通过单元测试、集成测试、仿真环境测试（会员单位测试系统接入环境）、内部联调测试（业务人员参与的集成测试）、全市场测试、冒烟测试（变更上线过程中）等多方式、多层次、多场景的测试，对系统潜在风险做到早发现早处置；四是生产系统每日闭市操作时，多环节冒烟测试；五是业务系统应用层的异构部署、并行运行。典型防范应对策略总结如下图 2 所示。

综上所述，为了提升业务连续性保障能力，软硬件方面均采取多项措施，其中，硬件故障方面主要通过冗余备份和容灾能力建设来有效避免；应用软件缺陷方面，由于核心业务系统的错误容忍度低、业务中断影响范围广、应急处置措施难统一，其安全运行仍面临风险和挑战，难以根除，是核心业务系统安全运行的重要风险隐患之一。

3 核心业务系统的异构实践

异构并行提供了一种行之有效的解决思路和实践路径，通过差异化的运行环境、处理流程和

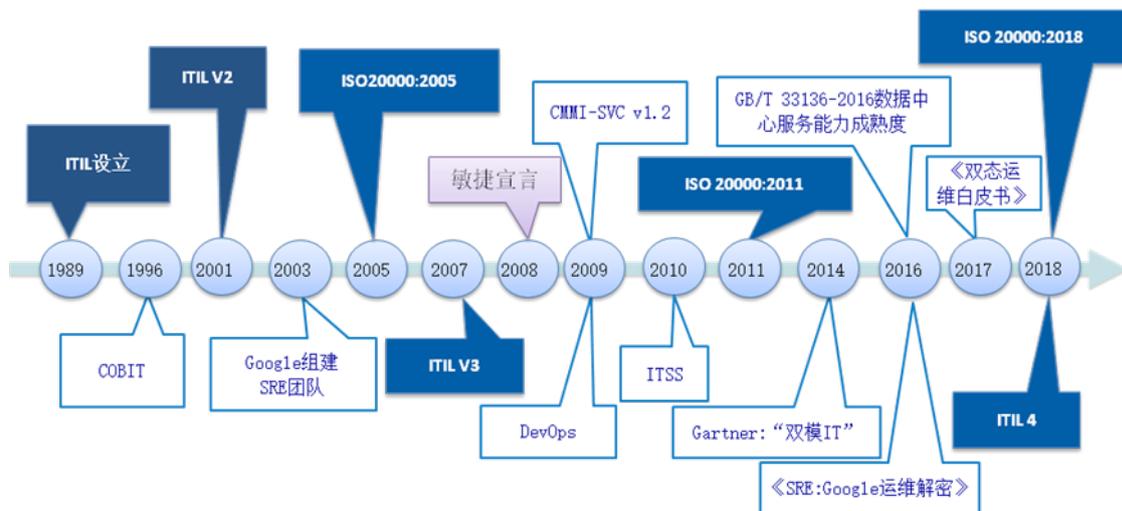


图 1：典型运维体系和方法论

表 1：常见运维事件及其处置方法

| 故障类型 | 故障点 | 解决方案 | 预防措施 |
|-------|--------------|---|--|
| 硬件类事故 | 硬件单点故障 | 多活、双活无影响，主备系统切换，时效性不强的温备替换 | 冗余部署、规避单点故障，确保系统的高可用 |
| | 同一功能模块硬件同时故障 | 评估中断影响和快速恢复方案耗时，采取异源温备设备替换或部分业务切换至备中心运行 | 硬件设备版本缺陷导致，跟踪并使用官方推荐的稳定版本的固件或系统。加强日常巡检，及时获取供应商更新的运维知识库 |
| | 硬件大面积故障 | 主备中心切换 | 容灾能力建设，两地三中心，数据同步机制 |
| 软件类事故 | 软件版本升级变更 | 按照应急预案进行版本回退或其他应急处置 | 重大变更多层次测试、制定并测试回退方案和应急预案 |
| | 容量问题 | 动态扩容、横向扩容 | 可扩展的系统架构、容量监控 |
| | 未知软件功能缺陷 | 根据日志排查故障原因，评估后修改代码，直接跳过部分特定业务环节 | 多层次测试 系统冒烟测试 异构并行 |
| 网络攻击类 | 恶意网络攻击 | 使用有效的防护手段和设备；针对各业务特点定制专属防护策略；安全设备的统一日志分析，形成联防联控机制；合理设置网络隔离，阻止潜在的横向渗透风险；把好新系统新设备上线的安全检查关 | 不断增强安全防护措施，上线 HIDS、动态防火墙、态势感知平台、蜜罐等防护设备，细化防护策略，完善防护体系，持续提升防护水平 |
| 操作类事故 | 业务人员误操作 | 根据具体情况评估后处理 | 两录一核、数据质量监控 |
| | 运维人员误操作 | 根据具体情况评估后处理 | 自动化运维工具 |

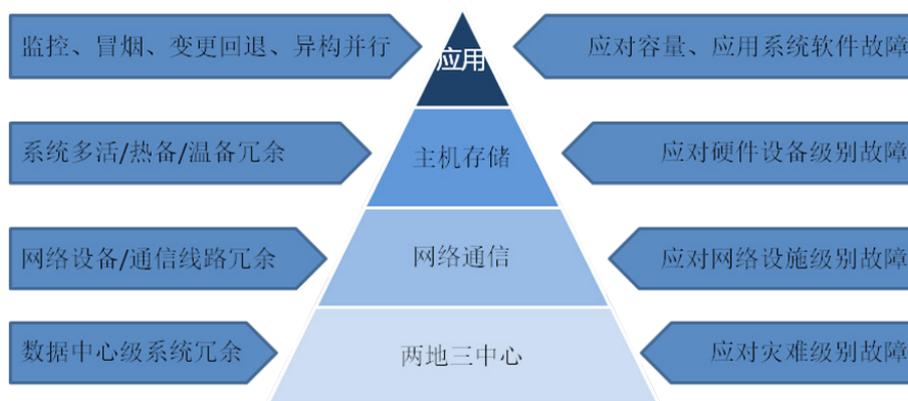


图 2：典型防范应对策略

执行顺序，可以降低软件缺陷导致的业务服务中断风险，有效提升核心业务系统的连续性。其代价是开发、测试、运维多套系统，维护成本较高。异构设计的核心是将功能模块进行拆解、归类、重构、组合，通过改变系统框架、软件架构、编程语言，多组开发团队等方式，最终实现“技术

异构、业务兼容、并行运行”。下面重点介绍交易、结算等核心业务系统的异构实践。

3.1 交易网关异构设计

在会员端、交易网关（Trade gateway，以下简称 GW）、交易撮合系统（Trading Engine，以

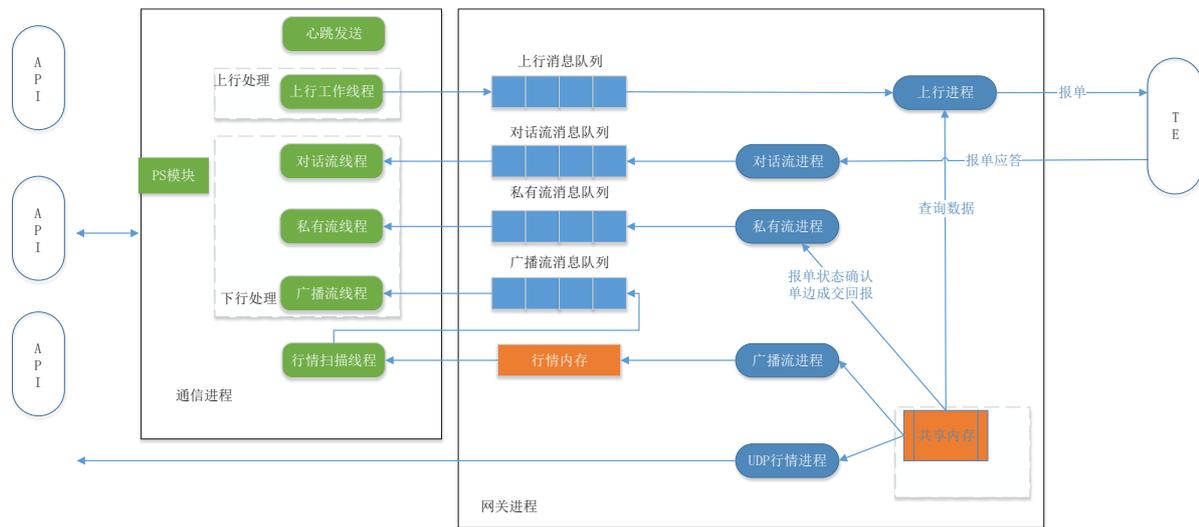


图 3：交易网关 A 系统架构示意图

下简称 TE) 三层结构中, GW 主要作用是减轻 TE 的压力, 减少 TE 对每个会员的查询请求的处理操作, 提高撮合速率。其主要功能包括: 接收来自会员端的请求, 将委托等相关请求发送给 TE (上行处理); 从 TE 接收处理结果信息, 并向会员端发送处理结果信息 (对话流); 响应会员端的查询请求, 返回需要查询的信息 (对话流); 生成成交回报信息并向会员端发送 (私有流); 生成行情信息并向会员端发送 (广播流)。

交易网关 A 采用 C 语言多进程的设计, 系统架构如图 3 所示。上行进程负责接收来自会员端的请求, 根据请求类型决定是自己处理还是交给 TE 处理, 其中事务处理 (入单, 撤单等) 转交 TE 处理, 查询请求自己处理; 对话流进程负责接收来自 TE 的事务处理结果, 转发给相应会员端; 私有流进程负责查询共享内存数据, 生成报单状态确认等私有流数据, 发送给会员端; 广播流进程负责查询共享内存数据, 生成行情等广播流数据后放到行情共享内存; 通信进程采用多线程, 利用多路复用机制, 负责与会员端的连接管理和网络通信。其中, 所有消息队列均采用了无锁环形队列技术, 具备低时延、高并发等特性。

交易网关 B 使用 C++ 语言, 基于 ACE (Adaptive Communication Environment) 通信库的反应器模式, 采用单进程多线程的架构来实现, 系统架构如图 4 所示。连接接收线程, 负责处理链路协商, 登录数据包处理, 登录链路分配等连接管理; 对话流线程负责处理对话流数据的收发以及与 TE 的数据通讯; 查询处理线程负责处理会员端查询请求; 私有流处理线程负责处理成交回报等私有流数据的发送; 私有流数据监控线程, 负责监测新生成的私有流数据, 并及时通知私有流处理线程将数据发送给会员端; 广播流处理线程负责处理登录到广播流的连接的数据发送, 包括行情和广播信息的发送。

交易网关 A 和网关 B 对外面向会员端 API 兼容, 会员端无需修改可无感切换, 对内与 TE 的通信协议一致、接口兼容。两者的主要区别在于: 开发语言不同、技术架构不同、功能模块划分不同。对于定单报送、定单状态回复和业务查询等三个核心功能部分, 网关 A 将定单状态回复单独处理, 另一个进程负责定单报送和业务查询, 网关 B 将业务查询单独处理, 一个线程负责定单报送和定单状态回复, 另一个线程负责业务

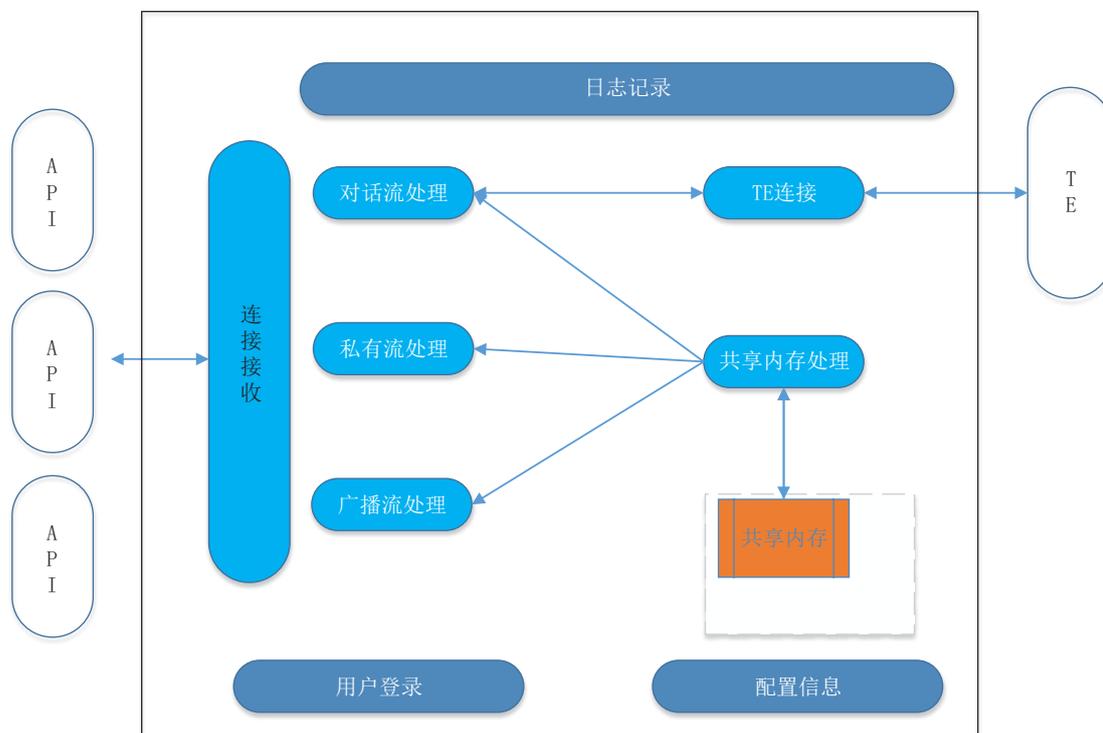


图 4 : 交易网关 B 系统架构示意图

查询处理。通过采用统一心跳源，保障行情数据切片的一致性。交易网关 A 和网关 B 输入入口和输出出口均一致，在应急处置过程中，通过自动切换脚本，在交易网关 A 或 B 发生软件故障时，可快速双向变更切换，丰富了应急处置的工具集。

3.2 交易撮合系统主备中心异构部署

交易撮合系统 TE 主要负责定单的集中撮合，由于体系架构限制，目前 TE 尚未实现异构并行。TE 的撮合主进程（以下简称 T 进程）负责具体逐笔定单的撮合处理，业务逻辑变化及技术架构变动均会修改 T 进程，引入故障风险的可能性较大。T 进程一旦发生未知软件功能缺陷故障，现场根据交易日志进行故障排查，评估后修改代码，通过紧急变更进行解决，相关应急处置耗时一般以小时为单位计量，给交易连续性带来巨大挑战。

通过分析 T 进程的运行机理，该进程主要运行逻辑为循环处理每一条报单数据，如图 5 所示。

其软件故障高概率出现在处理某一条报单数据时触发缺陷，故障表现为 T 进程异常退出，造成交易中断。T 进程发生此类软件缺陷时，可通过删除引发故障的最后一条报单数据，重新启动系统，重撮交易日志至故障前位置后，继续进行交易；同时采取针对性限制措施，防止引发故障的同类报单进入系统。但重撮耗时时间较长，相关故障排查、会商决策、应急处置等很难在 5 分钟内完成。

通过主备中心交易系统异构部署的方式，利用灾备中心日志中转系统缩减重撮耗时、提升异常报单导致系统故障的快速恢复能力。交易系统主备异构部署方式如图 6 所示。核心思路是通过灾备中心缓存一笔交易日志（含心跳日志），当主中心 T 进程故障时，灾备中心交易备机及网关均处于正常状态，此时舍弃缓存的最后一笔交易日志，再进行主备中心切换，省去了重撮耗时。同时，主备中心异构部署使得缺陷指令类的软件

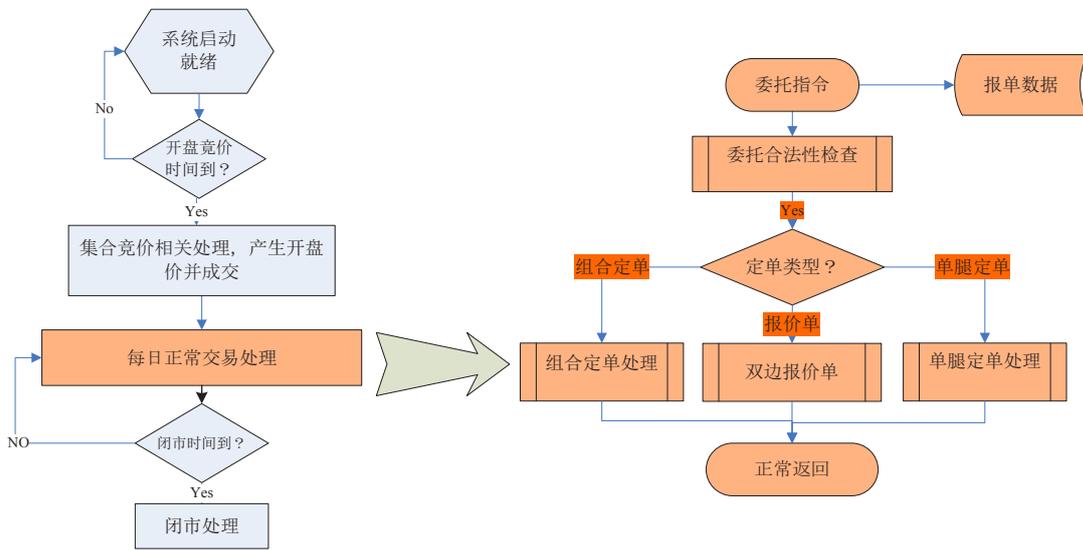


图 5：撮合主进程 T 进程运行逻辑示意图

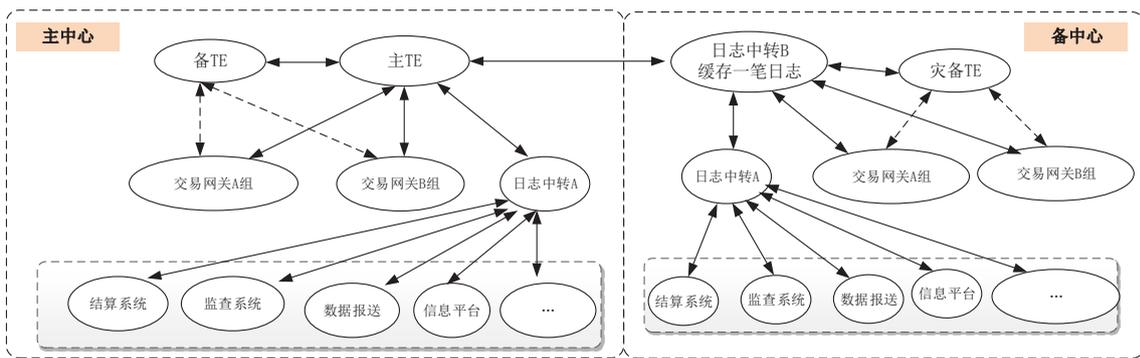


图 6：主备中心异构部署示意图

故障提前暴露，减小故障误判可能性，提升处置效率。

主备中心异构部署的关键在于备中心缓存了一笔交易日志，其具体实现方式如下：日志中转 B 在发送交易日志时缓存最后一笔（含心跳日志），当收到新的一笔日志时，才将缓存的日志发送，并继续缓存新的一笔日志。灾备 TE 及灾备网关连接日志中转 B 拉取交易日志，不含最后一笔日志。循环缓存逻辑原理如图 7 所示：0. 计算缓存一笔日志后的文件大小；1. 判断缓存标志是否不为 0，更新目前文件最新大小；2. 判断最新大小是否超过缓存一条日志后的文件大小，如否，delay 后转步骤 1，如是，转步骤 3；3. 发送

缓存的一条日志；4. 更新缓存新一笔日志后的文件大小，转 1。其中缓存标志采用共享内存方式创建及修改。

3.3 结算系统异构并行

期货市场的夜盘交易使得结算系统变更时间窗口、异常事件处理、应急处置等时间紧张，通过两套结算系统异构并行，提升结算业务的连续性保障能力。结算系统是典型的“输入 + 计算 + 输出”系统，通过多路输入、异构计算、对比输出，确保两套结算系统“技术异构、并行运行、业务兼容”。

结算系统 A 采用数据库存储过程的方式完

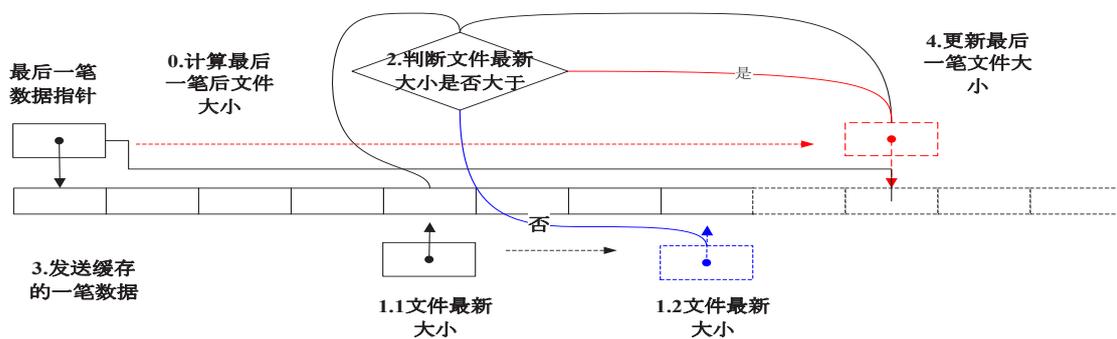


图 7：循环缓存逻辑原理示意图

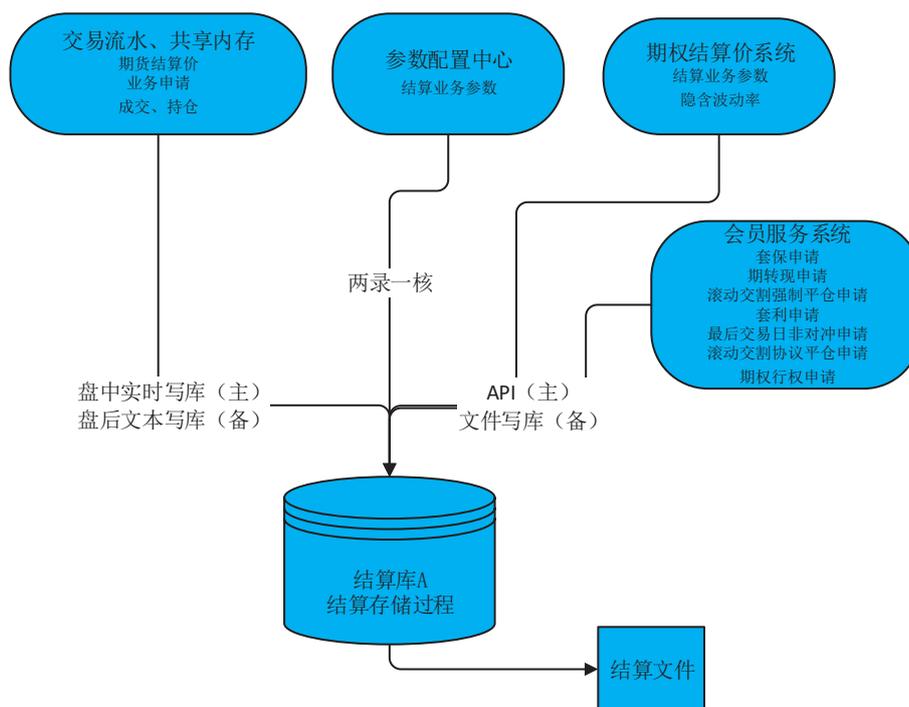


图 8：结算系统 A 运行方式示意图

成结算，其运行方式如图 8 所示。通过盘中实时写库和盘后文本写库，主、备两种异构方式，将成交、持仓、行情等数据写入结算库 A；结算参数经过两录一核写入结算库 A；业务申请数据、期权结算价及波动率通过 API 和文件写库两种异构方式，写入结算库 A；调用结算库 A 的存储过程，执行结算操作；结算操作将结算结果直接写入数据库中，用于结算文件生成和下一交易日交易系统初始化。

结算系统 A 将结算动作分为结算前准备、

交易结算和结算整理三个模块，均由存储过程实现。三个模块以数据库表为接口串行执行，其中交易结算模块支持多进程并发执行。其运行方式如图 9 所示。结算前准备模块负责进行持仓调整等业务。交易结算模块负责进行盈亏、手续费和保证金的计算。该模块通过对成交、持仓、会员资金等数据进行分组，分组个数(参数 N)可配置；在结算数据库 A 上创建 N 个任务，由存储过程触发任务运行；每个任务运行的参数不同，处理的数据不同，启动的进程也不同；通过界面监控

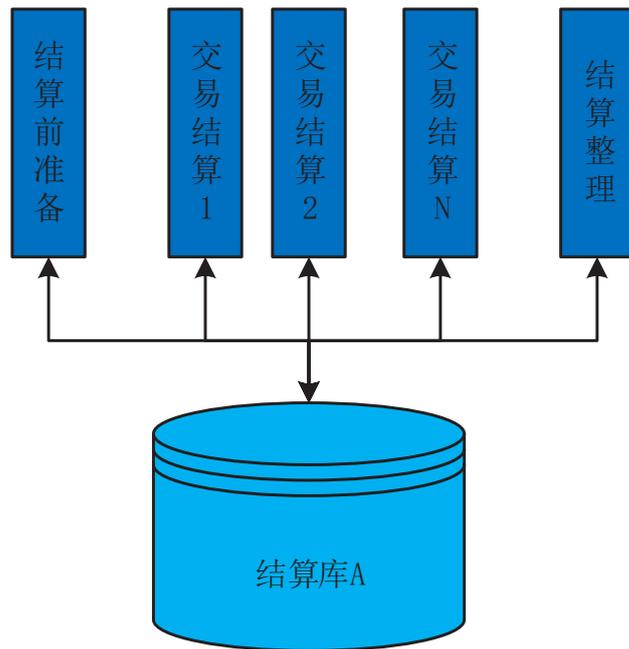


图 9：结算系统 A 功能模块示意图

所有任务的运行情况，所有任务都正常执行成功后，交易结算模块处理完成。结算整理模块负责合约和资金整理等业务。

结算系统 B 采用 C++ 和 Java 语言，通过分品种组多线程的方式在内存中完成结算。其运行方式如图 10 所示。相对于结算系统 A，其获取成交、持仓、业务申请指令、期货行情及结算价等主要结算源数据的方式有两种模式共三种方式，其中，主模式通过交易流水直接写入内存，备模式复用结算系统 A 的盘中实时写库和盘后文本写库，确保结算依赖的源数据“主备多路可行”。通过内存结算，将结算结果写入结算库 B 中，生成与结算系统 A 结果一致的数据，通过两个数据库之间的数据对比，确保结算正常。

结算系统 B 主要包括交易结算管理子系统、任务调度子系统、数据服务子系统和内存结算子系统四个子系统，子系统间关系和功能概述如图 11 所示。交易结算管理子系统负责结算计划的制定，管理子系统还需要从结算库同步必要的用户信息和品种信息，以及对业务库中的结算业务数据进行读写操作。任务调度子系统负责从管理

子系统获得结算计划以及结算的指令（包括计算前检查指令、交易结算指令以及生效入库指令），从内存结算获得结算的进度详情，并将进度详情反馈给管理子系统。数据服务子系统包括交易数据中心的基礎服务和结算数据快照服务组成。其中交易数据中心基础服务从交易系统获取交易事件流，并为本地其他进程提供交易事件数据访问接口；结算数据快照服务从数据服务子系统分别从数据库获得相应的结算数据以及结算计划中的配置信息，接收到任务调度的快照请求后完成快照并将结果反馈给任务调度。内存结算子系统从任务调度获得结算指令，从数据服务获得结算数据进行结算，向任务调度反馈结算进度信息，结算完成后将结算结果写入到结算库中。内存结算子系统从任务调度获得结算指令，从数据服务获得结算数据进行结算，向任务调度反馈结算进度信息，结算完成后将结算结果写入到结算库中。

两套结算系统的主要区别在于编程语言、运行环境、源数据获取等方面。此外，我们通过每日进行两套系统结算结果对比、盘中交易间歇期间进行结算冒烟、发布分时结算参考价等措施，

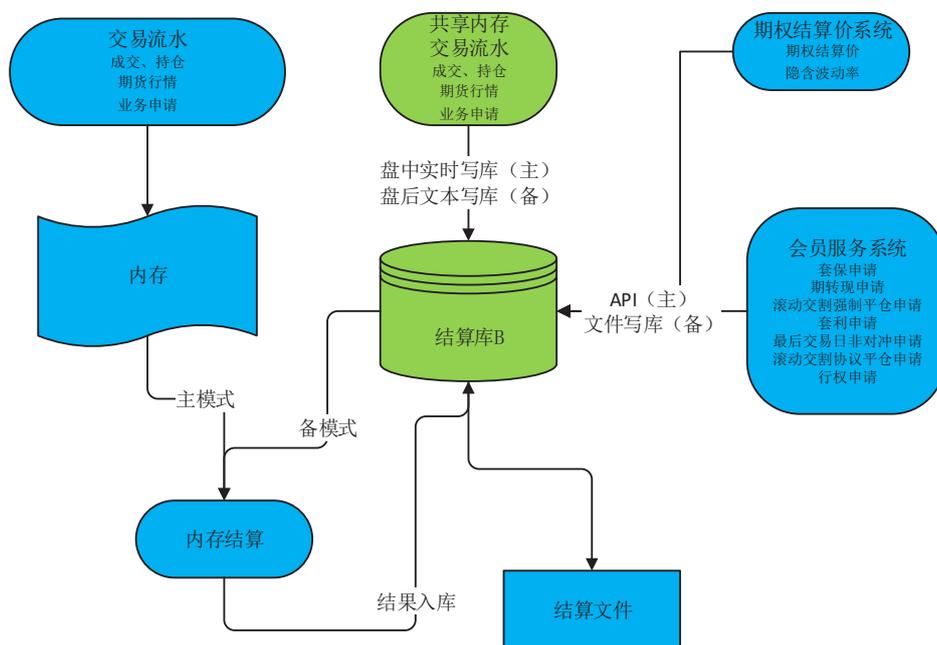


图 10：结算系统 B 运行方式示意图

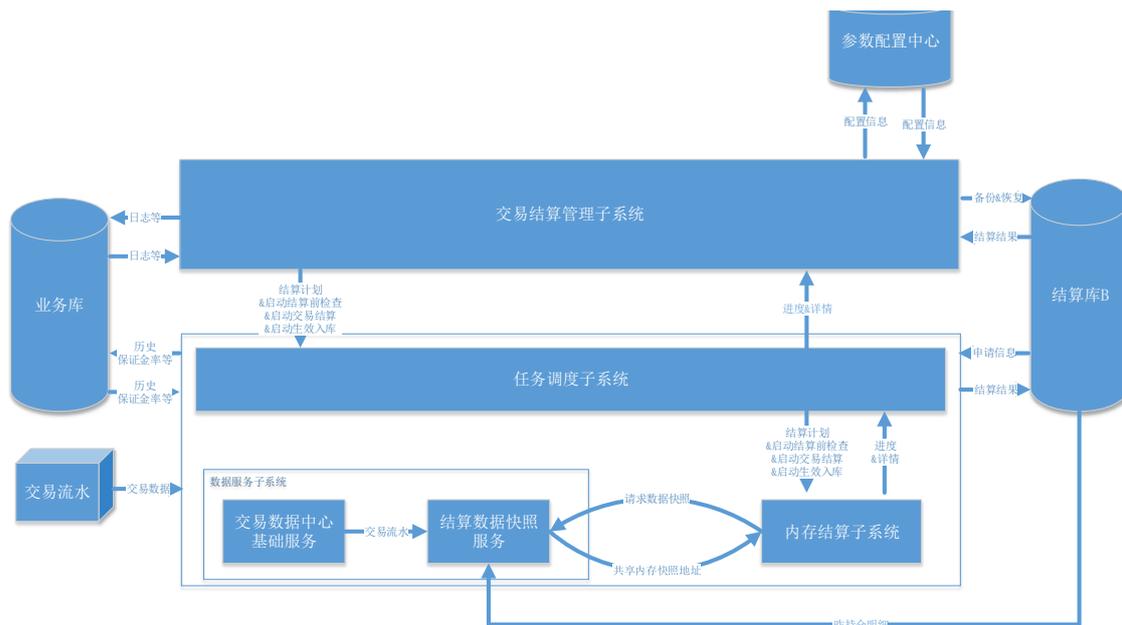


图 11：结算系统 B 功能模块示意图

提升结算业务连续性保障能力，守牢结算风险隔离底线。

4 总结

核心业务系统连续性保障是个系统工程，需以“行百里者半九十”的精神，坚持系统观念，

强化风险意识、树牢底线思维，聚焦关键环节，在发现问题、找到症结、消除瓶颈、提升韧性上下足“绣花”功夫。不断完善应急预案和处置流程，持续丰富切换、回退、冒烟、异构等处置工具箱，努力提升应急响应、应急处置、快速恢复能力，确保核心业务关键时刻“多条腿走路”，以做好核心业务系统的连续性保障工作。

构建邮件纵深安全体系，提升互联网重要入口安全运营水平

——海通证券安全运营实践分享

沈俞超、徐文韬、秦军、郑熊、马冰、王东 / 海通证券数据中心



一、海通证券邮件安全体系演进

在近年来举行的国家级攻防演练和常态化攻防对抗中，邮件系统和互联网暴露面成为攻击队撕开缺口的重点方向。“外科手术式精准打击有时比地毯式饱和轰炸更有成效”。相对于正面战场的硬刚，一封针对重点人群例如核心运维、公司高管的精准钓鱼邮件，往往成为企业防线的“阿克琉斯之踵”，极有可能被撕开缺口，损失惨重。中外多个披露的 APT 行动中，相关技战术展现的淋漓尽致。因此，我们认为关基单位针对钓鱼邮件的防护体系，包括安全意识宣贯，不能把他“娱乐化”，而是应该高度重视，全方面构建对抗体系，从管理、技术两个方面实现：重点人群抓严、心理博弈抓细、技术对抗抓实，情报共享出

口抓牢，形成全体关基单位一盘棋的局面，编织起全国关基行业的严密防钓鱼“天网”。

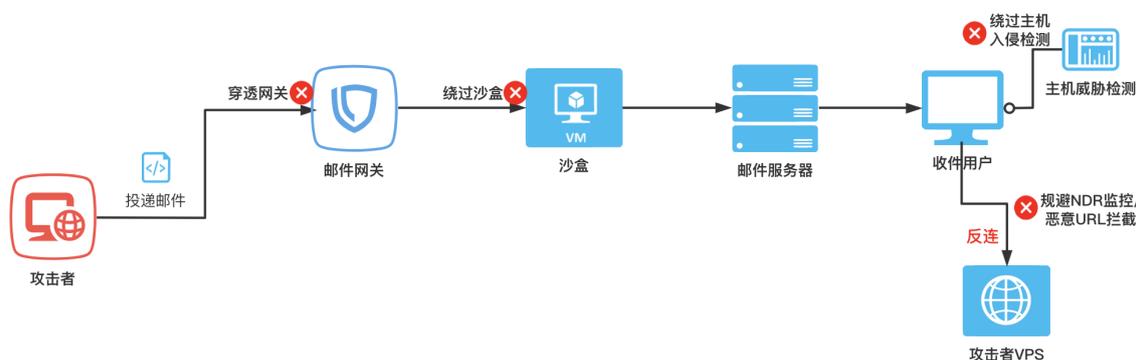
海通证券的邮件系统防护理念和水平是通过不断的对抗逐步提升的。我们在邮件安全攻防对抗过程中，逐步总结出邮件安全演化所经历各个阶段，从而针对邮件安全体系进化做到了有的放矢。

二、邮件系统两大重点威胁

2.1 攻击者攻击过程

Step1. 绕过邮件网关：

通过新型的钓鱼邮件、商业欺诈邮件、附件样本变形的恶意邮件、APT 邮件等多种形式，直接穿透以规则过滤的邮件网关。



Step2. 绕过沙箱检测：

对攻击邮件的附件部分进行压缩加密、加壳、内容混淆、远程加载宏等方式处理，使得通过静态查杀或简单动态查杀的沙箱无法发现附件中藏有的恶意威胁文件。

Step3. 绕过主机入侵检测：

恶意文件调用底层函数进内核，或是使用偏僻语言的小众工具，在主机入侵检测的盲区内进行运行。欺诈用户点击精钩邮件的 url 则不在主机入侵检测的防范范围内。

Step4. 绕过 NDR/ 恶意 URL 拦截：

当文件被收件用户点击后，以全程加密的方式和攻击者的 VPS 建立连接，使得 NTA 无法发

现传输内容中的命令执行；或是在最初注册全新域名，使得恶意钓鱼域名不在威胁情报库的范围内，绕过针对 DNS 解析的恶意域名拦截。

海通证券总结得出在邮件维度，攻击者的攻击主要可分为邮件本体攻击和邮箱资产攻击，将在下文 2.2 及 2.3 中做详细介绍。

2.2 邮件本体攻击

通过攻防对抗复盘，恶意邮件的利用方式通常是通过两种载体：附件和正文（正文包含文本、图片、链接）。

2.2.1 附件载体

在邮件中暗藏的附件往往具有迷惑性、隐蔽

| | |
|------------------------|---------------|
| 附件：Office FRAMESET漏洞利用 | 附件：LNK文件 |
| 附件：Office 图标伪造 | 附件：COMDOS文件 |
| 附件：Office OLE加密 | 附件：URL快捷方式文件 |
| 附件：Office 溢出漏洞利用 | 附件：IQY文件 |
| 附件：Office 恶意宏公式 | 附件：恶意EXE可执行文件 |
| 附件：Office RTF结构异常 | 附件：恶意JAR可执行文件 |
| 附件：扩展名不一致(done) | 附件：恶意CHM可执行文件 |
| 附件：扩展名未知(done) | 附件：恶意LNK可执行文件 |
| 附件：扩展名欺骗 | 附件：恶意网页快捷方式文件 |
| 附件：扩展名逆名欺骗 | 附件：恶意图标 |
| 附件：扩展名相似字符欺骗 | 附件：恶意IMTA |
| 附件：密码保护 | 附件：恶意字符串 |
| 附件：加壳 | 附件：恶意PACKER |
| 附件：变量混淆 | 附件：恶意IMSC |
| 附件：内容加密 | 附件：恶意SCMH |
| 附件：恶意JavaScript脚本 | 附件：内嵌EXE |
| 附件：恶意VbScript脚本 | 附件：恶意UKPE |
| 附件：恶意WSH脚本 | 附件：恶意PDF文件 |
| 附件：恶意WSF脚本 | 附件：诱骗下载 |
| 附件：恶意HTA脚本 | 附件：钓鱼链接 |
| 附件：恶意BAT脚本 | 附件：SPAM链接 |
| 附件：恶意HTML脚本 | 附件：可疑链接 |
| 附件：恶意PowerShell脚本 | 附件：xn域名链接 |
| 附件：恶意MHT脚本 | 附件：短网址 |
| 附件：VBS脚本 | 附件：对象混淆 |
| 附件：WSH脚本 | 附件：对象嵌入 |
| 附件：WSF脚本 | 附件：恶意文件-Yara |

图：海通证券安全团队对恶意附件所做分类

性、危害性三大特点。离线附件通过具有诱惑力的命名使人们麻痹大意沦为受害者，再通过隐藏变换的多种文件类型，可以是可执行文件、具有恶意宏的 Office 文档及混淆后的脚本程序。利用沙箱自动分析，通过动静两种模式进行检测：

a) 静态检测技术可以对通过文件头对文件类型进行识别，主要对文件结构、是否引用恶意代码实现检测。

b) 动态行为检测技术依托于虚拟机技术，根据文件类型，自动选择匹配虚拟机环境，监控进程运行期间系统 API 调用行为，并通过抓包获

取不同协议的网络连接行为。

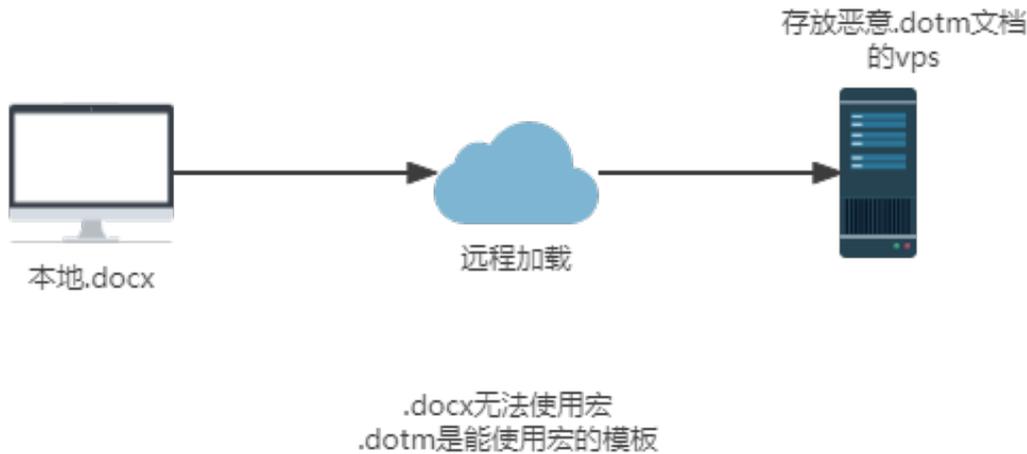
▪ 加密附件

加密压缩包却是最常见的 APT 攻击手法之一，通过将威胁附件进行压缩加密，并将解压密码写进正文的方式来绕过常规网关、沙箱等邮件防护设备。面对这种攻击，就需要安全人员对这些重点邮件多进行关注，配合网关将所有的加密压缩包邮件单独提取出来进行检测。

2.2.2 文本

▪ 诈骗文本

诈骗文本通常是定向发送给安全意识较为



图：Office 宏病毒绕过检测的方式



图：一封真实绕过网关的攻击邮件

薄弱的企业员工。目前阶段最需要引起重视的诈骗就是 BEC（商业欺诈邮件），此种攻击方式通常会冒充企业高管或是合作金融单位，向财务人员发送需要挪用资金或者向供应商打款的邮件文本。

▪ 勒索文本

攻击者通过描绘模糊性事件或是获取收件人一定的信息后进行扩展捏造，使得收件人似是而非的认为攻击者掌握了确实可以要挟自己的信息，慌乱和害怕中被敲诈成功。勒索文本在 WannaCry 病毒爆出的那段时间内非常流行，很多时候企业因为无法快速理自身到底是否存在机器沦陷，从而在危急情绪下接受了勒索团队的要挟。

▪ 广告垃圾

各大邮件服务商都会引入 DNSBL,RBL, 等技

术，使得垃圾邮件都会被打上 SPAM 标签被拒收，所以广告之类的垃圾邮件现阶段对安全的威胁性较小。

2.2.3 链接

▪ 伪造的 URL 链接

伪造 URL 一般都会结合诈骗文本使用，在邮件中的 URL 是可以编辑的，从而使得邮件真正的跳转 URL 和文本中显示的 URL 不一致。

▪ 短链接跳转

短链接是攻击者为了绕过 URL 检测的一种新思路，毕竟正常的邮件中很少会出现短链接。但是没那么优秀的网关就会把这些威胁 URL 放过，导致企业邮件防护阵线出现安全隐患。

2.2.4 图片

攻击者把恶意的内容制成图片放到正文中来绕过安全设备的检测。又或是把钓鱼链接或由其

I am aware, 810304 is one of your pass words. Lets get straight to point. absolutely no one has compensated me to check about you. You may not know me and you are probably wondering why you are getting this email?

我是威胁者 810304是你的一个密码 让我们直入主题

In fact, I installed a malware on the xxx video clips (adult porn) site and guess what, you visited this site to experience fun (you know what I mean). While you were viewing videos, your internet browser initiated working as a RDP having a keylogger which gave me accessibility to your display screen and also web camera. Right after that, my software program collected every one of your contacts from your Messenger, social networks, as well as email . after that I made a double video. First part shows the video you were watching (you've got a fine taste lol . . .), and 2nd part displays the view of your cam, and it is you.

当你浏览xx网站时，我访问了你的摄像头和浏览器，并获取了你的联系人

You do have two possibilities.. We will review these types of possibilities in aspects:

First option is to just ignore this e mail. in this case, I am going to send your actual tape to each one of your contacts and thus think regarding the embarrassment you will get. Furthermore if you are in an affair, precisely how it will eventually affect?

第一个选择是忽略，我会发送给你每一个联系人关于你xx的视频

2nd alternative will be to give me USD 979. We will think of it as a donation.. Consequently, I will quickly delete your video recording. You could go forward your life like this never took place and you will never hear back again from me.

第二个选择是向我支付979美元，我会将视频立即删除

You'll make the payment by Bitcoin (if you don't know this, search 'how to buy bitcoin' in Google search engine).

BTC address: 1A1zP1eP5QGTHiSb2fBcvDRbKwvaoHb3m

一封常见的比特币勒索邮件

[CaSe SeNSITive so copy & paste it]

比特币账户

If you may be thinking about going to the cops, surely, this e mail can not be traced back to me. I have taken care of my actions. You now have two days to pay. I have a specific pixel within this email message, and now I know that you have read this email. If I do not get the bitcoin, I will send your video to all of your contacts including relatives, co-workers, and so on. Having said that, if I receive the payment, I will erase the video right away. If you need evidence . reply Yes then I definitely will send out your video recording to

图：一封真实的比特币勒索邮件



图：一封真实伪造的 URL 链接邮件

他恶意链接生成二维码插入到邮件正文当中，利用收件者会拿出手机扫描二维码的习惯来绕过安全防护措施中的检测机制。

2.3 邮件账号安全

大多数邮件运维人员通常将更多精力放在能否从邮件接受层面来防御邮件攻击，通过识别邮箱载体是否存在威胁进行防护。然而，攻击者早已经转换了攻击思路，通过从邮箱账号层面进行更深入有效的攻击。

2.3.1 获取邮箱账号权限

- 爆破

攻击者测试通过 web 界面以及 smtp、pop3、imap 登录的方式，判断是否存在爆破的可能性，此类手法的判断主要看输入错误时提示是否有

“用户名不存在”还是“输入密码错误”之类的有效信息返回。

邮件运营者应当设置一定的防爆破机制，对于账号或者 ip 进行相应锁定或者灰名单机制来抵御此类攻击。

- 撞库

很多邮箱使用者在使用不同网站或不同账号时使用的是相同的密码，因此攻击者可以通过社工库先获取邮箱使用者在他处使用的密码，进而对邮件系统进行撞库攻击，这类将大部分人使用密码习惯融入攻击的手法成功率往往不低。

- 劫持嗅探

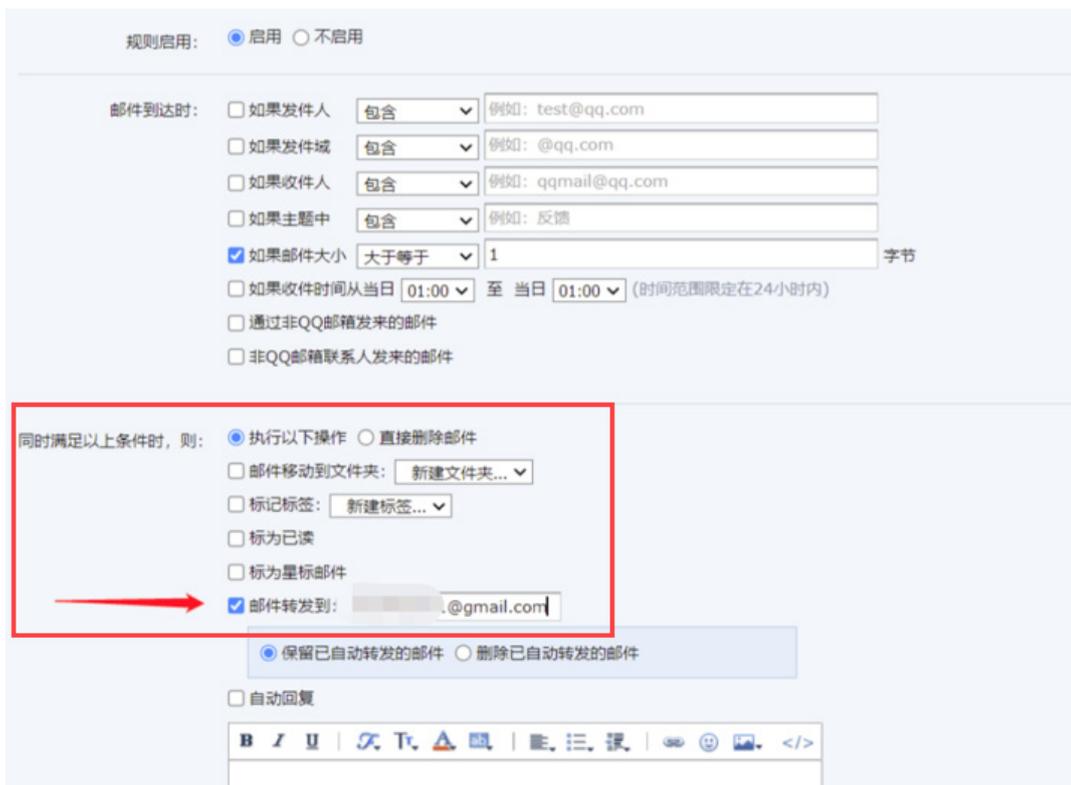
该种攻击方式适用于获取相关网络设备的权限后进行，如果邮件发送或收取时使用未进行加密成 smtps 和 pop3s/imap，而是单纯通过 base64

```
220 smtp.qq.com Esmtp QQ Mail Server
EHLO LAPTOP-0S0PEKRH
250-smtp.qq.com
250-PIPELINING
250-SIZE 73400320
250-STARTTLS
250-AUTH LOGIN PLAIN
250-AUTH=LOGIN
250-MAILCOMPRESS
250-8BITMIME
AUTH LOGIN
334 VXNlcm5hbWU6
[redacted] eH16 账户
334 UGFzc3dvcm06
[redacted] wUzMjc1YTdkMGI= 密码
235 Authentication successful
MAIL FROM: <[redacted]@004.xyz> SIZE=1340
250 Ok
RCPT TO: <[redacted]@.com>
250 Ok
DATA
354 End data with <CR><LF>.<CR><LF>
Date: Thu, 19 Aug 2021 10:57:41 +0800
```

图：SMTP 中的 base64 加密流量

```
+OK POP3
USER [redacted]@t.com 账号
+OK Send your password
PASS [redacted] 密码
+OK Mailbox locked and ready
STAT
+OK 288 3569640
LIST
+OK 288 messages (3569640 octets)
1 390
2 390
3 390
4 390
5 390
6 390
7 390
8 390
```

图：pop3 中明文流量



图：邮件转发示例

编码或直接明文发送，攻击者便可以直接利用劫持的网络流量，从中提取出对应的邮箱账号及密码等信息。

2.3.2 权限维持

攻击者通过撞库或者社工等方式拿到某一个账号的密码时可以设置代收接管这个账号的所有邮件。

有两种方式来长期获取目标邮件，pop3 收取和邮件转发。

- pop3 收取

如果是公开邮件服务，通过此方式基本可以无痕的收取所有目标账号上的邮件

- 邮件转发

三、构建邮件纵深安全体系

海通证券通过不断的对抗总结与运营反馈，针对邮件纵深安全体系总结出了一些应对经验和问题排查技巧。

3.1 邮件服务器基本核查

1) 注意邮件客户端或者是邮件服务器自身的漏洞，及时的更新服务器版本和客户端版本是非常重要的，同时安全管理人员需要时刻的关注有没有最新的 0day 信息，做好第一时间处理事件的准备；

2) 关闭外部通过 SMTP、POP3、IMAP 直接登录的功能。很多企业邮件服务器 25 端口是直接暴露在外的，直接 telnet 连接就可以对内伪造发邮件了。

3) 限制同 IP 和同账号一定时间的登录频率；

4) 如果条件允许，内对内的邮件使用独立的邮件系统，外到内和内到外使用另一套独立系统。这样一旦出网的邮件系统发生了账号失陷问题，对内无法横向移动，进而可以缩小影响；

5) 尽量不对公网开放 web 端登录邮件系统的功能，如果有移动办公的需求应该使用企业自身的邮件终端。

3.2 邮件服务器安全协议

- 使用邮件三大安全协议：

对于邮件伪造引起的安全问题，SPF、DKIM、DMARC 三大策略如果都配置正确，基本可以完全杜绝伪造发件人，其中尤其需要注意的是 SPF 的配置过程中 ~all(软拒绝)和 -all(硬拒绝)的区别。通常默认是软拒绝，这样不能完全保证安全。

同时查询 SPF 记录的过程中同样会暴露服务器的真实 IP，也需要对这些资产做安全性排查。

- 使用 RMX 协议：

RMX 类似于 RDNS，一个 MX 服务器收到了邮件之后，会去反向查询发件 IP 地址是否有对应的 MX 记录解析。

- 开启 SSL/TSL：

对邮件流量进行加密是保证邮件数据安全、邮箱资产安全的基本，应做到流量加密。

- 设置归档时间：

定期对邮件进行归档，可以防止攻击者拿到收件密码后，过多的收取邮件内容。

- 配置用户发件权限：

```

C:\>nslookup -q=txt spfl.baidu.com
服务器: bogon
Address: 172.16.120.16

非权威应答:
spfl.baidu.com text =

"v=spf1 ip4:111.202.115.85 ip4:220.181.3.85 ip4:111.206.215.185 ip4:220.181.50.185 ip4:180.76.13.18 ip4:12.0.24
.41 ip4:220.181.18.241 ip4:61.208.132.13 ip4:220.181.27.29 ip4:180.149.143.0/24 ip4:180.76.13.7 ip4:180.76.13.8 ip4:111
202.115.64/26 -all"
    
```

图：某度的 SPF 配置

除了有确定群发需求的账号可以开启无限群发功能以外，其他的账号都应该限制最大收件人数量。

▪ 配置单用户发件上限：

为了确保在某一账号失陷后能造成的影响最小化，应该把单用户的发件上限设置一个合理的阈值。

3.3 邮件资产安全防护

二次认证：

登录的时候尽量避免直接使用密码，可以通过密码 + 短信验证码 / 令牌的二次认证码等方式来解决。

▪ 限制登录口：

限制外部通过 telnet 25 端口连接邮件服务器，很多情况下攻击者可以直接通过命令 SMTP

来实现邮件内对内伪造发件人发送。所以公网建议仅开启 25 端口收件功能。

▪ 禁用 web 端：

公网 web 端收取邮件始终是存在风险的，如果不是刚需的情况，建议企业内使用移动端配合客户端的方式来收取邮件。

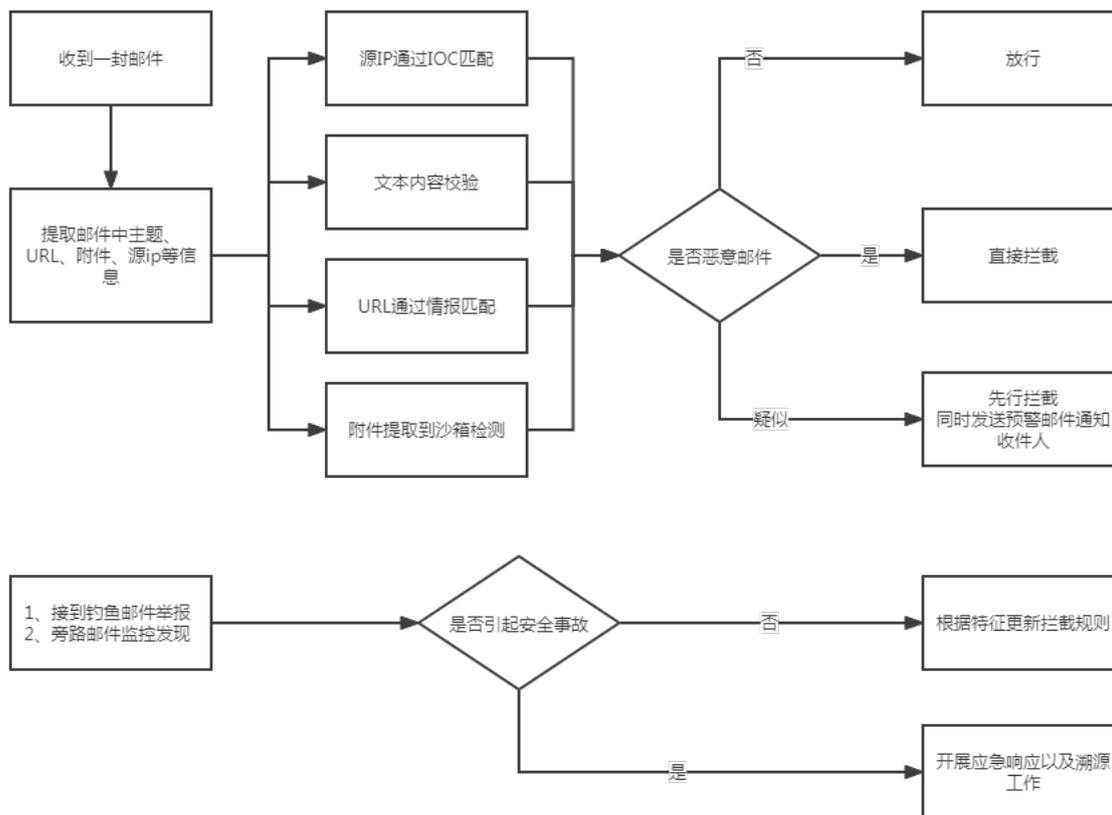
▪ 异常登录锁定：

登陆锁定是防止撞库和爆破的最佳方法，低阈值短间隔，附加单个 IP 单日内的上限错误次数，既避免了恶意锁定行为，也避免了撞库爆破攻击。

3.4 邮件运营体系流程

海通证券通过运营累积了完整的一套邮件处理流程，包含邮件拦截、拦截成功之后的隔离、预警、未拦截情况下的应急处置和溯源等。

3.4.1 一道拦截



图：海通证券邮件安全运营体系

网关拦截能力是有限的，需要不断的去调整规则的松紧度来保证在不影响业务的情况下去最大限度的拦截威胁邮件。如果为了保证业务邮件的正常发送，把网关拦截策略调成最低，放过大量的高级威胁邮件。这样做非常不可取。

3.4.2 二道隔离

当在邮件中检测到威胁、垃圾邮件、限制的附件和 / 或限制的内容时，客户端可以将此邮件移动到服务器中的隔离文件夹中。此过程将作为邮件 / 附件删除的备选处理措施，并可防止用户打开受感染的邮件和扩散威胁。

3.4.3 三道预警

网关检测能力有限，很多存疑的邮件运维人员也无法即时性做到逐封详细排查，这个时候预

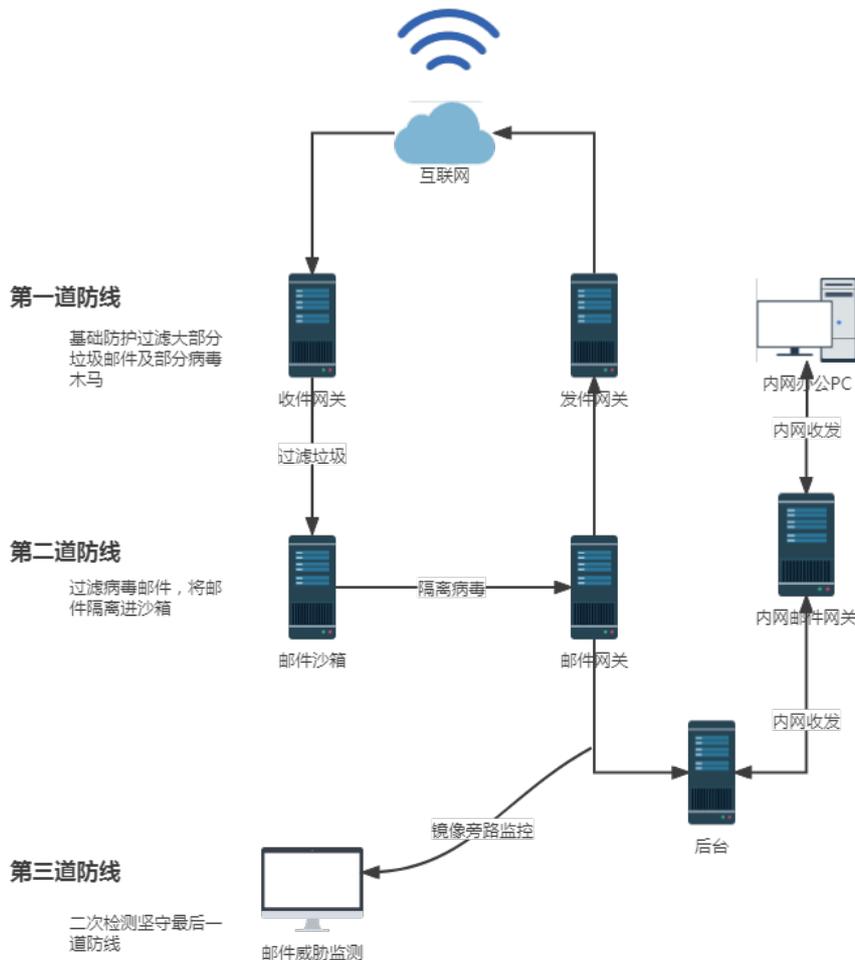
警功能就显得非常必要。在邮件服务器中可以设置一个专用的邮件预警账号，将对存疑邮件的怀疑和预警发送给该封邮件的收件用户，这样用户在收到存疑的邮件之后便会因为预警邮件的提示引起警觉，避免直接打开存疑邮件。

3.4.4 应急响应

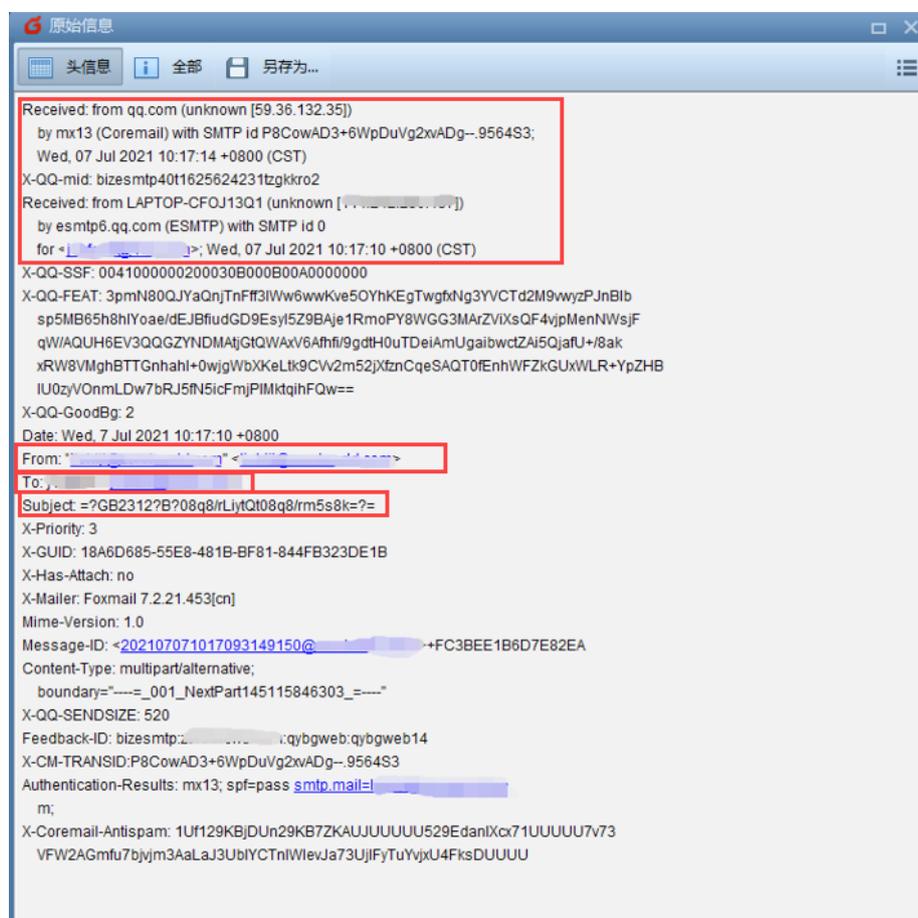
企业应该有完备的应急预案，包括单账户沦陷、内网横向发送钓鱼邮件等多种情况，避免在发现问题的时候束手无策，能做到安全事件发生的十分钟内定位到账号，半小时内定位到威胁邮件。

3.4.5 威胁溯源

安全人员、一线运维员工需要具备手工排查邮件头信息的能力，能分辨邮件头中真实的标签



图：海通证券邮件纵深安全体系架构



和伪造的标签。建议通过安全教育进行全员普及。

邮件头中需要重点关注的主要字段为 Received、From、to、subject、cc 和 bcc 五个字段，可使用这些字段来追踪邮件传输路线。

3.5 邮件安全意识建设

3.5.1 钓鱼演练

企业内应当定期组织钓鱼邮件演练，搭配不同的攻击形式，涵盖钓鱼、诈骗、恶意附件等，可以使用 GoFish 搭建大规模的发送平台，也可以使用 SWAKS 来进行小规模发送。

3.5.2 安全意识培训学习

安全意识培训学习主要面向其他部门以及邮件安全不熟悉的员工，讲清楚威胁邮件的载体和可能被用来钓鱼攻击的手短，能做到存疑邮件不点击、不输入、不运行三大要素。也建议在

完培训之后再做一次钓鱼演练，来对比体现安全意识培训是否取得了预期的效果，这同样也是对培训的必要性做一个数据证明。

3.6 邮件安全运营情报

通过一段时间的运营数据累积，可以构建自己的 IOC 情报库。在从流量层提出邮件中的发件 IP 后可以对 IP 进行 IOC 的匹配和预警，把从被动收件发现攻击转化成主动匹配情报，及时的发现恶意的攻击组织，避免危害形成或者是扩大。

3.7 邮件自定义拦截策略

钓鱼攻击背后往往隐藏着对人性弱点的洞察，邮件内容需要具有重要性、合理性和紧迫性三个特点，对邮件内容的策略选择一般有如下几点：

- a) 人性弱点：人物画像，痛击软肋（兴趣，忌惮，生活必需）
- b) 热门话题：关于目标或者目标利益的相关方
- c) 实时热点：节假日、重大活动、国际事件
- d) 工作相关：招聘、涨薪、密码过期、福利

我们通过在邮件网关中进行配置，对指定白名单账号配置允许发送包含上述特点的邮件，当其余账号接收含有此类关键字的邮件时，一律封禁，并且发送提示邮件给收件人如遇误拦请联系邮件运营者，最大程度的提前阻断危害。

3.8 主机控制权限

百密必有一疏，实战中，难免会遇到攻击者成功钓鱼导致主机失陷的情况发生。攻击者获取权限后会进行主机信息收集。通过禁止调用 CMD、PowerShell 等程序，保证主机信息不被攻击者看到。

四、海通证券邮件安全运营数据

海通证券在建设和丰富上文中介绍的邮件安全运营体系建设后，取得了实际落地的检测成效和闭环提升。

首先是防护的架构方面，通过邮件网关的过滤规则、沙箱对于携带附件类邮件的静态动态查杀，将攻击者无差别群发的垃圾邮件、威胁邮件分别进行两道防线的过滤。

再通过邮件溯源系统的深度检测，从已经被过滤的邮件中发现精钓或是采取了绕过手段的定向攻击事件，将这类攻击事件的告警通过邮件预警的机制发送给收件人和管理员，提醒收件人不要点开的同时也让安全团队可以从中深度分析采用的攻击手法、攻击者的攻击意图。

再将这类定向攻击事件中的行为特征进行提

取，加入到邮件网关的拦截策略中，形成从阻断到检测再到阻断的闭环防护，不断巩固邮件安全的防线。另外，常规的邮件告警实现了流程简化和自动化，更多的是以 syslog 汇总、自动的预警邮件机制来统一通知安全运营一线，使得安全工作从多台设备来回切换查看告警转换为处理汇总日志以及聚焦安全事件即可即可。

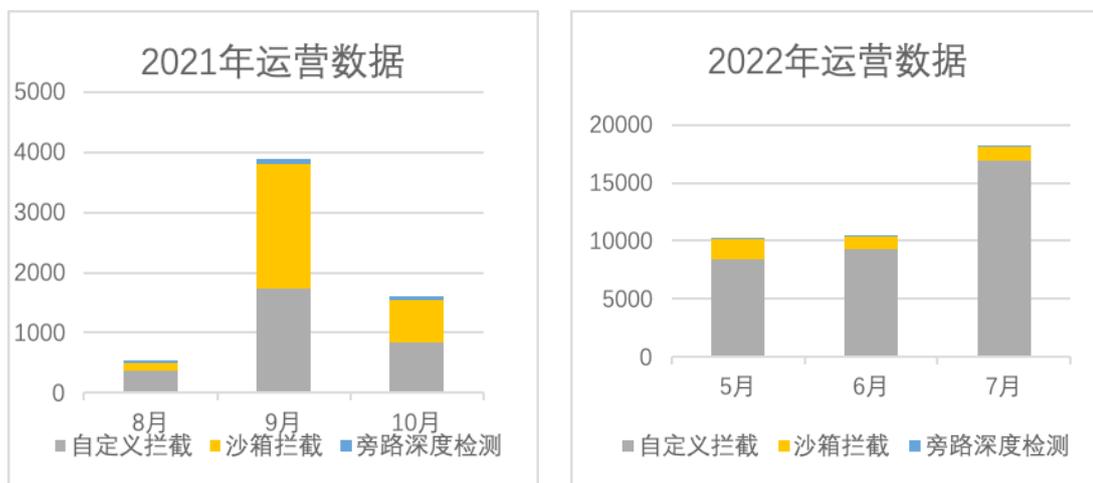
公网的 smtps 协议（465）供客户端发信时可以连接至邮件系统进行身份认证和发送邮件。但由于 smtp 协议本身即支持发送邮件（发件人是域内用户），也支持接收外部邮件（发件人不是外域用户），即 mx 功能，因此可以被外部系统或用户直连进来，并尝试向域内收件人发信。通过把 smtp 和 mx 进行分离，关闭邮件系统外网前端服务器的 mx 功能，禁止外域用户直连到邮件服务器发信，要求必须是域内用户才能进行发送邮件。同时，正常的外域邮件经由反垃圾网关投递到邮件系统，而网关设备被信任，因此接收正常外域邮件不受影响。

在完整配置 SPF、DKIM、DMARC、RMX 协议后有效的杜绝了发件伪造的行为，实现了向内投递邮件零伪造，同时配合二次认证和异常登陆锁，彻底的避免了账号被爆破、撞库的风险。

互联网侧关闭了 web 端访问，仅通过 EMM 收缩后的 APP 登录 web 邮箱，降低了 web 端暴露受攻击的风险。

对于 2021 年连续三个月各道防线发现的威胁邮件情况进行分析，在总邮件量差异不大的情况下发现，9 月的邮件攻击中存在恶意附件、需要深度检测才能发现以及被加入到定制化策略的比例明显相较上月增多。在此分析基础上我们结合威胁情报得知，9 月关于“新冠肺炎疫情持续扩散”相关的 APT 邮件攻击活动持续流行（单月此类攻击情报统计达 2090 万封），所以针对此类邮件发出通告，在安全意识角度提醒全体同事请勿点击。

经过一段时期的安全运营，月均数六十万封



图：邮件纵深防护体系中各阵线运营数据

的总邮件中，垃圾邮件占一半。已被安全网关拦截。随着自定义策略的丰富，自定义策略的拦截占总邮件的比例从2021年运营初期的3%提升到了5.5%，能无告警投递到用户手中的威胁邮件比例可以做到低于十万分之一。

在安全意识提升环节，2021年上半年对公司本部全员进行钓鱼邮件演练时发现，较多员工有打开钓鱼邮件链接的情况。对此我们进行了针对不同种类钓鱼邮件的花名册宣讲活动，详细讲解对钓鱼邮件的分辨手段以及防范措施，

在下半年对相同人数人员的钓鱼演练中，中招人数下降了75.76%。通过数据可以直观的反映定期开展安全意识活动的必要性以及进行大型钓鱼邮件演练和安全意识宣讲的显著成效。在2022年度的钓鱼邮件首轮演练中也发现中招人数略有回升，说明安全意识培训是需要持续不断进行的，来保持全员的反钓鱼意识维持在一个理想水平。

以上是海通证券邮件系统纵深防护体系的实践分享，欢迎行业交流指正。

0.12



信息技术创新观察

- 8 基于 FPGA 异构加速平台的高维美式期权定价
- 9 基于图技术的基金抱团识别方法的研究



基于FPGA异构加速平台的高维美式期权定价

喻伟 / 东方证券股份有限公司 上海 200010
邮箱: yuwei1119@sina.com



作为衍生产品市场上交易频繁的产品，美式期权一直备受资本青睐。受限于其自由的行权日期和弹性的交易策略，高维美式期权的定价极具挑战。经典的 Longstaff-Schwartz 算法通常在蒙特卡罗模拟路径上应用最小二乘线性回归法实现定价。本文对经典 Longstaff-Schwartz 算法进行改进，通过逆向执行流程，提出一种新的高维美式期权定价方法。方法基于异构平台，实现了可重构的位算法转换，消除了内存瓶颈，降低了访问延时，提高了系统吞吐量。实验结果表明基于异构加速平台的高维美式期权定价方法相比传统软件定价方法速度加快了近 22 倍，功耗降低了近 336 倍。

1 前言

金融世界中场外交易衍生品市场的年均交易量逐年增加，激烈的竞争和严格的法规引发了计算需求的稳定增长。为了满足这些计算需求，金融机构不得不运营庞大的设备集群，其高昂的硬

件开销和运维费用使得金融业在速度提升和降本增效方面动机强烈。FPGA 以其独有的计算体系，成为金融系统的硬件加速的不二选择。

虽然大部分数值算法可以映射到 FPGA 中，但仍有一些算法更适合在 CPU 上执行，这为混合架构提供了需求，如基于 x86 内核的 Intel 可

编程片上系统，就可以在 CPU 上配置 FPGA 芯片。这种异构计算系统因硬件来源的不同而带来诸多挑战，如系统通信、时间配置、存储带宽、FPGA 面积及 CPU 负载平衡。

在衍生品市场，期权是一种特殊的产品。从定义上来说，期权是一种赋予权利但没有义务以固定价格和固定日期买卖标资产。它的吸引力之处在于给买方带来有限风险的同时，赋予了与合同相关的潜在收益，即购买时支付的溢价，而这种溢价（期权价格）的计算正是金融机构所关心的核心。与只能在固定日期行权的欧式期权不同，美式期权允许持有人在起始日和到期日之间的任何日期行使期权，这种自由使得定价更具挑战。

经典的 Longstaff-Schwartz 算法 (LS 算法) 可以使用策略推导美式期权价格，即 LS 算法通过使用最小二乘回归法，从到期日到起始日逆向回溯蒙特卡罗模拟路径，计算出多个基础资产的多维期权价格。考虑到多维美式期权定价的复杂性，本文采用蒙特卡罗方法。蒙特卡罗方法中模拟路径的质量取决于用于描述基础资产演变的数学模型。对于高维美式期权，Black-Scholes (BS) 模型因其计算结果与市场真实价值接近而被广泛使用。

LS 算法的实施存在 2 个问题：(1) LS 算法没有区分软硬分区，作为一个计算密集型算法，模拟路径上某些基础函数的选择，可能会影响到最终的执行价格。(2) 最小二乘过程会影响整体运行时间，蒙特卡罗模拟路径数量以及行权天数的多少，决定了生成的数据量，尽管可以将部分临时数据存储到片外存储芯片中，但仍面临着一定的带宽限制和能源消耗。

为解决以上问题，本文提出了一种基于经典 LS 算法的自定义计算方案，使之能够运行于 Xilinx 异构硬件平台。本文还提出了一种改进的逆向 LS 算法，在无需存储所有路径的情况下实施计算重构，减少内存访问耗时。

2 相关概念介绍

2.1 美式期权

期权合约是一种合同，它从实体资产的业绩中获取价值。期权合约赋予买方以指定行权价和行权日买卖标的资产的权利，主要包括两种：

- 欧式期权，只能在到期日行权。
- 美式期权，可以在到期日之前的任何日期内行权。

期权赋予持有人以下权利：

- 卖出基础资产：看跌期权。
- 买入基础资产：看涨期权。

针对给定美式看涨期权，考虑不同行权价和到期日的模拟组合，期权持有人需要在各时间节点上决定是行权还是继续持有。

$s(\tau)$ 为 τ 时刻的基础资产价值或市场价， K 为行权价，则 τ 时刻的期权内在价值可用下式表示，其中 call 为看涨期权，put 为看跌期权：

$$\begin{aligned} \text{payoff}(\tau) &= \max(s(\tau) - K, 0) \quad \text{call} & (01) \\ \text{payoff}(\tau) &= \max(K - s(\tau), 0) \quad \text{put} & (02) \end{aligned}$$

如下式所示，当看涨期权的行权价低于市场价或看跌期权的行权价高于市场价时，这种情形被称为价内期权 (In the Money, ITM)，期权价值为正：

$$\begin{aligned} s(\tau) > K & \text{ for a call option} & (03) \\ s(\tau) < K & \text{ for a put option} & (04) \end{aligned}$$

当期权价格满足 ITM 条件时，持有人可以选择提前行权或继续持有该期权，以获取最大利润。

美式期权的买卖也需要权利金，美式期权的权利金由下式给出：

$$P = \sup_{\tau \in T\{t_1, \dots, t_m\}} E(e^{-r\tau}(S(\tau) - K)^+) \quad \text{call} \quad (05)$$

$$P = \sup_{\tau \in T\{t_1, \dots, t_m\}} E(e^{-r\tau}(K - S(\tau))^+) \quad \text{put} \quad (06)$$

其中， $(x)^+ = \max(x, 0)$ ， K 是行使价， T 是

期权的到期日, r 是无风险利率, $T\{t_1, \dots, t_m\}$ 是行权日期组合, $\{t_1, \dots, t_m\} = \{\frac{T}{m} \times 1, \dots, \frac{T}{m} \times m\}$ 是该期权的潜在行权日期。

对于多维期权, 其价值来自于多个基础资产, 当期权行权日期如公式所示时, 该期权被称作百慕大期权。

美式期权的主要复杂性在于其定价, 如前所述, 这种风格的期权不仅可以在到期日 (如欧式) 那样行权, 还可以在中间时间节点执行。期权的卖方 (通常是银行或金融机构) 必须在最坏的情况下估算其价格, 利用合理的策略带来良好投资回报率, 而这正是 LS 算法发挥作用的地方。

2.2 布莱克 - 斯科尔斯模型

期权市场价 $s(\tau)$ 采用布莱克 - 斯科尔斯模型 (Black-Scholes, BS 模型)。BS 模型假设, 除其它因素外, 股价遵循随机波动, 即未来任何时刻的股价均服从对数正态分布。我们可通过随机微分方程来描述股票价格 $S(t)$:

$$ds(t) = S(t)(r - q)dt + S(t)\sigma dW(t) \quad (07)$$

其中, r 是无风险利率, q 是股息收益率, σ 是股票收益率的恒定波动率, $W(t)$ 是相关布朗运动。

BS 模型通常假设波动率恒定, 但实际市场可能并非如此。不过, 由于其使用简单、容易扩展且与期望利润近似, 如今仍有很多研究人员使用。

2.3 蒙特卡罗方法

蒙特卡罗方法 (Monte Carlo, MC) 是一种随机抽样的统计模拟方法。为了对 BS 模型进行模拟仿真, 需要对时间进行适当离散化, 在本文中, 我们采用欧拉离散法, 即以相等的步长将时间离散化为 $\Delta t = \frac{T}{m}$, 为此我们得到 t_i 时刻的期权贴现收益值:

$$\hat{S}_{t_{i+1}} = \hat{S}_{t_i} e^{\left((r-q) - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\Delta W_i} \quad (08)$$

其中 ΔW_i 为独立的标准正态随机变量。

经典的 MC 算法将价格 P 估计为贴现收益值 $g(\hat{S})$ 的模拟实例样本均值。由于 MC 方法的复杂度取决于维数的大小, 这使其成为解决高维问题的最佳方案。此外, 由于误差的标准偏差仅随模拟次数的平方根而减小, MC 的收敛结果在很大程度上取决于模拟路径的数量, 路径数越多, 其产生的结果越准确。根据 LS 算法, 一个基于 BS 模型包含 2 个关联股票的美式看涨期权的最佳预期贴现收益值表示为:

$$P = \frac{\sup}{\tau \in T} E[e^{-r\tau} (\max\{S_1(\tau), S_2(\tau)\} - K)^+] \quad (09)$$

对于给定示例, 模拟路径数量 N 对期权定价准确性影响较大。如果配置相关参数:

$$S_1(0) = S_2(0) = 100, K = 100, r = 0.05, q_1 = q_2 = 0.1, \sigma_1 = \sigma_2 = 0.2, \rho = 0.1, T = 1, m = 365, T = \{\frac{T}{m} \times 1, \dots, \frac{T}{m} \times m\}, N = 10000。$$

那么根据运行 100 次 LS 算法的不同模拟路径下权值的离散分布情况, 可以看出, 与基于二叉树方法的基准期权价值 10.12 相比, 本文期权定价估值更准确, 且所要求的最小路径数约仅为 10000。

2.4 路径生成步骤

BS 模型需要一系列正态分布随机数来生成路径。受限于基础资产共存于同一市场, 且随机数之间的关联度, 本文拟按以下步骤顺序执行:

(1) 执行梅森旋转演算法 (Mersenne Twister, MT): MT 是一种广泛使用的伪随机数生成器, 本文采用 MT-19937 版本生成一个 32 位无符号整数序列。MT 算法分为两部分:

- 基于一组状态集 (624 个内部状态) 生成随机数。内部状态的初始化通过种子完成, 并且每隔 624 个输出数完成状态修改。
- 实现退火功能。算法执行一系列异或运算, 输出最终数字。

可以使用 FPGA 流水线优化此算法, 以便在每个时钟周期获得一个输出。实际上, 优化工作

是在实现过程中完成的，因此每个状态在当前循环的最后一次使用后都会立即实现。

(2) 引入逆累积分布函数 (Inverse Cumulative Distribution Function, ICDF): 前述 MT 模块生成的随机数服从均匀分布，而 BS 模型则需要正态分布的随机数。该领域的先前工作提供了 ICDF 的有效实现和所需的标准正态分布。ICDF 方法虽不能精确地保证每个时钟周期的有效输出，但可在硬件利用率和性能之间达到良好的折衷。

(3) 引入对称变量：如前所述，由于 MC 方法的收敛速度较慢（模拟运行时间较长），除了可以通过减小其方差来加速外，最简单的方法是通过引入对称性来实现对称变量。本文将对称处理方法放在 ICDF 模块之后实现，使之可以处理正态分布的随机数。在此情况下，可以证明，对于单个随机数 z ，则 $-z$ 也是有效随机数，这将使总随机数减少一半。此外，当使用基于布朗运动的模型生成路径时，高维期权的收益通常可以写成：

$$P = h(Z_1, \dots, Z_k) \quad (10)$$

假设 h 在每个变量上都是单调的，则 P_1 和 P_2 负相关，这意味着对称变量可以减少随机变量的个数。

$$P_1 = h(Z_1, \dots, Z_k) \quad (11)$$

$$P_2 = h(-Z_1, \dots, -Z_k) \quad (12)$$

(4) 执行关联分析：在二维选项的情况下，关联程度可以通过两个独立随机数 $y \sim N(0,1)$ 、 $z \sim N(0,1)$ 及其系数 ρ 获得，如下式：

$$w = \rho z + y\sqrt{1-\rho^2} \quad (13)$$

通过以上步骤，将之前生成的随机数代入到 BS 模型方程式中，就可以获得所需的路径。

3 美式期权定价方法

3.1 基于 LS 算法的美式期权定价

最小二乘算法通过模拟路径逼近美式期权价格，模拟的 MC 路径表示基础资产随时间变化的

行为。MC 路径可通过构建不同复杂度的数学模型（如 BS 模型）获得，一旦生成路径，持有人就可以通过其策略利润大小来评估期权价格。如果到到期日，则持有人必须行权。如果还没到到期日，则持有人可以选择行权或继续持有。

若当前行权收益高于延续价值（继续持有收益）时，则可以立即行权。但是，延续价值往往取决于继续持有该期权的期望价值。一般来说，为了执行最佳策略，LS 算法会基于当前时刻的所有生成路径来评估此期望价值。更进一步地，LS 算法会使用最小二乘线性回归方法来找到最佳执行策略，基本步骤为：

(1) 使用所选随机数生成器和所选数学模型，在所有可能的行权日为每个股票生成 N 条独立路径。对于多维期权，还需将这些生成的随机数关联起来；

(2) 使用到期贴现收益初始化现金流；

(3) 从当前节点的前一节点回退到起始节点，迭代执行如下过程：

(a) 执行线性回归：要确定是立即行权还是继续持有，需要使用回归方法进行估算，将当前的折现收益（立即行权）与未来的预期收益（继续持有）进行比较。

(b) 更新现金流量：对于当前时刻的每条路径，将预期收益与当前贴现收益进行比较，取较大的收益并更新相应的现金流量值。

(4) 在起始日，对现金流量中的所有值求平均值，获得期权价值。

LS 的挑战之处在于回归基函数的选择，基函数在很大程度上取决于期权定价的确切形式，且需要与收益函数的特征相匹配。

3.2 行业相关工作分析

FPGA 已成为金融加速仿真的热门选择，一些机构也提出了有效的随机数生成方法和路径生成方法，但大多集中在 BS 市场模型的研究上。基于 MC 方法，通过比较 FPGA/CPU 的加速

比, De Schryver 等人证明, Heston 市场模型的 FPGA 加速比可以达到 33 倍; Thomas 等人也证明, GARCH 模型的 FPGA 加速比可达到 80 倍, BS 模型的 FPGA 的加速比达到 313 倍。Sridharan 等人继续将这些工作扩展到基于 BS 模型的多维期权中, 从而继续使 FPGA 的加速比提高到 350 倍。尽管这些研究工作实现了加速效果的提升, 但仍然无法精确地实现美式期权定价。

基于 LS 算法, 一些学者使用 MC 方法对美式期权定价, 与 CPU 相比, FPGA 的处理速度提高了 20 倍。尽管他们利用了高效的并行架构和外部存储芯片来存储模拟的 MC 路径, 其设计仍局限于 26/32 位的定点算法, 这也给后来的改进提供了想象空间:

(1) MC 仿真存储的路径数量只有 4K 条, 相较于前文阐述的最小 10K 大大减少;

(2) 外部存储芯片带来的额外功耗和带宽限制, 需要引起关注。

为了完成以上两点改进, 本文采用了一种称作逆向 LS 的算法来重新计算, 从而解决技术限制。

3.3 逆向 LS 算法

在第 3.1 节 LS 算法的执行过程中, 首先要是在步骤 (1) 中生成所有路径, 然后在步骤 (3) 中以相反的顺序遍历。这意味着所有路径每个时间节点的每支股票价格都必须得到保存和传输。这个过程产生了 $d*m*N$ 个值, 其中 d 为资产的维度数, m 为时间节点个数, N 为 MC 路径的数量。我们将此标准方法称为路径存储解决方案。对于容量仅有几十兆 BRAM 的 FPGA 而言, 这带来了巨大的挑战。如果需要使用多个外部高速存储设备, 那么设计就会更加复杂。因此, 我们提出一种基于重构计算的新颖思想, 以避免大量数据存储。

我们没有存储中间过程各条路径每个时间节点的价格, 而只存储了到期日的股票价格, 结合 LS 算法步骤 (3), 重新计算所有其它节点价格。

为此, 我们需要找到一种基于未来价格 $\hat{S}_{t_{m+1}}$ 计算当前价格 \hat{S}_{t_m} 的方法:

$$\hat{S}_{t_m} \rightarrow \hat{S}_{t_{m-1}} \dots \rightarrow \hat{S}_{t_1} \rightarrow \hat{S}_{t_0} \quad (14)$$

我们发现, 基于相同的随机数, 离散 BS 方程是可逆的, 从而有:

$$\hat{S}_{t_i} = \hat{S}_{t_{i+1}} e^{\left(\frac{\sigma^2}{2} - r + q\right)\Delta t - \sigma\sqrt{\Delta t}\Delta W_i} \quad (15)$$

在本文工作中, 我们使用 MT 19937 算法生成随机数序列, 想法是构建一个逆向的随机数序列 (Reversed Number Group, RNG), 该 RNG 从最后一个数字开始逆序输出。幸运的是, MT 是一个线性 RNG, 这意味着其状态转换功能是可逆的。实际上, 在保持模拟回火功能不变的同时, 仅有 MT 的内部状态被重新计算。基于此, 通过构建逆向 MT 流程, 逆向 LS 算法存储和传递的数据量被降低到 $d.N$ 个。

4 系统架构设计

本节展示了用于路径生成、LS 算法执行及期权价格计算的整体架构方案, 描述并比较了外部芯片存储路径与逆向 LS 算法这两种方案的特点, 对整体架构中的各主要模块进行了详细描述, 对体系架构实现高吞吐量操作的原理进行了解释。

CPU 可看作一种具有固定结构并以预定义指令运行程序的通用设备, 而 FPGA 则是一种面向应用灵活配置并保证专用模块高效运行的可编程硬件。基于 Xilinx U50 的 CPU/FPGA 异构系统, 结合两者优势, 提供了足够的软硬件分区资源, 使得异构芯片之间的通信延迟非常之低。通过流水线设计多个并行实例, 该架构能以较低的功耗实现较高的吞吐量。

LS 算法的一个特征是, 一旦所有 MC 路径生成完毕, 它就只能开始工作 (从到期日向起始日回溯), 此时负责生成此路径的模块将返回空闲状态, 进而不必要地消耗 FPGA 上的资源。通过设置 Xilinx U50 器件中的配置选项: 我们可以

对 FPGA 进行全部或部分动态重编程。

基于前面的解释，我们将体系结构的执行分为三个步骤：

- 步骤 1：正向生成 MC 路径；
- 步骤 2：FPGA 重编程；
- 步骤 3：执行 LS 操作。

4.1 整体架构描述

总体架构如图 1 所示。如前所述，在步骤 1 中，正向路径生成模块会产生多个实例，从而增加总体带宽。在步骤 2 中，重编程意味着前面和后面的步骤都可以访问 FPGA 上的全部资源，重编程 FPGA 所需时间可以根据设置来分摊。在步骤 3 中，LS 算法以流水线方式工作，使得每个时钟周期都能计算一个值。流程一旦到达起始日，现金流的价值将被平均，从而产生期权价格。

可以看出，正向过程会将所有生成的路径存储在片外 DRAM 中，而且采用了边生成边写入的方式，一旦 DRAM 被写入并且 FPGA 被重新编程，那么正向 LS 算法就会调用所有从到期日到

起始日的路径。此方法显而易见的存在诸多限制：

- 容量限制：受限于路径、步骤及维度的数量，可用内存容量可能超出；
- 带宽限制：受限于内存类型、数据总线宽度及时钟频率大小，可用带宽也可能不足；
- 功耗限制：数据的读写可能会产生大量的功耗。

如图 1 所示，方案使用正向 LS 结合逆向 LS 算法，从行权日到起始日回溯，并行计算路径，即可突破上述限制。正向路径生成过程仍会计算所有 MC 路径，但只会将行权日的路径存储起来，存放至部分重配置的 FPGA 区域，以供步骤 3 的 LS 算法调用。图 1 中的体系结构也适用于高维期权，其中路径生成模块的每个实例（正向或逆向）都能够同时为每个基础资产生成路径，LS 的组成部分也可以进行相应调整。

4.2 正向和逆向路径生成

正向路径生成模块如图所示，其步骤如章节

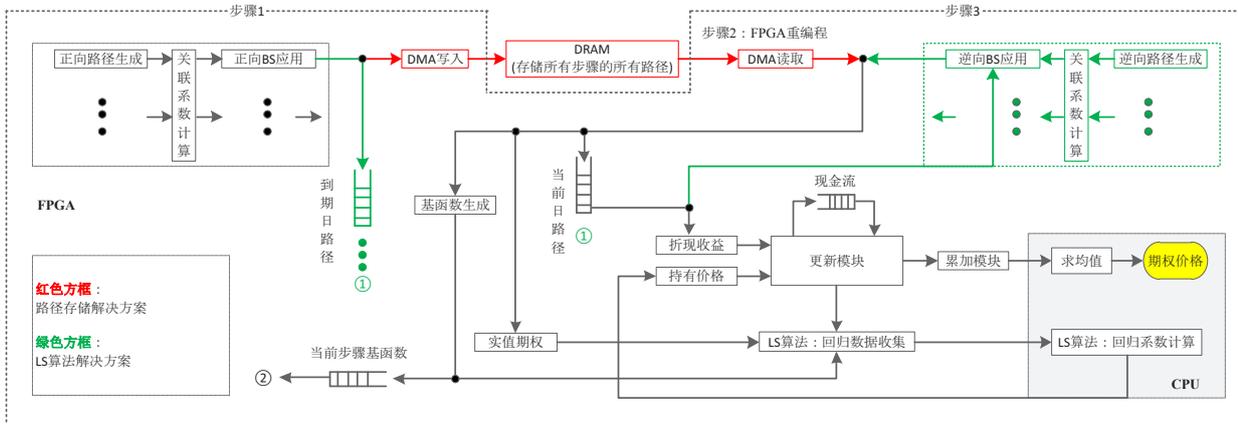


图 1：本文系统整体架构图



图 2：面向二维美式期权正向路径生成流程图



图3：面向二维美式期权的逆向路径生成流程图

所述，由于各底层路径彼此独立，因此完整模块的多个并行实例是可行的。图仅展示了针对二维期权的配置，其中每个 BS 模块都服务于二维期权的路径生成，开发者可以调整并行模块的数量以实现更多维度的期权应用。

逆向路径生成模块如图 3 所示，其步骤如章节所述，也同样面向二维美式期权。逆向过程有点类似于正向过程，不同之处在于，新路径会从到期日到起始日逐步逆向生成。

如前所述，BS 模块的逆向过程比较容易实现，MT 模块的逆向操作也仅需要修改其内部状态的更新过程（退火功能仍保持正向）。为了获得逆向条件下的相同随机数序列，只需利用正向条件下最终状态和最终索引的副本，即可实现。

4.3 LS 实施

LS 回归过程的核心在于找到在每个步骤生成条件期望函数所需的回归系数。这些系数是通过求解线性方程组得到的：

$$Xb = y \quad (16)$$

其中每行 X 包含了价内期权路径基函数的值，y 是现金流值，b 中的系数个数与基函数的个数相等。

使用硬件解决回归过程要么在资源上过于浪费（完全并行实现），要么在延迟上较为明显（受限于序列化版本）。就所使用的方法和要计算的系数个数而言，它也变得无比僵化。为了消除这些限制，通过计算 CPU 上的系数，引入了智能软硬分区概念。同时，为了减少 FPGA 和 CPU 之间的通信开销，减小了矩阵的大小。重写式为：

$$(X^T X)b = (X^T y) \quad (17)$$

对于 k 个基函数， $X^T X$ 和 $X^T Y$ 的大小分别

为 $k \times k$ 和 $k \times 1$ 。已经证明，可以使用累加器对这一过程进行流水线处理。对于单项式基函数 x^0, x^1, x^2, \dots ，这些累加器形如：

$$X^T X = \begin{pmatrix} \sum_n x_n^0 & \sum_n x_n^1 & \sum_n x_n^2 \\ \sum_n x_n^1 & \sum_n x_n^2 & \sum_n x_n^3 \\ \sum_n x_n^2 & \sum_n x_n^3 & \sum_n x_n^4 \end{pmatrix}; X^T y = \begin{pmatrix} \sum_n y_n \\ \sum_n y_n x_n \\ \sum_n y_n x_n^2 \end{pmatrix} \quad (18)$$

可以使用不同方法来求解方程，如 Cholesky 分解法或如式所示的矩阵逆方程法。考虑到 Cholesky 分解效率更高，并且操作是在软件中执行的，本文使用如下 Cholesky 分解法。

$$(X^T X)^{-1} = \frac{Adjoint(X^T X)}{\det(X^T X)} \quad (19)$$

4.4 高吞吐量操作

为了实现高吞吐量，图 1 架构中的每个模块被设计为流水线工作方式，以便在每个时钟周期处理一个新值。几个模块并行工作，保证彼此之间的等待时间最小。高吞吐量的实现包括以下两步：

(1) 正向路径生成和直接内存访问：在 DMA 第一次存满之前，数据一旦生成，就会以最快速度被写入 DRAM。

(2) LS 算法和反向路径生成：CPU 计算回归系数并发送到现金流更新模块，现金流模块中的第一条路径更新后：

(a) 下一次回归数据收集调用新值。

(b) 反向路径生成模块或 DMA 模块更新当前步骤中已使用的路径值。路径新值生成后，不仅被引入到相应向量，而且同步参与到价内期权的判定条件，以及生成基函数。

通过以上步骤，整体架构即可实现高吞吐量。

4.5 FPGA 重配置开销

FPGA 重配置意味着一定的时间消耗和功率消耗，极易超过单个期权定价所需的运行时间和能耗限制。

当为多维期权定价时，使用路径存储方法和逆向 LS 算法的组合可轻松满足开销限制，在这种情况下，系统将为每个期权生成所有路径，且只存储到期日的路径。一旦过程完成，FPGA 只需重配置一次，通过初始化外部存储芯片和调用逆向 LS 算法计算路径，逐个完成期权定价。

| | |
|-----------|--------------------------|
| 数据类型 | 单精度浮点数 |
| 数据容量 | 27.85MB |
| 开发平台 | Xilinx U50/Vivado 2018.2 |
| 操作系统 | Linux Redhat 7.4 |
| DRAM 类型 | DDR4 |
| DRAM 数据位宽 | 32 位 |
| DRAM 时钟频率 | 533.33MHz |
| DRAM 带宽 | 4266.64MB/s |
| FPGA 时钟频率 | 133.33MHz |
| FPGA 带宽 | 4266.64MB/s |

5 系统性能测试

5.1 工程参数设置

在常见设置中，在对路径存储方法和常规重计算方法进行比较时，我们发现，DRAM 芯片具有带宽上限（由存储器类型、时钟频率及其数据总线宽度定义），FPGA 带宽也仅依赖于可用资源的数量（并行实例的数量），FPGA 可用资源因设备和供应商的不同而差异较大。两种实现方式都设置了最大 DRAM 带宽下的能耗对比条件。我们约定，带宽越大，能耗越低，那么方法就越有效。下表详细罗列了完整的工程设置以及硬件资源，其中，FPGA 时钟频率是 DRAM 时钟频率的约数，目标带宽值为 4266MB/s。需要说明的是，尽管我们的测试用例针对的是二维期权，但所提结构也完美适用于高维美式期权。

表 1：工程参数设置表

| | |
|---------|---------------------------------------|
| 设置内容 | 设置描述 |
| 期权种类 | 美式期权 |
| 行权类型 | 买入期权 |
| 期权特征 | 最大值 |
| 期权维度 | 2 维 |
| 基函数类型 | 单项式 |
| 基函数细节 | $1, \max(S_1, S_2), \max(S_1, S_2)^2$ |
| 单维期权路径数 | 10000 |
| 步骤数 | 365 |

与此同时，通过实施基于二叉树的期权定价方法，本文工程也得到了交叉验证。我们设置， $S_{1,2}(0)=100, K=100, r=0.05, q_{1,2}=0.1, \sigma_{1,2}=0.2, \rho=0.1$ ，通过试验可以得到，逆向 LS 方法期权价格 $P=9.85 \pm 0.12$ ，二叉树方法期权价格 $P=10.00$ 。

就通用美式期权定价来说，本文所选基函数对应实验结果够好但还不最好，这取决于基础函数的类型和数量，也需要实施多次测试。

5.2 测试平台和估算方法

我们基于 RTL(Verilog) 语言开发，以最少的 FPGA 资源到达了 7.5ns(133.33MHz)的时钟周期，通过在 Xilinx Vivado 开发平台使用基准测试案例，将布局布线的资源利用情况导入到 Xilinx 功耗估算器，获得了各个模块的功耗估算值。我们以类似的方式，基于同一开发板，也测算了不同带宽下的 DRAM DDR4 功耗值。整个测试过程遵循了完整架构的相同访问模式，实测的读写带宽，分别达到了其最大理论峰值的 75%和 66%。

5.3 测试结果

对于给定的参数设置，FPGA 按照模块分组的资源利用率如表 2 所示。

如前所述，本文架构中的各模块流水线配置时间间隔为 1 个时钟周期，这意味着每个模块在每个时钟周期的开始即可处理一个新的数据。如表 3 所示，在 4266MB/s 的情况下，27.85MB 的

表 2 : FPGA 资源利用情况表

| 步骤 | 模块 | 查找表 | 寄存器 | 数字信息处理 | 块存储 |
|----|----------|-------|-------|--------|-----|
| 1 | 正向路径生成模块 | 19120 | 18345 | 234 | 70 |
| | 路径触发行权模块 | 1654 | 1832 | 0 | 45 |
| 2 | 重配置模块 | - | - | - | - |
| 3 | LS 算法模块 | 34987 | 45875 | 421 | 256 |
| | 逆向路径生成模块 | 28029 | 31404 | 218 | 103 |

表 3 : 架构各模块运行时间、功耗及功率情况表

| 模块 | 运行时间(ms) | 功耗(mJ) | 动态功率(mW) |
|---------------|----------|--------|----------|
| 正向路径生成模块 | 4.24 | 7.15 | 1024 |
| 路径触发行权模块 | 0.02 | 0.01 | 262 |
| 路径存储写入模块 | 4.24 | 6.08 | 1196 |
| 梅森旋转通信 1/2 模块 | 1.21 | 0.12 | 133 |
| 重配置模块 | 32.00 | 62.00 | 1740 |
| 梅森旋转通信 2/2 模块 | 0.62 | 0.06 | 128 |
| 路径存储读取模块 | 4.24 | 7.25 | 1278 |
| 逆向路径生成模块 | 4.24 | 7.01 | 906 |
| 回归数据收集模块 | 4.24 | 4.66 | 845 |
| 回归系数计算模块(CPU) | 2.03 | 0.24 | 114 |
| 现金流更新模块 | 4.24 | 1.88 | 336 |
| 现金流模块 | 4.24 | 1.01 | 150 |
| 当前路径生成模块 | 4.24 | 1.97 | 262 |
| 累加模块 | 0.02 | 0.00 | 75 |

数据处理时间大约只需 4.24ms。与此同时，考虑到 CPU 和 FPGA 之间的通信延时，排除掉 FPGA 的重配置耗时外，一个期权定价的总运行时间仅为 12.36ms。

作为对比，我们在 Intel Xeon (3.50GHz) CPU 平台上，使用 Matlab 语言，实现了本文算法工程。对于给定参数设置，基纯 CPU 架构的测试，一个期权定价的总耗时达到了 270ms，而能耗也

将近 12.70J。可以看出，相较于纯 CPU 架构，本文基于 FPGA 和 CPU 的混合架构，整体延时仅为 12.36ms，整体功耗仅为 37.8mJ，速度提升了近 22 倍，而功耗降低了近 336 倍。

在动态能耗分析方面，参考表 3，可以计算出整体架构每个模块的动态能耗。图 2 展示了基于新的逆向 LS 算法的整体能耗组成，其中，梅森旋转通信能耗包括了初始化前向 MT 模块内部

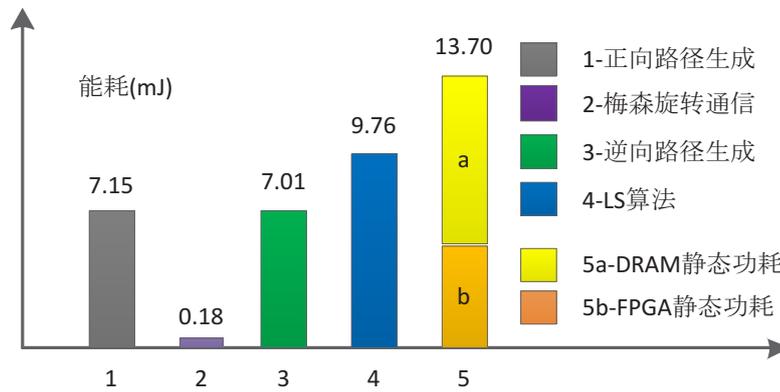


图 2：本文架构各模块动态功耗图

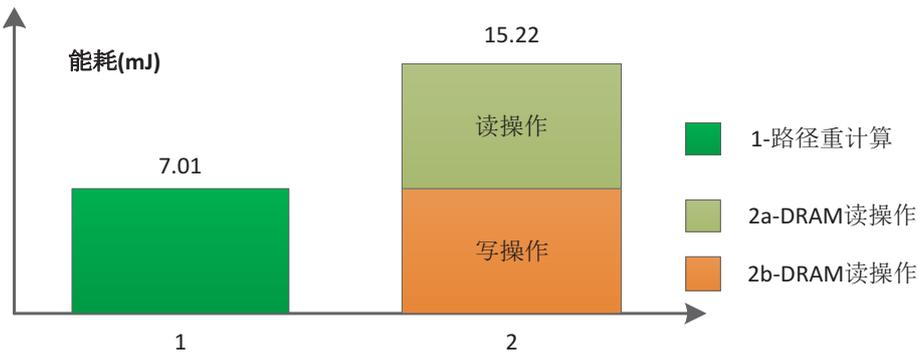


图 3：基于逆向路径生成方法与 DRAM 路径存储方法的能耗对比图

状态的能耗、读取最终状态索引的能耗以及初始化反向 MT 模块的能耗。图中的 LS 项能耗包括了 FPGA 能耗 (9.52mJ) 和 CPU 能耗 (0.24mJ)，其中 CPU 能耗又包括了计算回归系数的能耗，读取累加矩阵的能耗，以及并写入回归系数的能耗。

在逆向 LS 算法与路径存储方法的能耗对比方面，考虑到混合 CPU/FPGA 架构运行在 Linux CPU 内核，而 DRAM 已被操作系统调用，为了进行公平的比较，我们仅考虑了额外的动态能耗。如图 3 所示，可以看出，当两种方法都基于相同的参数设置和操作步骤时，逆向路径生成方相较于 DRAM 路径存储方法的能耗降低了近 1 倍。

6 总结

美式期权定价给金融机构大型计算集群

带来了严重挑战。本文提出了一种基于混合 CPU/FPGA 器件，兼具低延时和高能效的美式期权定价系统。与当前技术相比，这是一个基于 FPGA 实现高维美式期权定价加速的典型实例。

本文主要贡献是利用重计算和位运算实现了逆向 LS 算法。通过重计算，得以精简存储路径，消除带宽限制，提高能源效率。通过利用运行时重配置并利用优化的调度算法来分摊重新配置时间，实现了更高的能源效率。在这方面，与 Matlab 中经过优化的纯 CPU 平台架构相比，本文体系架构在速度方面提高了 22 倍，在能耗方面降低了 336 倍。

本文工程实施得到了德国凯泽斯劳滕大学 de Schryver 教授的支持，我们对其深表感谢。

基于图技术的基金抱团识别方法的研究

武凯, 刘宏, 王洪涛 / 国金证券股份有限公司 信息技术部 上海 201204 邮箱: xh_wukai@126.com



本文与市场上常见的基金抱团识别方式不同, 采用了图相关技术知识, 建立基于基金持股网络的基金抱团识别模型, 可供基金投资者和管理者参考使用。首先, 通过构建基金持股网络图, 利用图嵌入算法获取基金持股网络图中的个股结构化的图特征, 同时引入几类个股定量特征, 利用综合评分模型合成个股的抱团得分。其次, 构建基金网络抱团得分分子图, 利用社区发现算法对基金进行社区划分。最终, 给出本期抱团基金和主要抱团的行业板块。

1 概述

从业务场景的角度分析基金抱团, 可知不同基金在配置持股时, 存在投资理念的相似性。这种投资理念的相似性主要体现在持股的关联性, 即不同基金所持个股的重叠关系。当市场上绝大多数基金配置相同个股或相同板块的个股时, 就是所谓的基金抱团行为。然而图技术是天然的处理数据关联关系的工具, 能够很好的刻画出市场上基金所持个股的重叠关系, 这种图上的重叠关系就可以理解为图结构特征。通过提取基金持股关系的图结构特征, 可以弥补传统基金抱团识别模型的定量特征单一性的缺点。同时, 应用关系

型的技术去处理关系型的问题, 能够进一步提升基金抱团识别模型的准确性, 为证券行业持股关系研究带来了新的思考。具体模型方案如下图 1 所示。

2 基础知识

2.1 基金持股网络图

基金持股网络图是由节点和边组成, 图中的节点有两类, 分别为基金和股票, 图中的边为基金与股票之间的持股关系, 边的属性为持股占比值。描述的是基金与股票之间某一种联系。如下图 2 所示。文中子图是由基金持股网络图中分离

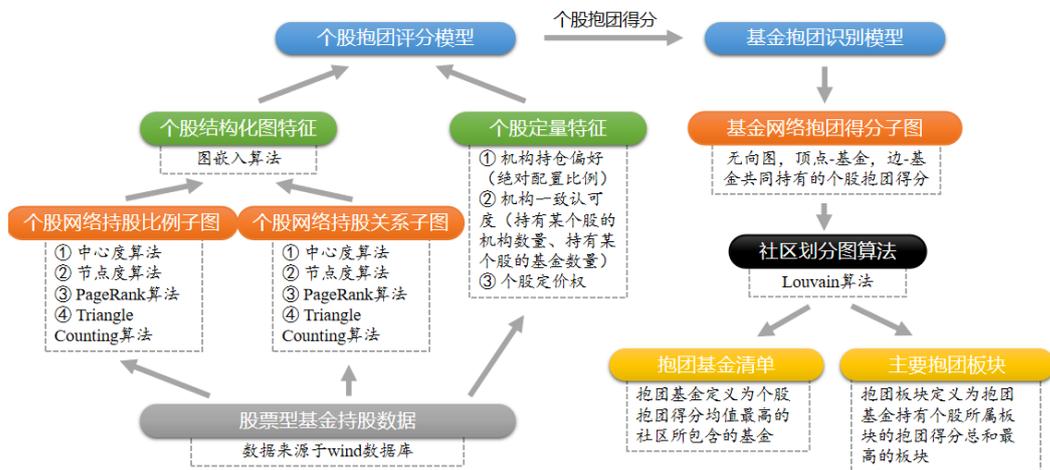


图 1：模型方案

出来的一部分小图。本文共构建了三个子图，分别为个股网络持股关系子图、个股网络持股比例子图和基金网络抱团得分分子图。

个股网络持股关系子图，是指从基金持股网络图中抽出个股之间有持股关系的子图，图中节点为个股，边为两支个股被基金共同持有的关系，边的属性为两支个股被多支基金持有的基金

个数，该子图为无向图。如下图 3 所示。

个股网络持股比例子图，是指从基金持股网络图中抽出个股之间的持股比例关系子图，图中节点为个股，边为两支个股被基金共同持有的关系，边的属性为两支个股被基金持股占比的最小值。若两支个股被多支基金持有，边的属性值取上述多支基金持股占比最小值的平均值，该子图

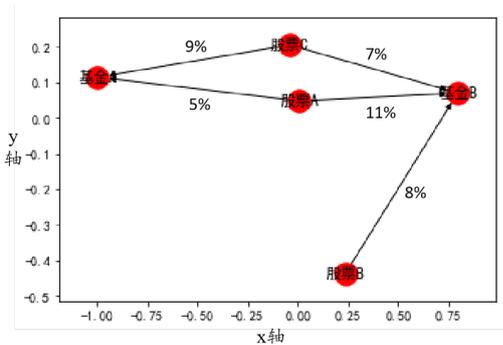


图 2：基金持股网络图

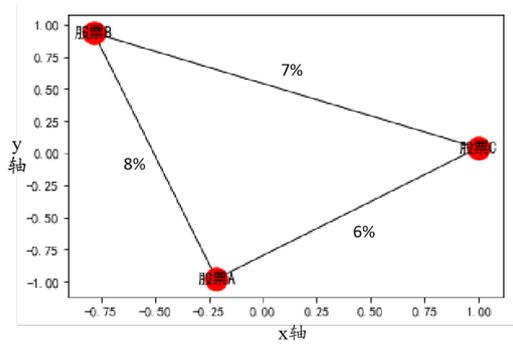


图 4：个股网络持股比例子图

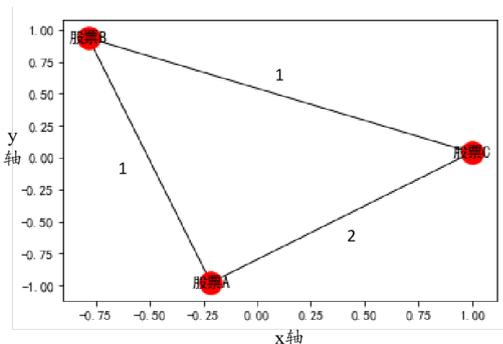


图 3：个股网络持股关系子图

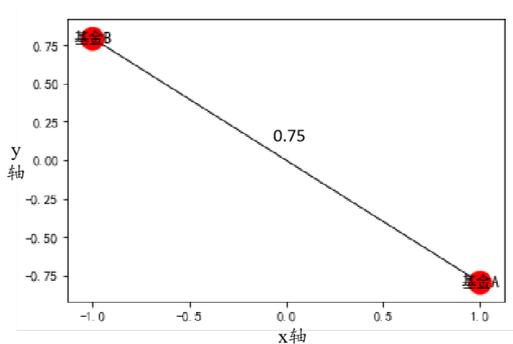


图 5：基金网络抱团得分分子图

为无向图。如下图 4 所示。

基金网络抱团得分子图，是指从基金持股网络图中抽出基金之间的持股关系子图，图中节点为基金，边为两支基金共同持有个股的关系，边的属性为个股的抱团得分，该子图为无向图。如下图 5 所示。

2.2 模型特征

本文在建立个股抱团评分模型和基金抱团识别模型时，区别于市场上常见的基金评价模型的方法。针对入模特征变量，创新地应用了图领域相关的技术。因此，本文的基金抱团识别模型特征变量的设计，在常规基金抱团识别模型的定量特征基础上，充分地结合了图特征变量。

本文的图特征是基于上述个股网络持股关系子图和个股网络持股比例子图通过图嵌入算法计算出来的，图嵌入算法主要应用了节点度算法、中心度算法（主要计算了度中心性和中介中心性两个指标）、Triangle Counting 算法和 PageRank 算法。

由于机构掌握基金市场的筹码相对较多，其增减仓行为对个股价格的影响较大。因此，本文采用的个股定量特征是从机构角度量化描述个股的抱团行为，机构维度的抱团识别特征主要包括机构持仓偏好、一致认可度和个股定价权。其中，机构持仓偏好是指绝对配置比例，是用机构持有某个股的市值除以机构持有的总市值，如公式（1）所示。一致认可度是用持有个股的基金数量和基金公司的数量来衡量。个股定价权是指个股配置占流通股本的比例。如公式（2）所示。

$$\text{绝对配置比例} = \frac{\text{fund_}MV_i}{\sum_{i=1}^N \text{fund_}MV_i} \quad (1)$$

$$\text{个股定价权} = \frac{\text{fund_}MV_i}{\text{market_}MV_i} \quad (2)$$

其中， $\text{fund_}MV_i$ 为机构持有个股 $\sum_{i=1}^N \text{fund_}MV_i$ 的市值， $\text{market_}MV_i$ 为机构持有的总市值，为机构持有个股 i 的流通市值。

3 实例分析

3.1 个股抱团评分模型

3.1.1 数据处理

本研究采用的数据主要来源于 WIND 数据库，共采用了 5785 支基金数据。基金筛选逻辑如下：（1）基金投资风格包括普通股票型、股票型、成长型、强化收益型、保本增值型、增强型、收益型、价值增长型、优选稳健型、积极成长型、成长收益复合型、优化增强型、增值型、稳定型、混合型、平衡型、优选增长型、增强回报型、增强型、增强收益型、保本混合型、稳健增长型、主题型、稳定增值型、灵活配置型、积极配置型、创新型、平稳型、价值型、稳健型、积极型；（2）基金存续状态为正常和发行；（3）基金到期日期和退市日期均晚于 2021 年 07 月 27 日。

本文基金持股数据采用 WIND 数据库中截止于 2021 年 07 月 27 日的最新持股数据，同时对持股数据进行数据处理，删除缺失数据和无效数据，最终加工得到共计 54217 条有效的基金持股数据。

3.1.2 建立基金持股网络图

1) 个股网络持股关系子图

针对上述采集到的基金持股数据，按照基金持股网络图构建方式。定义个股为节点，两支个股被基金共同持有的关系为边，两支个股被多支基金持有的基金个数为边的属性。在软件 Python 中利用 networkx 模块构建个股网络持股关系子图，如下图 6 所示。

2) 个股网络持股比例子图

同理，定义个股为节点，两支个股被基金共同持有的关系为边，边的属性为两支个股被基金持股占比的最小值。若两支个股被多支基金持有

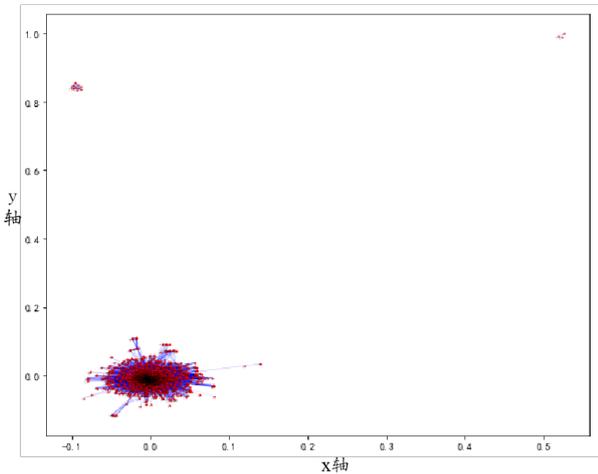


图 6：个股网络持股关系子图

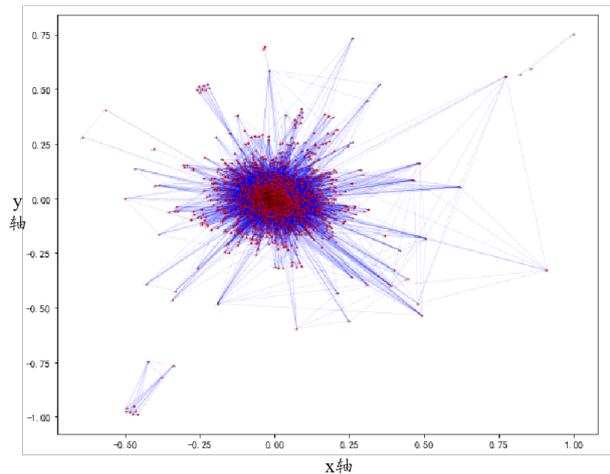


图 7：个股网络持股比例子图

时，边的属性值取上述多支基金持股占比最小值的平均值，如下图 7 所示。

3.1.3 建立个股抱团评分模型

1) 加工模型特征

个股抱团评分模型共采用了两方面的模型特征，一方面是个股网络结构化图特征，另一方面是个股定量特征。其中，个股网络结构化图特征是通过上述两个子图的图特征加工得到。首先，利用图嵌入算法，分别计算出个股网络持股关系子图和个股网络持股比例子图的相关图特征，即 Triangle Counting 指标、节点度指标、度中心性指标、中介中心性指标和 PageRank 指标。其次，计算每个指标的平均值作为最终的人模图特征变量。

同时，由于不同的评价指标往往具有不同的量纲和量纲单位，为了消除指标之间量纲的影响，需要对得到的两方面入模特征变量进行数据归一化处理，如公式（3）所示。处理后的各指标处于同一数量级，适合进行综合对比评价。

$$x_i' = \frac{x_i - x_{\min} + 0.001}{x_{\max} - x_{\min}} \quad (3)$$

x_i' 为特征变量 x_i 归一化处理后的数值， x_{\max}, x_{\min} 分别为一组特征变量的最大值和最小值。其中，为了避免一组变量的最小值在归一化

处理后为 0，公式中分子加了常数 0.001 作为调整参数。

2) 特征筛选

个股抱团评分模型特征变量直方图分布情况如下图 8 所示，通过对 9 组特征变量进行数据相关性分析，可以得出持有个股机构数和持有个股基金数两个特征变量相关性很强，度中心性和节点度变量相关性也很强。因此，入模特征变量去除了持有个股机构数和度中心性两组特征变量。

3) 建立评分模型

综上，构建了 2 个大类，7 个小项的个股抱团衡量指标体系，用来定量描述个股抱团行为，采用综合评分模型来加权合成最终个股抱团综合得分。综合评分模型的变量权重采取专家权重的方式，具体入模变量和权重如下表 1 所示。

表 2 展示了个股抱团评分模型输出结果中，排名前 20 支股票的抱团得分，可见抱团个股主要为酒类、电力设备新能源和医药等行业的龙头股票。其中，抱团得分最高的个股包括贵州茅台、宁德时代、五粮液、招商银行和药明康德等。

3.2 基金抱团识别模型

3.2.1 建立基金网络抱团得分分子图

同理，基于基金持股数据，按照基金持股网络图构建方式。定义基金为节点，两支基金共同

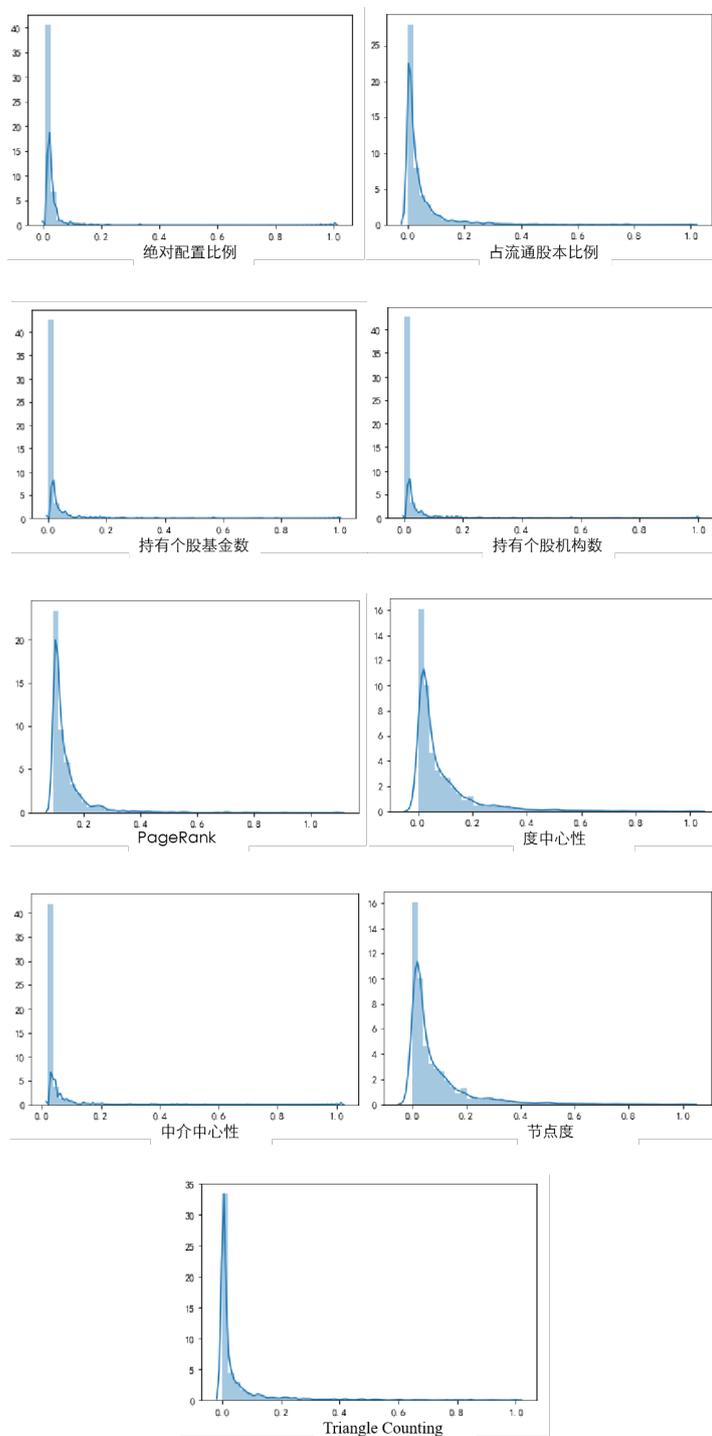


图 8 : 个股抱团评分模型特征变量

持有个股的关系为边，所持有个股的抱团得分为边的属性，若存在两支基金共同持有多支个股时，边的属性值为多支个股抱团得分的求和。在软件 Python 中利用 networkx 模块构建基金网络抱团得分分子图，如下图 9 所示。

3.2.2 基金社区发现

从模型角度分析，可以得知基金网络抱团得分分子图是用来描述基金抱团关系的图表示。其中，抱团基金在图中分布会相对聚集，或者可以理解为抱团基金在图中的距离相对更近，关联关

表 1：模型变量与权重

| 变量 大类 | 个股定量特征 | | | 图特征 | | | |
|----------|----------------|---------------------|---------------------|---------------|-------------|-----------------------|----------|
| 变量 名称 | 绝对 配置 比例 | 持有 个股 基金 数 | 占流 通股 本比 例 | 三角 形个 数 | 节 点 度 | 中 介 中 心 性 | pagerank |
| 变量 权重 | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 |

表 2：个股抱团得分（排名前 20 支股票）

| 股票 WIND 代码 | 股票简称 | 绝对配置比例 | 持有个股基金数 | 三角形个数 | 度值 | 介数中心性 | 占流通股本比例 | pagerank | 抱团得分 |
|------------|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 600519.SH | 贵州茅台 | 0.02370993 | 1.000000485 | 1.000000028 | 1.000001316 | 1.019171698 | 0.002333713 | 1.091827963 | 0.616308926 |
| 300750.SZ | 宁德时代 | 0.02774678 | 0.784779932 | 0.915083416 | 0.881580263 | 0.700937653 | 0.006650892 | 0.904481224 | 0.504043776 |
| 000858.SZ | 五粮液 | 0.022493667 | 0.730974794 | 0.845258686 | 0.797369737 | 0.462427567 | 0.004806487 | 0.807420744 | 0.442902663 |
| 600036.SH | 招商银行 | 0.019414224 | 0.570044111 | 0.846589489 | 0.827632895 | 0.55184218 | 0.001882885 | 0.771180254 | 0.417992726 |
| 603259.SH | 药明康德 | 0.022679601 | 0.567620456 | 0.765580344 | 0.710527632 | 0.390385931 | 0.010724501 | 0.704822818 | 0.377336584 |
| 601012.SH | 隆基股份 | 0.026770652 | 0.492971886 | 0.812526574 | 0.757896053 | 0.441904161 | 0.007020835 | 0.703885929 | 0.376973946 |
| 002415.SZ | 海康威视 | 0.024042134 | 0.428987397 | 0.793187445 | 0.747369737 | 0.481691632 | 0.009741746 | 0.669724134 | 0.36175155 |
| 300059.SZ | 东方财富 | 0.020720878 | 0.321861852 | 0.694849534 | 0.675001316 | 0.375898422 | 0.005967024 | 0.564271931 | 0.300712071 |
| 000001.SZ | 平安银行 | 0.015771747 | 0.307319922 | 0.677039414 | 0.668422368 | 0.37379783 | 0.003386223 | 0.561240711 | 0.293345611 |
| 601888.SH | 中国中免 | 0.018588499 | 0.371304411 | 0.648696124 | 0.625001316 | 0.298393916 | 0.005091185 | 0.544928832 | 0.290698838 |
| 601166.SH | 兴业银行 | 0.014615002 | 0.245759089 | 0.582807175 | 0.652632895 | 0.491330019 | 0.002730659 | 0.54488288 | 0.279786247 |
| 601318.SH | 中国平安 | 0.019211151 | 0.265148328 | 0.547809865 | 0.600001316 | 0.461550181 | 0.000633187 | 0.510649425 | 0.268999612 |
| 300760.SZ | 迈瑞医疗 | 0.024439209 | 0.373243335 | 0.60058898 | 0.555264474 | 0.164002571 | 0.016920684 | 0.517696729 | 0.266675921 |
| 002475.SZ | 立讯精密 | 0.019756658 | 0.234610276 | 0.660531784 | 0.630264474 | 0.2896031 | 0.008267367 | 0.499196569 | 0.260486453 |
| 300014.SZ | 亿纬锂能 | 0.025271729 | 0.249152206 | 0.609649772 | 0.568422368 | 0.170815728 | 0.017397111 | 0.481158095 | 0.241368806 |
| 600905.SH | 三峡能源 | 0.014775958 | 0.17692729 | 0.539117167 | 0.597369737 | 0.38493427 | 0.004822445 | 0.447119552 | 0.236159211 |
| 689009.SH | 九号公司 | 0.010260015 | 0.010179835 | 0.044596115 | 0.105264474 | 0.021588314 | 1.000046992 | 0.140922014 | 0.23533446 |
| 002142.SZ | 宁波银行 | 0.016722007 | 0.257392632 | 0.514086729 | 0.531580263 | 0.191733809 | 0.009627848 | 0.465308049 | 0.227019382 |
| 603501.SH | 韦尔股份 | 0.027955825 | 0.190015027 | 0.602599343 | 0.568422368 | 0.200955481 | 0.008254583 | 0.431525683 | 0.225595374 |
| 300197.SZ | 节能铁汉 | 1.005901523 | 0.000969947 | 2.83433E-05 | 0.001317105 | 0.019171698 | 0.011325188 | 0.158933033 | 0.22158435 |

系相对更紧密。针对基金在基金网络抱团得分分子图中的分布结构，结合图相关技术，可以很好的识别出抱团基金。本文利用社区发现算法对基金网络抱团得分分子图中的基金进行社区划分。在软件 Python 中利用 Louvain 算法基于模块度发现基金网络抱团得分分子图层次性的社区结构。采

用 Louvain 算法的最优社区划分标准，本文划分基金社区为 5 个社区，其中，基金最大社区含有 2651 支基金，基金最小社区含有 2 支基金。划分后的基金社区分布情况如下图 10 所示。

3.2.3 确定抱团基金

基于上述基金社区划分结果，分别计算各社

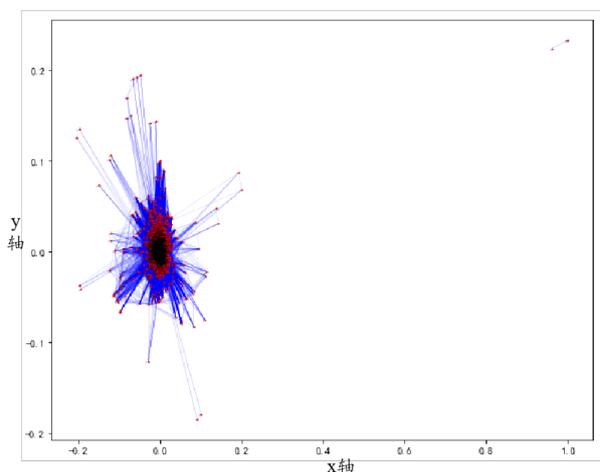


图 9：基金网络抱团得分分子图

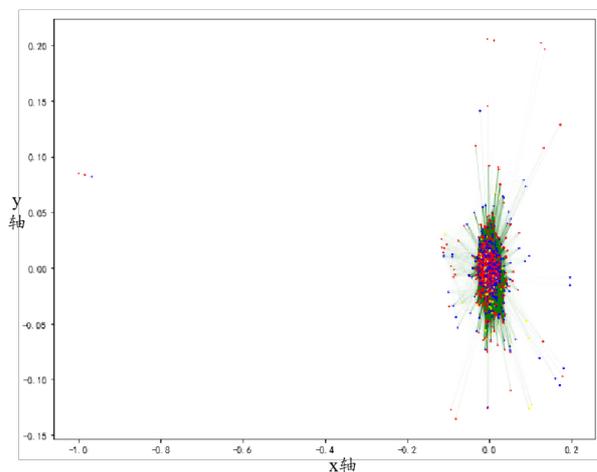


图 10：基金社区划分情况

表 3：社区划分情况

| 社区编号 | 基金数量 | 社区内个股抱团得分均值 |
|------|------|-------------|
| 3 | 1397 | 0.231705086 |
| 2 | 1711 | 0.17687791 |
| 1 | 2651 | 0.162879806 |
| 4 | 3 | 0.150680891 |
| 0 | 2 | 0.01990606 |

区内基金持有个股的抱团得分平均值，计算结果如下表 3 所示。选取社区内个股抱团得分平均值最大的社区确定为基金抱团社区。定义抱团基金为抱团社区内所包含的基金。因此，得到本期抱团基金数量为 1397 支基金，占同期股票型基金总数的 24.14%。表 4 展示了抱团基金排名前 20 支基金的信息。

3.3 确定抱团板块

针对模型确定出的抱团基金，将抱团基金所持个股按照 WIND 四级行业分类标准进行板块分类，再计算出各行业板块的个股抱团得分总和作为板块的抱团得分。板块抱团得分排名前 15 的板块如下表 5 所示。我们定义板块抱团得分最高的板块为本期基金抱团板块，由此得出本期基金抱团板块为白酒与葡萄酒板块。

4 总结

从基金持仓角度分析基金抱团行为，是市场上绝大多数基金配置持股时，集中配置热门股票或热门板块的股票。因此，抱团基金的持股存在很强的重叠性。图技术是天然的处理数据关联关系的工具，能够很好的刻画出市场上基金所持个股的重叠关系。本文通过构建基金持股网络图，结合图相关技术，建立个股抱团评分模型和基金抱团识别模型，通过模型研究最终给出本期抱团基金名单，并确定本期基金抱团板块为白酒与葡萄酒板块。本文研究的基于基金持股网络的基金抱团识别模型，在基金市场投资和管理考核方面，可供基金投资者和管理者参考使用。此外应用研究还表明：本文将图相关技术应用到基金持股网络研究中，为证券行业持股关系研究带来了新的思考。

表 4：抱团基金排名前 20 支基金

| 基金 WIND 代码 | 基金名称 | 基金简称 | 管理人 | 投资风格 |
|------------|----------------------------|--------------|--------|-------|
| 000001.OF | 华夏成长证券投资基金 | 华夏成长 | 华夏基金 | 成长型 |
| 007548.OF | 易方达 ESG 责任投资股票型发起式证券投资基金 | 易方达 ESG 责任投资 | 易方达基金 | 股票型 |
| 007550.OF | 东兴未来价值灵活配置混合型证券投资基金 | 东兴未来价值 C | 东兴基金 | 灵活配置型 |
| 007569.OF | 南方安福混合型证券投资基金 | 南方安福 C | 南方基金 | 混合型 |
| 007570.OF | 方正富邦红利精选混合型证券投资基金 | 方正富邦红利精选 C | 方正富邦基金 | 混合型 |
| 007574.OF | 宝盈新价值灵活配置混合型证券投资基金 | 宝盈新价值 C | 宝盈基金 | 灵活配置型 |
| 007632.OF | 华润元大安鑫灵活配置混合型证券投资基金 | 华润元大安鑫 C | 华润元大基金 | 灵活配置型 |
| 007534.OF | 格林创新成长混合型证券投资基金 | 格林创新成长 C | 格林基金 | 混合型 |
| 007663.OF | 平安安享灵活配置混合型证券投资基金 | 平安安享灵活配置 C | 平安基金 | 灵活配置型 |
| 007686.OF | 东方价值挖掘灵活配置混合型证券投资基金 | 东方价值挖掘 C | 东方基金 | 灵活配置型 |
| 007777.OF | 中邮研究精选混合型证券投资基金 | 中邮研究精选 | 中邮基金 | 混合型 |
| 007781.OF | 天弘弘新混合型发起式证券投资基金 | 天弘弘新 | 天弘基金 | 混合型 |
| 007863.OF | 长信利泰灵活配置混合型证券投资基金 | 长信利泰 C | 长信基金 | 灵活配置型 |
| 007884.OF | 易方达恒盛 3 个月定期开放混合型发起式证券投资基金 | 易方达恒盛 3 个月定开 | 易方达基金 | 混合型 |
| 007893.OF | 平安估值精选混合型证券投资基金 | 平安估值精选 A | 平安基金 | 混合型 |
| 007894.OF | 平安估值精选混合型证券投资基金 | 平安估值精选 C | 平安基金 | 混合型 |
| 007669.OF | 太平睿盈混合型证券投资基金 | 太平睿盈 C | 太平基金 | 混合型 |
| 007533.OF | 格林创新成长混合型证券投资基金 | 格林创新成长 A | 格林基金 | 混合型 |
| 007524.OF | 汇添富内需增长股票型证券投资基金 | 汇添富内需增长 C | 汇添富基金 | 股票型 |
| 007523.OF | 汇添富内需增长股票型证券投资基金 | 汇添富内需增长 A | 汇添富基金 | 股票型 |

表 5：基金抱团板块得分

| WIND 四级行业分类 | 板块得分 |
|-------------|-------------|
| 白酒与葡萄酒 | 1327.123381 |
| 电气部件与设备 | 198.7415383 |
| 生命科学工具和服务 | 148.175068 |
| 多元化银行 | 135.5997408 |
| 半导体产品 | 115.3851945 |
| 酒店、度假村与豪华游轮 | 111.162987 |
| 电子设备和仪器 | 103.0073121 |
| 医疗保健设备 | 76.25211105 |
| 家用电器 | 72.36765096 |
| 生物科技 | 65.37373794 |
| 电子元件 | 63.88502166 |
| 食品加工与肉类 | 45.98080434 |
| 投资银行业与经纪业 | 44.84417366 |
| 区域性银行 | 34.18164034 |
| 保健护理设施 | 32.82084325 |



I 信息资讯采撷 Information

监管科技全球追踪

监管科技全球追踪

5月31日至6月1日，人民银行与国际清算银行（BIS）、欧洲中央银行、央行与监管机构绿色金融网络（NGFS）联合举办第二届“绿天鹅”视频会议，围绕“央行与绿色转型：下一步工作”的主题，回顾了各方自首届“绿天鹅”会议以来在绿色金融领域的积极进展，并就完善国内外合作、绿色通胀、未来一年的重点工作等问题进行了探讨。

6月1日，欧洲央行执行委员会成员法比奥帕内塔在法兰克福举行的泛欧洲金融基础设施欧洲网络弹性委员会（ECRB）第七次会议上发表演讲：瞬息万变的网络威胁环境不容自满。

6月6日，新华三发布新闻称，后疫情时代，无接触、线上化、云端化成为金融业务发展的主流趋势之一，也使得在线支付、远程支付、非接触式支付的需求爆发式增长。新华三集团为农信银资金清算中心打造了一套具有自主知识产权的全栈式“农信云”，构建农村支付互联网，搭建支付清算系统，助力金融行业数字化发展。

6月9日，国家网信办联合市场监督管理局开展数据安全认证工作，鼓励网络运营者通过认证方式规范网络数据处理活动，加强网络数

据安全保护。

6月14日，网信办发布新修订的《移动互联网应用程序信息服务管理规定》，并自2022年8月1日起施行。新规定旨在进一步依法监管移动互联网应用程序，促进应用程序信息服务健康有序发展。

6月16日，工行深圳分行与深圳市科技创新委员会联合举行科技金融创新发展战略合作框架协议签约暨高新区科技金融创新服务中心揭牌仪式。该中心旨在推动深圳科技与金融深度融合，探索科技金融创新发展新路径，形成可复制、可推广的科技金融创新产品和服务模式。

6月22日，英国政府宣布，将继续加强针对“先买后付”（Buy-Now Pay-Later, BNPL）服务的监管。在新的监管要求下，提供BNPL的机构需要对借款人进行还款能力检查，确保借款人能够负担其所用额度，并修改广告宣传监管规则，要求BNPL广告公平清晰，且不会误导消费者。

6月23日，国务院印发《关于加强数字政府建设的指导意见》，就主动顺应经济社会数字化转型趋势，充分释放数字化发展红利，全面开

创数字政府建设新局面作出部署。

7月4日，经国务院批准，人民银行与香港金融管理局签署常备互换协议，将双方自2009年起建立的货币互换安排升级为常备互换安排，以进一步深化内地与香港金融合作，更好支持香港国际金融中心建设，推动香港离岸人民币市场稳健发展。

7月5日，金融稳定协会发布报告：《大型科技公司的依存关系—关键政策盲点》(Big tech interdependencies a key policy blind spot)。报告指出，金融行业内大型科技公司的依存关系为运营弹性造成风险，因此需要为其制定基于实体的运营规则。

7月7日，2022数字金砖论坛在线上举办。本届论坛落实《金砖国家领导人第十四次会晤北京宣言》，以“金砖国家携手拥抱数字化转型”为主题，邀请金砖国家和国际电信联盟有关代表，

共话数字经济全球发展、共商产业转型交流合作、共享技术创新成果经验。

7月11日，金融稳定委员会发布关于加密资产活动的国际监管合规声明，重申金融稳定委员会致力于促进监管合规方法的国际一致性，并促进国家当局和国际标准制定者之间的跨境和跨部门合作。

7月15日，为进一步规范商业银行互联网贷款业务经营行为，促进互联网贷款业务平稳健康发展，银保监会印发《关于加强商业银行互联网贷款业务管理提升金融服务质效的通知》。

7月20日，欧盟中央银行(ECB)、欧盟银行管理局(EBA)、欧盟证券和市场管理局(ESMA)及美国证券交易所(SEC)、美联储(FRB)、美国联邦存款保险(FDIC)等机构在欧盟-美国联合金融监管论坛上展开交流，聚焦数字金融等六大领域。

2022年三季度《交易技术前沿》征稿启事

《交易技术前沿》由上海证券交易所主管、主办,以季度为单位发刊,主要面向全国证券、期货等相关金融行业的信息技术管理、开发、运维以及科研人员。2022年二季度征稿主题如下:

一、云计算

(一) 云计算架构

主要包含但不限于:云架构剖析探索,云平台建设经验分享,云计算性能优化研究。

(二) 云计算应用

主要包含但不限于:云行业格局与市场发展趋势分析,国内外云应用热点探析,金融行业云应用场景与实践案例。

(三) 云计算安全

主要包含但不限于:云系统下的用户隐私、数据安全探索,云安全防护规划、云安全实践,云标准的建设、思考与研究。

二、人工智能

(一) 应用技术研究

主要包含但不限于:语音识别与自然语言处理,计算机视觉与生物特征识别,机器学习与神经网络,知识图谱,服务机器人技术。

(二) 应用场景研究

主要包含但不限于:智能客服、语音数据挖掘、柜员业务辅助等。

主要包含但不限于:监控预警、员工违规监控、交易安全等。

主要包含但不限于:金融预测、反欺诈、授信、辅助决策、金融产品定价、智能投资顾问等。

主要包含但不限于:金融知识库、风险控制等。

主要包含但不限于:机房巡检机器人、金融网点服务机器人等。

三、数据中心

(一) 数据中心的迁移

主要包含但不限于:展示数据中心的接入模式和网络规划方案;评估数据中心技术合规性认证的必要性;分析数据中心迁移过程中的影响和业务连续性;探讨数据中心迁移的实施策略和步骤。

(二) 数据中心的运营

主要包含但不限于:注重服务,实行垂直拓展模式;注重客户流量,实行水平整合模式;探寻数据中心运营过程中降低成本和提高服务质量的途径。

四、分布式账本技术(DLT)

(一) 主流分布式账本技术的对比

主要包含但不限于：技术架构、数据架构、应用架构和业务架构等。

（二）技术实现方式

主要包含但不限于：云计算 + 分布式账本技术、大数据 + 分布式账本技术、人工智能 + 分布式账本技术、物联网 + 分布式账本技术等。

（三）应用场景和案例

主要包含但不限于：结算区块链、信用证区块链、票据区块链等。

（四）安全要求和性能提升

主要探索国密码算法在分布式账本中的应用，以及定制化的硬件对分布式账本技术性提升的作用等。

五、信息安全与 IT 治理

（一）网络安全

主要包括但不限于：网络边界安全的防护、APT 攻击的检测防护、云安全生态的构建、云平台的架构及网络安全管理等。

（二）移动安全

主要包括但不限于：移动安全管理、移动互联网接入的安全风险、防护措施等。

（三）数据安全

主要包括但不限于：数据的分类分级建议、敏感数据的管控、数据共享的风险把控、数据访问授权的思考等。

（四）IT 治理与风险管理

主要包括但不限于：安全技术联动机制、自主的风险管理体系、贯穿开发全生命周期的安全管控、安全审计的流程优化等。

六、交易与结算相关

（一）交易和结算机制

主要包含但不限于：交易公平机制、交易撮合机制、量化交易、高频交易、高效结算、国外典型交易机制等。

（二）交易和结算系统

主要包含但不限于：撮合交易算法、内存撮合、双活系统、内存状态机、系统架构、基于新技术的结算系统等。

投稿说明

1、本刊采用电子投稿方式，投稿采用 word 文件格式（格式详见附件），请通过投稿邮箱 ftt.editor@sse.com.cn 进行投稿，收到稿件后我们将邮箱回复确认函。

2、稿件字数以 4000-6000 字左右为宜，务求论点明确、数据可靠、图表标注清晰。

3、本期投稿截止日期：2022 年 9 月 30 日。

4、投稿联系方式 021-68607129, 021-68607131 欢迎金融行业的监管人员、科研人员及技术工作者投稿。稿件一经录用发表，将酌致稿酬。

附件：投稿格式（可通过电子邮件索要电子模板）

标题（黑体 二号 加粗）

作者信息（姓名、工作单位、邮箱）（仿宋 GB2312 小四）

摘要：（仿宋 GB2312 小三 加粗）

关键字：（仿宋 GB2312 小三 加粗）

一、概述（仿宋 GB2312 小三 加粗）

二、一级标题（仿宋 GB2312 小三 加粗）

（一）二级标题（仿宋 GB2312 四号 加粗）

1、三级标题（仿宋 GB2312 小四 加粗）

（1）四级标题（仿宋 GB2312 小四）

正文内容（仿宋 GB2312 小四）

图：（标注图 X. 仿宋 GB2312 小四）

正文内容（仿宋 GB2312 小四）

表：（标注表 X. 仿宋 GB2312 小四）

正文内容（仿宋 GB2312 小四）

三、结论 / 总结（仿宋 GB2312 小三 加粗）

四、参考文献（仿宋 GB2312 小四）

电子平台

欢迎访问我们的电子平台 <http://www.sse.com.cn/services/tradingservice/tradingtech/sh/transaction/>。我们的电子平台不仅同步更新当期的文章，同时还提供往期所有历史发表文章的浏览与查阅，欢迎关注！

联系电话：021-68607129
021-68607131
投稿邮箱：ftt.editor@see.com.cn

ITRDC

证券信息技术研究发展中心（上海）



中国上海市杨高南路388号

邮编：200127

公众咨询服务热线：4008888400

网址：<http://www.sse.com.cn>

内部资料 免费交流

本资料仅为内部交流使用，本季度印500册，编印单位为上海证券交易所，面向证券期货行业发送，印刷日期为2022年10月，印刷单位为主人印刷厂。

部分图片或文字来源于互联网等公开渠道，其版权归属原作者所有。如有版权相关事宜，请发送邮件至ftt.editor@sse.com.cn