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 第三期 总第50期

本期主题

数字化转型

No.3

交易技术前沿
THE FORELAND OF TRADING TECHNOLOGY

内部资料 2022 年第三期（总第 50 期）
准印证号：沪（K）0671

NO.3

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

党的十八大以来，习近平总书记高度重视信息化建设和数字经济发展。国家“十四五”规划明确提出“加快数字化发展,建设数字中国”,要求金融业“稳妥发展金融科技,加快金融机构数字化转型”。党的二十大报告指出，实施创新驱动发展战略，要面向世界科技前沿、面向经济主战场、面向国家重大需求。资本市场作为我国经济金融领域的重要组成部分，以数字化发展推进资本市场的质量变革、效率变革、动力变革，助力资本市场实现高水平开放和高质量发展，是新时代资本市场更好服务实体经济的重要路径。

金融业正处于传统业态模式重构和数字化转型变革的交汇时点，这对金融机构如何维护金融安全、规范市场运行提出了挑战。日前，在会科技局统筹部署下，“证券期货业数字化转型”主题宣传月顺利召开。为响应宣传月主题，进一步促进行业数字化高质量发展，《交易技术前沿》推出本期“数字化转型”主题专刊，收录行业数字化转型及其相关应用的优秀文章，旨在针对金融机构，特别是证券期货业机构在数字化转型中存在的体系建设、实施路径、技术方案等共性问题进行讨论研究，其中。

《证券公司数字化转型实践与研究》以证券公司视角深入研究分析证券行业数字化特点与现状，并结合发展趋势，设计了证券公司数字化转型能力体系，提出数字化转型的实施建议。

《数字化转型下的技术合规建设》聚焦于技术合规建设领域，主要从实战经验和解决方案出发，探讨了合规建设，特别是安全和数据合规建设的基本流程。

《浅议数字化转型下如何做好信息科技的需求管理方法》旨在协调监管需求、业务需求和信息科技部门现有的资源之间存在的矛盾，重点论述了引进项目管理工具、建立符合实际情况的需求制度、建立需求跟踪制度、建立制度修复机制等手段，提出针对需求的管理办法。

《东方证券研发运行一体化平台探索与实践》基于 Devops 方法论及自身实践，利用敏捷开发技术，建设了运行一体化平台，并制定了从需求接受到生产交付的全流程研发体系，实现需求交付效率及研发质量的快速提升。

《数字化转型下的东吴证券数据中台建设实践》介绍了东吴证券围绕数据“易得、易懂、易用、易开发”建设的数据基础设施和数据应用系统的顶层设计、主要架构和技术路线，分享了建设自主可控的数据类系统的实践经验。

《交易技术前沿》编辑部

2022年12月23日

目录 Contents

本期热点

探索与应用

信息资讯采撷

本期热点

数字化转型下的技术合规建设

王志玮、裘岱、张晓军、宋良夏、周蒙 / 上证所信息网络有限公司

本文关注合规建设中的技术部分，从行业实际情况出发，描述了合规建设的基本流程，内容较为实战化和偏解决方案，从安全合规、数据合规等描述了一些重点技术合规领域，不一定完善，希望可以为企业在建设技术合规过程中提供一些思考。

1 背景

依据《中央企业合规管理办法》，合规是指企业经营管理行为和员工履职行为符合国家法律法规、监管规定、行业准则和国际条约、规则，以及公司章程、相关规章制度等要求。很多人理解为合规就是按照规定来做就好了，但是具体遵守哪些规定，且规定大到法律小到技术指引，有偏大方向也有偏细节类型的要求，如何结合企业实际进行落地也不会给出非常详尽的解释和产品适配性，具体的落实需要企业按照自己的情况和业界的情况来把握，或索性找一个权威机构出具某方面的证书，但覆盖面毕竟有限。

从技术角度考虑，每个企业面临的技术阶段不同，技术团队架构也不同，技术管理流程也肯定存在差异，如何去切实地做到技术合规，是本文讨论的主要问题。

2 “不以规矩，无以成方圆”

2.1 确定业务范围、数字化转型方向

首先，企业需判断自己的业务范围和数字化转型方向，通过业务范围确认企业的定位以及企业适用的法律法规、监管规定、行业准则和国际条约、规则，通过数字化转型方向确定企业需重点参照的技术合规要求以及对应的技术标准、行业最佳实践和合规案例。假设企业涉及跨境，则需确认涉及境外地区、业务类型，传输的数据类型等，需研究国际法律法规以及对应的解读内容和案例。一般证券行业涉及的技术相关的合规内容主要包括知识产权、技术类型的项目管理、安全合规、数据合规等以及其他证券行业监管相关要求。

近年来，我国出具了很多的法律法规。如《网络安全法》《个人信息保护法》《证券法》《密码法》

《关键信息基础设施安全保护条例》《网络安全审查办法》《中央企业合规管理办法》等。法律法规是强制需要履行的，技术标准则分为强制执行和参照执行，所以具体的细节如何落地，还需要企业根据自己的实际情况，在满足大方向的情况下，参照标准和企业实际情况进行落实，业界也针对性得产出了各项总结条目供参考，专业的团队力量在这部分工作中会发挥较大作用。

2.2 识别相关方

企业有必要识别相关方并建立和维护相关方清单，以此管理和控制来自相关方的合规风险，包括但不限于电力、电信提供商、机房维护方、技术项目合作方、涉及数据传输的相关方、技术人力外包方等服务资质、监管处罚案例、依赖性、合作需求等以及其他任何可能影响业务、导致事件、争议或案件的风险。该项工作在形成工作机制或统一流程后可不断积累，不建议花费大量的时间用于年度、季度整理，应持续性维护并评估风险。

2.3 反面教材

搜集业界不合规的案例，尤其是公示的对企业影响较大的案例，以及结合企业历史以来遇到的一些案例，组成“避坑”合集即合规风险案例合集，可用于评审、建设、审计等的重点参考内容。就像司法领域中，已有的案例的审判情况对后续案例的审判过程有较大借鉴和参考意义，甚至公众对审判的结论对应了导致舆论影响的负面案例。而且，多了解相关技术合规案例的司法、监管判断也有利于公司了解监管红线。

2.4 建设制度、合规风险库、合规需求清单

对于部分法律法规、行业监管要求，企业不一定适用于其全文的要求，需要结合企业现有业务、技术发展情况进行适用性分析，比如企业具体什么业务涉及该项条款或者如何不涉及该项条款、企业需要履行的具体义务等等。具体企业应怎样履行其要求的合规义务，是制度体系建设者需要重点考虑的事情。

技术合规风险库中风险对应的解决方案即为技术合规需求，技术合规需求如安全需求可以和业务需求同时出具，在技术建设的需求分析阶段即进行介入，系统应充分评估业务需求和技术合规需求，整合成较为合规的技术解决方案，该做法可防止系统上线前后出现合规问题而导致重复开发，或因涉及架构问题无法更新而导致企业承担合规风险。

制度建立需要时间，且需要厘清各部门职责，即由谁来负责落实该项合规要求及制度内容，通常需要较多的时间落实，而且有一些细节的内容比如等保要求等不一定全部落地在制度中，可通过表单的形式进行维护，避免增加制度建设和执行的压力。因此，通过建立技术合规风险库可以帮助企业尽快完成合规任务，类似于等保安全要求的 checklist。

3 统一化管理流程和模板

企业应统一流程和模板，尤其是对外的一系列如合同、隐私政策、数据保护协议、客服回复语、甚至招聘口径等内容，既统一对外口径，又可以防止任何合规模块的遗漏，也可以节省审批流程，因为统一化的模板已经过严密的评估、审核和批准流程，后续补充上个性化内容后，一般不需要走全量流程或审批起来相对容易，仅需针对个性化的内容进行审批。而对外的协议是用户、合作企业及监管最关注的内容，也是出现争议时最强有依赖的证据。

可由权威机构出具模板，且根据适用情况不断更新模板，以适用于更多的实际场景。后续通过实践经验，整合个性化内容，逐步完善后会出现不同类型场景的各类模板，可使得个性化的内容逐步减少，合规风险进一步降低。此外应建立统一的应急处置流程且定期维护及培训，统一流

程也是企业数字化转型的必经之路。

4 完善技术合规规范

4.1 安全合规

目前安全合规已逐渐成熟，企业可参照行业、国际标准如 ISO27001/27002 进行管理，本文仅列举较为影响合规风险的几项内容，且梳理了技术分布图供参考（详见图 1）。

（1）完善基线管理：比如防火墙、操作系统等需要按照统一配置基线进行设置，像是关闭一些不必要的端口、启动已有安全模式等，此类统一的操作降低了技术人员的工作量，也增加了对现有技术系统的掌握程度，是很有必要的技术合规内容。

（2）建立系统上线发布审批流程：不单是走流程，而是通过流程审核各项合规内容是否落实，核对需求阶段的合规需求部分以及数据安全、个人信息保护等方面的内容，必要时需要法务介入审核对外公布协议等内容。

（3）外网 or 内网如何平衡：内外网的选择有时并不明朗，需要结合上层文件要求、系统涉及的数据敏感性、人员权限配置等进行判断，应以数据不泄漏为底线要求。

（4）完善操作留痕、审计工作机制：除了留存日志，企业还需补充集中管控、设置自动告警策略、定期审计和分析等工作。考虑存储空间有限，集中管控日志不仅是把日志搜罗到一起，而是选择性的筛选对事件发生有记录作用的日志，至少包含“4W”信息（Who\When\Where\What）。

（5）日志数据的敏感性：日志信息中有可能留存个人信息如用户名、手机号、操作数据等，有些是外部用户的个人信息或内网 IP。考虑日志并非数据库中的数据，未像数据库中的数据一样得到保护如加密等，所以日志中的敏感数据容易发生数据泄漏，故应引起重视对日志服务器进行保护，如加强权限控制、隔离测试日志和生产日志、尽量避免外网可访问日志内容，以及仅记录低识别度的信息（比如若订单号即可定位交易内容则不需记录账号和手机号）。

（6）API 接口规范：随着数字化转型的进行，技术架构也越来越轻量化，需建立 API 接口数据传输规范，统一加解密过程和流程，并建议加入审核流程中。

（7）应急处置：应定期完善应急预案，包含报送和处置流程以及具体的处置步骤和指令等，且应覆盖可能遇到的场景类别，并定期演练验证有效性。

4.2 数据合规

安全合规中较为重要的内容为数据合规，数

图 1：安全技术合规领域细分

据安全也包含个人信息保护，且分别颁布了数据安全法和个人信息保护法，故重点阐述。

（1）采集和使用的授权：尤其是个人信息（特别是敏感个人信息如人脸）的授权，为目前监管重点关注的事项，也是容易出现合规风险甚至纠纷的事项，该部分内容涉及对外发布的公开信息，需法律专业人士介入并出具专业意见，如隐私政策的内容、出现的时间、获得同意的记录，以及协议等。

（2）加密管理：企业应具备密钥管理系统（软）、加密机（硬）双重产品以维护不同的数据加密需求，如核心交易、个人信息等敏感信息采用加密机，内网数据采用密钥管理系统，以及通过设置电子证书等形式加强前后端的权限控制，也可以保障数据加密的安全进行，对于行业普遍存在的内部信息泄漏也是一个有效的阻断方式，当然也可以通过增加数据获取的认证因素保障数据获取的安全性。

（3）冷备热备温备：数据库的冷备热备选择在最开始是业界比较纠结的问题，随着监管环境、行业实践和数据产品越来越多，企业也基本都形成了自身行之有效的一套选择模式。但是，个人信息的冷备热备保护存在特殊性，首先个人在个人信息处理活动中存在一定的权利，这要求企业需在规定时间内完成个人信息响应，因此冷备个人信息的获取时间需评估，个人信息的存储时间也需按个人权利、监管要求（如反洗钱法、个人信息保护法）等进行明确，此外需要建立个人信息响应流程，以保障反馈时间满足要求。

（4）数据保护协议及责任界定：数据的权属问题若存在争议，对数据保护整体的工作都是很大的挑战，数据共享或转移应通过合同、数据保护协议、数据安全保护评估流程、数据保护问卷等工作机制进行保障，了解双方的数据保护技术能力，以及权利和义务，若数据保护技术能力存在缺失，是否合作则需要重新考虑，此外，数据保护协议需技术人员联合法律专家一同出具。

（5）数据销毁：数据删除≠数据销毁，物理销毁、消磁、覆写均是解决方案，应针对不同的销毁需求，采取不同的方案，如采用销毁密钥的方式销毁加密数据，采用覆写的方式以重复利用存储敏感性较低数据的磁盘等，企业应综合分析相关风险后决策。

（6）全生命周期管理：数据收集、存储、使用、加工、传输、提供、公开等流程应得到记录和梳理，有条件的企业可建立数据流转地图，记录 X 等级的某数据（如交易数据、密码等）从 A 系统前端采集后传输到 B 系统，在 B 系统后台管理可脱敏展示，以及存储在 C 数据库中，N 年后从热备转为冷备，使用 D 算法进行加密，仅限 E 岗位员工读取，以及拟留存 F 年（如 10、15、20 年）或接到用户需求即销毁等的一系列情况。

5 其他技术合规关注点

5.1 技术类合同中的注意事项

（1）关注源代码所有权、保密以及是否需要额外的安全要求、数据保护要求等，可通过建设各业务场景模板的方式解决；

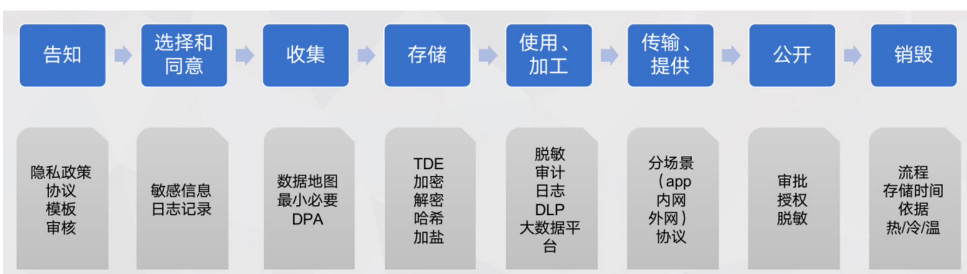


图 2：数据全生命周期及关注点

（2）关注技术需求的细致程度：技术合同中建议明确和细化技术需求的内容，减少对厂商自觉性的依赖，明确交付材料、验收标准、项目计划，并补充服务满意度评价及可导致后果的部分内容，如满意度不合格可能导致合同款项的部分扣除等。

（3）资质校验：若企业需要获得某项证书，第三方是否有资格做该项工作或有资格发放该项证书，需要明确在合同中，避免影响项目进度，产生纠纷。

（4）加强沟通：需求人员应加强与法务人员、外部专业律师团队的沟通，多听取专业人士对于技术合同、协议中技术需求的提示和建议。

5.2 企业内部技术合规审计注意事项

（1）技术类制度及操作规程的制定以及符合程度：比如数据安全合规中，数据安全法要求建立健全全流程数据安全管理制度，个人信息保护法要求制定内部管理制度和操作规程，所以企业一般都存在相关制度，制度可能存在如“每年做一次数据安全评估或数据应分级分类”，“某类数据需要存储加密、传输加密“，”某类数据不可在互联网传输或需加密传输“等等要求。技术部门和人员在管理数据的时候是否满足制度及操作规程的要求，是内部技术合规的重点，同时也是行业内常见的不足，原因是很多技术人员缺乏对制度内容的理解甚至不知晓企业存在相关制度，需要增加技术类制度的培训和宣贯。

（2）技术合规审计人员要求：根据《关键信息基础设施安全保护条例》要求，运营者应当设置专门安全管理机构，并对专门安全管理机构负责人和关键岗位人员进行安全背景审查，具体的审查点建议包括工作经历及能力、相关管理和技术能力、违法犯罪情况、安全从业过程中需要说明的事件事项等。

5.3 合规建设者的服务意识

技术合规建设者比一线的技术人员要更了解合规要求的具体审核点，需要帮助一线技术人员思考并找到合适的满足要求的技术方案，所以技术合规建设者应较多的了解各项技术，但不要求特别精通，建议具备一定的理工科背景或在技术合规领域工作经验较为丰富的人员承担合规建设工作，可避免合规要求和具体落实出现偏差，以导致外部审计时出现不合规问题，给企业带来合规风险。

5.4 合规的技术界限

不是所有的合规要求都需要做到极致。类似不同业务程度的信息系统有对应的等保要求，为了满足合规，有一些做到基线就可以，有一些则需加强防护，企业需根据实际情况进行合规风险评估。

比如数据安全评估中，如出现 APP 收集用户个人信息却没有隐私政策，则为高危，需要立刻整改，若采用加密算法时，未采用国密算法，使用的国际算法，考虑技术上更新算法的可行性等判断风险是否需要接受或列制整改计划进行跟进，以将有限的资源用在需要的业务中。

5.5 技术培训和人员风险

（1）人是最危险的：风险评估中不会缺少对人员管理的评估，人为因素因不可控给业务、技术带来了较大的风险，人员管理在合规建设中较为重要，所以企业可能会采用背景审查、双人双锁、最小化权限、二次审批、出入口监控等方式保障企业和数据安全。技术合规也涉及技术人员的合规，所以对于人员管理也需考虑在内。

（2）培训和考核：随着数字化转型的进行，技术架构不断迭代，企业不可能随时招聘到最精尖、最合适的人才保障技术开发和运维，而且很多技术并不像科学研究需要长时间的积累和培养，短期或阶段性得培训和考核也可以帮助企业获得需要的技术人才。

（3）对岗位和外包的依赖性：企业可通过建立操作手册、外包培训手册、专家知识库等方式降低对人员的依赖性。

5.6 权威机构的作用

（1）获得解决方案：随着等保的普及，企业可通过测评访谈、现场查看以及最终出具的测评报告中了解到很多合规要求及企业现状与要求的差异，并获得很多必要的解决方案和建议，通过采纳建议部署相关安全防护措施后进一步满足要求，企业技术获得了提升，合规也得到了保障。

（2）个人信息保护和数据安全等新兴合规领域已出现权威的评估报告或认证，通过评估和认证以帮助企业识别自身的技术成熟度和后续的发展方向，并对企业的符合情况进行认证，输出的证书和报告在外部检查或者审计中是较为有效的证明材料。此外，一些国际证书 ISO27001（2022）、ISO27701（2019）、ISO37301（2021）的权威性较高，可以有效完善企业的合规建设。司法案件需要专业律师和公司法务团队处理，专业的人做专业的事情，避免企业走弯路。

5.7 保障企业自身的权利

合规不只是对企业自身的高标准高要求，也是企业自我权利的体现。企业可通过免责条款、数据保护协议、各项技术认证等方式保护自身权益，也可利用合规要求约束相关方，达到通过监管、合规的力量维护自身权利的目的。

5.8 加强与业界、监管的交流

在合规条款出来后，业界专家一般会启动条款解读等工作和服务，监管后续也可能会予以回复，这部分内容较为重要。和行业的交流可以帮助企业了解行业协会、论坛等的信息内容，以及获取到业界比较成熟的、先进的思想和实践经验，通过交流沟通甚至合作，便于摸索出对企业发展和实践都更合适的技术合规落地方案，对合规发展有很好的促进作用。

企业应保持与相关法律、监管等职能机构的信息沟通，以及知晓什么情况下需要报告何种类型的信息，以便于借助监管力量限制不合理的相关方行为，且有助于了解监管的趋势，以及法律法规、监管要求有可能发生的变化等。

6 总结

很多技术专家对合规要求存在偏见，觉得增加了技术负担，但其实合规做得好对企业来说则推动了技术完备性，降低了事件发生概率，提高了事件追溯效率，增加了各部门的有效沟通且统一了流程和模板，将会少走较多弯路，从长远来看则是减轻了负担。

随着监管环境愈发严峻，合规风险愈发引起业界的重视，本文从合规建设的角度提出了一些解决方案，希望对业界起到抛砖引玉的作用，合规建设保障了企业安全、行业安全，是行之有效的安全保障，且是能引起从监管至企业再至用户集中关注的一项工作，影响面可想而知。

参考文献：

[1] ISO/IEC JTC 1/SC 27 Information security, cybersecurity and privacy protection. ISO/IEC 27001:2022Information security, cybersecurity and privacy protection — Information security management systems — Requirements[R]. 2022.

[2] ISO/IEC JTC 1/SC 27 Information security, cybersecurity and privacy protection. ISO/IEC 27002:2022 Information security, cybersecurity and privacy protection — Information security controls [R]. 2022.

[3] 孟洁，薛颖，朱玲凤 . 数据合规 – 入门、实战与进阶 [M]. 北京：机械工业出版社，2022.

证券公司数字化转型实践与研究

刘汉西、吴士荣、丁一、陈镇光 / 国信证券股份有限公司 广东 深圳 518001
E-mail：chenzguang@guosen.com.cn

随着金融科技的发展，数字化转型已是证券行业变革的大势所趋；实现动力变革、效率变革、质量变革，成为我国证券行业高质量发展的必由之路。面对数字技术带来新的机遇和挑战，如何解决证券行业数字化转型痛点，加快公司数字化转型进程，成为摆在证券公司面前的重大课题。但是，证券业如何有效实现数字化转型，目前暂未有统一的共识方案。因此，本文对国信证券的数字化转型实践进行阶段总结，同时深入研究分析证券行业数字化特点与现状，结合未来发展趋势，设计了证券公司数字化转型能力体系，并提出实施建议。

1 方兴未艾，证券公司加快数字化转型步伐

“数字化转型”是指以数据为核心，充分运用新一代的数字化技术，对管理模式、业务模式、商业模式进行创新和重塑，进而实现企业成功转型。

作为证券市场的重要参与主体，证券公司数字化转型的程度、效果，在很大程度上反映了整个行业的数字化转型程度。自 2020 年以来，包括国信证券在内，大部分证券公司都在半年报和年报中明确提到“数字化”。当前行业证券公司数字化转型整体呈现如下特点：

一方面绝大多数证券公司已开启数字化转型进程并已取得阶段性的成果。证券公司普遍意识到全面数字化转型是推动公司高质量发展的必要

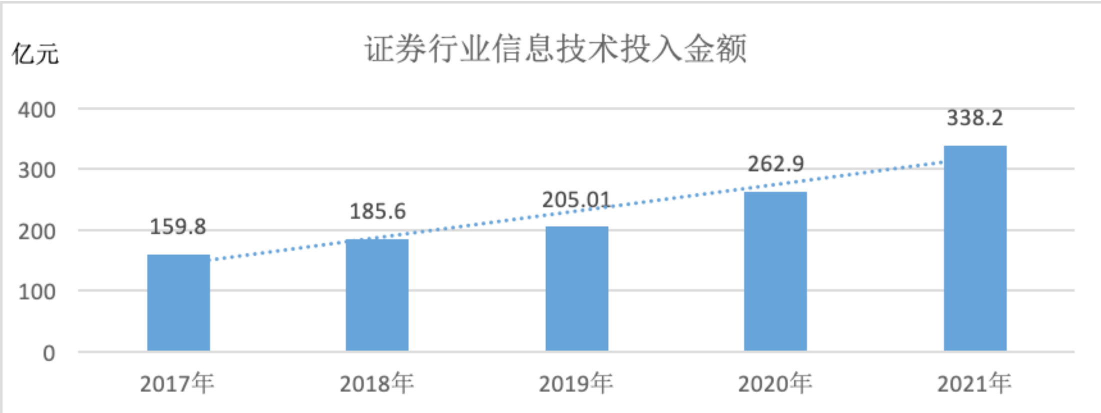


图 1：证券行业信息技术投入金额

途径，因此先后都开启了数字化转型的进程；并且每年加大 IT 投入，确保数字化转型必要的资源保障（见图 1）。在某些关键领域取得了显著成效，催生了智能投顾、智能投研、智能客服等新型服务或产品。

另一方面不同类型的证券公司在推进数字化转型过程中表现出不同的进度与特征。大型券商更多的是采取多点投入全面推进的方式，引领行业数字化转型发展方向，注重科技创新驱动业务发展，挖掘新的需求增长点。中小券商在数字化建设上则更加聚焦，围绕特定领域的业务定位，量身定制有助于解决发展的关键痛点，助推公司深耕核心业务的差异化发展。

2 知易行难，迎接转型实践中的困难与挑战

由于历史原因，证券公司在数字化转型实践中不可避免普遍会面临各种类型的问题和挑战，集中体现在下面几个方面：

挑战一：如何达成全员共识？数字化不仅仅是一场技术变革，更是一场思维革命。数字化转型不仅要技术先行，还需上升至公司战略层面，达成高度共识，充分调动和促进公司各部门的主动性、积极性和创造性，才能加速实施的进程。

挑战二：数字化人才缺乏。当前行业数字化人才队伍距离全面数字化转型仍有差距，包括信息技术条线的技术专业人才，也包括各业务部门和职能部门的业务产品经理、数据分析师等复合型数字化人才。大部分公司都存在人才引进困难、人才激励不到位或人才队伍结构不合理等问题。

挑战三：跨条线资源整合不够。由于跨条线的资源整合需要公司层面领导者和相应配套机制的统一统筹与协调，对当前大部分证券公司来说仍是一个难题，部门之间仍存在本位主义，没有动力推进共享协同。

挑战四：在安全和风控上，一方面数字化转型将带来更多的数据应用和数据开放，对数据安全和保护、网络安全防护上提出了更高的要求；另一方面，数字化转型必将带来业务新模式和新变革，需要在业务合规、业务质控、业务风控上同步进行转型。

3 高屋建瓴，设计券商数字化转型能力体系

数字化转型是一项系统工程，需要加强顶层的规划和设计，自上而下的推进落实。从实践看，目前行业大部分证券公司开展数字化转型都已取得阶段性成果。数字化成果来源于多维度能力体系的综合运用。国信证券根据自身的实践经验，并结合对行业的深入分析，总结出构建数字化能

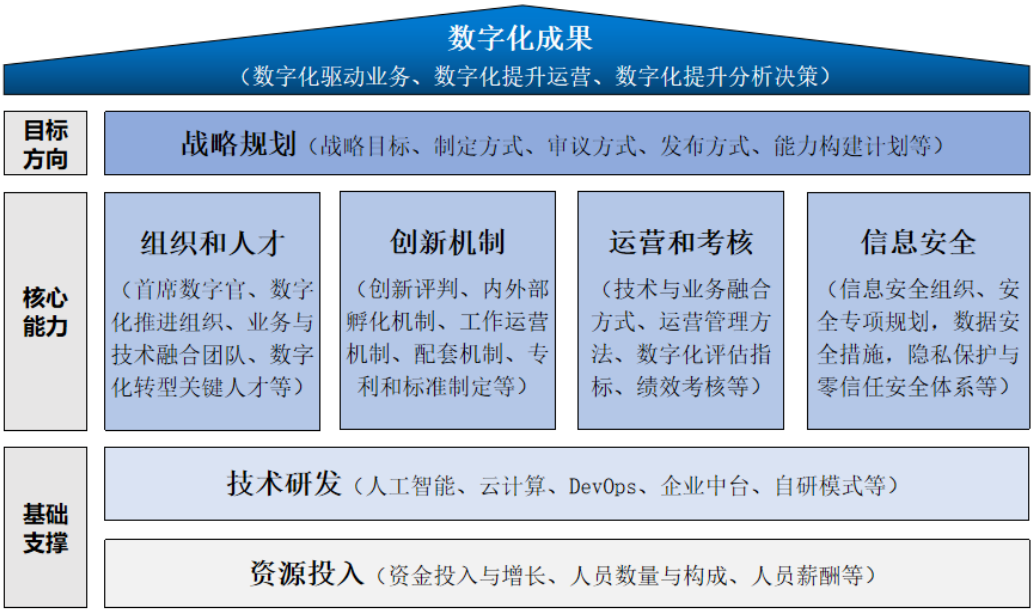


图 2：数字化转型能力框架

力体系的六大核心主线，包括战略规划、组织和人才、创新机制、运营和考核、信息安全、技术研发，具体如图 2 所示。

根据数字化转型能力框架，证券公司可以围绕以下六大主线推进转型，构建核心数字化能力，打造面向未来的长期竞争优势。

（1）战略规划的统筹能力。战略规划指明了企业数字化转型的目标和方向，可以使企业的各级人员都能够知晓企业的共同目标及努力方向，进而增强企业的凝聚力和向心力，有效指引数字化转型目标得到实现。

（2）组织和人才的协调能力。组织与人才相辅相成，组织架构决定人才分工与协作，人才素质影响组织效率和发展。因此企业有必要根据数字化转型要求、结合现状进行必要的组织架构变革。

（3）创新能力。创新是企业发展的动力，是数字化时代尤其成为企业的重要竞争力。创新机制通过激发企业员工的积极性和创造性，促进企业资源的合理配置利用，从而推动企业数字化转型不断取得新的成果。

（4）运营和考核的管理能力。数字化运营目的在于解决企业数字化转型过程中“责权利”的有效分配，优化企业资源配置效率；而数字化指标评估和考核有利于调动各方面推进数字化转型的积极性。

（5）安全保障能力。数字化转型推动企业信息和数据越来越多的实现线上化，并将更多地与外部机构进行数据互联，如何保障信息安全越来越成为企业数字化转型过程中的一项不可或缺的重要能力。

（6）技术研发能力。技术是数字化转型的基础能力，平台是对技术能力和业务能力的抽象与沉淀。企业如何通过数字化技术开展数字化转型，为管理变革和业务突破提供基础支撑，是衡量数字化转型能力的重要因素。

4 全面部署，推进数字能力落地与价值释放

根据数字化转型能力框架中的六大主线，并结合行业目前普遍面临和亟需解决的问题，建议

证券公司在未来数字化转型进程中采取如下重点举措。

4.1 明确数字化转型战略目标，制定可实施计划

证券公司要在数字化转型过程中取得好的成效，首先必须在公司层面制定明确的数字化战略规划。

一方面，在战略规划中明确数字化转型的目标。一是战略规划的目标应当是具备一定前瞻性，代表了公司未来数字化发展方向。二是战略规划的目标是在公司基于诸如 SWOT（内部优势、劣势和外部机会、威胁）等分析后得出，符合公司自身特点与定位。三是战略规划的目标应当是可执行、可衡量的，经过公司全体人员的努力可以达到，具有落地可行性。

另一方面，制定实现目标的落实计划。数字化战略不仅要有方向和目标，还要有配套的落实计划，也就是通过使用什么手段、什么措施、什么方法，以及在什么时间段来达到制定的目标。一个有效的落实计划应当是跟公司的资源、环境、管理等相匹配的，是多个维度综合考量下的结果。

2022 年 7 月，国信证券正式发布“数字化转型工作方案”。该方案向上承接公司“十四五”发展战略规划和“各条线十四五子规划”，向下对接金融科技“十四五”规划，是疏通金融科技赋能业务、精准匹配业务需求的有效机制。

“方案”中以建设“数字国信”为总目标，以深入融合业务与技术并加快金融科技与数字化建设为总抓手，切实发挥“创新驱动、科技引领”作用，加速实现公司“十四五”战略目标”的指导思想，并制定了：加强数据信息底座构建、深挖数字化业务与管理场景并建设、不断巩固网络安全、夯实数字化转型基础等 4 大领域 53 项重点工作。按“方案”设定的目标与计划，各项工作正在紧锣密鼓地推进落实中。

4.2 针对数字化进行组织调整，培养关键性人才

在数字化转型过程中，证券公司必须开展与

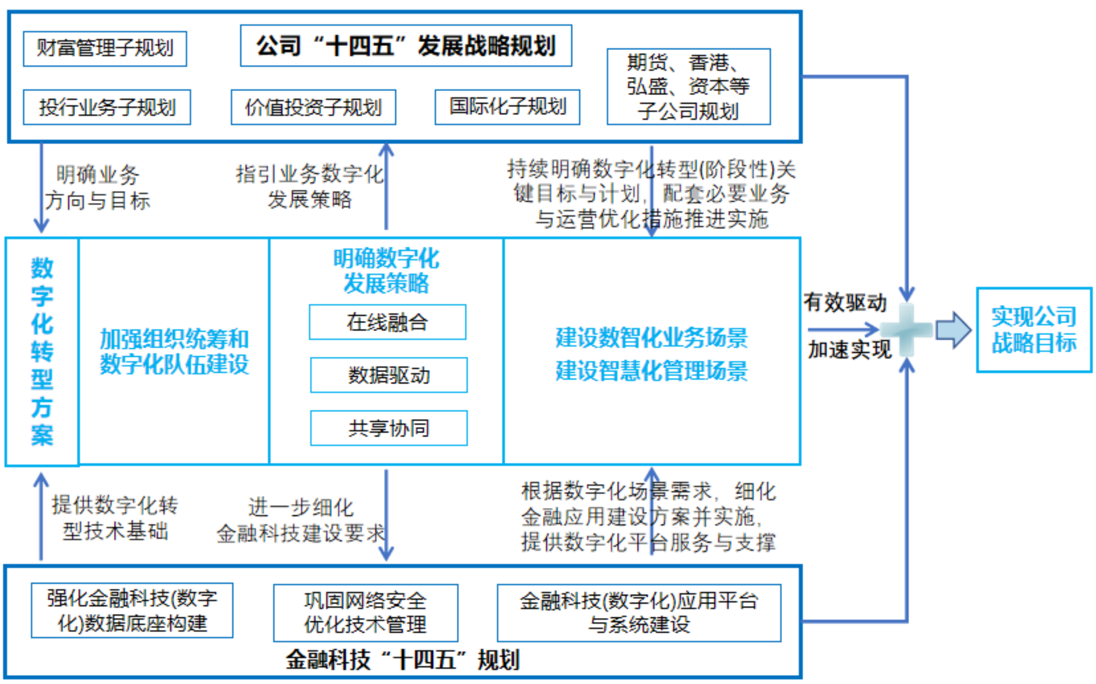


图 3：国信证券数字化转型工作方案

数字化转型要求相配套的必要组织变革、并大力推进数字化关键人才培养，才能有效推动转型目标的达成。

第一，要明确数字化领导者。数字化转型的决策领导必须是公司的“一把手”，通常是 CEO(或董事长、总经理)。只有站在全局视野的“一把手”才能从公司整体的数字化转型目标出发、突破部门圈子，带领公司真正从用户的角度出发，推进和实现未来业务演进的蓝图。

第二，要成立数字化推进组织。数字化转型需要实现公司跨条线的资源整合，因此需要构建数字化推进组织以推动现有组织在业务、系统等多个领域创立有效连接与协同的接口，促进不同业务领域之间的信息和资源共享、业务协同，进而促进数字化业务与管理在跨部门间的有效运作。

第三，要培养具有数字化思维的人才。一是要建立符合公司自身的数字化人才标准，为数字化人才的选用、培养、管理奠定坚实基础。二是要落实数字化人才培养，根据数字化人才的标准培养符合数字化转型需求的核心人才，如技术业务复合人才、平民化数据分析师、数字化产品经理等。

国信证券在现有组织和人才基础上，明确目标，继续加大数字化转型组织架构与人才队伍建设，已经开展的工作包括：

（1）在组织机制上，在公司党委会、董事会、经营管理层的领导下，明确“一把手”负责制，由公司金融科技委员会负责确定数字化转型工作目标、工作方案等重大事项，组织领导和督促指导公司各条线的数字化转型落实工作。

（2）在组织架构上，公司战略发展总部和金融科技总部作为公司数字化转型统筹部门，协同公司各部门成立数字化转型推进工作组，统筹推进数字化转型重点工作。

（3）在数字化人才培养上，一方面技术部门持续加大引入和培养技术产品经理、架构师、数据（算法）专家等人员，不断提升数字化建设与金融科技能力；另一方面公司各业务部门与职能部门，设置数字化专员、业务产品经理、数据分析师等岗位或角色，协同主动挖掘数字化业务场景、开展数据分析并与业务过程紧密结合，促进本部门数字化转型工作走向主动性。

4.3 积极开展创新实践活动，建设形成长效机制

数字化时代公司的发展壮大更加离不开持续不断的创新发展，而只有配套行之有效的创新长效机制，才能保证公司创新活力的长盛不衰。

首先，建立数字化创新孵化机制。公司可设立科技创新基金，用于数字化创新类项目的建设；同时，制定配套的创新项目的筛选标准、验收标准，并为创新项目提供专家指导及资源支持。另外也可与头部科技公司或者高校合作，共建创新

图 4：数字化创新的关键要素

实验室，根据公司数字化创新需求共研与孵化创新产品。

其次，形成数字化创新工作持续运营机制。创新运营机制需要常态化，可通过设立创新实验室以及定期举办创新活动等多种方式，持续推动创新工作的落地实施。公司级的创新活动，有利于营造全公司创新氛围，形成全员创新的创新文化，并通过组织培训学习、成果分享、对外交流等活动，扩大创新成果的示范效应。

国信证券持续打造公司数字化主动创新文化，开展创新大赛、成果推广、树标立范等多种形式的活动，激发全员活力，营造勇于、乐于、善于数字化转型的数字化创新文化氛围。包括：举办“金融科技（数字化）文化节”和开展“数字化产品创意大赛”创新活动、设置“数字化发展奖”等。2022 年国信证券举办首届金融科技文化节，通过教练辅导 + 培训 + 沙盘演练 + 现场编程 + 业务技术混合组队 PK 的方式，结合设计思维、精益创业、产品功能规划、敏捷项目管理等创新思维方法，突破当前组织条框、参与者与志同道合的团队伙伴在 36 小时内将创意从 0 到 1 打造一个产品。活动极大激发了员工的创新思维和创作热情。

此外，在 2022 年各类金融科技评选中，国信证券已获得包括“深圳市金融创新奖”、“第八届证券期货科学技术奖”等在内的多个奖项。

4.4 推进数字化运营管理，配套数字化考核指标

数字化运营与考核是确保公司数字化转型朝着正确目标持续前进的必要保障，各公司需要根据数字化转型目标，设定符合自身发展要求和管理特色的数字化运营与考核机制。

一方面，要改善协作，推动业务与技术的充分融合。业务与技术一体化运作有三种方式，企业根据自身情况选择。一是采用业务合作伙伴模式，将科技部门进行前置，部分 IT 人员与业务部门合署办公，初步形成融合机制；二是推行产品管理制度，在 IT 部门或业务部门设置数字化产品经理，加强双方的进一步融合与协同；三是对于创新性业务或敏捷需求类业务，业务团队和技术团队可以组成新的团队或部门，进行快速开发和尝试，实现深度融合。

另一方面，要建立各条线的数字化绩效（考核）指标。设置绩效指标可进一步促进数字化转型工作在各组织内得到有效贯彻。首先，设置业务部门的数字化绩效指标，如线上客户开发指标、线上运营客群覆盖指标等，评估数字化目标推进情况和满意度评价情况。其次，设置 IT 部门的数字化绩效指标，如 IT 规划任务年度推进情况、业务系统建设情况等，提升 IT 部门数字化转型的驱动作用。

在融合工作模式上，国信证券以产品制的方

图 5：业务与技术融合的三种模式

式管理关键系统，并通过敏捷研发模式实现需求的快速迭代。根据各条线金融科技发展水平、产品线复杂度，通过成立专班、技术向业务派驻ITBP、技术培训产品经理等方式，全力打造融合工作模式。同时，国信证券周期性开展“首席信息官带队”的业务技术交流，及时把脉信息系统建设方向，实现从管理层至普通员工的立体式IT需求沟通。结合产品管理体系对产品经理职责的落实，以及数字化专员对典型场景的挖掘，形成需求主动挖掘、双向主动沟通的长效融合机制。

在绩效指标上，国信证券针对重大项目、系统安全运行及信息安全情况等设置了KPI考核指标，同时在探索目标与关键成果法（OKR）的落实方式。

4.5 加强信息安全管理，为实现转型目标保驾护航

在“万物皆线上”的数字化时代，信息安全的重要性更加不容忽视，尤其是数据安全的挑战更为突出。因此，保障信息安全将是公司可持续发展、提高整体竞争力的重要手段。

首先，要加强数据的全生命周期安全保护和管理。数据安全是数字化时代信息安全的核心内容之一，建设内外部数据安全及投资者信息保护的监测体系，对数据全生命周期的安全流动风险进行识别、评估和预警，防范数据泄露、数据篡改、非法访问等安全风险事件发生。

其次，积极引入信息安全新技术以应对新的不确定性。随着人工智能、区块链、云计算等新技术的快速发展，在推动公司数字化进程不断递进的同时，也催生了新的安全需求，证券公司应该建立更加主动的防御措施。例如，建立零信任防护体系，对每个交互过程进行动态权限控制和细粒度访问控制，以及全方位的安全审计和持续的风险与信任评估，达到对身份、设备、应用和数据的共同保护。

国信证券在实践中从应用系统的设计、开发、测试、部署、运维、技术服务等各个环节加强安全管控，做好威胁情报监控和漏洞闭环管理，力争从源头解决网络空间的安全隐患，有效防范和化解供应链带来的网络安全风险，补齐网络空间安全保障体系的防护短板。

同时，国信内部贯彻落实《数据安全法》，建立公司敏感数据的技术管控体系，加强数据安全保护关键技术的研发工作，规范数据处理活动，促进数据资源合规开发利用，保护客户隐私数据安全。建立有效识别敏感数据的机制，通过高强度算法与轻量级密码应用相结合的方式实现敏感数据的保护，以最小化授权的模式管控数据的使用范围，确保数据操作留痕可审计；基于漏洞挖掘和渗透测试技术方法，研究通过协议漏洞、系统入侵、侧信道攻击等方式对数据安全隐患进行探查，主动发现数据采集、汇聚、存储、处理、分析、共享、使用等各个环节存在的安全隐患。

4.6 推广金融科技应用，夯实数字化转型技术底座

证券公司需要从金融技术的普及性、可控性、高效性、可复用性等方面去推动数字化转型的有效实施。

一方面，加速推广金融科技应用，并提升自主可控性。首先，积极推广行业应用已经相对成熟的金融科技技术，同时要适时试验和引进新技术，包括人工智能、DevOps、区块链、物联网、5G等，提高公司的差异化竞争力。其次，对具备业务能力重要性、需求快速迭代等关键应用系统实施自主研发转型，实现“投入”、“产出”和“价值”之间的有效平衡。

另一方面，抽象并沉淀企业能力，逐步建设企业级中台。证券公司内部普遍存在“烟囱式”架构，系统之间无法集成或者交互协作成本高昂，并且各独立封闭系统之间出现功能交叉导致重复建设，因此证券公司应该推进中台建设，将公共能力进行抽象和沉淀，实现跨业务领域、跨用户、

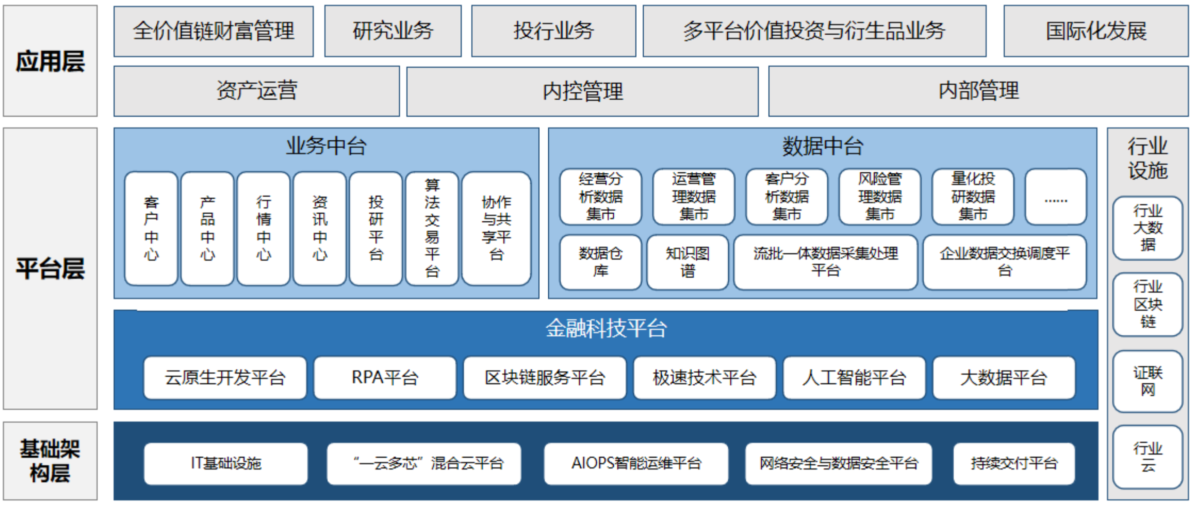


图 6：国信证券基础平台体系

跨场景、跨地域的协同与复用，支持业务的快速拓展创新，减少重复建设，同时为生态建设奠定基础。

经过多年建设与积累，国信证券已建立起较为完善的中后台基础平台体系，如下图所示。

其中包含了一站式数据中台，主要有经纪、风控、托管、资管等主题数据集市，及大数据挖掘建模工具等；并建立了可支持各业务条线应用快速研发的“企业级”业务中台，包括客户中心、产品中心、资讯中心和行情中心等。自研新一代技术平台架构，形成了应用架构域、共享能力域、前端应用域为主体，研发支撑域和服务治理域为两翼的“擎天”证券数字化技术平台。在云计算方面，已明确从云计算到云原生的演进路径，基于软件定义横向建设企业级云平台和云原生 PaaS 平台，打造全面支撑应用软件向云原生演进的云计算大后台。

5 卓有成效，国信证券数字化转型渐入佳境

数字化转型是一项长期工程，在每个阶段中，只有定期回顾与总结，才能更好的推进下一阶段的工作。回顾近几年国信证券在数字化转型道路上取得的阶段性成果，主要体现在以下几个方面。

在客户财富管理领域，国信证券成功打造了以金太阳 APP 为核心的多渠道一体化互联网金融平台，以及“鑫投顾”证券投顾服务、“国信优选”基金代销服务、“国信智投”公募基金投顾业务等，实现客户与产品的精准匹配，为客户提供个性化的产品和服务建议。

在大投行业务领域，国信证券从 2020 年起，陆续实现了发行人风险监控、银行流水分析、底稿智能识别、文档智能撰写与审核，并将数字员工在投行领域广泛应用，数字化、智能化转型有效助力了大投行业务驶入快车道。

在机构综合交易服务方面，国信证券目前已打造全链路低延迟交易解决方案，向顶尖高频私募、成熟私募、策略爱好者等提供各种类型的定制策略交易服务，满足不同投资者的差异化、专业化需求。

在投资交易支持体系方面，2020 年以来，国信证券持续打造“可转债、权益、固收、量化、衍生品”五轮驱动的大类资产体系。充分利用科技手段打造投研一体化服务、创新型衍生品一体

化服务、做市业务一体化平台等，持续完善投资交易支撑技术体系，实现投资智慧化。

在运营管理方面，2020 年国信证券在行业内首批实现“7 × 24 小时无间断业务审核”模式，同年还率先实现机构及产品一站式智能见证开户，全流程开户时长缩短 75%。经过深入推进，公司已建成网点、业务全覆盖的 VTM 智能终端，向全面建设智慧网点迈出了重要一步。

未来，国信证券将继续遵循回归金融服务实体经济的本源逻辑、坚持以客户为中心的服务逻辑、构建以创新为宗旨的战略逻辑、打造以数据为核心的风险管理逻辑，在建设“智慧国信”的同时，切实履行全国性大型综合类券商的责任，为资本市场高质量发展贡献国信力量。

参考文献：

[1] 中国证券业协会 . 证券公司 2021 年经营业绩排名情况 [ER/OL]. https://www.sac.net.cn/hysj/zqgsyjpm/202206/t20220617_168421.html. 2022,06.

[2] BCG 波士顿咨询 . 券商数字化转型破局之道 [ER/OL]. https://mp.weixin.qq.com/s/sKbnqVSOq5Z−lNVGM2aCJg. 2022,02.

[3] 麦肯锡 . 知易行难：探索券商数字化转型成功之路 [ER/OL]. https://www.mckinsey.com.cn/ 知易行难：探索券商数字化转型成功之路 /. 2021,12.

数据中台数据标准化设计与实践
——上交所数据主题域概念模型

贾晓雪、秦胜楠、洪绍歆、张恒、陈闯霞、王智宏 / 上交所技术有限责任公司 技术开发总部 上海 200120 E-mail：xxjia@sse.com.cn

在《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》一文中，介绍了上交所数据中台的模型体系，阐述了数据标准化设计思路，文章主要介绍了数据中台主题数据层的建模过程，首先是对本所全部业务数据进行了抽象与归纳，形成了业务数据的八大主题域，然后对各个主题进行了逐层分解，列出了每个主题包括的主要业务实体。本文将在前文基础上继续深入分析每个主题中的业务实体概念，并描述实体与实体间的主要业务关系。

1 背景介绍

随着时间推移和各类创新业务的开展，上交所沉淀了大量的业务数据，形成了一定规模，自上而下越发重视数据价值，希望让数据更好的赋能业务、赋能行业，数据中台应运而生。

数据被复用的能力是数据中台最核心的价值，这就需要做好数据架构体系：主题数据体系、指标体系、标签体系的建设，本文在《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》的基础上，详细剖析主题数据体系中各主题的概念和实体。

2 参与人主题

参与人是指金融市场活动中的参与主体，根

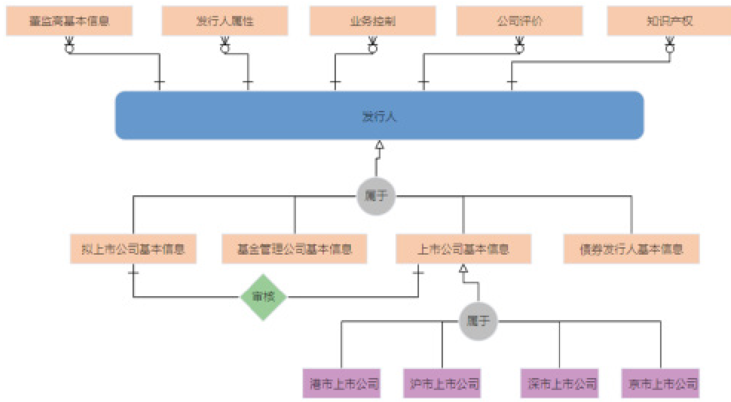


图 1：参与人主题 – 发行人层面概念模型

据在市场活动中承担的角色，可以将参与主体拆解为发行人、机构、参与人关系和业务控制四个子主题。

2.1 发行人层面

发行人指为筹措资金而发行债券、股票等证券的政府及其机构、金融机构、公司和企业，根据发行证券的品种不同，发行人分为上市公司、基金管理公司、债券发行人和拟上市公司，拟上市公司经报批审核通过后可转为上市公司，存在一对一的审核关系。根据发行人选择的上市地点不同，上市公司下分为沪市上市公司、深市上市公司、京市上市公司和港市上市公司。

董监高信息、公司评价、知识产权和发行人其他附加属性都是围绕发行人主体展开的，与发行人存在归属的关系。

2.2 机构层面

除发行人外的金融市场其他参与人主体均归入到机构子主题下，包括外接机构、一般企业信息、监管执行机构和金融中介机构。金融中介机构可以分为储蓄机构、投资机构、保险机构和信息资讯机构；外接机构是由会员上报的一些外接机构的信息；监管执行机构是依法制定有关证券市场监督管理的规章、规则，并依法对证券市场的参与者进行监督管理的部门，包括中国证券监督管理委员会和地方证券监管部门、法院、自律性组织等；一般企业就是普通的工商注册公司，

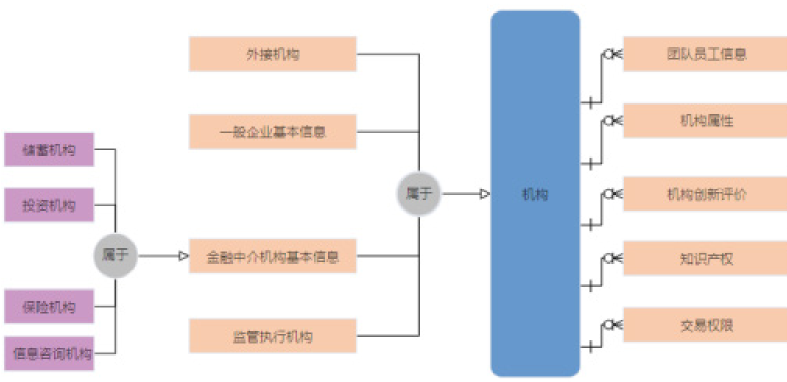


图 2：参与人主题 – 机构层面概念模型

此处重点关注与上市公司有关联的企业，如供应商、客户、子公司等。

团队员工信息、交易权限、知识产权、机构创新性评价和机构其他附加属性都是围绕机构主体展开的，与机构存在归属的关系。

2.3 参与人关系层面

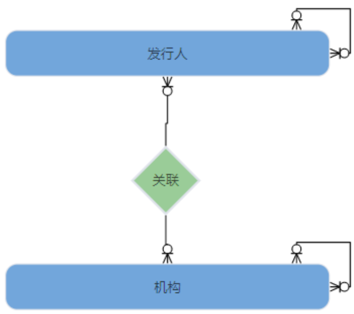


图 3：参与人主题－参与人关系层面概念模型

参与人关系子主题记录参与主体间的关联关系，主体间均存在多对多的关联关系，包括发行人和发行人、发行人和机构、机构和机构之间的关系，比如承销商与上市公司、债券发行人和上市公司、证券公司和外接机构之间均存在一定的关联关系。

2.4 业务控制层面

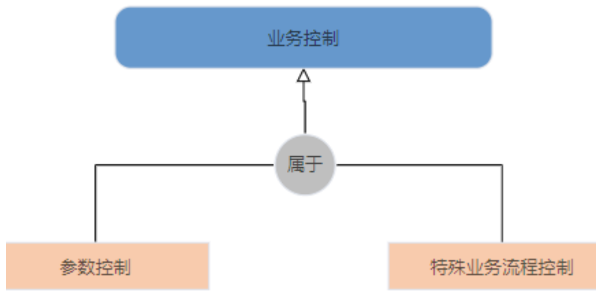


图 4：参与人主题－业务控制层面概念模型

业务控制子主题分为参数控制和特殊业务流程控制。

参数控制是对参数阈值的设置，如基金通买入额度、期权会员特殊限仓等。

特殊业务流程控制指为实现特定业务功能对流转流程进行控制，如计费所需的证券公司新设注销变更 pbu 的流程数据。

3 账户主题

账户是投资者持有资产及参与各类证券行为

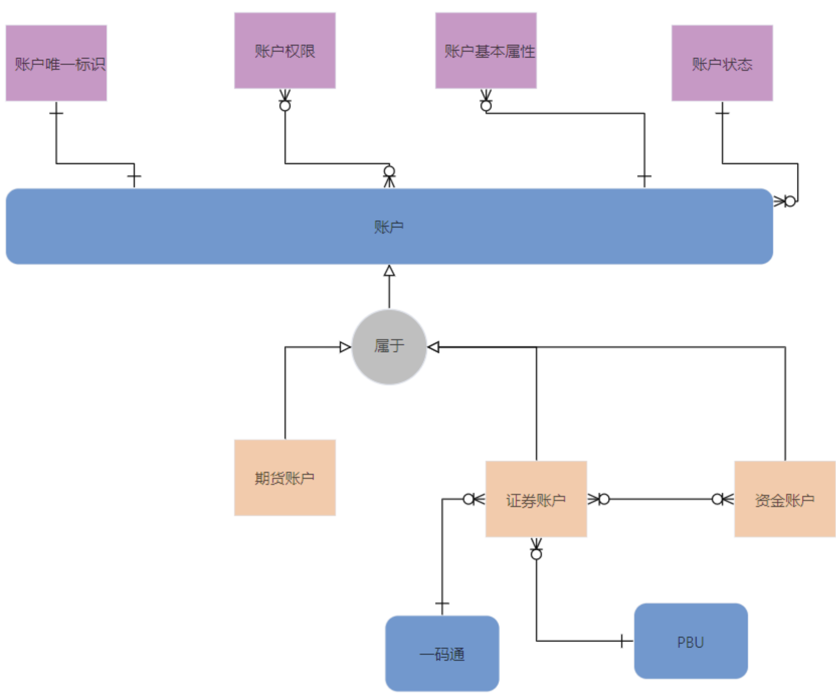


图 5：账户主题－账户信息层面概念模型

（如委托、交易、结算、过户、分红等）的载体。根据账户参与各类市场活动的不同限制，可分为账户信息、账户分类、业务控制三个子主题。

3.1 账户信息层面

账户信息是对账户相关附加属性的描述，账户基本信息包含账户唯一标识、账户权限、账户基本属性、账户状态等要素。账户唯一标识即账户ID，账户基本属性包含姓名、性别、年龄等信息，账户可以有多种状态，包括正常状态、挂失状态、冻结状态、休眠状态、注销状态等等。

账户包含期权账户、证券账户以及资金账户等。期权账户用于期权交易，开通期权账户首先需要开通证券账户；证券账户主要用于证券交易，证券的交易持仓也需要以账户为载体；资金账户是用于证券交易资金清算的专用账户，结算参与人在中登开通资金账户，用于结算参与人的资金结算，而个人投资者资金账户则于券商开通，用于券商内部的二级资金清算。

单个投资者可开通多个账户，为方便识别，引入一码通账号作为投资者的唯一标识，通过一码通与账户的对应关系可以了解投资者所有账户的交易情况以及持仓情况。

3.2 账户分类层面

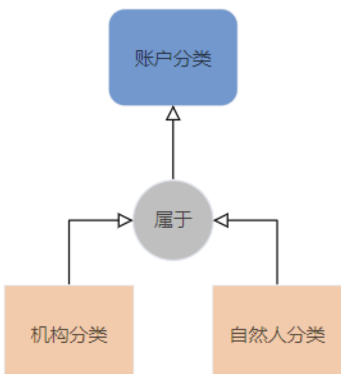


图 6：账户主题 – 账户分类层面概念模型

按照不同维度，账户可分为不同的类别。按市场划分，账户可以分为沪市账户、深市账户、港市账户等。按账户用途分可分为人民币普通股票帐户、人民币特种股票帐户、证券投资基金帐户以及其他帐户。按证监会分类标准，A 股账户可以分为自然人账户和机构账户，自然人账户根据持股市值可进一步划分为超大户、大户、中户、小户等，机构账户可细分为一般机构账户、特殊机构账户以及陆股通账户，一般机构包含国有股东、外资股东等，特殊机构含券商自营账户、券商资管账户、保险机构账户等，陆股通账户目前分为两类，沪股通账户和深股通账户。

账户分类的主要应用是为了方便统计各个类型的投资者，特别是大户和超大户的交易与持仓情况，方便重点监管。

3.3 业务控制层面

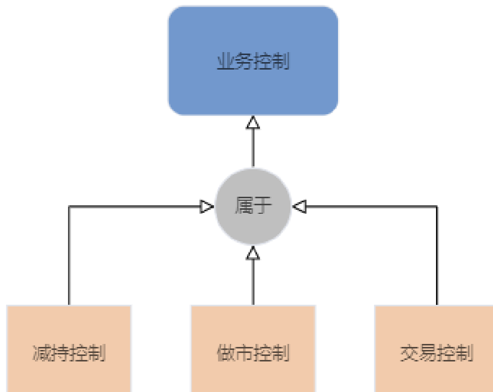


图 7：账户主题 – 业务控制层面概念模型

业务控制描述了对账户权限的控制信息，包括减持控制、做市控制以及交易控制。减持控制是对股东减持证券的控制，例如大股东在一定期限内只能减持限定比例的股份，而有一些账户，比如创投基金账户、沪股通账户均不在减持控制范围内；做市控制在做市商评级与收费中应用广泛，含做市豁免时点、时段信息等信息；交易控制含账户交易限制信息。

4 资产主题

资产指参与人进行证券产品投资所拥有的

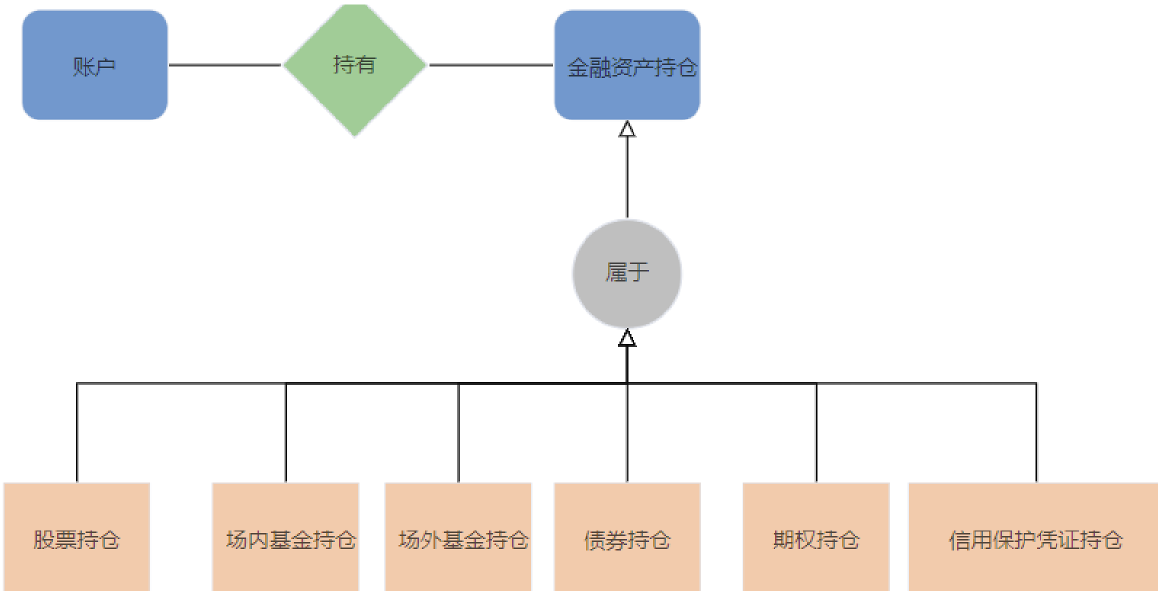


图 8：资产主题－金融产品持仓层面概念模型

资产信息。根据资产属性可划分为金融产品持仓、资金持仓、商品持仓、业务控制四个子主题。

4.1 金融产品持仓层面

金融产品持仓子主题描述了投资者对金融证券的持仓情况，包括股票持仓、场内基金持仓、场外基金持仓、债券持仓、期权持仓、信用保护凭证持仓等。

4.2 资金持仓层面

投资者除了可以持有证券，也可以持有资金，因此新增资金持仓子主题。账户与资金为持有关系，资金可包含货币类持仓以及其他非货币类持仓，例如票据。资金持仓包含资金的余额以及资金的流向两个要素，资金流向可分为资金流入与

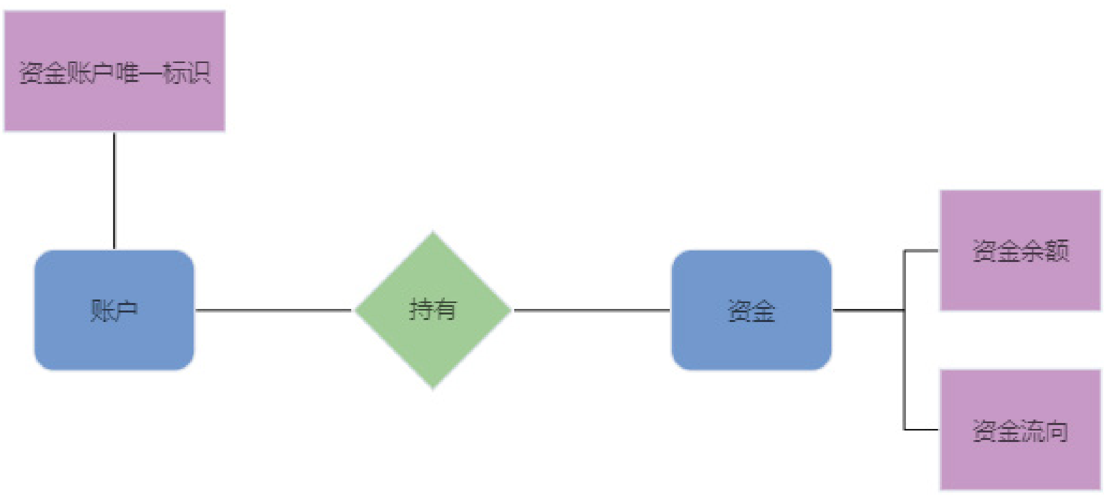


图 9：资产主题－资金持仓层面概念模型

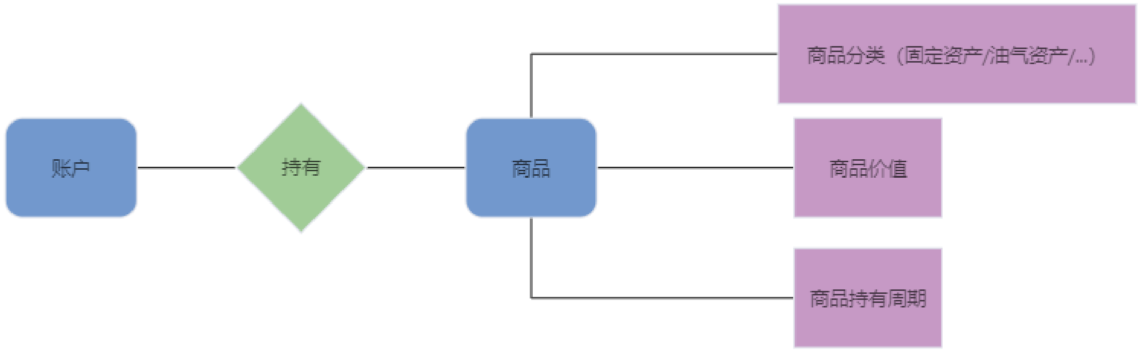


图 10：资产主题－商品持仓层面概念模型

资金流出。

4.3 商品持仓层面

除了资金和证券持仓，未来可能引入固定资产、油气资产这样的商品类持仓，因此设计了商品持仓子主题。商品持仓包含商品类别、商品价值以及商品的持有周期等信息。

4.4 业务控制层面

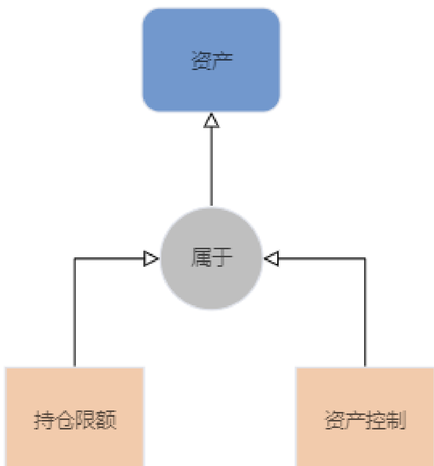


图 11：资产主题－业务控制层面概念模型

资产的业务控制描述了对持仓的限制，包括持仓限额以及资产冻结（控制），如固定收益平台账户证券冻结记录、账户每日 C 类大宗受让控制数据等。

5 产品主题

产品是金融市场活动中的参与客体，根据客体描述信息的不同，可以将产品主题拆解为产品信息、产品行情、业务控制三个子主题。

5.1 产品信息层面

产品信息描述了沪深京港市场的股票、债券、基金、衍生品、指数产品、特殊业务产品、以及其他市场产品的基本信息和相关信息。

产品信息、产品规模、产品行情、产品评级、产品分类是产品的核心概念。其中产品信息、产品规模、产品行情在特定时间点具有唯一性，产品评级和产品分类根据不同的评级机构和分类方式可以在特定时点存在多份信息。

产品发行上市阶段形成产品发行上市信息和发行中介机构信息，如债券发行阶段的会计事务所信息、律师事务所信息、承销商信息等。

不同产品拥有独特的业务属性，形成独有的基本信息，如股票分红送配、除权除息信息，债券应计利息信息、债券回售信息，基金分红信息，指数基本信息、指数成分股信息、指数权重信息等。

各产品包含了一系列的子产品，子产品之间、子产品与其他父产品彼此之间也存在复杂的关系，如：股票与债券中的可转债、可交换债存在对应关系，指数成分股包含股票或债券，ETF 基金可作为期权的基础合约标的，一只指数可以作为不同 ETF 基金的跟踪指数。

股债基基础产品经过一定的业务规则或标的测算可生成特定业务标的清单被用于特殊业务，如沪深港通产品、沪伦通产品、B 转 H 产品、债

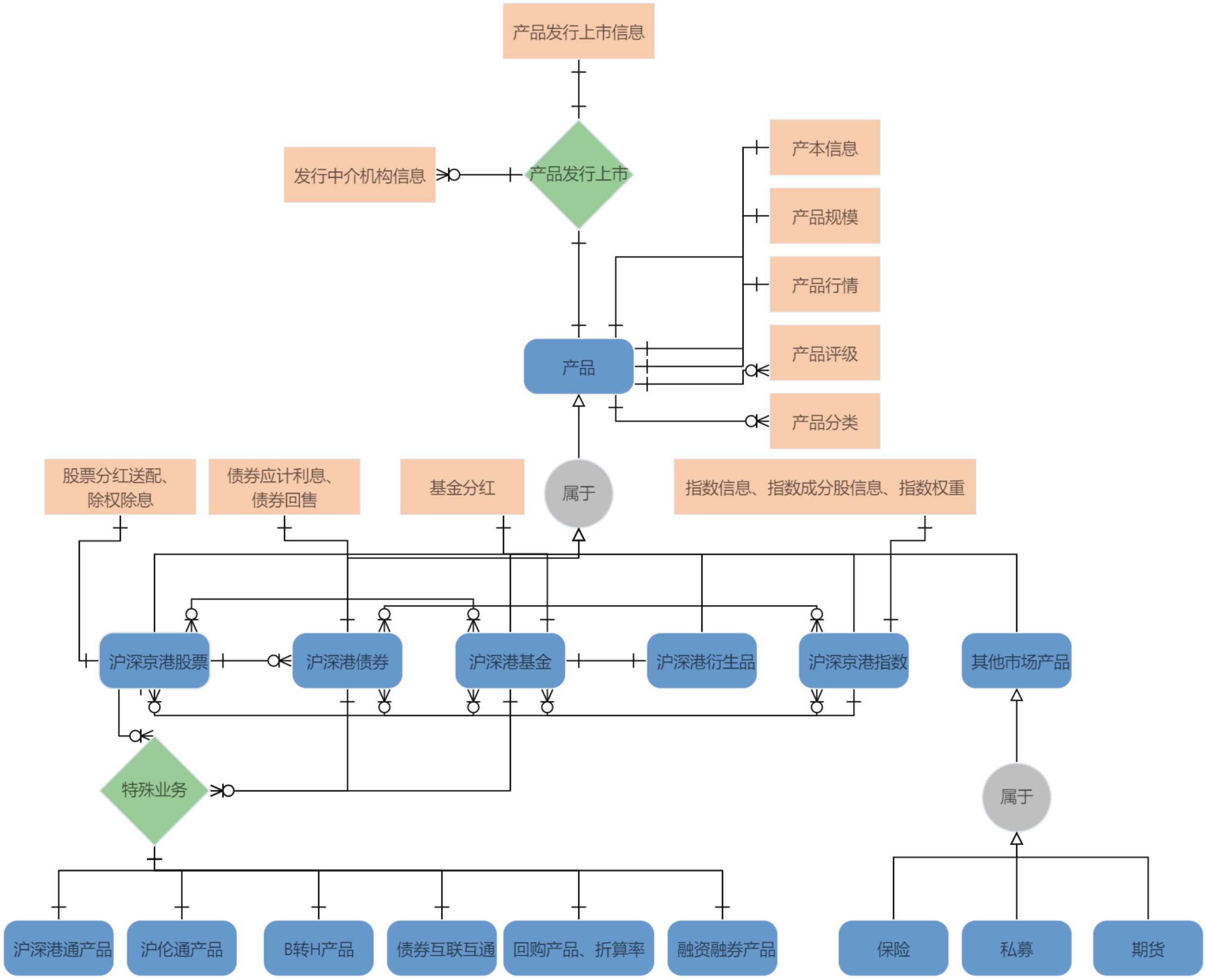


图 12：产品主题－产品信息层面概念模型

券互联互通产品、回购类产品折算率、融资融券产品等。

5.2 产品行情层面

产品行情描述了沪深京港市场的个股行情和指数行情、特殊业务或特定平台行情、其他市场行情。

交易日期、产品代码、开盘价、最高价、最低价、昨收盘价、今收盘价、成交量、成交金额、利率（同业拆借利率）是行情信息的核心要素。

行情按照发布频率可以分为盘后行情和高频行情。

其他市场行情包括非沪深京港证券市场行情和境外市场行情，如外汇行情、中国商品期货行情、HiBor 行情等。

5.3 业务控制层面

业务控制描述了产品停复牌信息和证券级投资者适当性控制信息。

产品停复牌如股债基衍停牌信息，固收平台、基金通平台临时停牌信息等；证券级投资者适当性控制信息如：债券投资者适当性限制清单。

6 事件主题

事件是指金融市场活动中的参与主体对客体

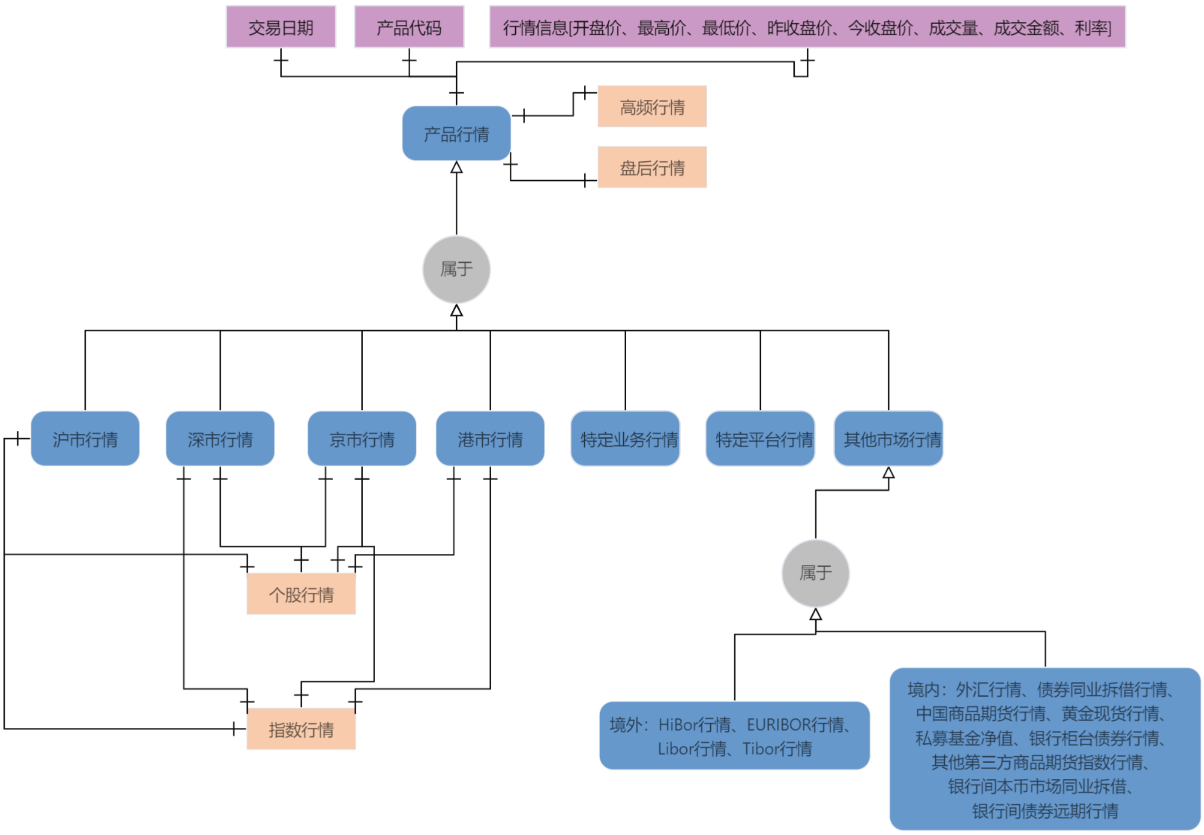


图 13：产品主题－行情信息层面概念模型

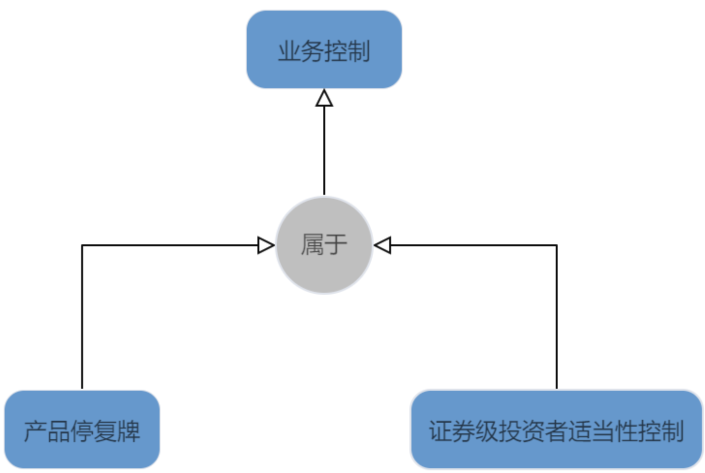


图 14：产品主题－业务控制层面概念模型

所进行的一系列活动，包括过程中所产生的交互等。根据事件的不同阶段，可以将事件主题拆解为发行、交易、权益变动三个子主题。

6.1 发行层面

发行层面主要描述了发行过程及其相关概念的信息。

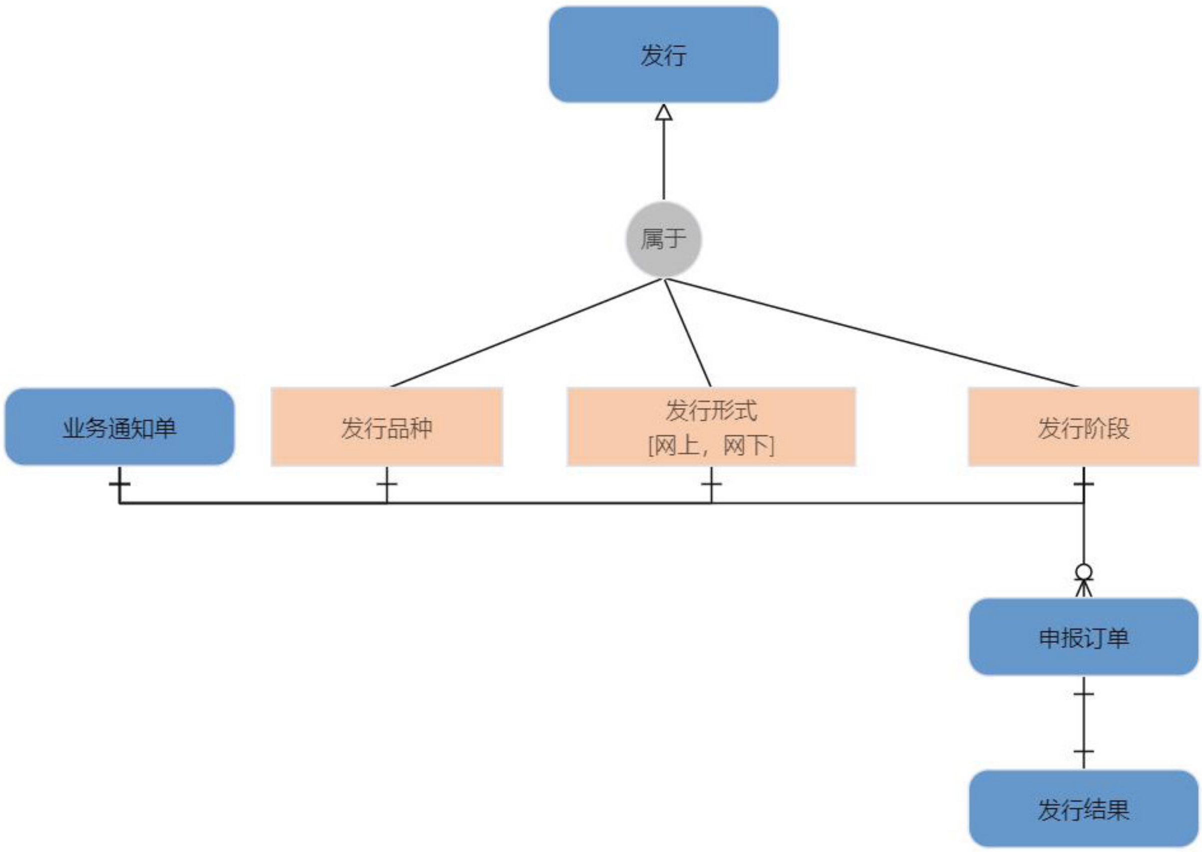


图 15：事件主题－发行层面概念模型

发行包括发行品种、发行形式、发行阶段。根据不同的发行品种，其相关的形式及阶段也可能不同。其中：

发行品种，包含不同证券品种，如股票、债券等；

发行形式，包含网上发行及网下发行；

发行阶段，包含申购、中签、认购、询价等阶段。

在各阶段的过程及流转推进中，往往伴随着一个重要的承载体，即业务通知单。业务通知单的发起、生效是相关阶段的一个重要依据。如公开发行优先股上市通知单、增发新股价格及配售情况通知单等。

发行阶段的各个过程也会产生相应的申报订单，如申购订单、认购订单、询价订单。它们的成交结果即相应发行阶段的最终结果。

6.2 交易层面

交易层面主要描述交易过程中的重要载体及其属性、分类的相关概念信息。

整个交易过程中主要围绕着申报订单及成交订单。订单一般情况下都有共同的属性，如都拥有订单唯一识别码、买卖标志、订单时间等。对于申报类订单，由于可撤销，还有状态标识其状态。

同时，由于交易类型、品种类型的不同，订单的具体内容也会有相应变化：

从订单类型的角度而言，可以分为普通交易订单及非交易订单；

从品种类型的角度而言，可以分为股票、债券、基金、衍生品、场外交易品种等不同类型的订单。不同的证券品种交易特征不同、业务不同，因此订单也可能存在不同的属性。

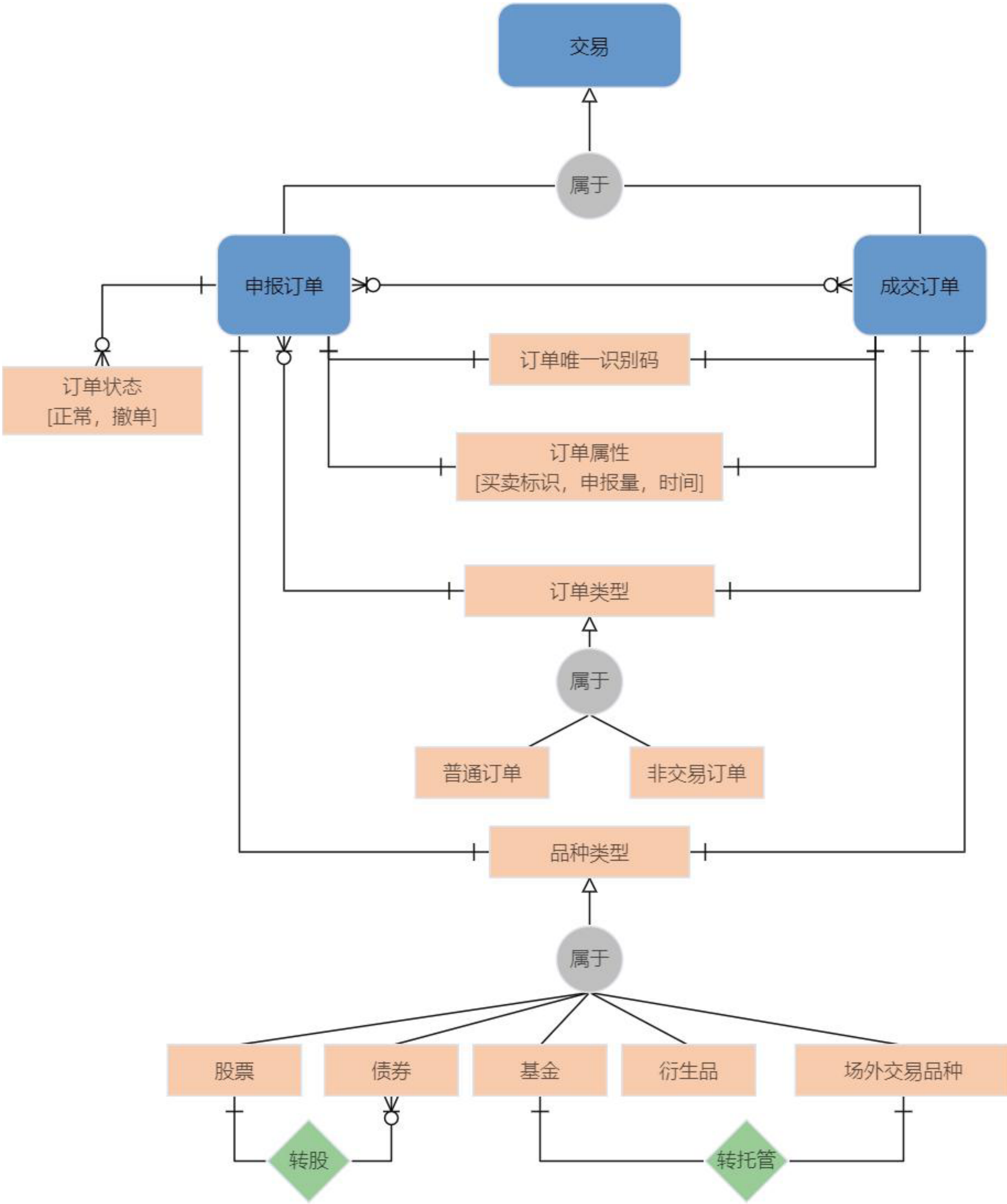


图 16：事件主题－交易层面概念模型

此外，品种之间也可能存在联系，如债转股、场内场外转托管等。各个过程中都会被相应特性的订单所承托。

6.3 权益变动层面

权益变动层面主要描述了权益变动的分类及其相关概念的信息。

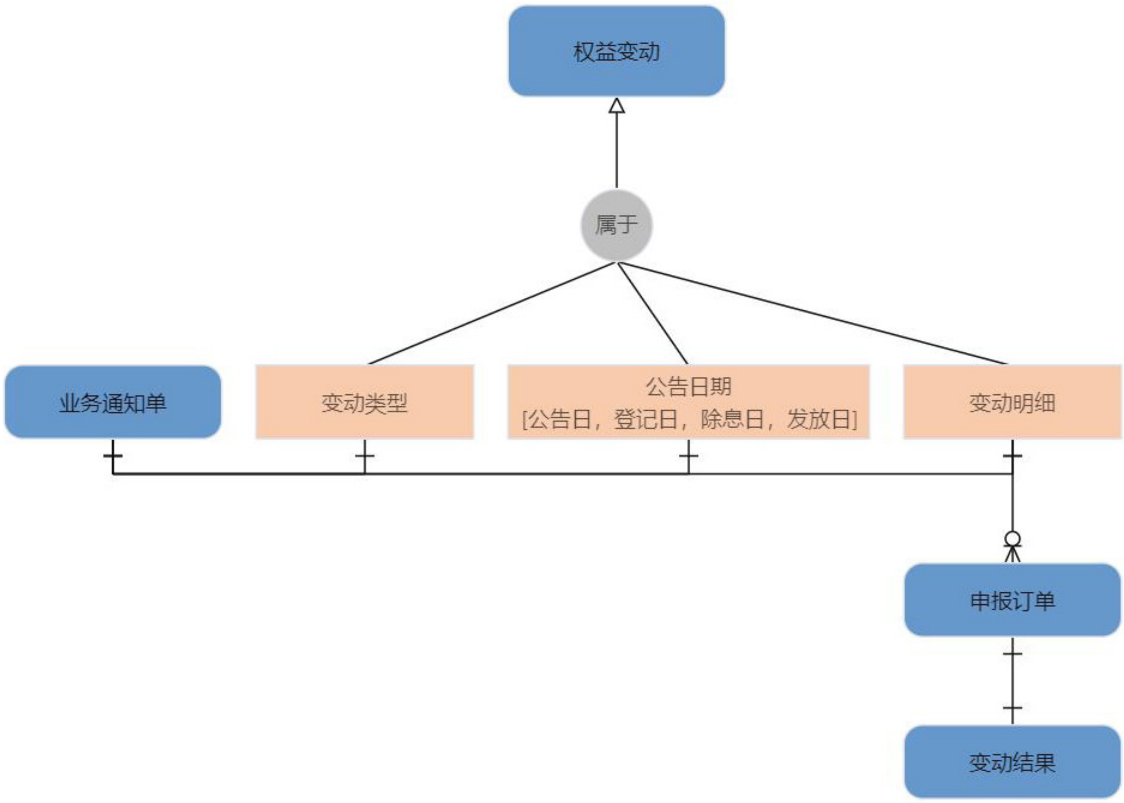


图 17：事件主题 – 权益变动层面概念模型

权益变动包括变动类型、公告日期、变动明细。不同的变动类型所重点关注的日期节点不同。其中：

变动类型，包含红利登记、送股（转增）等;

公告日期，包含公告日、登记日、除息日、发放日等；

变动明细，为变动类型的具体明细及相关比例和结果。

同发行层面，在权益变动的各阶段同样伴随着业务通知单。如配股实施通知单、优先股股息派发通知单等。

权益变动也会产生相应的申报订单，它的变动结果即相应变动类型下的最终结果。

7 监管主题

监管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。根据监管单位的不同分为司法监管、行政监管、行业监管、内部监管四个子主题。

7.1 司法监管层面

司法监管是指由国家司法机关依据法定职权，依照法定程序，对法律实施的情况进行监督。司法机关主要包含了公安机关、检察院、法院三类。贯穿于整个司法案件流程的数据有：立案调查数据、审判过程数据、公告数据、司法执行情况数据、监管对象未履行判决结果而产生违法失信的数据。

7.2 行政监管层面

行政监管层面主要是行政机关或其他行政主体依法对行政违法企业所实施的监察及处罚。

该主题下的数据的范围主要是国家下属行

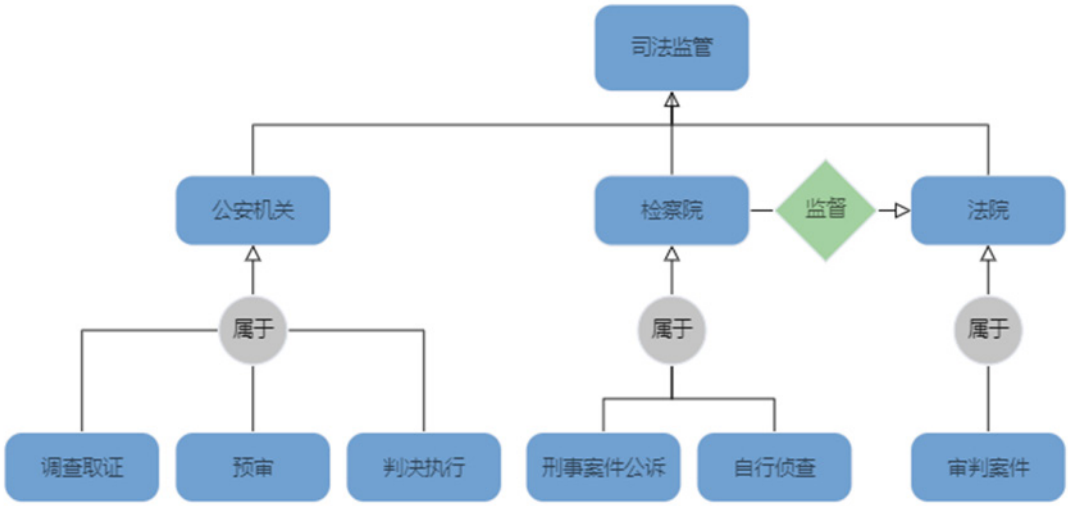


图 18：监管主题－司法监管层面概念模型

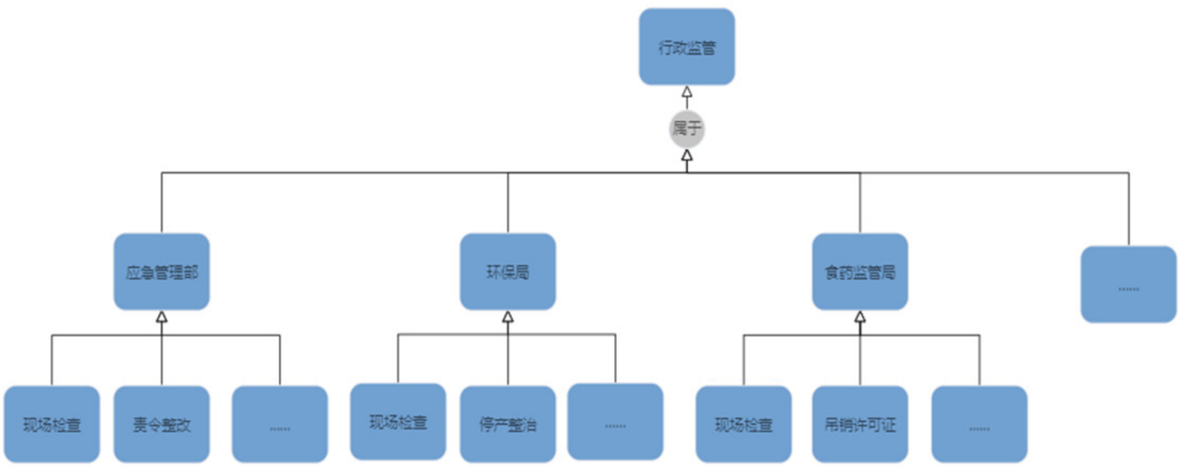


图 19：监管主题－行政监管层面概念模型

政机关对金融行业相关公司及机构的行政处罚数据。核心要素为处罚主体、判决机构、监管措施。比如应急管理部门、食药监管局、环保局对违法公司实施了现场检查、责令整改、停产整治、吊销许可证等监管措施。

7.3 行业监管层面

行业监管主要分为交易所的监管、中国证券管理委员会的监管、中国证券业协会的监管。在概念层面上，交易所监管机构目前主要包括上交所与深交所，但依据数据的重要程度、使用特点，将上交所的自律监管单独作为内部监管的子主题。

交易所监管的内容主要包括公司监管、债券监管、交易监管、会员及其它参与人的监管。

证监会监管的主要内容有发行监管、公司机构的监管、交易监管等。其中，发行监管主要是主板、中小板、创业板的发行审核，科创板、创业板的注册审核。公司机构的监管包括上市公司监管、基金、期货、债券发行公司及相关中介机构的监管。

证券业协会监管的主要内容是会员监管和对

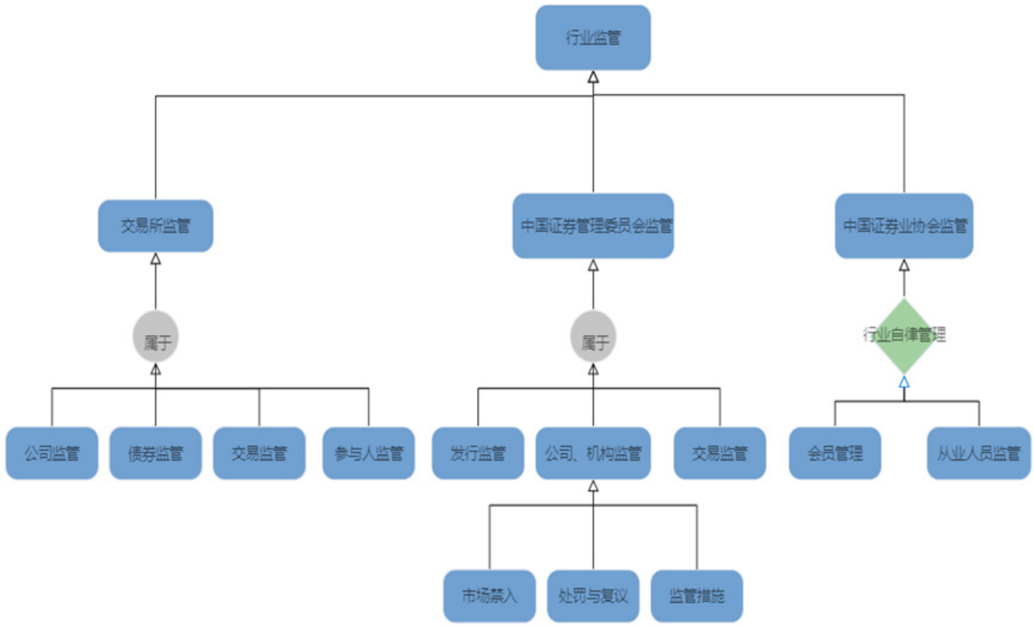


图 20：监管主题－行业监管层面概念模型

从业人员的监管，具体为对会员及会员间开展与证券非公开发行、交易相关业务活动进行自律管理，对网下投资者、非公开发行公司债券、场外市场及场外衍生品业务进行自律管理。

7.4 内部监管层面

内部监管层面主要是交易所内部的债券监管、公司监管、交易监管、参与人监管等。监管日志是开展日常监管、专项检查活动的记录数据；

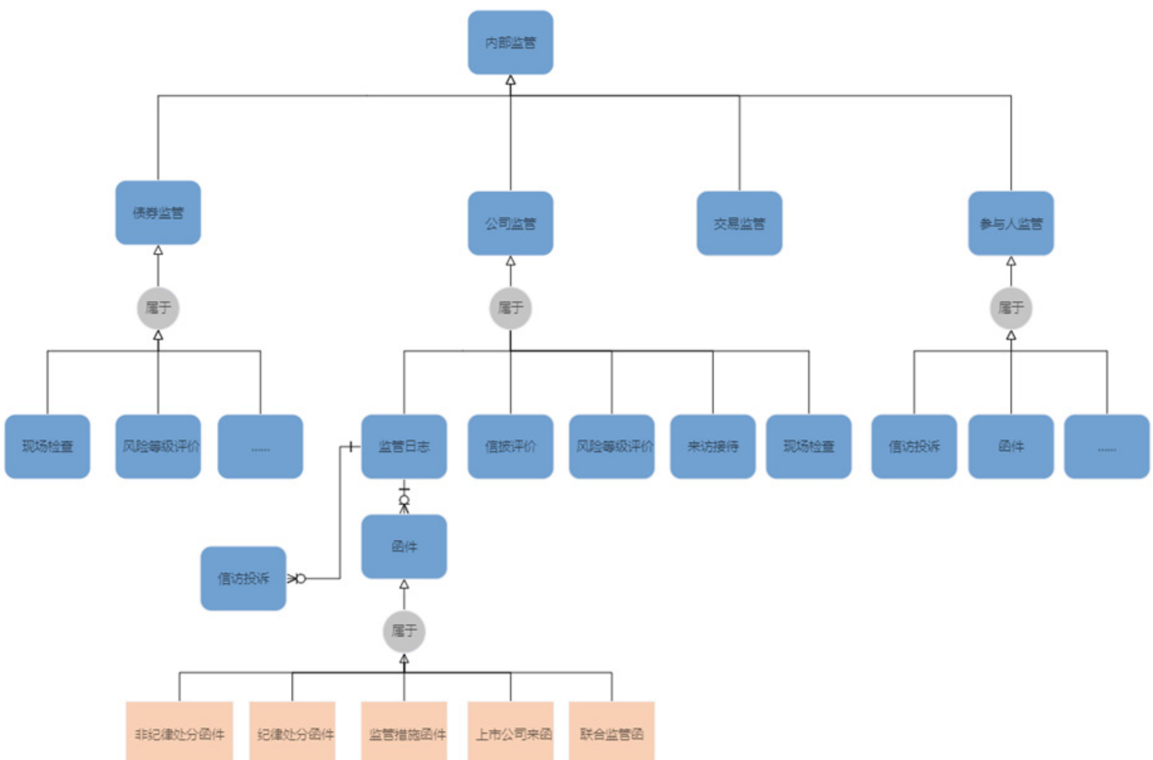


图 21：监管主题－内部监管层面概念模型

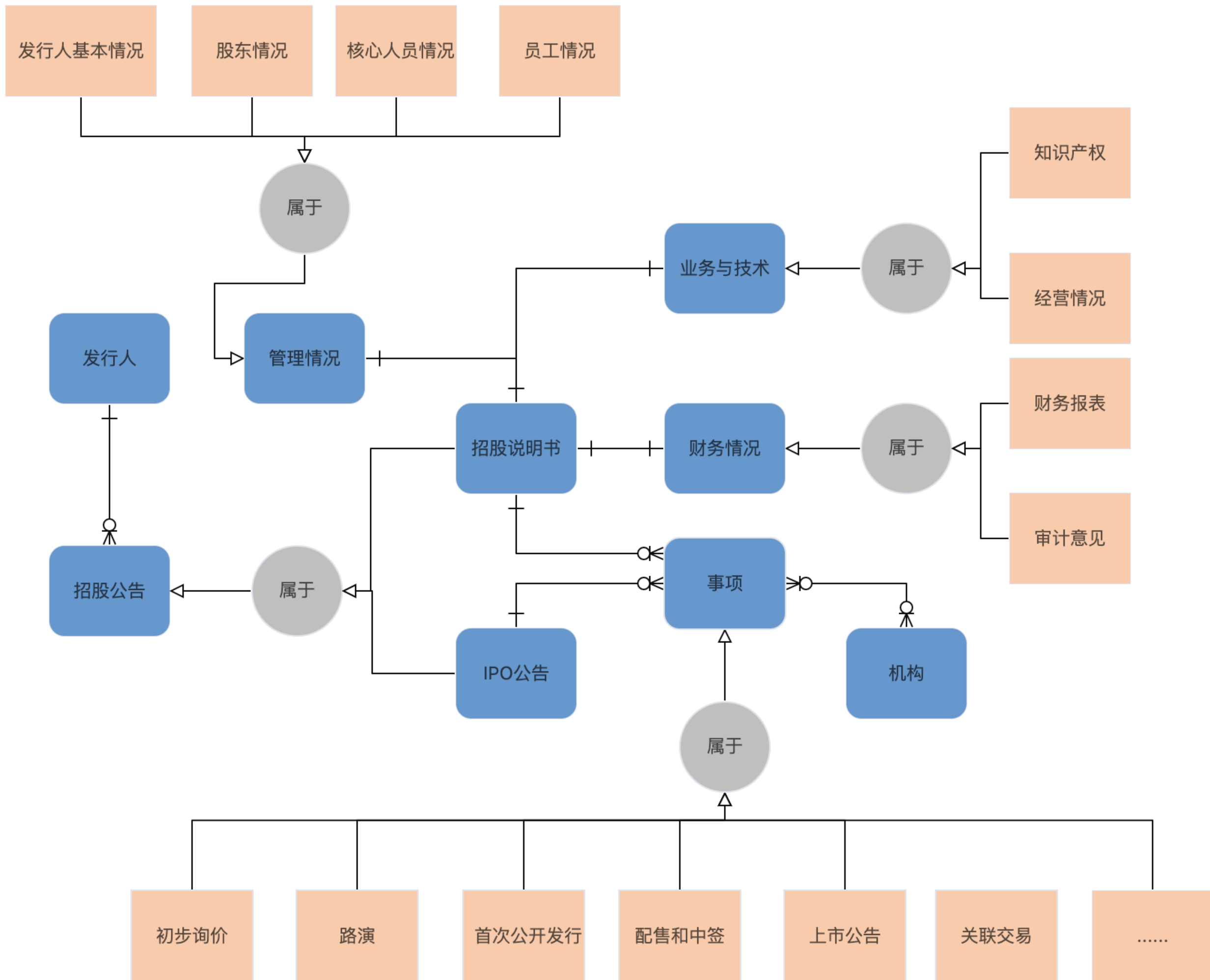


图 22：披露主题 – 招股公告层面概念模型

监管函件是监管工作的载体，分为纪律处分函件和监管措施函件，纪律处分函件比监管措施函件警示处罚的程度更重。

8 披露主题

披露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，披露内容包括招股公告、定期公告和临时公告。

8.1 招股公告层面

招股公告层面主要描述了发行人在上市前披露的相关公告，包括招股说明书和 IPO 公告。其中，招股说明书包含发行人的管理情况，如发行人基本情况、股东情况、核心人员情况和员工情况等；业务与技术情况，如知识产权、经验情况等；财务情况，如财务报表及审计意见；同时还会披露发行人的相关事项信息，如关联交易、资金投向等，事项和机构之间存在多对多的关系，如披露方可以和多家企业存在关联关系，一个企业也可以出现在多个关联事项中。IPO 公告指上市公司上市前披露的临时公告，主要包含的招股事项为路演、询价、配售、中签等，即上市申请通过后到正式上市之前这一阶段披露的相关公告。

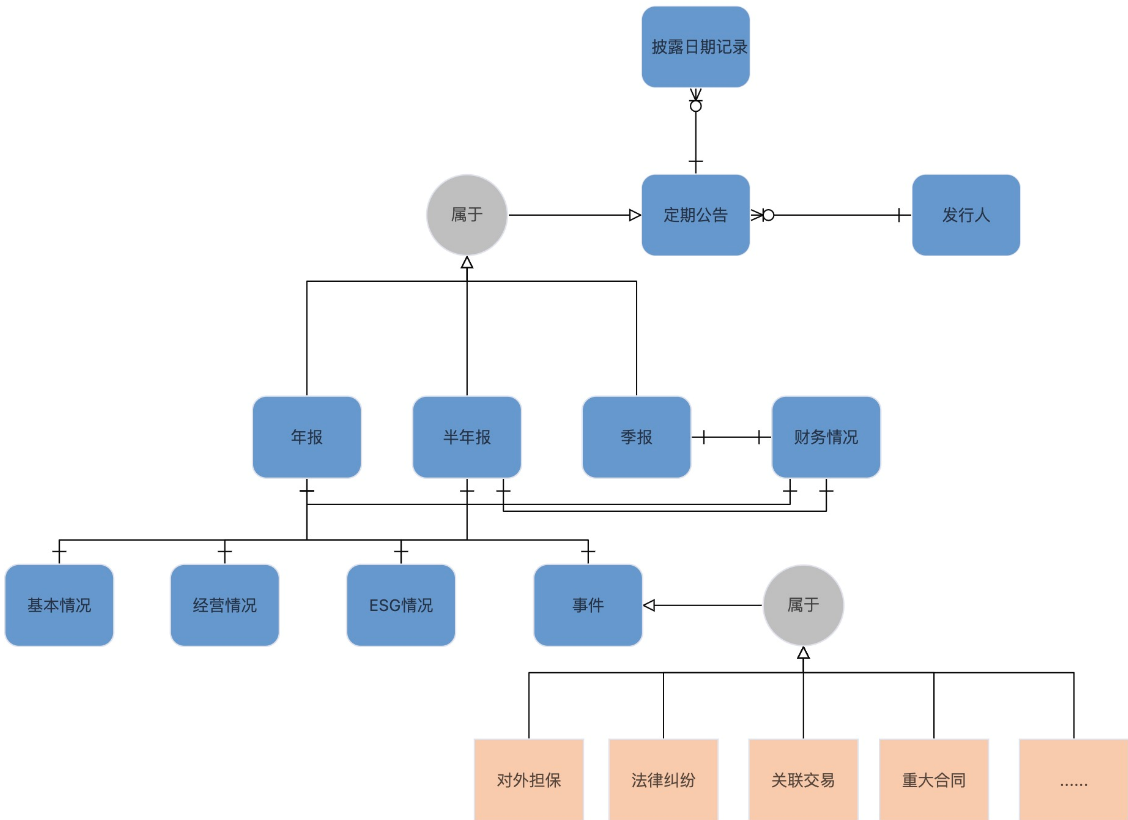


图 23：披露主题 – 定期公告层面概念模型

8.2 定期公告层面

定期公告层面主要描述了发行人在特定报告期间披露的定期公告，包含季报、半年报以及年报。其中，季报分别在一季度和三季度披露，内容为未经审计的财务报表；半年报和年报披露内容一致，均包含公司基本情况、经营情况、ESG情况、财务报表、以及重大事件等。

公司基本情况包括公司信息、联系人、股票简况等相关资料；股东情况如控股股东、实际控制人、持股在百分之十以上的法人股东相关信息。经营情况包括所处行业情况、从事业务情况、核心竞争力分析等。ESG 情况包含公司治理、环境与社会责任相关事项。财务情况包含审计报告、四大财务报表以及相关财务附注。重大事件包括发行人需披露的对外担保、法律纠纷、关联交易、重大合同等事项。

8.3 临时公告层面

临时公告层面主要描述了发行人上市之后不定期披露的相关事项，是业务部门进行公司持续监管的重要依据。临时公告主要可分为重大事项、股权股本事项、交易提示事项以及一般事项四类。

重大事项类公告是上市公司对拟发生或已发生的重大事件进行披露的公告，包括利润分配、权益分派、股份增减持、资金投向、并购重组、重大合同、股权激励、关联交易、抵押担保、委托理财、法律纠纷、政策影响、人事变动、配股增发、业绩预告、业绩快报等。

股权股本事项类公告是因上市公司股权或股本发生变化而进行披露的公告，包括股权分置改革、质押冻结、质押式回购、回购股权、约定购回等。

交易提示事项类公告是因上市公司发生交易特殊情况或者特别处理，需对广大投资者进行提

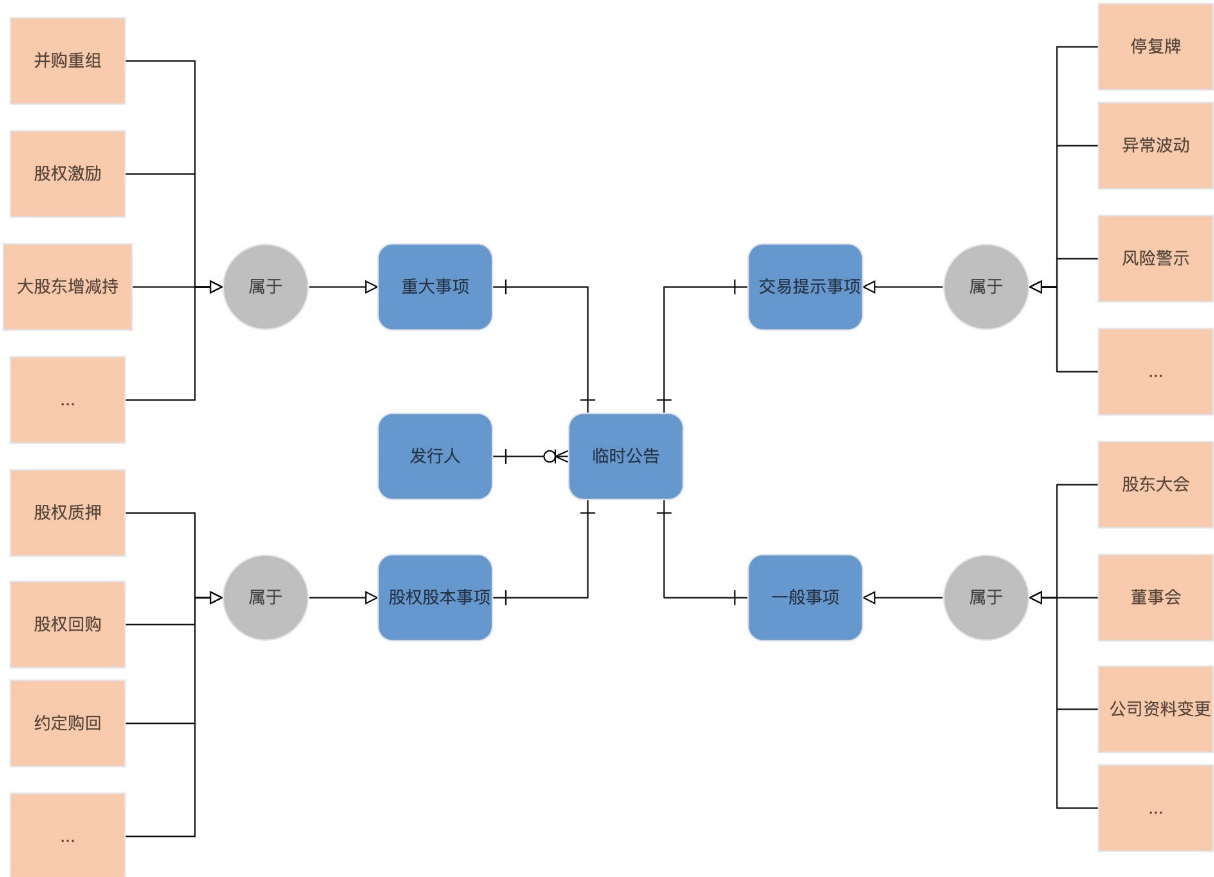


图 24：披露主题－临时公告层面概念模型

示而进行披露的公告，包括停复牌提示、异常波动、澄清公告、风险提示、上市状态变更等。

一般事项类公告是上市公司日常运营中发生的基本事项，包括股东大会、董事会、公司资料变更、中介结构、其它补充更正等。

以上四类公告之间存在一定的关联性，如股东大会或者董事会会对相关重大事项或者股权股本事项进行决议或投票，公司重大事项的实施会进一步影响股票交易，从而引发交易提示公告的发布和披露。

9 信息资讯主题

信息资讯是指第三方机构或自然人对金融市场的评价及其他与金融市场有关的数据，包括研究报告、宏观数据、舆情信息、法律法规四个方面。

9.1 研究报告层面

研究报告是指由参与人对特定主体进行具体调查、研究、分析后发表的产物，分为分析报告、调研报告、预测报告。参与人和研究报告主体之间存在发表关系，研究报告和参与人存在分析关系。

9.2 舆情信息层面

舆情信息指通过互联网传播的公众对现实生活中热点、焦点问题所持的有较强影响力、倾向性的言论和观点，分为新闻舆情和网络舆情。参与人和舆情信息主体存在发表关系，舆情信息和参与人之间存在围绕关系。

9.3 宏观数据层面

宏观数据指一系列的宏观经济学的统计指标经过一定公式计算得到的一个可以反映经济状况

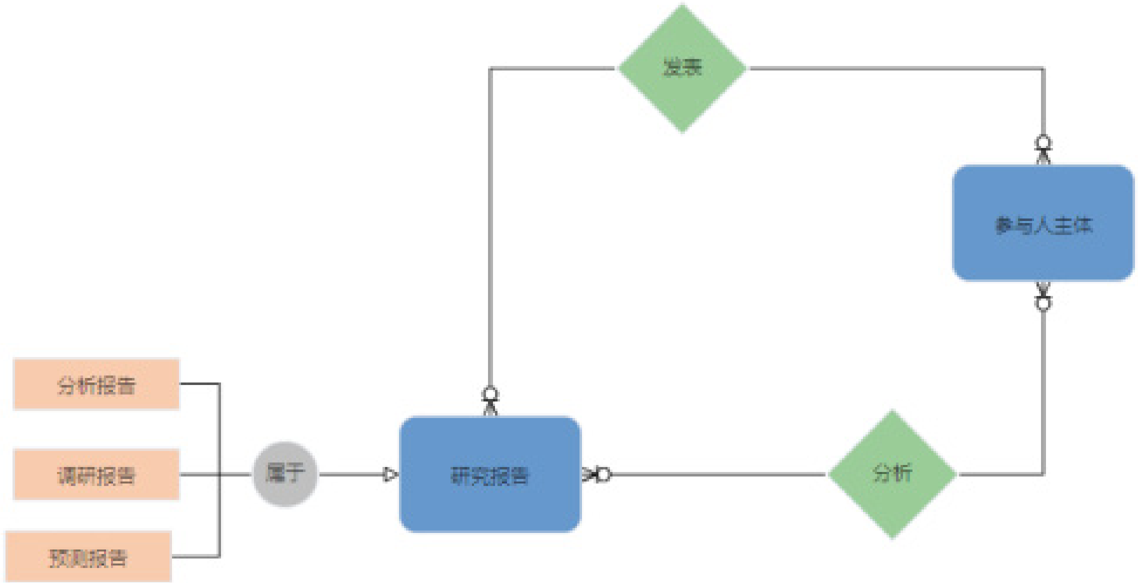


图 25：信息资讯主题－研究报告层面概念模型

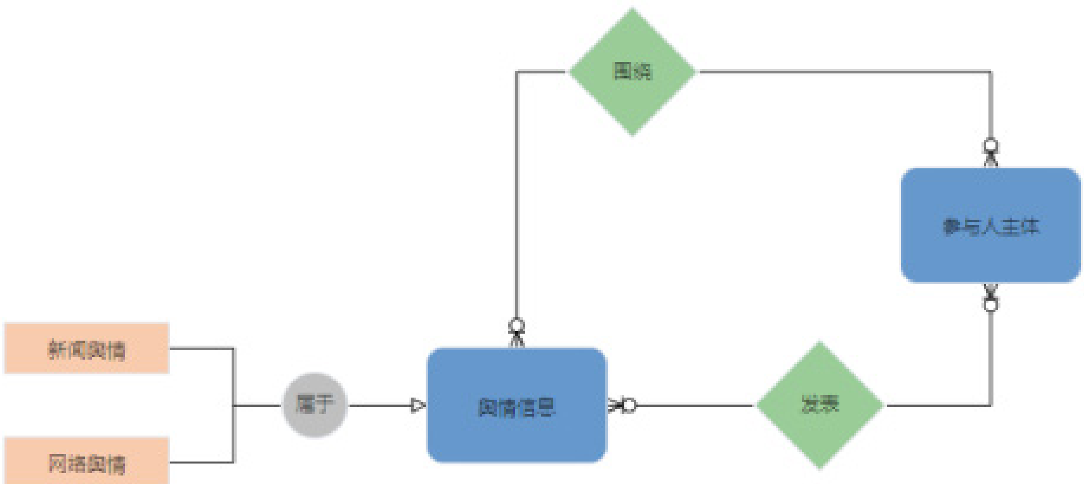


图 26：信息资讯主题－舆情信息层面概念模型

的综合指标，包括生产总值(GDP)、国民总收入(GNI)、劳动者报酬、消费水平等等。参与人和宏观数据主体存在产生关系。

9.4 法律法规层面

法律法规指中华人民共和国现行有效的法律、行政法规、司法解释、地方法规、地方规章、部门规章及其他规范性文件以及对于该等法律法规的不时修改和补充。参与人和法律法规主体存在制定发布关系，法律法规和参与人存在约束关系。

10 总结

以上是对八大主题概念模型的描述，是从业务视角分析了每个主题包含的主要业务实体以及

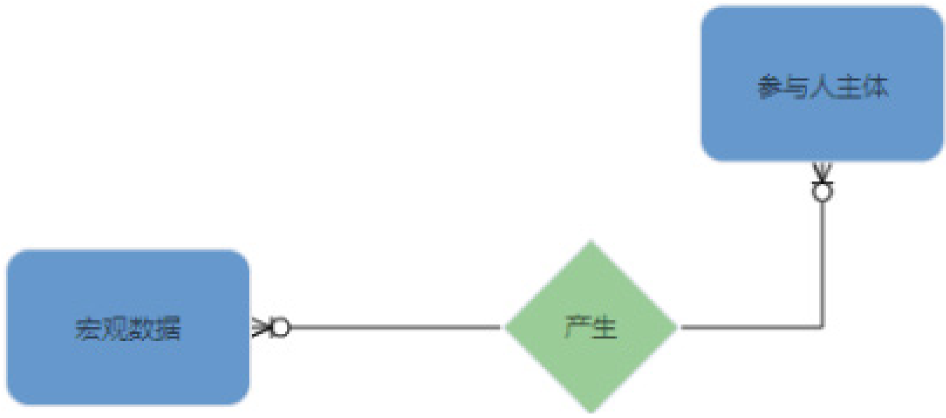


图 27：信息资讯主题－宏观数据层面概念模型

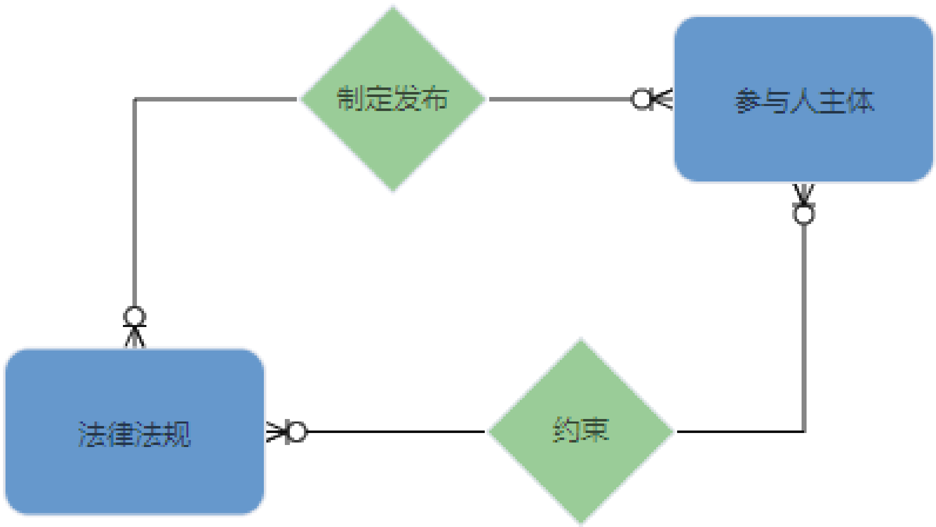


图 28：信息资讯主题－法律法规层面概念模型

实体之间的业务关系。在形成了概念模型之后，下一步的工作就是把这些实体与关系映射到数据库中，形成一张张数据库表，用于指标的统计和标签的分析。

以上内容是与本所多位数据专家反复讨论后形成的，需要说明的是，不同的人对同一个业务场景提炼出来的概念模型可能不完全相同，概念模型的设计颇受主观认识的影响，文章中部分内容或许仍有不妥之处，后续我们将在实践过程中持续进行推敲与打磨。

数字化转型下的东吴证券数据中台建设实践

葛菊平、潘金锐、陈震诞 / 东吴证券股份有限公司 运营中心 苏州 215100
E-mail：gejp@dwzq.com.cn

本文简要回顾了东吴证券 20 多年来数据平台从数据中心转向数据中台的建设历程，总结了东吴证券数据中台建设的背景和目标，通过介绍数据中台建设的成果，体现了数据中台在公司数字化转型中重要作用，同时综合这些年的经验和技术发展趋势，对于未来数据平台建设提出了一些粗浅想法。

1 背景

近期国务院印发的《“十四五”数字经济发展规划》中将数据资源作为数字经济发展的关键要素，将全要素数字化转型作为重要推动力，数据的重要性提升到前所未有的高度。在证券行业，数据作为最重要的资产之一，随着业务的飞速发展呈现爆炸式增长，传统的数据体系和支撑方式已经难以适应纷繁复杂的业务需求，如何唤醒沉睡的数据资产，把数据真正用起来，以支持自身业务数字化升级，成为摆在眼前的转型难题。

让数据真正发挥价值，在业务上发挥关键作用的关键在于让数据真正被用起来。东吴证券一直对数据的使用非常重视，多年来坚定地投入大量的人、财、物在数据上进行探索，逐渐形成了一支有技术、懂业务的数据人才队伍，建设了一系列的数据基础设施和数据应用系统。数据要被用好，需要解决“找数 - 识数 - 用数”的难题，为此公司在建设数据平台时围绕数据的“易得、易懂、易用、易开发”开展，最终期望形成数据生产和使用的闭环。基于上述原因，形成了公司数据类系统“共享共赢、自主可控”的基本建设思路，即建设公司统一的基础数据平台，让基础数据平台为公司各业务条线提供统一的数据服

务。为使业务部门需求得到快速响应，公司在与供应商保持合作的情况下，一直坚持一些重要的数据类平台和应用的自主研发。

2 数据平台建设历程

公司最早一代数据平台是在 2000 年建设的基于传统 OLTP 数据库的数据中心，该平台期初的功能主要是风险监控。在随后的多年时间里，公司持续对该系统进行升级改造，纳入了更多的数据，并把各种新业务的风险监控纳入该平台。同时反洗钱、异常交易监控也部署在该平台上，大量的统计报表也在该平台上进行了自主研发。通过持续的研发和应用，该平台逐渐成为公司那些年唯一的数据服务平台，内部的数据分析和数据应用都在该平台上自主研发出来，并在该平台上自主研发了公司第一个版本的非现场审计系统。虽然随着最近几年新技术的发展和需求的变化，部分数据应用已经从该平台剥离，采用新的平台来建设，但该平台上的很多功能依然在发挥着重要作用。

随着业务和技术的发展，公司对于数据的要求越来越高，所以公司加大了对数据基础平台的投入，并加快了建设的步伐。2017 年，公司建设了大数据平台，并开始基于大数据平台建设公司的数据中台，同时也开始了基于大数据平台进行的数据类应用系统建设，原先的数据中心也逐步进行了改造，把数据的采集和清洗任务交给了数据中台处理，新建的各类应用也把基础的数据采集、清洗及部分繁重的计算等功能交给了数据中台，而专注于应用本身功能的实现。

3 数据中台能力建设

在长期的数据平台建设实践中，我们逐渐总结出，要让数据用起来，建设一个强大的公共数据基础平台，对于公司各种数据需求和数据应用的实现非常必要，能够让数据便捷的实现“进得来，留得住，出得去”。我们确定的公共数据平台的建设方针是在数据治理框架下建设功能完备的数据中台，把数据中台的能力做扎实，让中台把脏活、累活都承接过来，使数据开发者轻松开发，数据使用者便捷获取。所以数据中台的规划框架是“强底座、厚中台、薄前台、全管控”，即数据中台应该具有数据的“获取、存储、计算、推送、报表、分析挖掘、服务 API、数据类应用建设、科学能力”九大能力。

3.1 顶层设计，以数据治理为切入提升数据管理质效

我们以数据治理任务为抓手，厘清数据脉络，贯通公司数据管理体系，提高全员参与度。数据治理作为数据中台建设的基础工作，只有定义明确、标准统一的数据才是构建数据资产体系的基础。近几年我们在数据治理的制度建设、组织建设、平台建设等多个方面上持续投入，我们首先完成以公司分管领导挂帅的数据治理的顶层组织架构设计，制定了数据治理基础制度，明确了数据治理工作的人员、内容和方向。通过持续开展的数据治理工作，我们对公司主要的 80 多

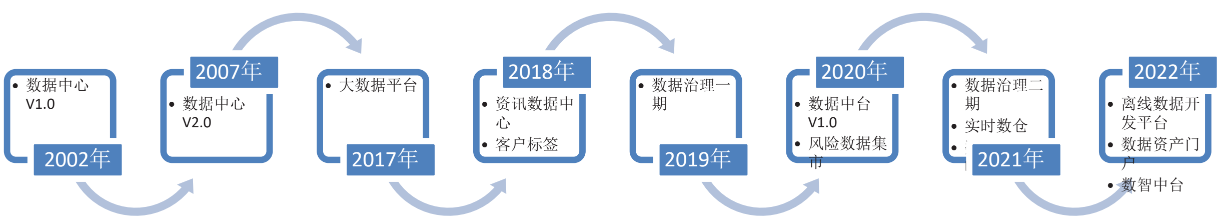


图 1：数据平台建设历程

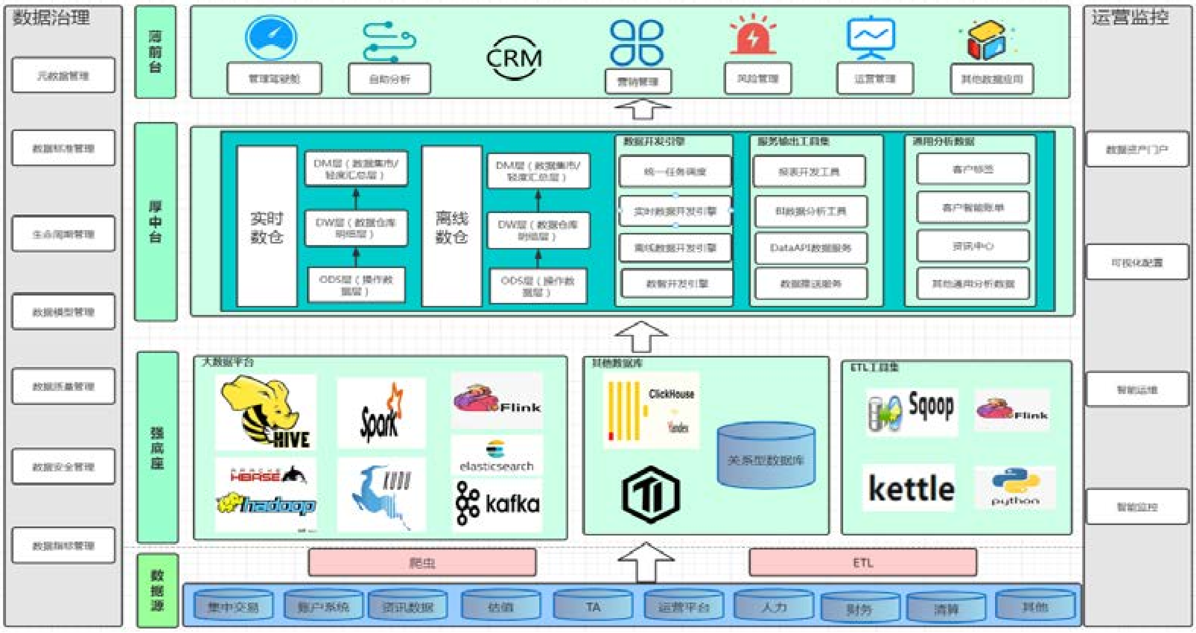


图 2：数据中台架构

套系统的元数据完成了梳理，形成公司的数据资产目录；通过对数据处理脚本和程序的解析，梳理出数据中台和风险数据集市的数据血缘关系和影响分析；联合各业务部门，梳理和发布了1200 多项数据指标；根据业务指标和基础数据定义，梳理出数百项数据质量校验规则并开展数据质量核检；针对重要系统的数据进行了分级分类梳理工作，并把分级分类结果纳入数据管控平台进行管理，实现 OneData（数据的标准与统一）、OneEntity（实体的统一）、OneService（数据服务统一）目标。

3.2 夯实底座，构筑坚如磐石的数据中台基础能力

数据的获取、存储、计算、推送这几项是数据中台的地基，所以我们首先从这四项基础能力

图 3：数据管控平台

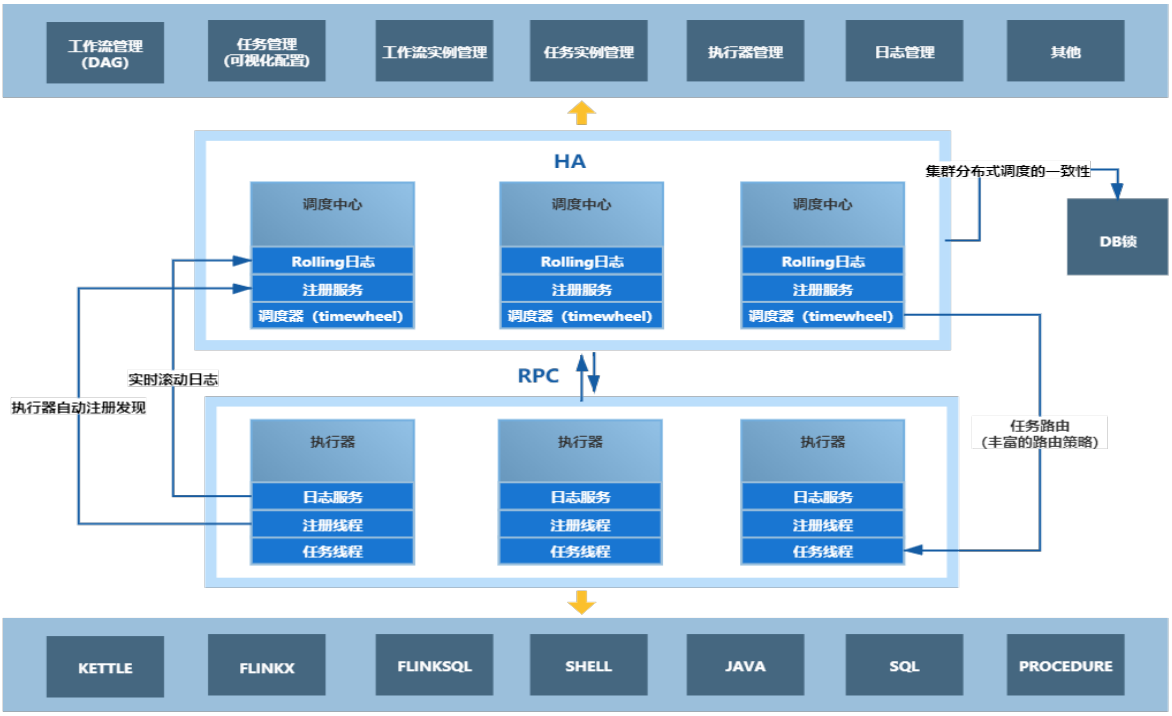


图 4：任务调度平台架构

开始打造数据中台。我们搭建了基于 hadoop 的大数据集群，集群的 CPU、内存、硬盘等资源为公司大量的数据提供集中存储和计算能力。根据不同的数据使用需求和数据处理层级，数据会先进入大数据平台，然后经过处理的数据会根据不同的使用场景进行存储；为便于管理，我们设计了数据抽取和推送的处理流程，把数据的获取和推送等操作都定义成一个个原子任务，然后通过自主研发的调度管理工具进行统一的调度管理。在数据获取和推送上目前我们主要采用 kettle 和 FlinkX，但我们并没有直接使用 kettle 和 Flinkx，而是把这两个工具作为采集和计算的底层执行器，由我们的任务调度管理工具进行调度管理。通过这样的设计，解决了原生工具在大量任务并发执行时的设置和管理问题，界面的可视化设计也极大的简化了采集和推送任务的配置工作。目前数据中台对接了有 80 多个数据库，每天执行的数据采集、计算、推送任务接近 5000 个。

3.3 开拓创新，打造丰富便捷的数据中台服务能力

让数据“高效率，多模式”的使用。我们统一数据服务窗口，打造数据共享能力，提高数据开发效率，降低数据开发门槛。

实时数仓：数据越早被使用，价值也越高。为提高数据的实时性，准确性，我们在数据中台的实时数据采集和落地上引入了 flink 流式数据处理能力，建设了实时数据仓库。通过采用流式数据处理实时数据，改变了传统的轮巡模式，实时数据延迟有秒级提升到毫秒级，数据的完整性，准确性得到了明显的提升。我们利用 flink 对流处理的优势，把公司原先的融资融券实时风险监控和实时异常交易采用流式计算进行了重构，对融资融券的风险管理和异常交易监控的质量提升提供了很大的帮助。

离线数仓：在采集和存储了大量的离线数据后，根据 SDOM 模型，我们规划并建设了数据仓

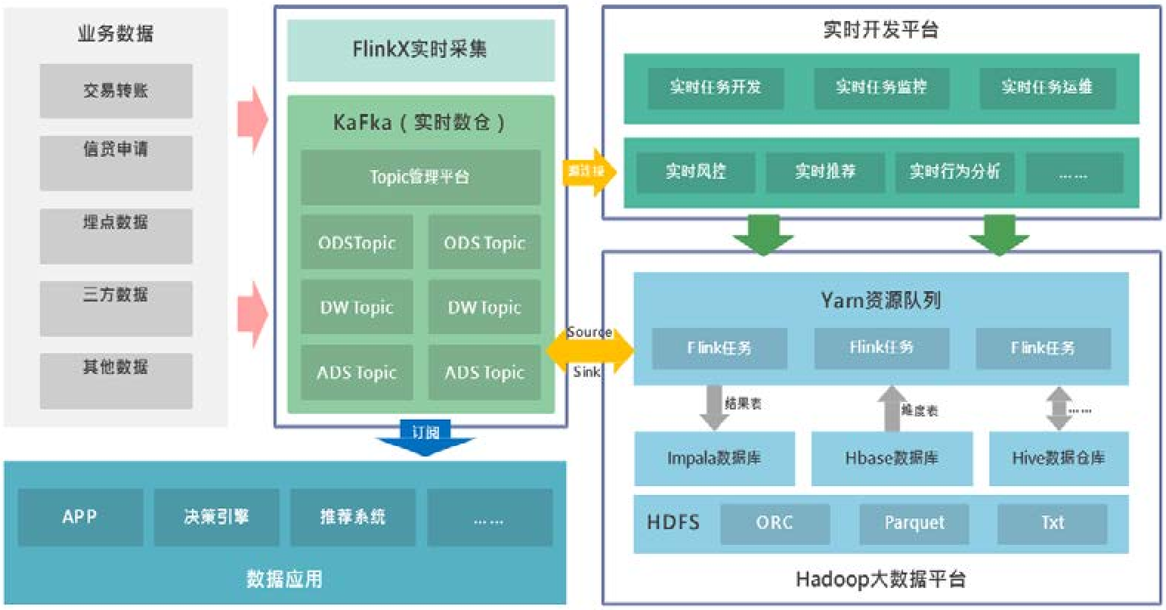


图 5：实时数据仓库构建模式图

库，并基于此建设了经纪数据集市一期和风险数据集市。经纪数据集市一期的建设为公司零售业务线的数据分析和业务运营提供了有力的支持，公司秀财事业部的数据分析人员在通过中台提供的 BI 工具对接集市数据，实现了日常运营数据的自助分析，并用大量的数据分析结果来指导运营工作。风险数据集市一期的建成为公司的全面风险管理建设奠定了坚实的数据基础，统一了风险监控的数据基础和数据标准，加快了风险监控系统的建设进程，公司基于风险数据集市公司已经建设了大量的风险监控类应用。

DataApi：为了提供更便捷的数据服务，减少数据烟囱的形成，我们建设了数据服务 DataAPI 平台。DataAPI 基本是个无代码集群平台，API 的开发只需要通过简单的配置或者写 sql 语句就可以快速生成，同时 API 的授权机制、流

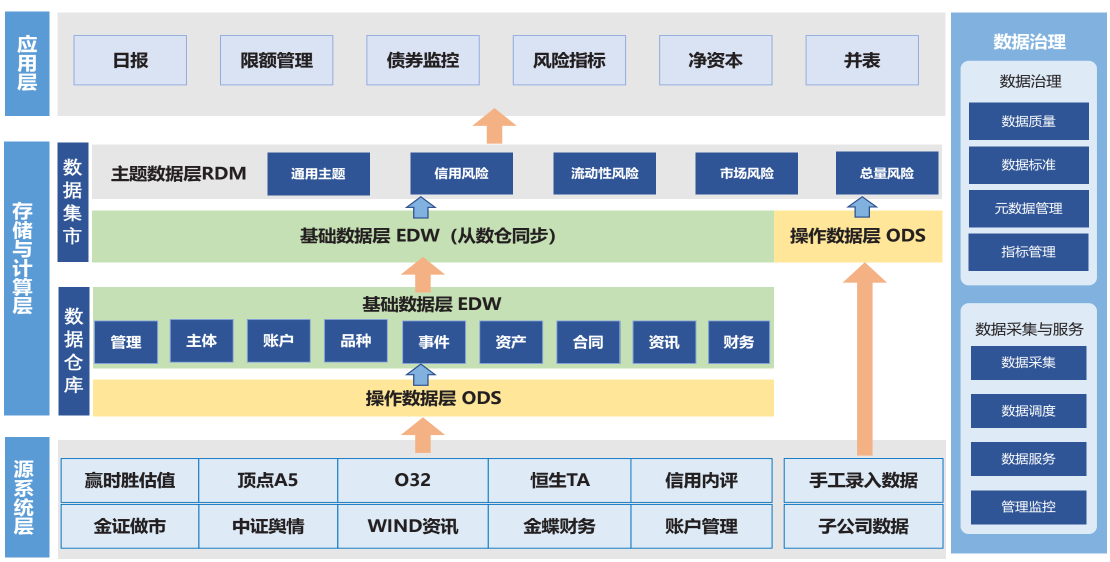


图 6：风控数据集市架构

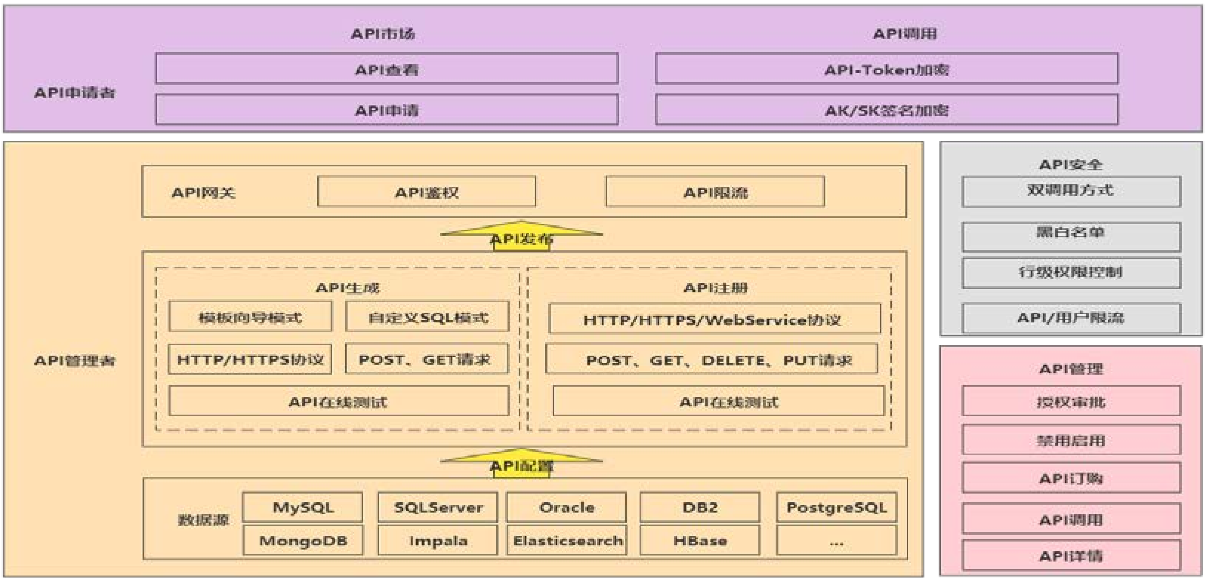


图 7：DataApi 系统架构

量控制机制、集群管理等保证了 API 的高可用可扩展能力。DataAPI 不仅提供了内部应用系统数据获取服务，也为公司的新版 APP 提供了数据服务。我们结合实时数仓的流式数据快速处理功能，为秀财 APP 的 F10 提供了个股咨讯数据 API，实现了完全自主的全链路资讯数据技术服务。DataAPI 的建成并投产，有效的降低了部分应用系统使用数据时的数据烟囱生成，且由于数据获取的安全、便利、同源，保证了应用系统获取的数据的准确性和及时性。DataAPI 目前已经为 6 个应用系统提供了 80 个 API 服务，最高时提供了 80 多万次的日调用。

3.4 赋能业务，建设数字化运营支撑下的新应用

依托数据基础底座能力，数据中台持续赋能公司数字化转型。数据的价值在于使用，数据中台通过数据推送给其他应用系统使用外，本身也提供多种数据使用模式。

通用数据分析：数据中台强大计算能力可以替代数据类应用系统的一些繁重计算任务，数据经过数据中台的计算预先处理好，应用系统只需调用计算结果即可，结果在不同系统间的一致性也得到了保证。我们在数据中台上建设了客户标签系统、客户智能账单系统、资讯中心等系统，提供一些公共的基础数据服务，这些公共的基础数据被 C 端和 B 端系统直接使用。

客户标签系统：公司为提供客户的统一画像，在数据中台上搭建了客户标签计算引擎，通过分析客户的基本信息、交易行为，为每个客户生成丰富的行为特征，然后根据客户画像需要，在客户特征上使用算法计算后生成每个客户的特定标签。数据中台提供的客户标签通过提供 API 接口，给包括客户分级分类、CRM、营销管理、MOT 等应用提供服务。

客户智能账单：为丰富给广大投资者提供的账户资产、交易相关数据服务，并精准及时的了解客户的投资能力，便于作出合理的客户服务决策，我们在数据中台上生成了客户智能账单。客户智能账单的数据生成后通过 API 和批量数据输出等服务方式提供给客户和客户运营人员使用。

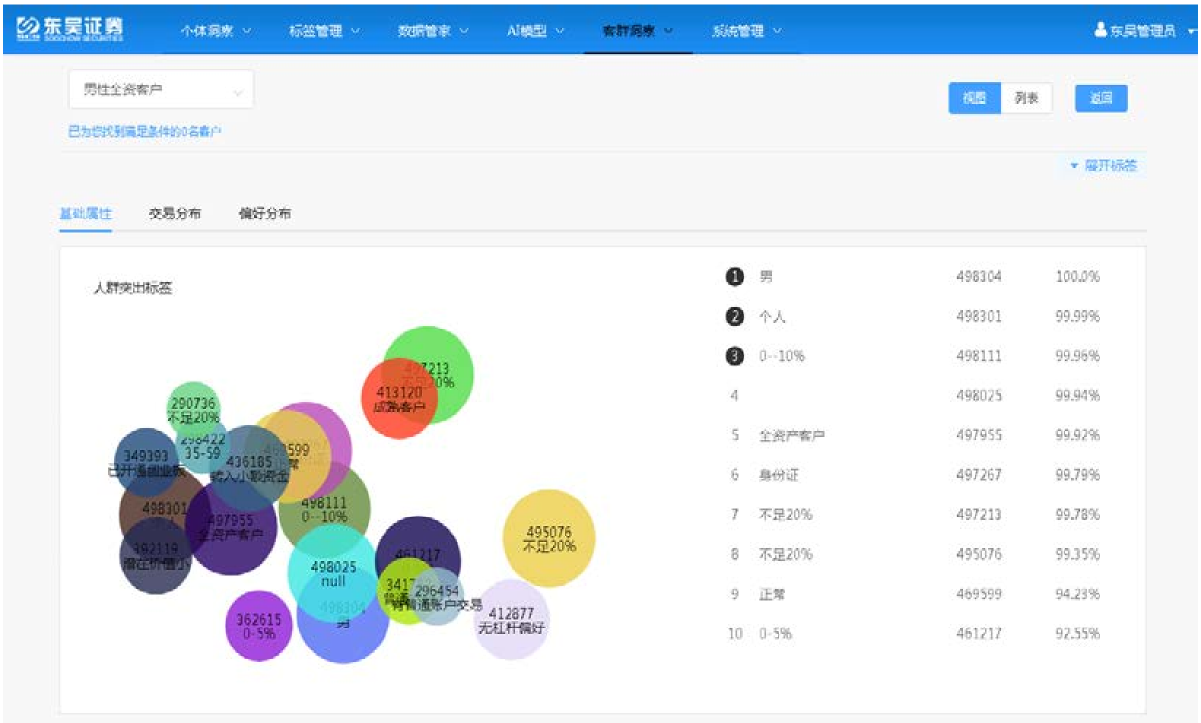


图 8：客户标签之人群划分实例

资讯中心：公司的资讯中心通过把采购的多套资讯数据库的数据通过实时采集的方式采集到资讯中心，资讯数据在资讯数据中心中通过一定的规则进行清洗转换，然后通过 DataApi 等方式给应用系统使用。目前资讯中心已为秀财 APP，CRM，全面风险管理，投研平台，CISP 数据报

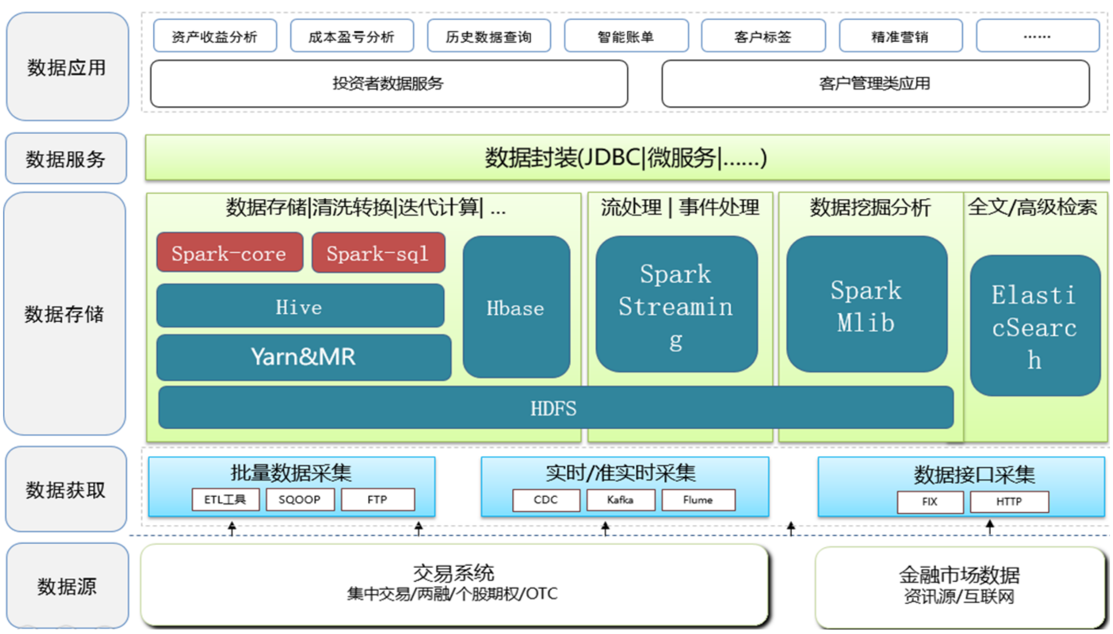


图 9：客户智能账单架构

图 10：管理驾驶舱示例

送等大量的系统提供了实时资讯数据服务。

报表及 BI 自助工具：数据的获取、存储、计算、推送能力的建设为数据中台提供丰富的数据服务能力奠定了坚实的基础，为践行让数据便捷的用起来这一理念，需要根据场景建设合适的数据中台服务能力。现阶段，我们在数据中台的数据仓库和数据集市基础上提供了数据报表，并基于此建设了管理驾驶舱系统，我们也提供 BI 工具给具有数据分析能力的一线业务人员进行自助分析。目前，报表系统的日均（含休息日）访问量达到 4500 次，月使用报表 500 个以上，月活跃用户接近 400 个；BI 自助分析工具的日均访问量达到了 120 次，月使用 BI 模板 325 个，月活跃用户数 31 个。BI 工具的引入，并对有关部

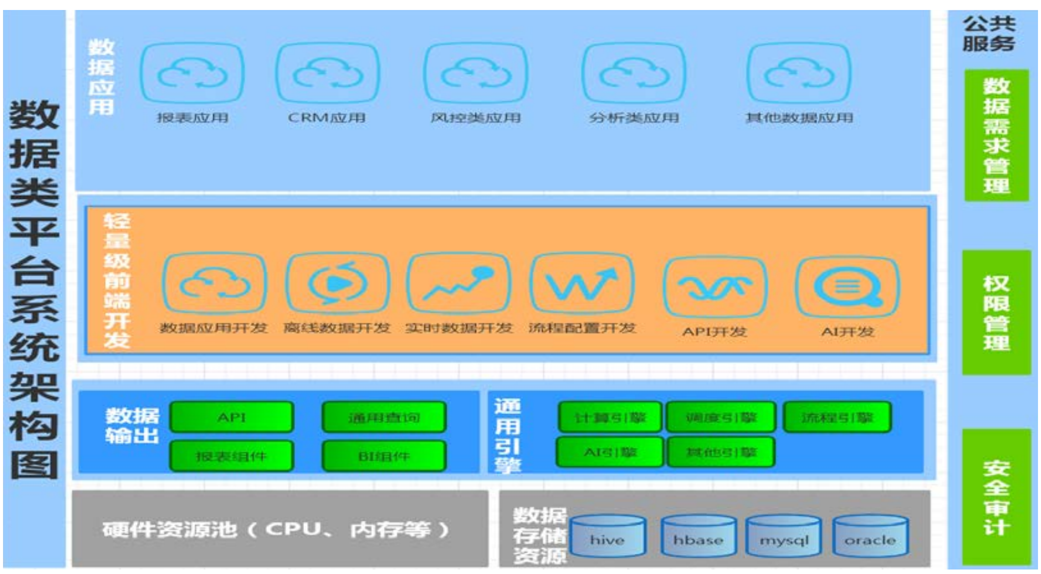


图 11：数据类平台架构设想

门和人员进行合理的数据授权，有效缓解了后台技术人员的压力，也极大的提高了前台业务人员获取数据的灵活度和自主性。

4 未来计划愿景

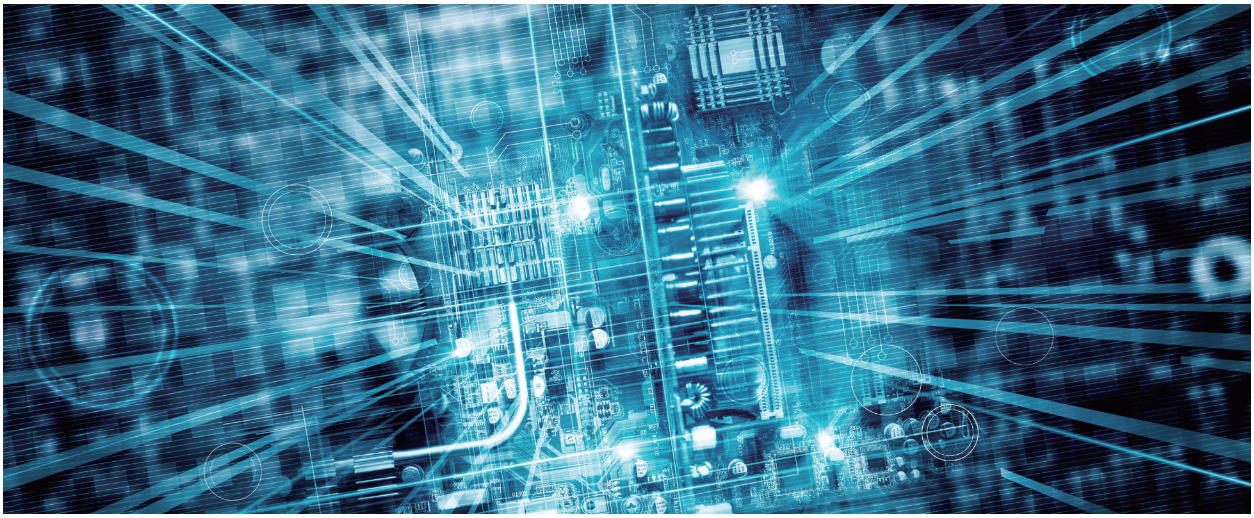
让数据需求开发“更规范，更高效”，数据使用“更智能”，公司持续强化数据中台的数据服务能力，通过引入和提升包括计算、调度、流程、AI 等引擎的基础能力，让数据类服务能够更加快速得到响应。公司将继续从两个广度和深度两个维度继续开展数据中台建设工作。

当前，我们正在建设离线数据开发平台，规划把离线数据的开发全部迁移到一个平台上管理，把现有不同的数据开发模式，多样的开发工具统一起来，采用统一的规则进行数据开发和版本管理，期望降低数据开发人员的技术难度，提高其开发效率及开发质量。同时我们也在数据中台中规划建设“数智中台”，计划在数据中台上引入数据智能开发和使用模块，让数据更深层次的价值被发现处理，为公司数字化转型提供更强有力的数据工具。

另一方面，我们要把积累的知识和工具用深、用透。前期由于知识受限，人员紧张，我们的主要精力是把各种基础做起来，现在在人员和技术上都有了一定的积累，在补短板的同时我们开始把重点转向精品化、产品化方向。如在数据治理知识和工具积累的基础上，我们计划建设数据资产门户，把数据相关的知识、工具、成果都作为产品放到资产门户中可以让用户便捷的查看和使用，为业务人员解决“数据在哪里、怎么用”等问题，实现数据需求、数据开发、成果发布、数据使用的一体化。我们将进一步对数据中台中的数据进行模型化、精加工，让数据更贴近用户需求，降低数据应用类系统的研发成本和研发周期。我们要把数据中台做厚做实，通过把更多地基础数据服务往中台上迁移，让数据类应用类系统的轻量化。

东方证券研发运行一体化平台探索与实践

樊建、赵泽 / 东方证券股份有限公司 系统研发总部 上海 200120
E-mail：fanjian@orientsec.com.cn、zhaoze@orientsec.com.cn



为满足证券业务需求日渐频繁和行业数字化转型的要求，自研及敏捷开发模式已经开始在证券科技部门进行了大量实践。技术模式改变的同时，对传统的科技项目研发与管理模式带来了巨大挑战。为此，东方证券基于 DevOps 方法论及自身实际情况建设了研发运行一体化平台，制定了需求接受、设计规划、版本任务、管理审批、生产交付的完整交付流程体系，纳管了项目整个生命周期各阶段标准和数据记录。打造了持续集成、持续交付的自动化流水线，实现了 IT 各岗位工作的自动化流转，增强了团队协作能力，从而快速提升需求交付效率及研发质量。

1 背景

为了推动业务创新，从而带来更高质量的金融服务，证券行业正在进行大规模的数字化转型，随着客户服务的线上化发展，各种个性化和定制化的需求都在推动产品研发的快速迭代，为此，证券行业纷纷引入敏捷开发模式解决个性化需求响应速度慢的难题，但敏捷型开发提高研发迭代速度的同时也带来了新的挑战：

1）研发管理标准，历史应用维护更新与新建项目并存，供应商模式与自研模式并进，不同项目开发方式与技术栈差异巨大，导致产品交付缺乏统一尺度的衡量标准，给交付质量和团队效率的提升带来极大难度；

2）基础设施建设，IT 部门自身整体规划和基础设施配套建设并未全面跟上技术的革新，滞

后于团队需求，导致各团队自行建设工具链，重复建设并互相隔离，缺乏统一基础平台使得规范无法落地。

3）岗位协同手段，在《证券基金经营机构信息技术管理办法》等合规风控要求下，岗位职责、办公地点、工作环境等各方面的隔离极易产生信息孤岛。研发和运维身处不同的职能部门的情况下如何对齐目标；网络环境复杂且互不联通的情况下研发人员如何了解服务的生产运行状况；各地机房和分公司的数据在异地的情况下如何进行中心化管理。

为解决以上痛点，东方证券以 DevOps 作为理论依据，建设研发运行一体化平台。在支撑业务需求的过程中，通过不断实践进行快速反馈及迭代，完成技术能力和管理能力的数字化升级。

2 DevOps 概念与实践

DevOps 是一套方法论体系，其延续了精益、敏捷的理念，强调以业务价值为目标，并打破原有团队间组织考核壁垒，进行统一化的合作和沟通。

其三步工作法的步聚为“以建立全局观和将工作可视化作为前提，实现从开发到运维工作的快速价值流动，实现问题报告后的快速反馈机制”，打破原有的组织考核壁垒实现团队间协作与沟通，将 IT 交付和运维服务团队统一起来，围绕一个统一的业务价值目标及业务交付范围加强沟通，通过频繁、快速地迭代交付和反馈，从而达到加快交付速度和提高交付质量的目的。

在东方证券 DevOps 落地过程中着眼于项目的实践，以贯通项目交付全流程的数据链路为目的，结合部门现状整合独立服务，统一自动化工具链。平台以流水线形式实现持续集成和持续交付的职能，并将交付标准化通过平台的流程管控落地。同时，从全局角度以可视化的形式，对研发效能过程进行观察和复盘，最终实现 IT 数字化管理与交付能力的提升。

3 架构设计

3.1 整体架构

软件开发生命周期中涉及多个阶段和多个岗位，使用不同的工具，既有信息同步的协作需要，也有权限隔离的执行要求。整体架构如图 1 所示，研发运行一体化平台以三层架构设计分离工具、功能与操作，既可以明晰角色权限边界，也可以灵活增减功能和更换底层工具。

1）界面层提供权限与环境隔离的研发前端与运维前端。其中，除运维人员外的其他角色通过研发前端提供的流水线功能完成开发、测试、审批工作。运维前端提供生产环境的各项操作功能，由运维人员单独使用，不同角色间以工单形式进行任务流转。

图 1：研发运行一体化体系

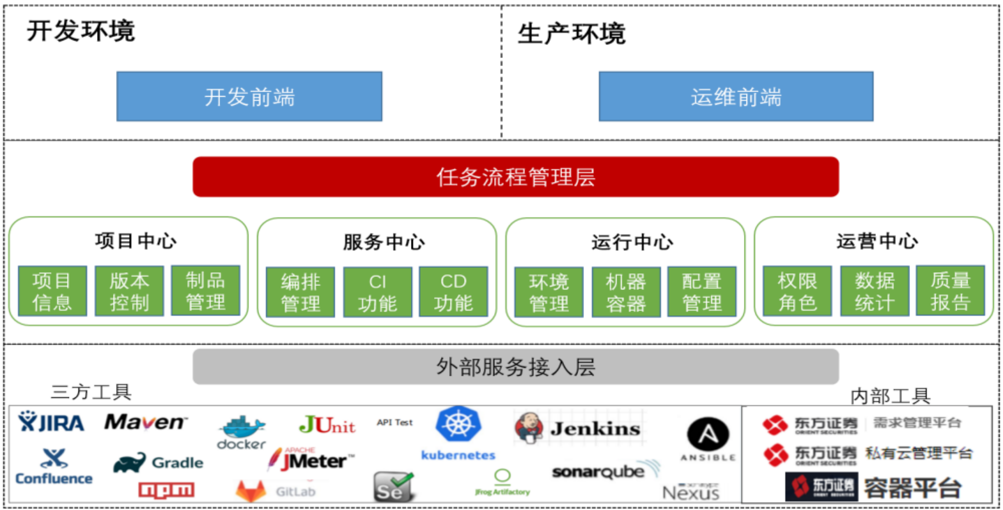


图 2：研发运行一体化平台整体架构

2）中台层设置不同功能中心管理模块，以流水线形式组合自动化功能模块，并通过任务流程管理层供前端调用。项目中心管理项目中的元数据如应用服务、团队人员、版本计划、制品状态等信息，对数据整合并保障跨环境同步；服务中心提供平台自动化功能，以流水线满足不同岗位工具需求；运行中心负责不同环境中机器、容器、中间件、实例部署的管理；运营中心负责生产发布的质量记录，平台使用的数据统计和平台人员、权限等设置等等。

3）资源层通过建设统一工具链，并纳管其它三方和内部工具，并通过以接口形式进行封装，屏蔽同类服务的差异，形成服务接入层，供中台调用，从而实现工具的快速接入和替换。

3.2 业务架构

东方证券研发运行一体化平台纳管需求发起到生产部署的整个流程，如图 3 所示，其主要业务功能如下：

1）项目配置，DevOps 平台以项目组作为逻辑上承接需求的最小单位。项目中具有从开发至上线所需的完备资源，为技术人员提供一站式的工具支撑。纳管任务目标、团队成员、职责分工、代码库、制品库、配置中心、自动化流水线、开

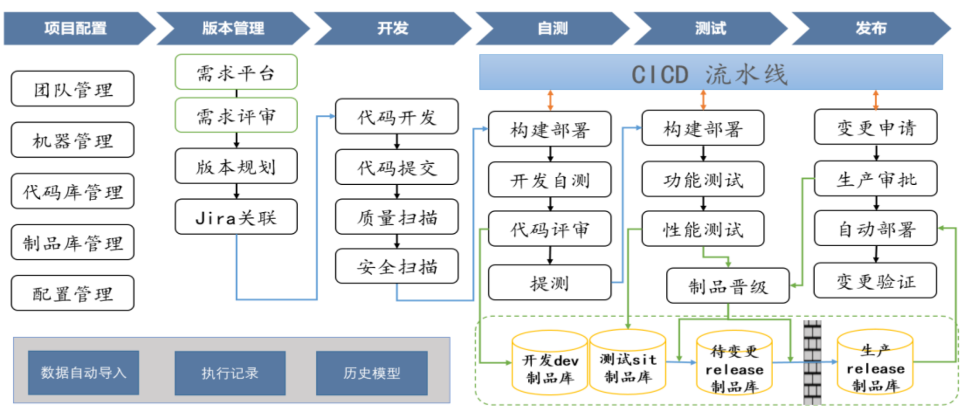


图 3：研发运行一体化业务架构

图 4：版本变更流程

发 / 生产环境的机器资源等等。

2）版本管理，版本代表一组待上线的需求，如图 4 所示，项目组成员依据版本对齐目标，该项目上线流程中各岗位产出的过程信息，无论是研发代码、质量数据，还是构建制品、测试报告都会按版本规则由平台统一存储。

3）开发阶段，研发人员完成提交后会触发代码质量扫描和安全扫描。通过流水线执行自动化构建和部署工作，在组内代码评审和功能自测通过后，通过提测单发起提测。

4）测试阶段，测试人员执行提测单包含的流水线进行构建，制品存入测试制品库并部署进行测试。测试通过后将制品晋升至待变更制品库，并通知项目经理发起变更申请。

5）发布阶段，变更审批通过后，制品由待变更制品库晋级至生产变更制品库。运维人员根据变更单信息核对版本号，执行流水线进行生产发布。

4 技术与功能创新介绍

在技术功能上，平台一方面为项目经理、开发、测试、运维各岗位提供全套自动化工具，实现从开发环境到生产变更部署的全流程自动化，需求开发结束后可在几分钟内自动化完成生产环境部署。

另一方面平台遵守开发 / 运维的环境、操作隔离与合规审批的规定，提供了开发端和生产运维端两个互相隔离的工作台，通过工单审批流将任务传递，以下进行具体功能介绍。

4.1. 研发端

研发端由产品、开发、测试等岗位使用，平台提供需求划分、版本建立、测试自动化部署、

图 5：研发端登录

图 6：需求管理

审批等功能，完成生产变更前所有工作。

4.1.1 需求管理

产品经理使用需求管理功能实现从业务需求到实际研发任务的转化，并起到追踪需求状态的作用。如图 6 所示，某项目根据业务需求建立了产品版本，业务分析师将该需求拆成两个任务分配给团队，项目经理根据各自的任务组织其项目版本开发和协调产品版本上线计划。需求方根据各项目版本状态变化可以追踪交付进度，多团队协作时能从该功能同步相关信息。

4.1.2 版本管理

项目经理使用版本管理功能进行团队内部开发任务跟踪。一个版本对应该团队的一次上线变更任务。如图 7 所示，项目经理会将 Jira 任务分配给开发人员，并将条目拖入该版本追踪开发进度。在开发结束后通过“提测单”或“变更单”推动版本状态改变，直至该版本所有内容完成上线。

4.1.3 开发流水线

实际研发及交付工作中需要使用众多的软件工具，平台以流水线的形式将不同软件提供的功能串连，提供一键构建、扫描、部署等研发必备功能。平台在流水线灵活装配方面做诸多改进，既能自动编排又能批量执行的流水线为技术人员提高整体工作效率带来很大便利。

1）流水线编排

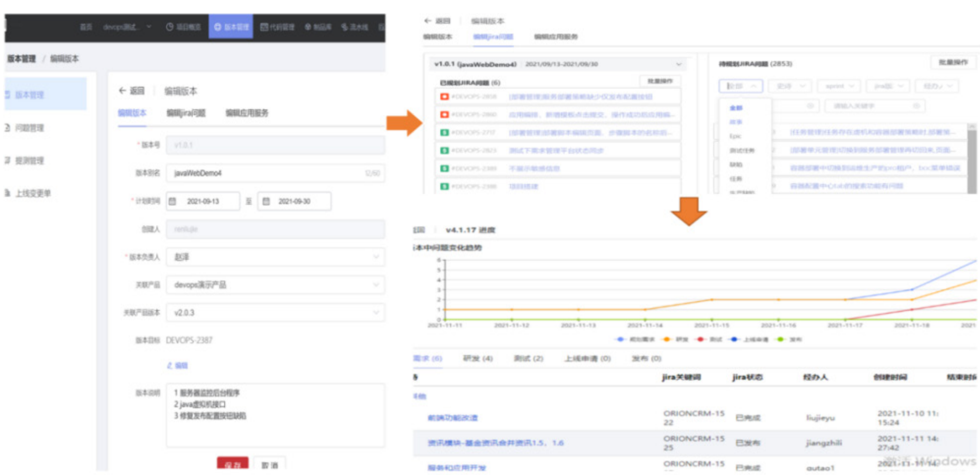


图 7：版本管理

图 8：流水线编排

流水线编排能力使得项目可以根据不同要求添加其所需模块，并对步骤进行自由组合。根据需要可以将多个服务按依赖顺序，在一条流水线中搭建同时执行串行、并行、混合的多种 CI/CD 任务，支持项目的不同要求。

在单流水线精细化定制的能力以外，对需要批量执行任务的项目提供进一步的扩展搭建功能。即在服务流水线的基础上继续组合，搭建多服务处理的应用级流水线。

2）自动化部署

平台将常用部署步骤模块化，在服务搭建部署脚本时可根据类型自由组合。用户通过简单勾选所需步骤和修改步骤参数可以快速生成所需脚本。并使用 4.2.1 节所述部署单元选择部署集群，并可选择灰度、串并行等不同发布策略。

部署执行过程由平台同步输出日志，运维、研发人员可随时观察部署状态去分析、解决问题。平台提供丰富的自动化功能减轻运维人工操作，比如健康检查步骤通过端口和接口调用来完成基本的验证工作。比如环境检查可以对新机器的部署环境进行检查和软件安装等。

4.1.4 测试管理

1）自动化测试

依托东方证券自动化测试平台，流水线中提供多种测试模块供项目组使用。实现了代码提交 -> 构建 -> 部署 -> 测试的全自动过程。研发人员提交代码后可自动执行回归测试或选择测试案例执行测试计划，极大的提升了研发效率和产品质量。

2）提测单

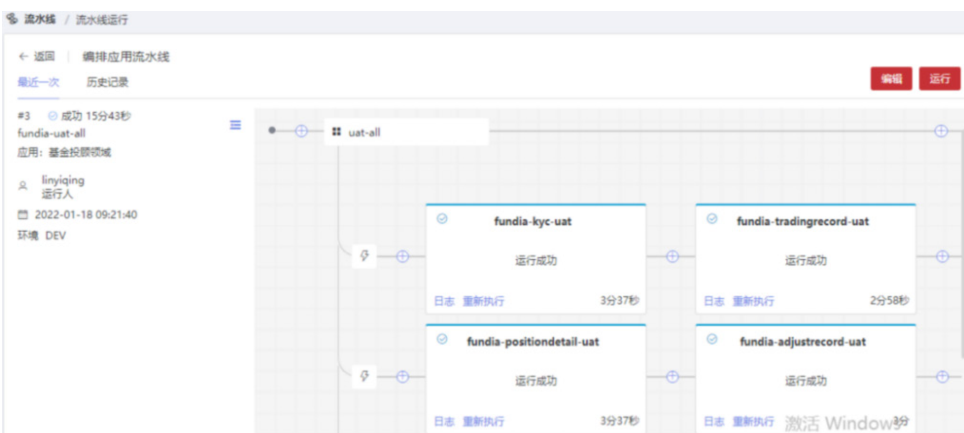


图 9：应用级流水线

图 10：模块可选的自动化部署

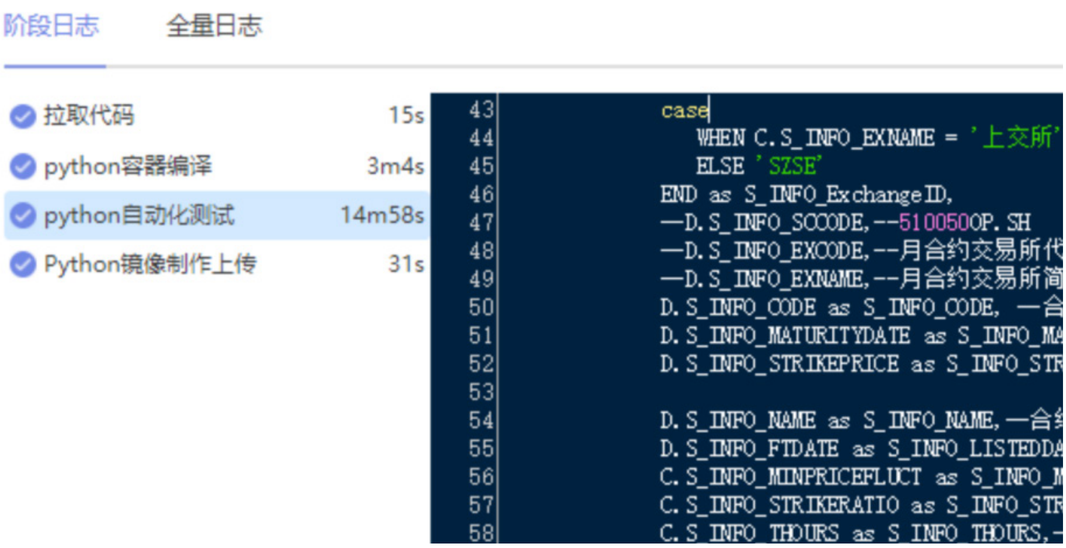


图 11：自动化测试

开发任务结束后，开发人员通过提测单将任务流转给测试人员验证。如图 12 所示，测试人员根据提测单中信息，执行流水线一键完成测试环境的搭建和部署。完成测试的制品和测试报告会由测试制品库晋级至预发布制品库，至此测试阶段结束。项目经理发起的变更单审批通过后，测试验证过的产品被推送至生产制品库。

4.1.5 制品管理与晋级

交付过程中产生的制品包、镜像、配置文件、测试报告、依赖包等作为各岗位的产出物需要进行验证、审批与记录保存。在东方证券研发运行一体化平台中，如图 13 所示，使用制品库作为唯一信息源，并为各项目提供开发、测试、预发布、生产发布四个制品库，通过制品晋级制度分岗位管理。制品通过流程控制，由开发向生产发布制品库一步步晋级，在交接节点进行自动的安全检查。最终审批通过后制品传至生产制品库节点并永久保存。

统一信息源多节点分布模式如图 14 所示，对制品进行全流程管理，在保证合规、安全的前提下极大地提升了生产上线的自动化水平，杜绝以往不规范操作可能产生的风险。此外，该模式具有可扩展性，以往因为距离、网络隔离原因无法统一管理的外地机房部署数据，可以按此模式

图 12：提测单

图 13：制品晋级

接入后由总部中心管理。

4.2 运维端

运维端为运行人员于生产环境专用，提供自动化部署、运维监控、运维操作。项目上线变更审批通过后，运维人员在运维端根据任务单执行生产部署。

4.2.1 配置与机器管理

平台提供多环境多集群的机器和配置管理。将机器归属于服务后，根据该服务不同环境的配置文件，使用部署单元概念将机器与配置组合管理。提供跨应用公共配置与服务私有配置管理，文件形式与热变更 K-V 形式等多种特性功能，满足各项目组技术要求。

4.2.2 实例管理

实例管理功能提供在各环境服务实例的查看和操作能力，查看功能包括实例状态与日志使得开发人员能够在问题发生时快速针对错误信息进行有效修复。对运维人员提供启动、停止、重启、备份、回滚等操作功能而无需远程登录机器。

备份和回滚对生产部署的安全性有很大帮助，平台提供自动、手动，部分备份、全备份等

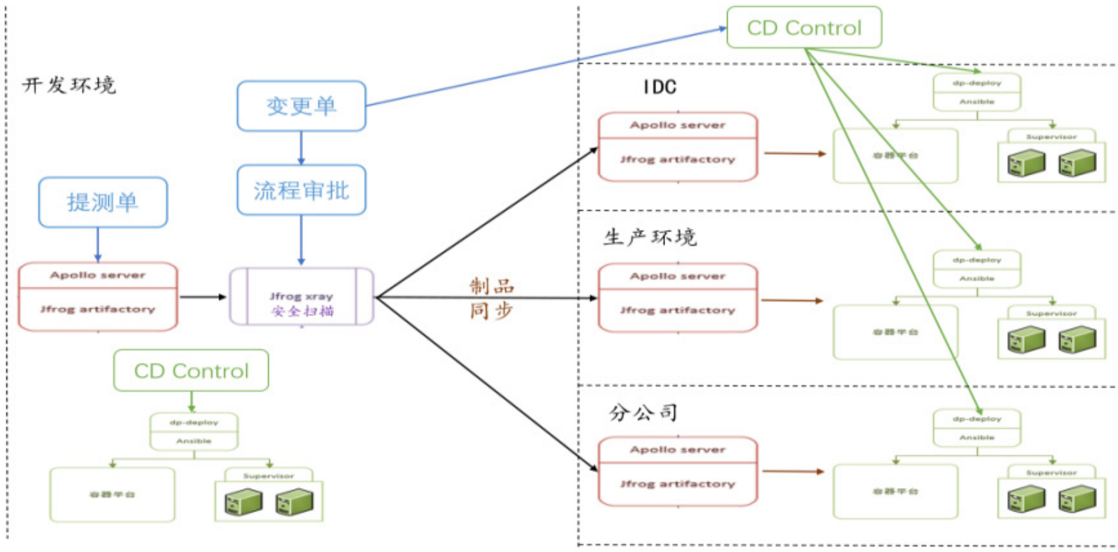


图 14：生产环境多节点分布

图 15：生产运维登录端

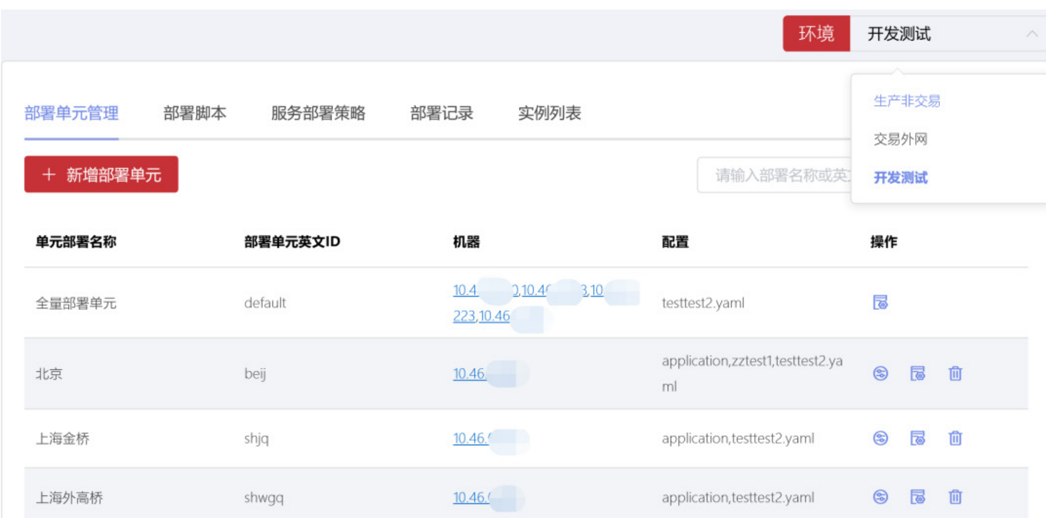


图 16：部署单元

图 17：实例操作

不同备份机制供用户选择。当遇到部署失败时，可以根据情况快速回滚至稳定状态。

4.2.3 生产变更

项目经理通过变更单将流水线和制品关联提交审批，审批通过后运维人员会在生产环境操作台收到任务单如图 18 所示。

运维人员进入任务单对提交信息进行校验，确认版本、机器、制品等信息。检查无误后使用变更单中的流水线一键部署，出现部署失败情况可按照 4.2.2 节所述进行回滚至变更前的状态。

部署操作完成后需确认服务执行正常，各项指标监控正常，返上线成功结果通知研发人员与需求发起人进行验收。

4.3 发布审批

依托上述研发全链路的自动化功能，平台整合了研发过程中的项目数据，通过提测单、变更单的流程将交付数据提交给交付审批流。

项目负责人、合规和管理人员在审批单中可以查看该版本需求内容、变更服务、质量情况等信息，从而判断该版本是否符合上线条件。驳回的版本修正后可重新发起变更审批。通过审批的

图 18：变更任务

图 19：执行反馈

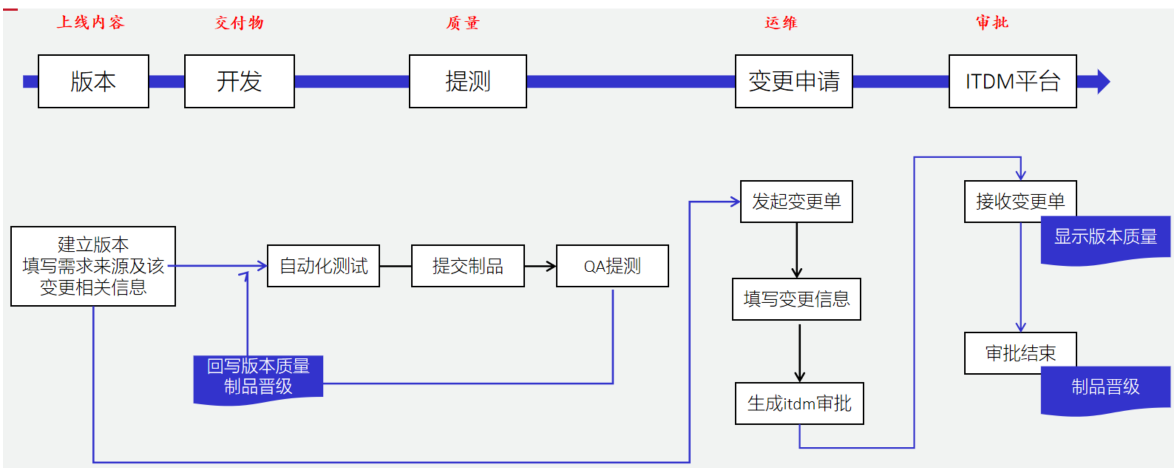


图 20：流程审批

版本制品会被推送到生产制品库保存。

5 实践效果

研发运行一体化平台打通了研发、测试、审批、生产部署的过程，在交付的过程中各岗位的任务和交付物都经由该平台流转，天然具有将规范制度落地的优势。由于其在自动化和标准化上的优势，上线后在东方证券使用比例逐渐上升。

当前，共接入各类项目 63 个，服务数 654 个，流水线数量 1638 条，流水线执行 4 万 5 千余次，生产发布 474 次，开发部署 3 万 6 千余次，将代码编译、测试到生产发布的整体流程时间从小时级降低到分钟级，实践成果非常明显。

6 总结

本文探讨了 IT 数字化管理的关键技术，并详细介绍了研发运行一体化平台在东方证券的建设成果与实践经验。东方证券制定企业架构转型目标战略，旨在通过架构转型为公司科技工作的长远发展打下坚实基础。作为企业架构转型的核

图 21：流程审批单中 DevOps 质量报告

心组成部分，研发运行一体化项目的建设，是公司提高金融科技核心竞争力的重要突破。研发运行一体化已经应用于东方证券业务中台（财富中心、交易中心、账户中心、产品中心等）、东方赢家 APP、证投业务管理系统、投资者服务平台、统一营运平台、固收债券业务扩展系统等数十个核心项目，完成研发工具链的统一纳管与升级，制定了需求接受 -> 设计规划 -> 版本任务 -> 生产交付的规范研发流程体系，并通过平台落地研发标准规范执行，从而保障并加速研发整体生命

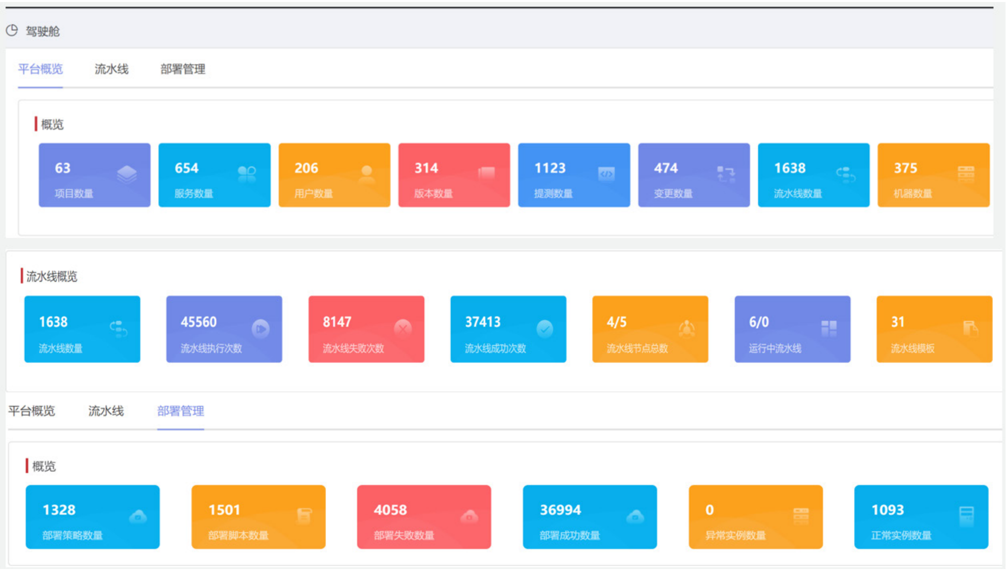


图 22：运营数据

周期，大幅提高交付效率与交付质量，随着平台生态的不断优化和发展，未来将在内部全面推广，服务于更多产品线和用户，为公司数字化转型做出更多贡献。

浅议数字化转型下如何做好信息科技的需求管理方法

彭辉 / 海通证券股份有限公司

本文针对数字化转型下，入额做好信息科技的需求管理进行阐述，首先梳理券商机构的需求现状，论述了有业务部门对接的需求，无业务部门对接的需求，创新业务的需求种类。针对监管需求、业务需求和信息科技部门现有的资源之间，存在缺口不能满足的矛盾，提出针对需求管理办法，重点论述了引进项目管理工具、建立符合实际情况的需求制度、建立需求跟踪制度、建立制度修复机制等管理办法。指出需求管理制度推进注重形式且不可流于形式，根据公司的规模灵活对待，大公司与小公司区别对待的需求规范制定和灵活的管理办法。

一 券商机构需求现状

近年来，数字化转型下，如何做好加快落实业务需求与信息技术等双向融合相关工作，首要工作要做好需求管理。尤其是科技发展及监管业务的爆发，券商信息科技部门人员逐步扩大，整个项目的生命周期进行流程化、正规化管理，逐步提高业务需求立项的门槛，形成需求说明书模板，要求业务部门把需求应用、需求场景、需求功能、需求接收、需求分管领导、需求验收等流程进行梳理和撰写，需求流程也逐步规划化，有甚至要求业务部门把表和字段都写在业务需求流程上。这样会造成业务部门提需求畏惧感，能不提的需求尽量不提，业务部门自己解决的，就自己解决。

信息科技部门的需求流程规范化对业务部门提需求也提高门槛，也会对业务部门造成“误伤”。信息科技部门与业务部门之间逐渐形成“裂缝”，很难形成合力，较难让技术更多赋能业务。

尤其是券商机构投入很大程度的自研系统，需要较多即懂技术又懂业务的复合型人才，也需要业务部门投入较多的时间来参与整个项目中来，“人和时间”很难保证的情况下，针对一些系统的需求专业化和详细化也面临考验。

由于券商的投入人员逐步增多，一些大的机构的信息科技拆分多个部门，各自部门的制度和人力投入重点的不同，面临跨部门的需求，协调和沟通也会遇到一些阻力和障碍。整合跨部门需求有时需要分管领导来协调资源的。导致一些项目面临要么被反复修改、耗时耗力，要么被迫“流产”、无疾而终。在业务部门看来，“同一个信息科技”视角下，都是信息科技部门带来的需求周期时间长。

信息科技部门与业务部门的协助，往往是一半汗水，一半苦水。但当前一个状态是人力投入仍是不足，疲于应付状态。主要是因为涉及需求不能满足的问题。

二 需求种类

目前券商机构的需求，有监管需求和业务需求等两大类，但从需求业务提出部门来分，以券商自营系统为例，自营系统的相关的需求主要分为三大类，一是有业务部门对接的需求，二是无业务部门对接的需求，三是创新业务的需求。具体如下图 1：

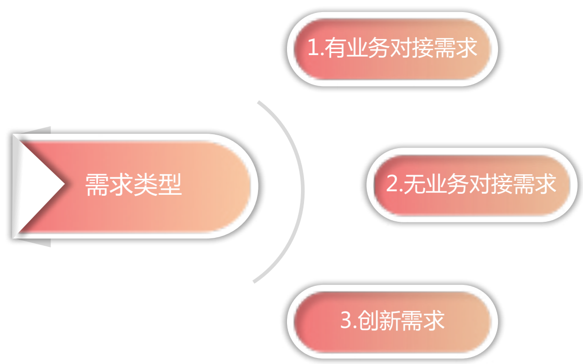


图 1：需求类型的分类

2.1 有业务部门对接的需求

针对有业务部门对接的需求，也有以下几种情况。

第一业务部门一句话需求。很长一段时间，都是业务部门一句话，我做这个功能，电话沟通后，就开始做了，尤其是自营系统，标准化产品，市场上自营系统比较集中化，多家需求提出、测试、上线，很大程度上冲淡这种一句话的需求风险。但随着 IT 的流程化建立，需求文档填写内容复杂而费时，需要业务部门投入较多的精力和时间，获得感很差，这是业务部门抱怨之一。

第二针对自研产品的需求。自研系统是目前一个趋势，把握核心代码，以防出现系统性风险，但是系统自研要求之高，一是需要业务部门既懂业务又懂 IT 的复合型人，二是复合型的人有时间配合业务部门投入进来，配合信息科技的人一起分析需求和需求验证等整个流程，而这部分时间投入对于业务部门来说，短期看收益不大，业务部门还是以“业绩论英雄”。曾经一个“业务和 IT 的复合型人”说，“我投入 IT 很多，但是我业绩不行，我也是要被领导批评，甚至于要走人的风险，我需要更多的时间投入业务上来。”同时 IT 人员缺少既懂 IT 又懂业务的复合型人。很多业务部门长时间历史认为，IT 需求是信息科技部门的事情，业务部门审慎投入 IT 人员。

第三个性化需求。个性化需求一般是业务部门为了提高系统使用方便或业务在实际工作中遇到一些难题，而提出一些解决办法。这类需求基本是业务部门能梳理清楚，IT 配合完成，业务部门需求高，愿意积极推动 IT 部门来完成该类需求。这类需求一般要求快速完成开发，尽快上线的功能。针对这类需求，业务部门和信息科技部门通过面对面沟通比较效果，尤其是做好会议纪要，信息的纷发和各方确认，对于信息科技开发和业务部门后期的需求验收是个约定的载体。

2.2 无业务部门对接需求

针对无业务部门对接需求情况主要有几点：

第一是监管类需求，中登技术接口改造、交易所字段优化等需求，这部分需求一般情况下，涉及相关交易系统，是要求配合改造。一般情况下在监管需求下发一段时间，很少有业务部门主

动认领需求，更少于业务部门主动撰写需求，提出需求流程，并参与需求验收等过程。针对类似需求，业务部门认为是系统层面的需求，跟业务部门关联不大，即使该需求给业务部门相关，如何提需求，如何验收等都会摆在业务部门面前一个问题。同时有个需求改造很大，涉及关联部门较多，费用承担也是不可避免的事宜。尽量能规避而避之的心态。

但此类需求，系统不改，则很大程度业务受影响，甚至业务中断的风险。现状是这类需求是信息科技部门最先知道需求，也是推着该类需求朝前走的现状。

2.3 创新业务的需求

上述两类需求，基本是对接外购型系统，针对创新业务的需求对接，这类需求是往往发生在自研系统上，信息科技针对行业竞品进行梳理，进行优化系统功能。这类需求是信息科技主导，业务部门一般也不确认，针对此类需求，有时会给业务部门眼前一亮的感觉；有时会面临信息科技自主开发。高密度响应开发出一堆功能后，会面临一些功能没人管、没人用，逐渐变成一个个“小土堆”、“小烟囱”。业务部门不用的“自嗨”尴尬局面。

如对公司具有战略意义，未来新技术的创业业务，信息科技部门对自研系统的创新业务需求，有时没需求，制造需求也要上。可以适时投入较小人力，不放在需求管理范围之内，进行孵化式创新，允许差异化管理，作为“期货的远期”来管理。但自研系统也是差异化自研，作为战略，分而计划，量力自研。

三 建立需求管理制度

对于信息科技部门来说，建立需求管理制度，是迫在眉睫，但制度推进第一步注重形式且不可流于形式，因为注重形式而不注重内容，将是更大的人力资源的浪费。尤其是近年来监管需求、业务需求和信息科技部门现有的资源之间，存在缺口不能满足的矛盾，将会存在一段时间。建立

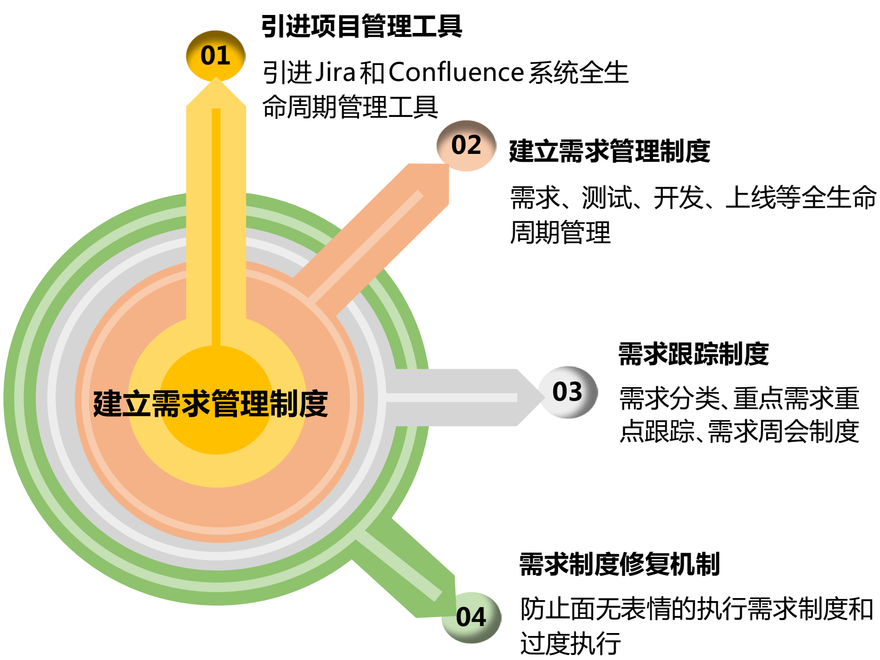


图 2：建立需求管理制度

需求管理制度一般涉及为引进项目管理工具、建立符合实际情况的需求制度、建立需求跟踪制度、建立制度修复机制等。具体如图二，建立需求管理制度。

3.1 引进项目管理工具

信息科技加大投入，尤其是大的券商或机构，需求管理及项目管理亟待提上日程，有的券商先行一步，引进 Jira 或 Confluence 之类的管理工具，从而建立需求制度及版本发布全生命周期管理工具。结合项目实际情况，逐步推进，逐步提升这个项目管理，提高自己管理项目能力，提高团队作战能力，是每个项目经理要不断挑战自己的事情，也是不断进步的动力。

3.2 建立需求管理制度

建立需求管理制度，要进行需求流程规划化和灵活化。需求制度建立后，不再是吼一嗓子，就能承接需求、测试、开发、上线了，而是要靠制度和流程来取约束和管理，这也是必然之路。但由于实行的过程中，各业务部门的 IT 投入多少、重视程度等不同，需求制度推行过程中会遇到很多“不舒服区”，比如推动业务部门写需求文档，这本身对业务部门来说是个不小的挑战，好在实行过程中，逐步完成公司的需求。

需求制度的建立，项目经理或部门经理是执行者，而需求制定的管理“顶层架构”设计，一般是出于咨询公司的意见，来结合机构实际情况来推行。对于一个项目是采用敏捷管理还是 CMMI 管理，取决于三点，一是 IT 公司的负责人，对这个认识和理解。第二项目团队具体负责人对对这个认识和理解。第三才是这个团队开发的项目是快速迭代性还是稳态性，是自主开发还是外购性等这三点。

3.3 建立需求跟踪制度

建立需求跟踪制度，首先要根据需求类型进行区分，找到业务主管部门，针对不同的需求，一般分为，监管需求、重要业务需求、跨部门需求、业务需求、内部需求、跟踪制度。

监管需求，一般为沪、深、北交所；中登结算公司创新业务和技术接口改造的需求。有业务需求必须柜台系统及周边系统需要改造的需求。这类需求大部分有详细的仿真测试、全网测试、系统通关、业务通关的时间点。如不进行系统改造，会面临业务中断的风险。

重要业务需求，为业务部门针对具体某项业务而提出的需求，这类需求一般是比较重要，还有时间限制。

跨部门需求，主要是在需求管理中，需要领导协调多部门，多系统同步上线时间，以防造成针对同一个需求，A 系统上线了，B 系统还没有上线，导致整个需求链条上不一致的情况。跨部门需求可以是监管需求，或业务需求、内部需求等。

业务需求是指一般业务部门提出需求，重要性不是那么急的需求。

内部需求是指信息科技部门 A 部门提给 B 部门的需求。一般是内部协调性的需求。

跟踪制度是指，需求流程建立后，要制定需求监督和提醒机制，需求跟踪制度，可以通过部门会议一个重要组成部分，需求跟踪制度在会前请项目管理人员更新需求进度，以及在周会中提醒需求版本计划，和需求存在的时间，如存在一定的时间将标记风险，促使项目经理更快得完成该需求的管理和版本上线等。

建立建议需求跟踪制度意义，主要是让项目经理和部门经理、部门负责人更好的需求的量化管理、需求时间的维度管理，作为监督需求和激活员工对需求的完成力度的把控和促进、监控。

3.4 建立制度修复机制

作为信息科技的需求规划者与相关干系人，设计相关流程与制度或采取哪种管理模式，是需要人力资源和时间资源投入作为“充足的弹药”，

来保证执行过程中坚强的战斗力和极强执行力，进一步将流程、制度的实践化，内容化。

但制定太多需求流程、太细的规则，往往会造成流程管理的困境：第一“面无表情”地执行相应制度，确保程序正确，以此来降低执行者的压力和监督者的责任，无疑就成为大家认为理性的选择。第二在管理过程中，逐步细化，逐步繁琐，而专业性太强的执行者往往看不到作为管理者该有的层次和格局，做起事情来放不开手脚。遇到突发事情，过于追求流程正常导致灵活性不足，效率低下。第三很多流程建立，缺少回头看的机制，建立需求管理制度的初心是为了提高整个项目质量和快速上线等。而不是为了选择某种管理模式而选择。

需求制度的规划者与相关干系人，心中常问问自己：该项目实施后会带来哪些负作用？（这一项也是目前大家极为忽视的）。给大家带来多少难点和不便。“知屋漏者在宇下，知政失者在草野”，作为信息科技的的规划者和执行者，一定要多调研，俯下身，听真话，执行需求制度修复机制。而不是流程百分之百正确，数据百分之百完美，系统百分之百有问题的流于形式而“荒腔走板”。

四 小结

目前整个行业有券商机构有 140 家左右，针对大中小不同的券商，信息科技投入的资源也不一样，在数字化转型的当下，选择需求管理的方法也不一样，如只有 50 个人的券商信息科技部门要学习头部优秀券商信息科技规划、需求管理之路，但如全“照搬照抄”，则会怨声载道，流于形式。而一个 400 人的信息科技部门，没有一定的需求管理制度，上层管理看不到具体的需求管理、问题单管理，不知道如何进行决策与规划，也会出现混乱情况。

同时，券商的信息科技部门的需求管理方法，不仅跟人力资源、资金投入相关，还跟目前该信息技术部门所处的阶段不同；还要跟业务部门发展情况有关。如信息科技部门规划太超前，业务跟不上，业务部门与信息科技投入不协调，科技赋能也是会大大打折。而业务发展很好，信息科技部门投入较少，也会拖业务部门发展的“后腿”。

可以确定的是，不论大券商还是小券商，都要进一步加快落实业务需求与信息技术等双向融合相关工作。做好业务和科技的两个轮子，共同打好公司高质量发展的数据之路。


CONNECTION
ANALYSIS
DATA
SEARCHING
VERIFICATION
CODING
SENDING

探索与应用

一种面向期权做市业务的FPGA快照实时合成方法与实现

刘垚、林子豪、马龙、杨树勇、张晓亮 / 中信建投证券股份有限公司 信息技术部 北京 100010
E-mail：liuyaodcq@csc.com.cn

期权做市业务是一种高频交易业务，对行情的获取速度有极致的追求。由于数据量的差异，交易所股票快照的更新速度通常无法匹配期权快照的更新速度。目前，上交所和深交所的 Level2 行情中都支持了逐笔订单行情，为高频交易客户在行情接收端自行合成快照提供了可能。本文对比分析了 3 种不同实现方案的优缺点，针对期权做市业务的需求特点，选择最低延迟方案进行了 FPGA 实现。文中给出了 FPGA 实现的具体思路，并给出该实现的延迟和吞吐量等性能指标。本文的 FPGA 实现方案在延迟性能上有明显优势。

1. 背景

高频交易是由强大的计算机系统和复杂的运算所主导的证券交易业务。为了争取竞争优势，证券公司必须持续不断地提升自己的系统性能，以寻求在激烈的竞争中赢取更多的获利机会。期权做市业务是一种典型的高频交易业务。行情获取是期权做市业务最重要的关键路径之一，需要持续不断地进行优化。

期权做市业务需要同时获取期权和标的物 2 种行情快照进行分析。通常，交易所提供的期权快照的更新频率是 500 或 250ms/ 次，而标的物

快照的更新频率一般不小于 3s/ 次。标的物快照行情的更新频率无法匹配期权做市业务对行情的要求。

中信建投证券是上证 50etf 期权和沪深 300etf 期权的主做市商之一。由于目前这两种期权都有对应的股指期货，股指期货行情的更新频率能够与期权行情相匹配，而且期货行情能够在一定程度商反应现货的价值，因此我们可以折中地使用股指期货行情作为主行情，而现货行情仅仅用作辅助和参考。但有一个问题是：在毫秒的测度下，期货与现货价格有一定的概率出现背离，影响定价的准确性。为了解决现货快照更新频率慢的问题，业务部门提供出利用 FPGA 技术在策略服务器端自行实时合成快照的需求。

2. 业务需求提炼

期权做市业务的策略系统需要从行情中提取的信息包括：证券代码（SecurityID）、时间戳（DataTimeStamp）、交易状态（TPC）、最新价（LastPrice）、成交量（TotalVolumeTrade）、成交额（TotalValueTrade）和买卖 5 档盘口，等等。

结合业务需求和 FPGA 片内存储特点，我们定义的合成快照的数据结构由图一所示的 26 个 64 位数据组成。

待合成快照的每一个字段都可以通过交易所提供的 level2 逐笔类行情直接获取或通过计算获取。我们在表一中列出每个字段的合成方法。

在快照合成过程中大部分字段的获取过程是比较容易的。最大的难点在于 5 档盘口的合成过程。虽然最终合成的快照只需要买 / 卖 5 档行情，但是我们必须在内存中记录下一天中所有出现过的档位。这是因为：一天中排在买 / 卖前 5 的价格是会时刻变化的。对于一只交易活跃的证券，一天中会出现超过千档的买 / 卖价格。为了保证快照合成的准确性，我们必须记录下所有出现过的盘口信息，并且实时维护所有档位的更新和排序，才能保证在各种极限情况下都能正确地获取前 5 档盘口。

字节位	7	6	5	4	3	2	1	0
0	DataTimeStamp				TPC	8'hcd	Market	MessageType
1	SecurityID							
2	TotalVolumeTrade							
3	TotalValueTrade							
4	LastPrice							
5	rsvd							
6	bid_price[1]							
7	bid_price[2]							
8	bid_price[3]							
9	bid_price[4]							
10	bid_price[5]							
11	bid_qty[1]							
12	bid_qty[2]							
13	bid_qty[3]							
14	bid_qty[4]							
15	bid_qty[5]							
16	offer_price[1]							
17	offer_price[2]							
18	offer_price[3]							
19	offer_price[4]							
20	offer_price[5]							
21	offer_qty[1]							
22	offer_qty[2]							
23	offer_qty[3]							
24	offer_qty[4]							
25	offer_qty[5]							

图 1：FPGA 合成快照的数据结构

表 1：目标字段的合成方法

目标字段	所需逐笔信息	操作过程
MessageType	--	固定为 0x00（表示消息类型为快照）
Market	--	固定为 0x00（表示上交所）
8‘hcd	--	标记表示快照为合成快照
TPC	TickBSFlag Type	若 Type 为 S（产品状态订单），用 TickBSFlag 修改 TPC；否则不操作。
DataTimeStamp	TickTime	用 TickTime 字段直接覆盖
SecurityID	SecurityID	用 SecurityID 字段直接覆盖
LastPrice	Price	用 Price 字段直接覆盖
TotalVolumeTrade	Qty	TotalVolumeTrade + Qty
TotalValueTrade	Price,Qty	TotalValueTrade + Price*Qty
BidPrice[n]	Price,TickBSFlag	若为新增买委托订单，则判断是否【插入新档位】or【在已有档位上的增加数量】；若为删除买委托订单或主动卖成交订单，则判断是否【删除已有档位】或【减少已有档位的数量】
BidQty[n]	Qty,TickBSFlag	
OfferPrice[n]	Price,TickBSFlag	若为新增卖委托订单，则判断是否【插入新档位】or【在已有档位上的增加数量】；若为删除卖委托订单或主动买成交订单，则判断是否【删除已有档位】或【减少已有档位的数量】
OfferQty[n]	Qty,TickBSFlag	

3 全档盘口的维护方案

我们以一支证券的快照合成为例，介绍三种常见的全档盘口的维护方法。

3.1 单向链表方案

单向链表方案的核心思想是：利用链表顺序进行档位排序。1 档放在链表的头部，末档放在链表的尾部。

单向链表的数据结构如图二所示。链表数据可以存储在片上 SRAM 或 DDR 中。链表维护需要用 3 个寄存器实时维护链表的状态，其中，最高价地址和最低价地址指示链表头和尾的地址，当前使用地址指示已经使用过的最新 SRAM 地

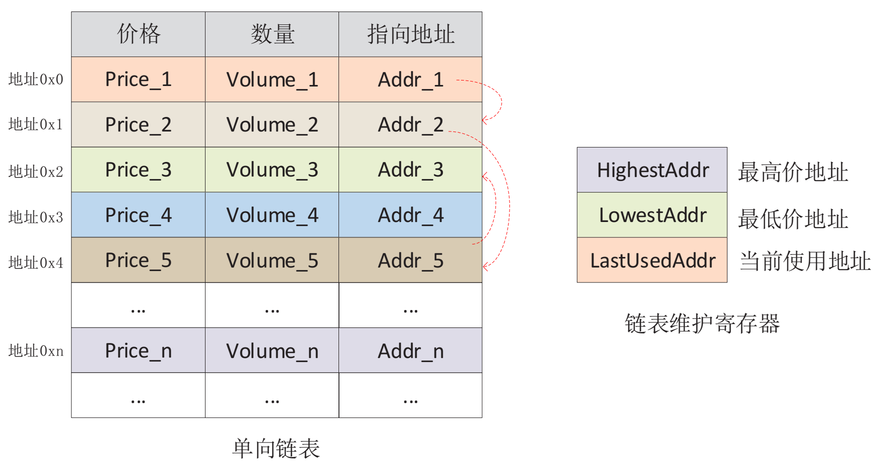


图 2：单向链表数据结构示意图

址。

FPGA 每收到一笔新增委托订单时，按顺序遍历链表中的价格，并同时进行价格比对。如果发现该价格已存在，则直接增加该档位的数量；如果发现该价格不存在，则将价格和数量填到 SRAM 当前使用地址的后一个位置（即 LastUsedAddr+1），同时修改链表相应的指针。

FPGA 每收到一笔删除委托或成交订单时，按顺序在链表中搜索与订单价格相等的价格档位，然后比较订单中的数量与链表中的委托数量，如果两者相等，则删除链表中该节点信息，修改链表相应的指针；如果订单中数量小于委托数量，则减少相应档位数量；如果订单中数量大于委托数量，则表示有乱序发生，需要延后处理。

五档盘口的输出只需要从链表头部顺序取出前 5 个档位信息即可。

3.2 AVL 树方案

AVL 树方案的核心思想是：利用自平衡二叉树的动态调整过程进行排序，二分法减少搜索次数。

AVL 树的构建必须满足每一个父节点的值大于左子树的值，小于右子树的值。树的构建过程中，如果左右侧子树深度之差（平衡因子）超过设定数值，则暂停接收数据，启动自平衡过程，通过“左单旋、“右单旋”、“左右双旋”和“右左双旋”四种操作将平衡因子调整到设定数值以下。

图三是使用 AVL 树维护档位的示例。FPGA 实现时需用 4 个寄存器实时维护 AVL 树的根节点位置、平衡因子、最高价位置和最低价位置。

FPGA 每收到一笔新增委托订单时，从根节点开始比较订单价格与树中的价格，如果小于树中的价格，则爬向左子树继续比较；如果大于树中的价格，则爬向右子树继续比较；如此往复，直到找到价格相等的树节点或者找到树末端，执行增加数量或插入子树节点的操作。

FPGA 每收到一笔删除委托或成交订单时，从根节点开始搜索与订单价格相等的价格。搜索到相等价格的树节点后，执行减少数量或删除子树节点的操作。

五档盘口的输出只需要从最高价或最低价位置开始反向爬树获取前 5 个档位信息。

3.3 地址直接映射方案

地址直接映射方案的核心思想是：采用面积

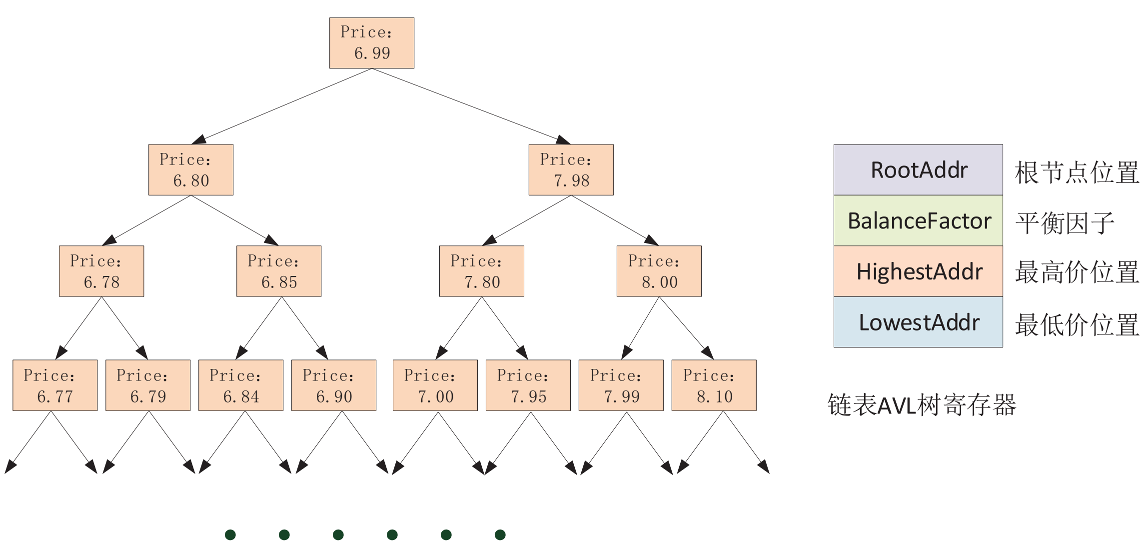


图 3：使用 AVL 树维护档位的示例

换速度的方法，价格与地址一对一映射，顺序隐含在地址中，无需排序和搜索。

由于国内交易所对证券的价格有涨跌停的限制，所有证券当天价格都有一定的范围。证券价格的最小单位是 0.001 元。为了保存一只证券所有可能出现的价格，占用存储空间必须大于等于可能价格的数量。比如，设一只股票的极限价格范围在 0~16.384 元，则至少需要预留 16384 个 64 位的存储空间。在记录档位信息的同时，需要实时维护前 5 档价格和最低档价格。

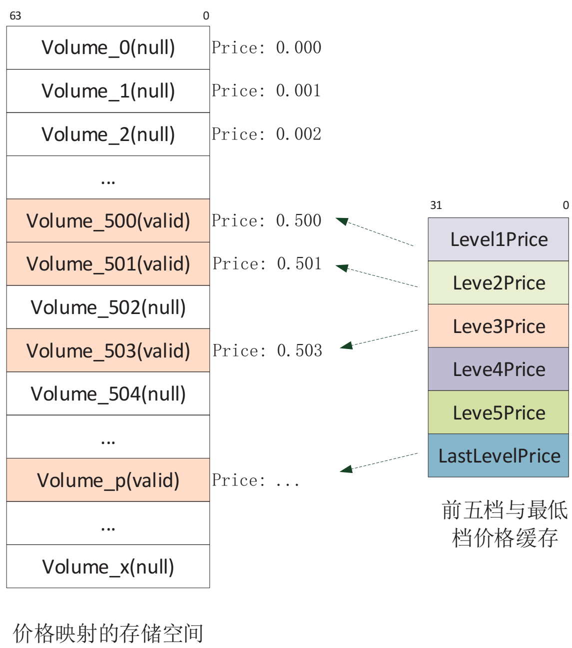


图 4：价格 / 地址一一映射的存储方案示意

3.4 方案的权衡

我们对不同方案权衡选择需要考虑延迟、抖动、存储空间、可扩展性和工程复杂度等多个方面的权衡。表二列出三种方案之间的权衡比较。

考虑期权做市业务是一种零和博弈，业务对延迟和抖动性能的提升会有不断的追求。方案首选应以性能指标为主，因此我们选择地址映射方案。

地址映射方案的问题是：存储资源耗费较大。考虑当前 FPGA 板卡一般都会集成较大的 DDR 存储，支持上千只证券的快照合成也是完全够用的。FPGA 片内存储不够时，可以用 DDR 扩展存储空间。与其闲置资源不如充分利用，我们不必为节省空间而折损性能。

4 FPGA 实现

本 FPGA 实现主要适用于上交所 level 2 行情的快照实时合成，其他交易所行情处理并不完全适用。

4.1 委托和状态订单的处理流程

由表一中列出的每个字段的合成方法可知，委托订单对合成快照字段的修改项包括：时间戳，交易状态和买 / 卖全档委托量。状态订单对合成快照字段的修改项包括：时间戳和交易状态。

表 2：三种方案的比较

方案	单向链表方案	AVL 树方案	地址映射方案
延迟	中等	中等	最小（best）
抖动	中等	中等	最小（best）
存储空间	中等	最小（best）	最大
可扩展性	较好（best）	较好	一般
工程复杂度	简单	复杂	简单（best）

图五是委托和状态订单的处理流程图，描述了 FPGA 收到一个的委托 / 状态订单后，对各字段进行操作的流程：

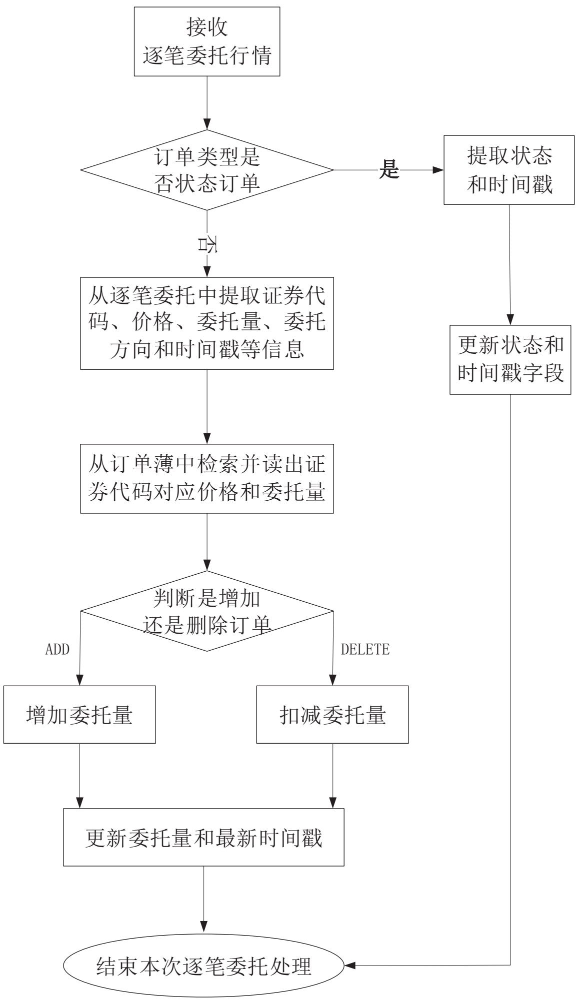


图 5：委托和状态订单的处理流程图

4.2 成交订单的处理流程

由表一中列出的每个字段的合成方法可知，逐笔成交信息对合成快照字段的修改项包括：时间戳、最新价、总成交量、总成交额和买卖全档委托量。

对于上交所集合竞价和连续竞价的逐笔成交行情，我们的处理流程是不同的。集合竞价阶段（9:30 以前），每接收一笔逐笔成交，需要同时修改买和卖双边档位。

连续竞价阶段（9:30 以后），每接收一笔逐笔成交行情，仅处理单边档位。选择处理买档位还是卖档位根据 BSFlag 标志判断。

图六是成交订单的处理流程图，描述了 FPGA 收到一支股票的逐笔成交行情后，对各字段进行操作的流程：

4.3 硬件架构设计

FPGA 快照实时合成的硬件结构如图七所示。根据图一定义的合成快照的数据结构，我们将 Order Book 分为 4 部分：Status book（用于保存图一中的前 2 行）、Trade Book（用于保存图一中 2–5 行）、Level Cache Book（用于保存图四所示前 5 档和最低档）和 Full Level Book（用于保存全档盘口）。其中，Full Level Book 占用空间较大，当片内 RAM 存储空间不够时，需要扩展使用 DDR 空间。

如图七所示，Order Book 每部分的存储空间都会有多个模块进行读、写操作，所以每部分内部都有一个多选一开关进行控制。

当 FPGA 接收到逐笔委托订单（含状态订单），由 OrderFixProc 模块和 OrderLvlProc 模块分别处理 Status Book 和 Full Level Book 更新。如果发现有新的价格档位出现，或者有现有价格档位被删除，则通知 LevelCacheProc 模块检查是否需要更新前 5 档和最低档。

当 FPGA 接收到逐笔成交订单，由 TradeFixProc 模块处理 Trade Book 更新，由 CallTradeLvlProc 模块处理集合竞价阶段的 Full Level Book 更新，由 ContiTradeLvlProc 模块处理连续竞价阶段的 Full Level Book 更新。如果发现有现有价格档位被删除，则同样通知 LevelCacheProc 模块检查是否需要更新前 5 档和最低档。

LevelCacheProc 模块随时接收通知，并行维护前五档和最低档的信息。

RestructOut 模块以条件输出或定时输出的方式从 Status Book、Trade Book 和 Level Cache Book

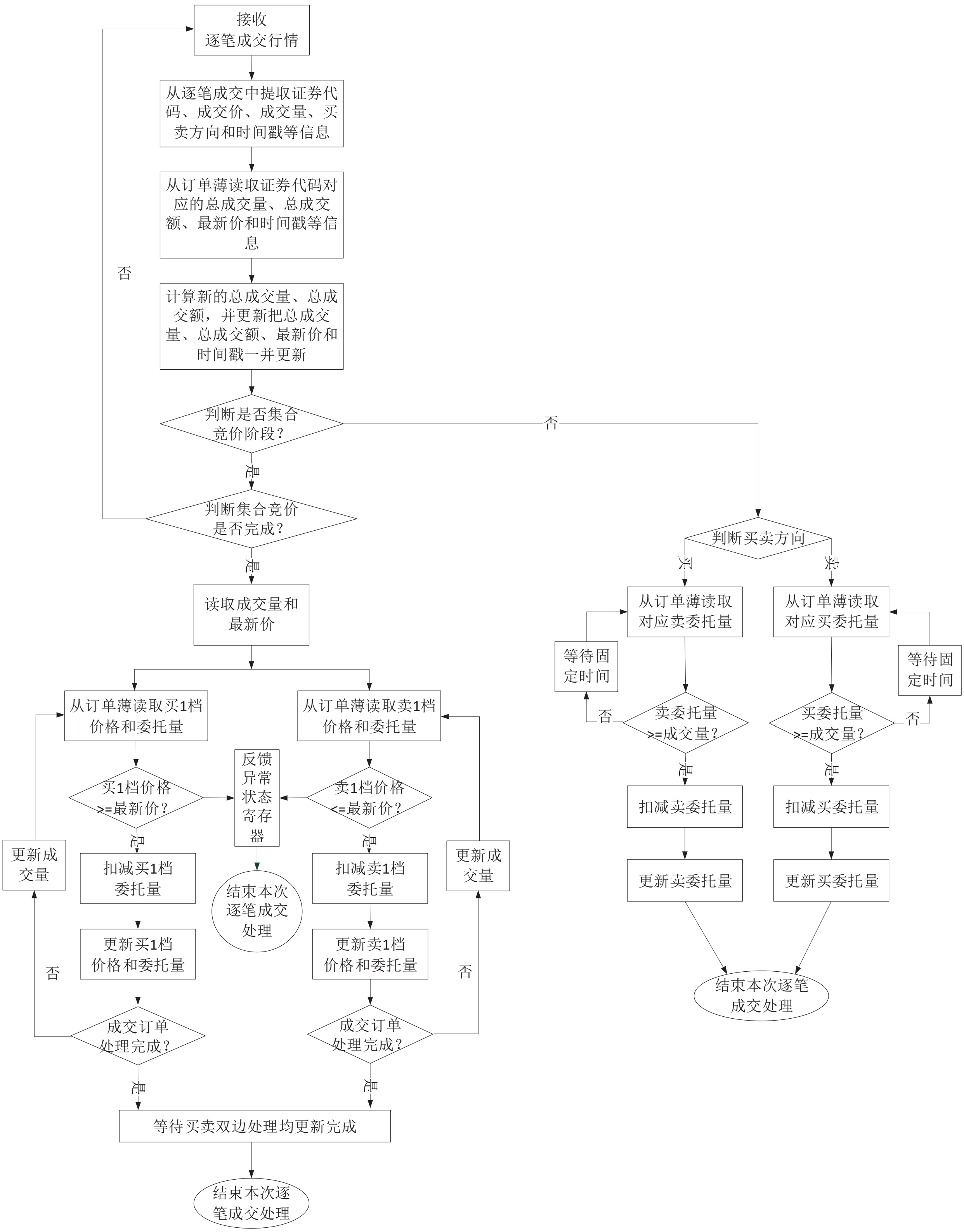


图6：成交订单的处理流程图

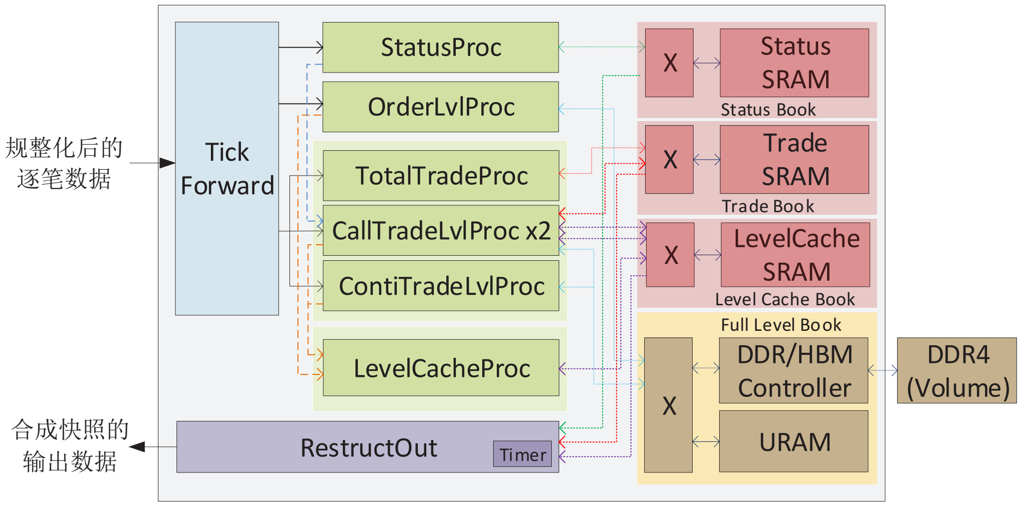


图 7：快照实时合成功能的结构图

中读取信息，将信息重组成图一中定义的数据结构，并持续输出给后级。

5 性能分析

5.1 延迟

快照合成的延迟主要分为四个部分：Order 处理、Trade 处理、Level Cache 处理和 RestructOut 处理。其中，有 2 个处理过程延迟是不固定的：

Trade 处理（集合竞价阶段）：为了保证合成的准确性，在集合竞价阶段，我们先等待收到所有成交订单，然后统一处理订单薄中已成交的档位。由于期权做市业务对集合竞价阶段的行情并不关心，因此此过程不在关键路径上。

Level Cache 处理：在前 5 档有档位全部成交（或撤销）发生时，延迟会有所增加，因为此时需要重新搜索第 5 档的位置。对于比较活跃的标的物，搜索一般很快就会得到结果，而且 Level Cache 的搜索过程与订单处理是并行运行的，阻塞传输的概率非常小。另外，搜索的延迟是有优化办法的：让 Level Cache 维护大于 5 档行情（如

表 3：快照合成模块的处理周期和延迟

延迟的组成		模块	周期	延迟	综合延迟
输入处理	Order 处理	OrderFixProc	11 拍	37 ns	43ns
		OrderLvlProc	13 拍	43 ns	
	Trade 处理	TradeFixProc	11 拍	37 ns	
		ContiTradeLvlProc	13 拍	43 ns	
		CallTradeLvlProc	不定	不定	
输出处理	Level Cache 处理	LevelCacheProc	21 拍	70 ns	63 -133 ns
	Restruct Out 处理	RestructOut	19 拍	63 ns	

6 档），则第 6 档也在 Level Cache 中，前 5 档更新不需要重新搜索，缺失的第 6 档在后台并行搜索。

除了以上 2 个处理过程，其他处理过程的延迟都是固定的。

在工程实践中，由于本方案占用存储资源较大，片内 SRAM 级联个数较多，会限制 FPGA 最终实现的主频。为了解决这个问题，我们设计了一种带存储转发的交叉开关，将片内 SRAM 组织起来，用以缓解时序问题。我们采用 Xilinx 的 Alveo U200 型号板卡，支持 8 只标的证券的快照合成，占用 56% 的 URAM 资源，最终主频可以实现 300MHz。

表三列出本方案实现的各个快照合成模块的处理周期和延迟。

5.2 吞吐量与抖动

吞吐量主要取决于整条通路的瓶颈位置，决定设计的处理能力上限。

抖动主要取决于吞吐出现瓶颈时，待处理行情的平均排队时间。如果逐笔行情的输入速度未达到处理能力上限，不会触发行情处理的排队问题，则可以认为抖动为零。

本设计的逐笔处理速度与前级解码速度相匹配，更新订单薄过程不会出现阻塞。本设计的瓶颈主要在行情输出位置。原因包括：1. 输出数据结构包含 26 个 64 位数据，完整输出一个快照至少 26 拍；2.Level Cache 处理一旦触发搜索，RestructOut 处理应等待 Level Cache 更新完成，才能启动输出（保证输出最新、准确的结果）。然而，在实际使用中通常采用定时输出或条件触发输出，输出过程不会阻塞处理通路。

行情一般是毫秒级突发的。吞吐量的测度不应该是秒，而是毫秒。

我们设每次输出行情时，Level Cache 都碰巧在更新中，则逐笔行情的吞吐量为 7518 笔 / 毫秒。换言之，如果交易所每毫秒发出的逐笔订单数小于 7518 笔，FPGA 快照合成能够做到零抖动。

6 结束语

本文针对期权做市业务需求，介绍了一种以面积换速度的方式进行快照实时合成的实现方案。与其他方案相比，本文实现方案的好处在于：由于在逐笔委托和逐笔成交订单的处理中不需要进行搜索，订单处理的速度是最快的，快照合成出错的概率也是最低的。而且，由于订单处理速度和前级解码速度相匹配，设计能够流水起来，不会对解码造成反压。本文所述实现方案的问题是占用存储资源较大，当支持快照合成的标的数量增加时，必须使用片外 DRAM 进行扩展，延迟指标会略有下降。但即便如此，与其他方案相比，本文所述方案在性能和稳定性上也具有一定的优势。

本文所述方案还有一定的优化空间。我们下一步工作将在以下两方面进行优化：1. 提高 Level Cache 的搜索速度；2. 提高设计的可扩展性，使单板卡可支持上千只证券的快照实时合成功能。

一种基于Kubernetes的Jenkins任务构建方法

王如迅[1]、任党恩[1]、王帅[2] /[1] 中国国际金融股份有限公司 信息技术部 北京 100020
[2] 浩鲸科技股份有限公司 云效业务部 北京 100020
E-mail : xiansheng410426@163.com

Jenkins 动态构建是基于 Kubernetes 动态 Slave 模式，Jenkins Master 接到构建任务后会动态在集群中的一个工作节点上拉起一个 Jenkins Slave Pod 进行构建。现有的动态构建技术存在构建初始化耗时效率低、构建不能排队、删除后不能自动重建问题。本文介绍了现有动态构建技术及存在的问题、国内有关研究进展。在此基础上提出一种基于 Kubernetes 的 Jenkins 任务构建方法并设计实现了构建资源池，包括技术方案、操作流程和监控告警，解决现有动态构建技术存在的问题，具有良好的创新性和借鉴价值。

DevOps（Development 和 Operation 的组合词）是一组过程、方法与系统的统称，用于促进开发、运维和质量保障部门之间的沟通、协作与整合，DevOps 强调整个组织的合作以及交付和基础设施变更的自动化，从而实现持续集成、持续部署和持续交付（CI/CD）。Kubernetes：简称为 k8s，是开源的容器集群管理平台，具有动态伸缩、资源调度、资源隔离、快速部署等优点。Jenkins 是一款著名的持续集成工具，在整个 CICD 的 Devops 中起到一个调度中心的作用。

Jenkins 的构建能力是 DevOps 能力的基石。Jenkins 构建采用 Server 和 Agent 模式。当前 Jenkins 构建存在以下痛点：1、当构建任务出现高峰期，会导致 Jenkins 服务不可用。2、构建服务器资源有限，不能随意调用空闲资源，资源利用率不高。3、构建环境不能充分隔离，多个项

目共用同一个构建机。当前有两种解决方案：静态构建和动态构建。

（1）静态构建分为虚拟机构建和容器化构建，构建时不需要初始化。静态构建虽然构建效率高，但是由于多个构建任务共享一个构建节点，导致构建环境不纯净。静态节点不工作也占用着资源，资源利用率不高。Jenkins 集群采用 Master-Slave 的形式，Jenkins Master 负责中心调度，如果 Master 是单节点，其发生故障后集群就失去了功能。容器化构建是指用 Kubernetes 部署 Jenkins 集群，Jenkins 以 Pod 的形式存在，使得 Master 具有自愈功能。

（2）动态构建是基于 Kubernetes 动态 Slave 模式，Jenkins Master 接到构建任务后会动态在集群中的一个工作节点上拉起一个 Jenkins Slave Pod 进行构建。动态构建需要创建构建节点。动态构建方便进行资源回收，保证构建环境的纯净。但是初始化构建 Pod 即便经过参数调优仍然需要消耗一定的时间，而且不能排队，这会降低构建效率。

本文第 1 节介绍现有动态构建技术，包括构建配置和性能优化。第 2 节介绍动态构建技术存在的问题。第 3 节介绍国内有关研究进展。第 4 节介绍一种基于 Kubernetes 的 Jenkins 任务构建方法并设计实现了构建资源池，包括技术方案、操作流程、监控告警。最后对全文进行了总结。

1、现有动态构建技术

通过构建配置使用动态构建技术，并且采用两种策略进行性能优化。

1.1 构建配置

传统的虚拟机构建模式由于需要占用大量的硬件资源，不同的项目采用同一执行机进行构建，构建环境不纯净，容易相互影响。采用动态构建会自动创建一个 Slave Pod，当构建结束之后会自动删除。保证每次都使用纯净的构建环境，多个不同的业务线构建不会相互影响，也能最大限度节省执行机的资源，不同的团队还可以根据基础构建镜像，自定义适合团队自身的镜像。

采用 Jenkins 的 Master Slave 分布式构建，是通过将构建过程分配到从属 Slave 节点上，从而减轻 Master 节点的压力，而且可以同时构建多个，类似负载均衡的概念。这种方式带来的好处有很多：1、服务高可用。当 Jenkins Master 出现故障时，Kubernetes 会自动创建一个新的 Jenkins Master 容器，并且将 Volume 分配给新创建的容器，保证数据不丢失，从而达到集群服务高可用。2、动态伸缩。合理使用资源，每次运行 Job 时，会自动创建一个 Jenkins Slave，Job 完成后，Slave 自动注销并删除容器，资源自动释放，Kubernetes 会根据每个资源的使用情况，动态分配 Slave 到空闲的节点上创建，降低出现因某节点资源利用率高，还排队等待在该节点的情况。3、扩展性好。当 Kubernetes 集群资源严重不足而导致 Job 排队等待时，可以很容易添加 Node 到集群中，从而实现扩展。这种方式的工作流程大致为：当 Jenkins Master 接受到 Build 请求时，会根据配置的 Label 动态创建一个运行在 Docker Container 中的 Jenkins Slave 并注册到 Master 上，当运行完 Job 后，这个 Slave 会被注销并且 Docker Container 也会自动删除，恢复到最初状态。

Jenkins 安装 Kubernetes 插件 [1]，配置 Cloud 和 Pod Template 实现动态创建 Slave 节点。其中 Pod Templete 中配置的是 JNLP 镜像，可以提前内置 Docker、Kubectl 等环境，然后安装 Maven、Nodejs 等编译环境。需要在 Jenkins 的 Configure Global Security 配置端口（例如，50000）。需要在 Jenkins master 节点的 Service 中增加如下配置：

```
- name: agent
    protocol: TCP
    port: 50000
    targetPort: 50000
    nodePort: 50000
```

如果同一个IP地址有不同的Jenkins应用，也可以设置其他端口，比如50002端口。然后在Sevice中进行配置。创建单独的命名空间，配置好角色权限。创建jenkins-slave.yaml、jenkins-slave-role.yaml、jenkins-slave-rolebinding.yaml，详见附录。为了能够正确连接kubernetes需要执行：Kubectl create clusterrolebinding per-missive-binding --clusterrole cluster-admin -serviceaccount=jenkins:default。其中凭据部分，可以在Jenkins的凭据管理中选择Kubernetes Service Account类型点击创建，Jenkins会自动去找Kubernetes的Service Account，然后测试是否连通Kubernetes。为了保证构建资源的独立性，需要建立单独的命名空间。为了使产生的Pod运行于指定的Node节点，可以在Pod Templete中设置“污点”和“容忍度”。例如在“Raw YAML for Pod”中配置：

```
apiVersion: v1
kind: Pod
spec:
  serviceAccountName: jenkins-slave
  imagePullPolicy: IfNotPresent
  tolerations:
   - operator: "Exists"
```

通过kubectl label node命令配置“节点选择器”的值，使得构建产生的Slave Pod运行于指定的Node节点，方便Slave Pod的统一管理。在Jenkins构建中选择对应的标签表达式，这样所有的构建通过Kubernetes的Pod实现动态的创建和删除。如图1所示为采用Slave Pod进行动态构建，Pod模板名称为“jenkins-slave”，产生的Pod名称为“jenkins-slave-cbnzk”。除此之外，还有静态配置Slave Pod或者根据Pod资源需求和Node资源情况，确定调度优先级的方式。

1.2 性能优化

与虚拟机作为Jenkins Salve不同，Kubernetes生成Jenkins Slave是个动态创建的过程，由于是动态创建，不能像连接虚拟机作为执行机一样点击构建可以立即排队，创建Pod需要消耗一定的时间才能开始进行构建。不能并发创建Slave Pod是Jenkins的调度策略问题，Jenkins调度策略尝试最大化利用每个Slave Pod，这涉及到效率问题。有两种解决效率问题的办法，一种是尽量利用已

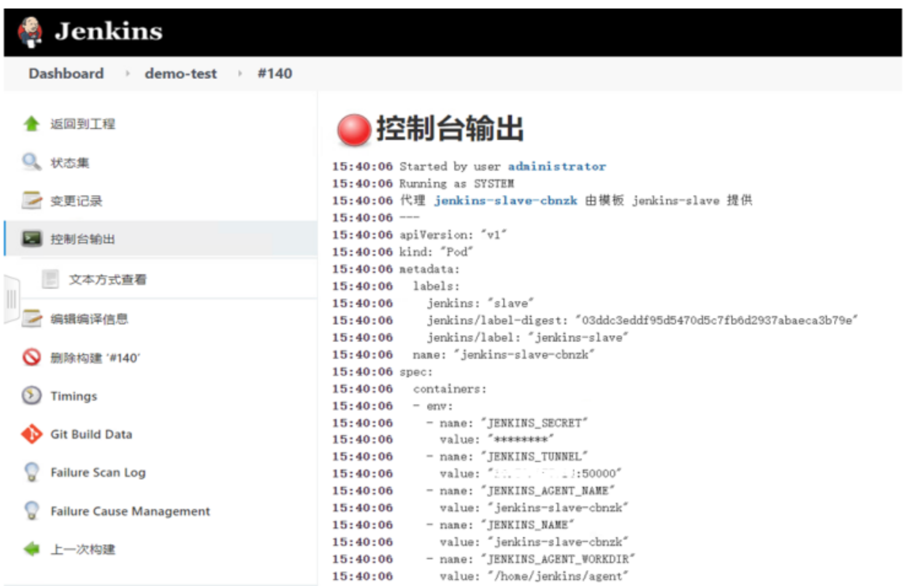


图1：采用Slave Pod构建

有的 Jenkins Slave，另外一种是提高产生 Jenkins Slave 的效率。

2.2.1 尽量使用已有的 Jenkins Slave

设置 idleMinutes 让 Jenkins Slave 在执行完测试 Job 后不被立即销毁，而是可以空闲一段时间，这段时间内如果启动了 Job，就可以分配到上面来执行，提高已有 Jenkins Slave 的利用率。但是这种方式会导致构建环境不纯净，不是严格的构建环境隔离。如果设置 idleMinutes 过短，会导致当构建完成之后其他构建任务使用此 Pod 进行构建时，一旦执行机被销毁会导致构建过程中止。可以根据项目的实际情况，进行保持时间的设置。如果需要保留到固定的时间点，则可以将 idleMinutes 设置为永久，然后在 Slave Pod 产生的 Node 节点上编写定时任务进行删除。在 Job 流水线中，可以根据构建需要使用相同的 label，前后启动的 Job 可以使用相同的 Jenkins Slave。比如，job1 使用的 jenkins Slave 的 label 是 Kubernetes-maven，当 job1 结束后，使用相同 label 的 job2 启动后可以直接使用 job1 用过的 Jenkins Slave。另外由于每次需要从私服拉 jar 包，可以在 Pod Templete 中挂载主机路径下的离线 jar 包，避免每次构建都拉取大量的 jar 包，从而节省构建时间。

2.2.2 提高产生 Jenkins Slave 的效率

Kubernetes 产生 Jenkins Slave 并连接 Jenkins Master 的流程是：Jenkins Master 计算现在的负载情况；Jenkins Master 根据负载情况，按需通过 Kubernetes Plugin 向 Kubernetes API server 发请求；Kubernetes API server 向 Kubernetes 集群调度 Pod；Pod 产生后通过 JNLP 协议自动连接到 Jenkins Master。后三步由于主要受网络影响，相对比较快。第一步，Jenkins Master 会经过一系列算法计算之后，发现没有可用的 Jenkins Slave 才决定向 Kubernetes API server 发起请求。这个过程在 Jenkins Master 的默认启动配置是不高效的。经常会导致一个新的测试 Job 启动后需要等一段时间，才开始在 Kubernetes 上产生 Pod。因此，需求对 Jenkins Master 的 Deployment 进行修改，增加如下配置：

```
- name: JAVA_OPTS
            value:
            -Dhudson.model.LoadStatistics.clock=2000
            -Dhudson.slaves.NodeProvisioner.recurrencePeriod=5000
            -Dhudson.slaves.NodeProvisioner.initialDelay=0
            -Dhudson.model.LoadStatistics.decay=0.5
            -Dhudson.slaves.NodeProvisioner.MARGIN=50
            -Dhudson.slaves.NodeProvisioner.MARGIN0=0.85
```

Jenkins Master 每隔一段时间会计算集群负载，时间间隔由 hudson.model.loadStatistics.clock 决定，默认是 10 秒，将其调整到 2 秒，以加快 Master 计算集群负载的频率，从而更快知道负载的变化情况。当 Jenkins Master 计算得到集群负载后，若发现没有可用的 Jenkins Slave，Jenkins Master 会通知 KubernetesPlugin 的 NodeProvisioner 以 recurrencePeriod 间隔生产 Pod。因此 recurrencePeriod 不能比 hudson.model.LoadStatistics.clock 小，否则会生成多个 Jenkins Slave。initialDelay 是一个延迟时间，原本用于确保让静态的 Jenkins Slave 和 Master 建立起来连接，由于使用 Kubernetes 插件动态产生 Jenkins Slave，没有静态 Jenkins Slave，所以我们将参数设置成 0。hudson.model.LoadStatistics.decay 这个参数用于抑制评估 master 负载的抖动，对于评估得到的负载值有很大影响。默认 decay 是 0.9，把 decay 设置成 0.5，允许负载有比较大的波动，Jenkins Master 评估的负载就是在当前尽可能真实的负载之上，评估需要 Jenkins Slave 的个数 hudson.slaves.NodeProvisioner.MARGIN 和 hudson.slaves.NodeProvisioner.MARGIN0, 这两个参数使计算出来的负载做整数向上对齐，从而尽可能多产生 1 个 Slave，以此来提高效率。

2、动态构建技术存在的问题

目前 Jenkins 动态构建针对构建后的 Pod 保留策略 Pod Retention 有 Never、Always、Default、 On failure 四种策略，其中 Never 表示构建后删除构建节点；Always 表示构建后保留构建节点；Default 表示使用保留时间设置；On failure 表示当构建失败时保留节点，其他情况删除节点。1、如果设置 Pod 保留时间为 Always，则构建后不删除节点。2、如果设置为 Never，则删除后下次构建仍需消耗一定的时间，不具备构建后立即重建的能力，多次的快速点击容易导致构建失败。3、如果设置为 On failure 是当构建失败时保留，构建成功则删除节点。4、如果设置为 Default，设置代理的空闲存活时间可以保留 Pod，但这样会导致构建环境不够纯净。比如不同的构建任务下载的 Jar 包可能有冲突或者修改了网络代理都会导致构建失败。Jenkins 动态构建目前存在以下问题：

（1）每次创建 Pod，即便经过参数优化提高产生 Jenkins Slave 的速度，也需要消耗一定的时间。

（2）构建不能排队，针对同一任务的连续快速构建容易失效。

（3）动态构建设置为 Default，设置保留时间虽然可以保证构建的连续问题，但是构建环境并非纯净。缺少删除 Pod 后自动重建的能力。

（4）动态构建的 JNLP 镜像中不支持修改自身 Agent 的标签。修改会报错“ERROR: anonymous is missing the Agent/Configure permission”，也即如果实现构建后对 Agent 打新的标签，采用目前的 Jenkins Kubenetes 插件是不支持的。采用当前的动态构建，如果在 JNLP 镜像中增加构建后删除 Pod 的操作会导致 JNLP 断开连接从而导致 Jenkins 构建失败，影响整个 Jenkins 的构建结果。所以采用当前的动态构建技术不具备构建后打新标签以及删除 Pod 的能力。

3、国内有关研究进展

当前有五个申请公开的技术专利与 Jenkins 构建技术有关。

1《基于 Jenkins 的任务构建方法与系统》[3]，该技术实现类似于 Jenkins 现有的动态构建技术，进行了一定的技术延展，不涉及提高构建的效率。

2、《基于 Jenkins 的设备管理方法、装置、存储介质及电子设备》[4]，该技术描述的第一容器、标识等概念类似于动态构建的 Slave Pod 和标签的概念，没有涉及现有动态构建方案存在的问题。

3、《基于 Jenkins 的分布式任务构建方法、装置、计算机设备》[5]，该技术是 Jenkins Master 和 Agent 连接的一种技术方案，是针对虚拟机环境的。

4、《一种基于 Jenkins 分布式的分配任务的方法及系统与流程》[6]，该技术方案描述了 Jenkins Master 和 Agent 连接，创建 Job 的时候采用 swarm-client 连接服务器，连接成功后为 Jenkins agent 绑定 Job。

5、《一种 Jenkins 任务管理方法、系统、终端及存储介质》。该技术方案通过判断所述流水线任务的运行环境下是否存在空闲 Jenkins 节点，是为了提高任务的执行效率，避免出现任务分配不均的情况，这是在 Jenkins Master 侧进行的调度，而我们研究的是在 Jenkins Slave 侧。

总之，以上五种方案不涉及解决动态构建存在的问题。国内采用的 Jenkins 构建，一种是静态构建，一种是动态构建，有的进行了性能优化。本文涉及 Jenkins Master 和 Agent 连接技术，不是只介绍 Jenkins 现有的动态构建技术，而是要解决现有动态构建技术存在的问题，提出新的 Jenkins 任务构建方法提高构建效率。

4、一种基于 Kubernetes 的 Jenkins 任务构建方法

提出一种基于 Kubernetes 的 Jenkins 任务构建方法，该方法有如下特点：

1、响应于接收到构建任务请求，确定资源池中是否存在可用的 Agent pod，资源池包括部署在 Kubernetes 节点集群中的 N 个 Agent pod 以运行 N 个 Jenkins 执行机，N 个 Agent pod 与 N 个 Jenkins 执行机一一对应；当确定资源池中存在可用的 Agent pod 时，使相应的可用的 Jenkins 执行机执行构建任务请求对应的构建任务；以及在构建任务执行完成之后删除构建任务对应的 Agent pod，并且创建新的 Agent pod 作为资源池的一部分，使得资源池中的 Agent pod 的数量恢复为 N 个。

2、创建 N 个 Agent pod，设置每个 Agent 的标签值和名称，并且获取与每个 Agent pod 对应的 Secret 值，所述标签值用于获取、筛选、组织 Agent pod，所述 Secret 值用于 Agent pod 与 Jenkins 服务器之间的通信。基于所述 Secret 值，将所创建的 N 个 Agent pod 注册到所述 Jenkins 服务器，使得每个 Agent pod 可以运行由所述 Jenkins 服务器调度的构建任务。

3、基于所述构建任务对应 Agent pod 的预定义标签值以及 Agent pod 中的 Jenkins 执行机的 idle 标志，确定所述构建任务是否完成，其中，所述预定义标签值用于指示是否需要将所述预定义标签值对应的 Agent pod 进行回收，所述 idle 标志用于指示对应的 Jenkins 执行机是否可用。

4、当所述资源池中的全部 Agent Pod 对应的执行机都处于繁忙状态，不存在可用的执行机时，通过 Kubernetes 插件的方式创建新的 Agent pod 来运行 Jenkins 执行机，以执行所述构建任务请求对应的构建任务。

5、当并发构建任务的数量逐渐减少时，减少所述资源池中 Agent pod 的数量。当并发构建任务的数量增加的时候，增加所述资源池中 Agent Pod 的数量。当所述资源池中可用的 Agent pod 的数量达到临界值时，向管理员告警。

基于上述任务构建方法的设计，本文技术实现了一种构建资源池。解决动态构建初始化耗时效率低、构建不能排队、删除后不能自动重建问题，提高构建的效率。

4.1 技术方案

（1）通过代码一键创建 Jenkins Agent，并从 slave-agent.jnlp 中获取 secret 值。这样可以替代通过 Jenkins 界面创建 Agent 的方式，获取到 slave-agent.jnlp 中 secret 值存到环境变量，获取 secret 值是为了保证 Jenkins Master 和 Agent 进行通信。

（2）编写 Kubernetes deployment.yaml 读取环境变量中的 Secret 值，由于每次创建 Agent 会造成 Secret 的值不同，因此需要进行每次读取。该功能可以有效实现一键创建可用的 Agent Pod。创建 Agent Pod 为 N 个，设置 Agent 的标签和名称，可以通过设置不同的环境变量来区分创建 Agent Pod 的不同方式，进而根据环境变量来设置标签值。示例性地，当采用动态构建的方式创建 Agent Pod 时，可以使用第一标记（例如，“dynamic”）作为 Agent Pod 对应的环境变量，而当采用脚本命令的方式创建 Agent Pod X 时，可以使用不同于第一标记的第二标记（例如，“static”）作为 Agent Pod 对应的环境变量。然后在构建后创建 Agent 脚本中进行判断，如果环境变量为第一标记则打“dirty”标签，创建新的 Agent。如果环境变量为第二标记则不进行处理，从而不影响原有动态构建的功能。

（3）通过预定义标签值（例如，“dirty”）和执行机的 idle 标志双重判断构建是否已经结束。当 Agent 的标签为预定义值且执行机的 idle 状态为 true 时，可以判定标签为第二标记的执行机已经停止运行，Jenkins Job 通过绑定其他标签去运行构建任务，从而保证每次构建资源纯净。维护进程通过读取所有包含有预定义标签的 Agent 节点，同时判断执行机是否空闲，定时删除。如果只采用 idle 标志，需要每秒扫描才能判断执行机是否停止工作，增加了资源池维护的工作负载。设计预定义标签是为了将 Pod 标记为不用。双重

判断是为了确认不用的 Pod 停止工作。

（4）重建的过程中，需要消耗一定的时间，期间如果有其他构建申请则从 N-1 个 Agent 中进行选择，不需要消耗连接时间。如果 N 个 Agent 都被使用，则会切换到具有同样标签的动态构建模式，之前的 N 个 Agent 相当于动态构建的排队队列。

（5）对资源池的 Pod 资源进行监控告警，根据业务繁忙程度进行自动扩容或缩容。

4.2 操作流程

通过代码一键创建的方式初始化资源池，创建 Agent Pod 为 N 个，如图 2 所示。同时增加判断，当资源池中繁忙的执行机个数等于 N 的时候，也即 Pod 中的 idle 为 false 的个数等于 N 时，自动创建 1 个新的 Agent Pod N+1，从而实现资源池的自动扩容。当构建并发量逐渐减少的时候，对资源池自动缩容。设置最大 Agent 个数为 M，则当资源池 Agent 个数等于 M 时，不进行扩容转而采用现有的动态构建。

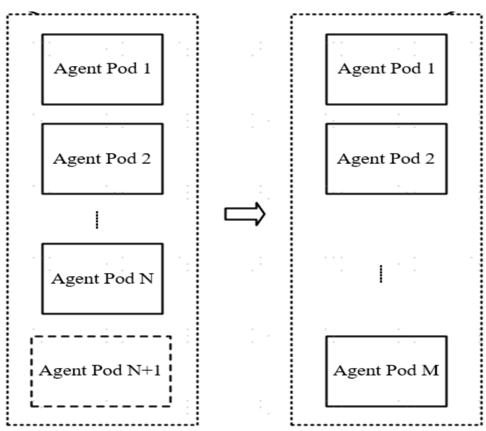


图 2：初始化资源池

发起构建任务之后，Jenkins master 进行任务调度，判断资源池是否空闲，如图 3 所示。如果资源池有空闲的资源，则资源池的执行机进行构建，当构建结束之后标记 Slave Agent 的标签为预定义标签值（例如，“dirty”）并生成新的 Agent。这样会保证下次构建的时候不采用已经打预定义标签的 Agent。如果资源池没有空闲的资源，则直接进行动态构建，设置 Pod Templete 的模板配置策略，一般为 Never，可以实现构建后立即删除 Slave Pod。设置维护进程根据预定义标签值和执行机的 idle 标志双重判断构建是否已经结束，如果包含预定义标签的 Slave Agent 已经构建结束，可执行删除操作进行资源回收。

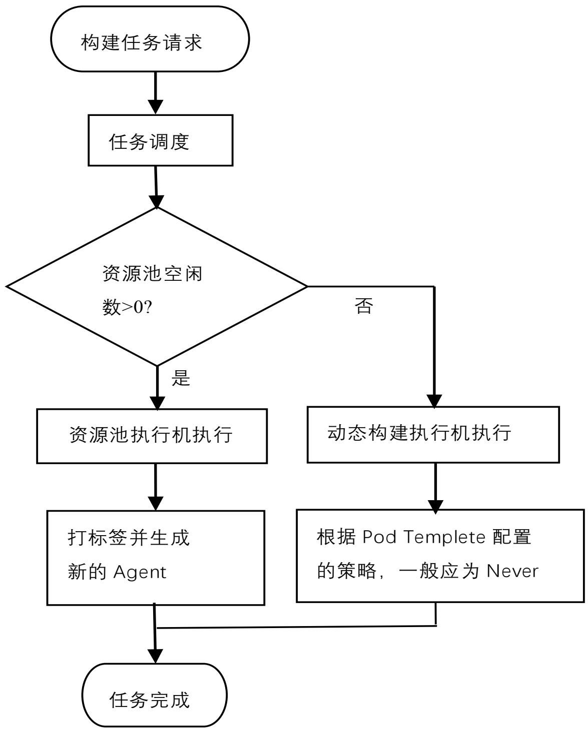


图 3：构建资源池操作流程

4.3 监控告警

将资源池中的 Agent 标签设置为特定值（例如，pool），监控结果如图 4 所示。通过 Prometheus（普罗米修斯）和 Grafana 配置 Jenkins Jobs Stats 模板中 Executors Busy per Label (Stacked) 标签为资源池中 Agent 的标签，然后配置告警信息：资源池中执行机的个数为 N，当繁忙个数等于 N-1 的时候，发送告警信息告知管理员并执行扩容操作。针对预定义标签的 Agent 进行监控，如果一定时间之内还存在预定义标签（例如，dirty）则说明资源没有被及时回收，发出告警通知并予以删除。采用特定标

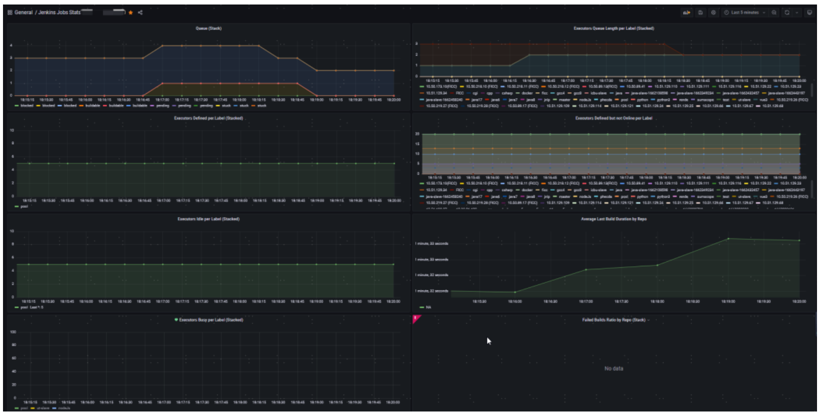


图 4：资源池监控

签值是为了监控本方案中的构建资源池，也可以针对构建机根据不同类别设置不同的标签进行分类监控。

5、总结

DevOps 作为一种软件开发和运维模式，它显著提升了团队的 IT 服务能力。Jenkins 作为 DevOps 平台的构建工具提供了强大的功能支持，能够进行代码的编译、打包、部署，可以处理任何类型的构建或持续集成。本文首先介绍了 Jenkins 构建技术背景，然后对现有 Jenkins 动态构建技术进行归纳总结，如何进行动态构建的配置，如何进行性能优化。指出动态构建技术存在的问题，分析对比了国内有关研究进展，指出其各自的特点。重点阐述了一种基于 Kubernetes 的 Jenkins 任务构建方法并设计了构建资源池，解决动态构建初始化耗时效率低、构建不能排队、删除后不能自动重建问题，并提供了技术方案、操作流程和监控告警。通过本文的分析、对比和实践，希望能够为行业在 Jenkins 构建技术领域带来有益的启发。

参考文献：

[1] Jenkins 集成 Kubernetes [ER/OL]. http://www.yaotu.net/biancheng/34368.html.

[2] Jenkins KUBERNETES Slave 调度效率优化 [ER/OL]. https://www.cnblogs.com/guguli / p/7827435.html，2017，11.

[3] 中国工商银行股份有限公司 . 基于 jenkins 的任务构建方法与系统 [P]. 中国 :CN202010020545.4，2020.06.09

[4] OPPO 广东移动通信有限公司 . 基于 jenkins 的设备管理方法、装置、存储介质及电子设备 [P]. 中国 :CN201910684091.8，2019.11.19

[5] 平安普惠企业管理有限公司 . 基于 jenkins 的分布式任务构建方法、装置、计算机设备 [P]. 中国：CN201811527233.1，2019.05.21

[6] 航天信息股份有限公司 . 一种基于 Jenkins 分布式的分配任务的方法及系统与流程 [P]. 中国：CN202011605122.5，2021.04.23

[7] 王如迅 . 商业银行云原生转型实践探讨 [J]. 金融科技时代，2021(10):46-51.

[8] 约翰•阿伦德尔 . 基于 kubernetes 的云原生 Devops[M]. 北京：中国电力出版社 ,2021.

浅析个人信息保护和数据安全

王志玮，裘岱，张晓军，宋良夏，周蒙 / 上证所信息网络有限公司

近几年，安全合规正逐步吸引着行业的目光，随着监管的不断加强，合规建设也成了企业关注的安全方向，尤其是个人信息保护法和数据安全法发布实施的背景下，重要数据、个人信息如何保护，如何满足安全要求，引发了行业思考。本文就数据安全建设给出了一些思路，从个人信息和数据的关系、两类评估的区别、数据分级分类建议等方面提供了一些想法，供各位看官思考，希望可以起到抛砖引玉的作用。

1 背景

2021 年 6 月 10 日第十三届全国人民代表大会常务委员会第二十九次会议通过《中华人民共和国数据安全法》，自 2021 年 9 月 1 日起施行；2021 年 8 月 20 日第十三届全国人民代表大会常务委员会第三十次会议通过《中华人民共和国个人信息保护法》，自 2021 年 11 月 1 日起施行。随着数据安全逐步活跃在公众视野，业界不乏数据安全白皮书、个人信息安全白皮书等指导意见出现，用于帮助行业建立数据安全保护策略，本文不同于白皮书，主要偏重于安全思路，希望可以给业界一些思考。

1.1 两法关联与区别

《个人信息保护法》和《数据安全法》出发点不同，但个人信息保护和数据安全又息息相关。《个人信息保护法》主要体现在组织、企业等在处理个人信息时的义务，主要为了保障自然人的个人权益，是在 GDPR 的背景下出台的法律法规，相较 GDPR 更符合国情且明确了业界一些争议，使得很多内容在法律层面得到了细化，对企业做个人信息保护有非常好的指导作用，在这之前，这部分工作主要依赖于国标、金标（金融行业）等。《数据安全法》主要体现在国家战略层面，数据（尤其是核心数据和重要数据）作为国家、行业、企业的重要资产，数据泄漏对国家、行业、企业、

个人等都可能有很严重的影响，为此数据安全相关法律法规应运而生。

1.2 个人信息归属于数据

在现行的国标、金标中，在枚举各类数据保护措施时均离不开个人信息安全的例子，当然数据也包含不涉及个人信息的部分，如无法定位到人的群体经济数据、行业内部研究报告等，泄漏了或者出境了对国家均可能存在影响。个人信息因其对自然人的重要性，使得个人信息在数据分级分类中的等级较高，尤其是个人敏感信息，所以在法律要求企业做个人信息保护影响评估和数据安全风险评估时，现行的一些数据安全评估与个人信息保护影响评估的表现形式和评估项存在较大关联。

1.3 个人信息保护影响评估和数据安全评估的意义不同

关注自然人：个人信息保护影响评估需要考虑对自然人的损害，它充分考虑了自然人的影响，更加偏向实施层，考虑一摸一样的业务很少，所以个人信息保护影响评估的频率一般较高，可能每有一个涉及个人信息的业务上线，则需要做一次个人信息保护影响评估。

侧重于架构：数据安全评估则从宏观角度，要求在国家、行业、企业、组织层面上关注到数据资产的安全性以及数据出境对国家、社会、行业以及国人的影响，更加偏向决策层、架构层，所以数据安全评估的频率不高。

1.4 黑灰产肆无忌惮

随着数字化时代的发展，也随着数据安全监管的严格，数据资产的价值得以攀升，这就吸引了黑灰产业的注意力，业界不乏有数据泄露在暗网被售卖的新闻，也有通过各种明目巧妙吸纳数据资产的手段，有些既涉及个人信息也涉及数据安全。举一个简单的例子，不法分子开发一些注册链接，从刷单群体入手，以注册新用户的返利为背景，吸引用户注册并填写个人信息，一次2-10元不等，有的注册需要填写身份证号、手机号等等，这些用户注册页面设计简陋，并不是以推广业务为主要目的，后续这些页面也难以找到，这些注册页面的开发者吸引了大批被诱导的用户自愿提供个人信息甚至个人敏感信息，而用户大多并不清楚把信息给了谁，后续信息被批量卖给需要开拓市场的营销、诈骗人员或组织，继而各种推销、诈骗电话随之而来，信息就这样泄漏出去，数据泄漏甚至引发了各种社工库活跃在外界市场，供消费查询使用。

2 简析数据保护实施方案

业界对数据安全的考虑普遍较大较全，本文从另外一个角度提供一些思考，即数据保护不一定需要一蹴而就，可以先解决核心问题、重点问题，然后逐个攻破周边问题，逐步描绘全面的数据安全地图。本文建议如下的数据保护实施方案。

2.1“统一标准”

首先意识到需要统一数据字典（包括个人信息），针对行业专有名词进行统一命名并识别，对同一个字段，不同的业务使用的字段尽量一致，且有依据可遵循。举例：某字段的中文名为“A字段”，代码参数层面为“abc”，源业务为”A业务”，这样可以防止后续各业务系统的数据无法交互或交互需要大量的梳理成本，以降低数字化转型的速度。

2.2“能不能做”

建立安全合规底线的把控流程，遵循个保法和数据安全法细化的强制性要求，在项目需求阶段即开始增加合规要求，引导项目经理、开发人员补充考虑安全合规需求，也类似于SDL的合规体现，可以集合在已有的SDL平台/流程中。这

部分之前主要参考国标的要求，在个保法出台后，核心参考数据安全法和个保法，个保法从法律层面上细化要求了一些内容，如个人主体同意、单独同意等等，对个人信息处理做了全面定义，这些对个人信息保护影响评估有核心指导意义。另外，个保法未详细列举但会重点影响个人权益或导致数据泄漏的需求也应该作为安全合规底线需求进行落实。本文列举了主要关注点，详见下表所示，供参考。

2.3“要怎样做”

为满足个保法、数据安全法的基本要求，即防止网络数据泄露或者被窃取、篡改、保障个人权益等要求，企业需要做一些内部管理基线要求。

（1）制度层面：公司制度相比较法律法规更贴近公司内部的业务发展情况，要求更加细化、可直接落地，比如《数据安全管理规范》、《个人信息保护管理规范》、《数据平台使用规范》、《敏感数据脱敏规范》、《数据安全应急预案》等等，需要结合企业的实际推行落地。

（2）流程层面：为配合履行个人信息保护和数据安全业务，也为落地制度要求，企业需要明确相关工作流程。比如对 C 端服务的企业需要制定响应用户需求及投诉的渠道与机制（如设置用户反馈页面、提供公共邮箱等）；设定数据泄露、个人信息泄露的应急演练流程并定期演练（建议每年至少一次）；设定数据泄漏报告机制，依照国家法律法规要求向监管机构汇报相关信息泄漏

表 1：建议安全合规需求举例

基本合规要求	参考
是否必要正当且符合最小化收集原则（即不收集无法完成业务、查看公开信息类页面，非登陆状态也应可以查看，不诱导/诱骗用户同意）	个人信息保护法
处理个人信息前是否告知用户相关信息（线上一般为隐私政策，线下一般为纸质隐私说明，隐私政策等需走审批流程才可上线发布，且用户登陆后可方便查看到隐私政策）	个人信息保护法
处理个人信息前是否获得同意（线上一般为用户主动勾选隐私政策，线下一般为签字）	个人信息保护法
是否提供撤回同意的入口（即线上提供用户注销入口，或其他便于注销的形式）	个人信息保护法
若涉及公开个人信息，是否取得个人单独同意（额外的勾选同意）	个人信息保护法
涉及在公共场所安装图像采集、个人身份识别设备时，是否为维护公共安全所必需，是否设置显著的提示标识。	个人信息保护法
处理敏感信息时，是否取得个人单独同意	个人信息保护法
处理敏感信息时，是否界面告知收集的必要性及对个人权益的影响（非隐私政策）	个人信息保护法
处理不满 14 周岁未成年人个人信息时，是否取得未成年的父母或其他监护人的同意	个人信息保护法
涉及自动化决策时，决策是否公平，是否避免大数据杀熟等差别待遇	个人信息保护法
涉及自动化决策时，向个人进行信息推送、商业营销时，是否同时提供不针对其个人特征的选项，或便捷的拒绝方式（可展示关闭按钮）	个人信息保护法
是否签订合同/协议约定双方责任和义务，以及个人信息处理的要求等	个人信息保护法
委托第三方处理个人信息时，是否签订数据保护协议	个人信息保护法
委托第三方处理个人信息时，是否获得用户单独同意	个人信息保护法
是否建立最小授权的访问控制策略，权限是否走审批流程，界面展示是否采取去标识化处理（如掩码、水印等），下载权限是否限制	GB/T35273 个人信息安全规范
处理敏感信息时，是否加密传输、加密存储	GB/T35273 个人信息安全规范
处理生物识别信息时，是否未存储原始信息，例如仅存储个人生物识别信息的摘要信息等	GB/T35273 个人信息安全规范

事件等；设置个人信息保护影响评估流程及触发节点（建议与已有管控流程结合）；设定特权账户的审批流程、隐私政策的审批流程等。

（3）数据内部使用层面：可设定脱敏数据库用于业务分析需要并与原始数据分开存储、设定动态脱敏规则、设定脱敏算法（建议使用国密算法）；为敏感数据的查询等操作划分专门的网段或使用虚拟桌面/云桌面使得数据无法落地到本地，并配合水印等保密手段；部署 DLP 相关终端安全策略。

（4）数据内部传输层面：可设定 API 接口规范、验签机制、加密算法要求、文件加密要求等；归口内部数据传输通道，防止数据不一致。

（5）数据对外传输层面：除上述内部传输的安全控制外，可考虑专线传输、互联网传输（https）等场景，以及关闭不需开放的端口、设置 IP 白名单、维护权威证书、缩短密钥有效期、签订数据保护协议等。

上述部分需求为新系统建设时需要补充的安全需求，而对于存量数据来说，数据字段级加密较为困难，那么可以选择文件加密、协议加密、接口严格验签、严格的访问控制、网络隔离、静态脱敏等措施，并早日迁移到数据集成类平台中（详见最后一章）。

（6）表单化 -> 工具化，常态化监测：后续可以把人工通过表单形式展现的需求，融合进工具中，比如通过使用统一的密钥管理系统而省去每次的加密流程设计，通过使用数据汇聚平台中的脱敏数据而省去调取生产数据进行分析以防止生产数据泄漏，通过设置统一接口完成加签验签而省去每次设计系统中思考如何做验签，通过态势感知类平台补充数据治理无法修复的弱点，通过 EDR 与 DLP 溯源数据泄漏事件等等，逐步节省人工操作依赖，达成常态化数据泄漏监测和处置。

3 数据分级分类怎么做

3.1 设定参考对象

证监会在 2018 年发布了《证券期货业数据分类分级指引》，中国人民银行在 2020 年发布了《金融数据安全 数据安全分级指南》，两个文章的后半部分均非常详细的罗列了各类数据的建议保护等级，有很全面的参考作用，本文分别截取了一部分内容如下图所示，全面的分级分类详见上述两篇指南。

3.2 设定底线原则

虽然有详细的分级分类目录，但企业数据纷

业务细分		数据归类和细分				最低参考数据级别	备注
业务条线一级子类	业务条线二级子类	数据一级子类	数据类型说明	数据二级子类	数据类型说明		
交易	交易管理	成交信息	一般指证券期货行业会管单位产品的成交相关信息，包含证券代码、成交金额、成交数量、成交时间、成交价格、成交日期、委托编号、会员编码、账户编码、品种合约等。			4	
		委托信息	一般指证券期货行业会管单位产品的委托相关信息，包含申请编号、证券代码、委托价格、委托数量、委托方向、委托类型、委托时间、委托价格、委托日期、会员编码、账户编码、品种合约等。			4	
		交易业务参数信息	一般指证券期货行业会管单位产品业务规则中定义的，完成交易所需的参数，如交易时段、涨跌幅、交易单元等。			4	
		交易日志信息	一般指证券期货行业会管单位产品订单及成交的流水和错误信息，可分为订单日志、成交日志等多类，包含订单编号、日期、时间、价格、会员编码、账户编码等信息。	订单日志	一般指证券期货行业会管单位产品订单流水和错误信息，包含订单编号、订单日期、订单时间、委托价格、会员编码、账户编码等信息。	4	

图 1：数据分级分类部分截图

数据归类和细分							安全级别	备注
一级子类	二级子类	定义说明	三级子类	定义说明	四级子类	内容	最低安全级别参考	
客户	个人	指金融业机构提供各种业务服务的自然人对象相关的各类信息。	个人自然信息	指个人的自然属性信息，即个人本身具有的属性，而非在金融业机构业务开展过程中产生的属性信息。	个人基本概况信息	指个人基本情况数据，如个人姓名、性别、国籍、《海外账户纳税法案》（FATCA）有关个人身份数据、民族、婚姻状况、证件类型、证件号码、证件生效日期、证件到期日期、家庭住址等。	3	
					个人财产信息	指个人的财产数据，如个人收入状况、拥有的不动产状况、拥有的车辆状况、纳税额、公积金缴存金额、个人社保与医保存缴金额等。	3	
					个人联系信息	指个人各类通信联系方式数据，如手机、固定电话、电子邮箱地址、微信号等。	3	
					个人健康生理信息	指个人生病医治过程中产生的相关记录数据，如病症、住院志、医嘱单、检验报告、手术及麻醉记录、护理记录、用药记录、药物食物过敏信息、生育信息、以往病史、诊治情况、家族病史、现病史、传染病史等，以及与个人身体健康状况相关的其他信息，如吸烟史等。	4	
					个人地理位置信息	指描述能具体定位到个人的地理位置数据，如包含所在国家、城市、区域、街道、经纬度等维度信息下的常在位置和当前位置等 。	3	

图 2：数据分级分类部分截图

繁复杂，数据逻辑各有千秋，把企业数据全部排摸清楚已实属不易，更何况还需要全部定好保护等级，并定好各等级对应的安全等级（保护措施）以及后续建设怎样的安全力量。这种思路很全面，但是试错成本很高，所以本文提出企业在数据保护一开始，不一定非要制定特别详尽的数据分级分类表格，而是先设定底线原则，即数据不泄漏，并对企业整体的数据保护力量进行评估，即已有的安全能力建设是否全面、缺失的部分对数据泄漏是否有影响，以及是否可以通过访问控制如二次审批进行加固。

3.3 划定“重点保护圈”，绘制“数据地图”

针对数据分级分类已有建议中的较高等级数据进行重点识别和保护，如个人信息（含个人敏感信息）、各类鉴别信息、交易管理信息、结算管理信息、监察与评价管理信息等。对这部分信息先进行重点的保护，如加密、备份、脱敏、隔离存储、设置销毁策略等，严格把控数据内部传输、下载和外发。为“重点保护圈”建造好“城墙”，这对还未建设数据安全合规的企业来说，可以理解是为企业合规工作“雪中送碳”。

此外，个人信息因其敏感度及监管的不断重视，已经在业界的“重点保护圈”中，而且，个人信息保护的要求普遍较为详细（详见国标、金标），这使得企业不得不先为个人信息建立一套完备的分级分类和保护措施，这些保护措施可以复用在非个人信息但同等敏感的数据中进行保护。具体可参考第二章中的第 3 节“要怎样做”。

3.4 扩充企业数据目录以及分级分类，逐步完善“数据地图”

企业在对这部分数据做重点保护的同时，也在积累着数据保护技术措施以及数据分级分类的经验，为敏感等级较低的数据分级分类奠定了基础，数据分级分类的详细内容得以逐步扩充和完善，安全等级也可以逐步的细化。数据保护措施对于企业的适用性也逐步明晰，各职责分工也有了基础，就像俗话说的那样“走的人多了就成了路”，后续的工作则可理解是为企业合规工作“锦上添花”。

4 数据集成类平台与数据治理

数据集成类平台可以集成企业各业务的数据资产，集中管控、集中治理、集中能力建设、集中提供服务，对数据治理可以起到比较好的作用。

4.1 统一数据源

为建造数据集成类平台，数据字段应已尽量进行整理并统一，这保障了数据的一致性，为各业务系统间获取数据提供了权威的数据源。为降低数据泄漏面，系统在留存大量“冷”（非动态获取）原始数据的同时，可选择留存近期（如一年）的“热”（动态获取）数据，其中“冷”数据加密存储在隔

离数据库中,仅允许“热”数据的有限交互。此外,“一数一源”也是行业乃至公众普遍的期望,统一数据源为“一数一源”打下了基础。

4.2 统一脱敏数据分析能力

系统可留存存量数据的脱敏数据,并对新系统采用动态脱敏的技术手段留存脱敏数据,然后对脱敏数据或原始数据(若必需)进行计算处理,继而输出数字化分析能力(同样是脱敏数据,如各种分析报告等),供业务分析使用。对于脱敏数据无法满足需求的部分业务,可以对“热”数据采用动态脱敏手段以返回业务需要的数据分析能力,对于不需要原始数据的业务则直接使用脱敏数据进行操作,这样也可以分散业务数据的计算压力。

4.3 统一出口控制

比较直观的好处是,数据接口规范在该平台可以实现统一,各业务系统依据不同场景和所需字段进行对接,降低了开发测试人员的压力。

4.4 统一数据操作权限控制

各业务系统若涉及运营、运维,难免允许内部企业员工接触个人信息(甚至敏感信息)或企业敏感数据,统一权限控制在加强管控能力的同时减少了各业务系统设定数据操作权限及规则的工作,而且,在推广全面的安全访问控制措施时(如双因素、强密码、密码更新、用户权限梳理等),也降低了业务和安全人员的工作压力。

这样,数据集成类平台即解决了数据治理的问题,也解决业务上需要数据结合与分析的需求,而且因为该类平台的存在,各系统之前的数据传输也降低了交互,一举多得,同步提高了数据安全能力、数字化服务业务的能力,以及数据传输的效率。不过数据集成类平台建设成功的挑战较大,需要大量的自研工作以及数据梳理、汇总的工作,所以可以不建设大而全的平台,采用分模块建设,数据之间设置策略互通即可,先试点几个相关业务的同时积累经验,为最终建设全面的平台夯实基础。

5 数据安全发展推测

数据安全在业界已不算陌生,但是如何有效的保障数据安全一直是热烈讨论的问题,本文从另外的角度提出了一些思考,希望可以起到抛砖引玉的作用,为证券行业数据安全发展尽一份力。

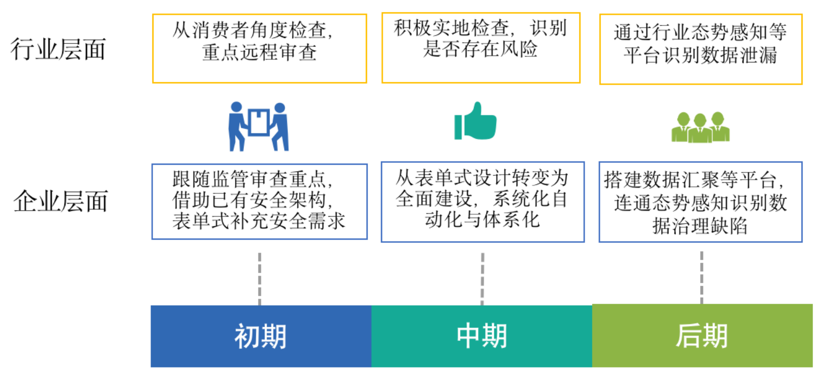


图 3:数据安全发展趋势图

参考文献:

[1] 中国人民银行,JR/T 0197-2020 金融数据安全 数据安全分级指南,2020.

[2] 中国证券监督管理委员会,证券期货业数据分类分级指引,2018.

基于TextRank的期货研报自动摘要方法研究

宁晓冬[1]、杨和国[1]、赵国平[2] /[1]郑州商品交易所 市场监查部 郑州 450000 [2]郑州易盛信息技术有限公司 郑州　　E-mail : xdning@czce.com.cn

期货研报是期货分析师针对各种期货品种、合约、宏观环境、行业动态和产业链等做出分析和研究的总结，是了解交易动向和市场行情走势的重要信息来源。如何快速、准确地提取研报中有价值的关键信息来研判期货市场行情和动态，从而更好地防范风险、保障期货市场稳定有序运行是期货市场信息管理部门面临的问题。本文提出了一种期货研报摘要的自动抽取方法，首先，基于自主构建的期货领域词表进行研报摘要的预抽取，并对预抽取结果采用融入 TF-IDF 的句向量表示;其次,采用 TextRank 算法抽取出候选摘要,保证摘要与原研报在内容上的高度相关性；最后，融入最大边界相关算法补充一些关键但“势单力薄”的关键句，以保证摘要内容的信息性。实验结果表明，该方法在期货研报自动摘要上是有效的，具备快速、准确的优点。

1 引言

期货研报是期货分析师在研究各种期货品种、合约、行业动态和产业链等信息基础上，形成的日评、周评、月报、季报、年报、热点分析等文档，是关于期货市场宏观环境、行业信息等的研究报告，是把握期货交易动向和市场行情走势的重要信息来源。这些信息每天都在以指数量级增长。人工阅读研报，提取其中的关键信息并对关键信息进行分类、汇总和研判，不仅工作量大、费时费力，还会因人的主观情绪不稳定导致出现理解偏差、遗漏关键信息等问题。

针对这一问题，脱水研报[1]通过人工收集材料，总结对比分析历史数据，得出未来一段时间内股票价格的预测结果。该方法需要人工在短时间内付出巨大的工作量。文献[2]采用基于规则

的要素抽取模板构建了研报要素抽取平台，抽取包括作者、机构名称、研报标题、个股标的、评级、目标价，营业收入、净利润、资产回报率和净资产收益率等要素。该方法抽取的是研报中的实体、短语、标题或具体数值，并没有关注期货研报中期货分析师关于当前期货行情的观点及论据。

自动摘要技术[3]通过相应算法从文本中挖掘信息，形成能概括文本内容且文本长度远小于原文本的摘要，以帮助用户快速、准确地把握原文本的中心内容。针对上述问题，本文提出一个期货研报摘要自动抽取方法。对一篇期货研报，首先，基于自主构建的期货领域词表进行候选关键句的预抽取，并采用融入词频 - 倒文档频率[4]（Term Frequency-Inverse Document Frequency，TF-IDF）的句向量表示预抽取结果中的句子；第二步，采用 TextRank[5] 算法获取候选摘要；最后，为避免遗漏关键信息同时保证摘要的简洁性，引入最大边界相关算法[6]（Maximal Marginal Relevance，MMR）对候选摘要进行补充，从而得到原研报的摘要。对于方法的有效性，采用ROUGE[7]进行评价。

2 问题描述

期货研报自动摘要是指从期货研报中抽取出期货市场监管人员关注的行情动态、交易动向和风险描述等方面的关键句，形成简洁、精炼并能全面表征研报原文的自动摘要过程，属于单文档抽取式自动摘要问题。其形式化定义为：设 D={w1，w2，…，wn} 为一个包含 n 个句子的研报文本，目标是得到一个由关键句集合 Y 组成的

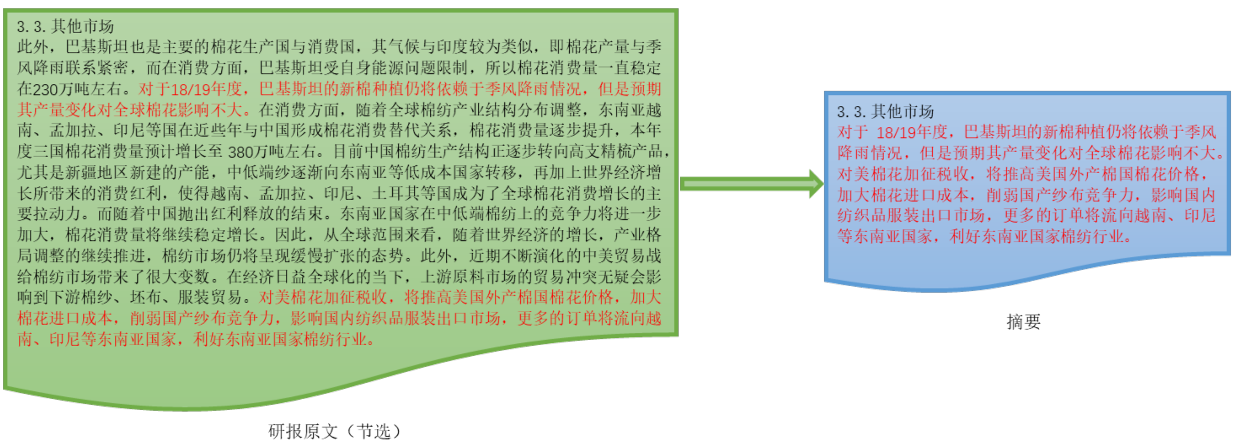


图 1：期货研报自动摘要问题描述

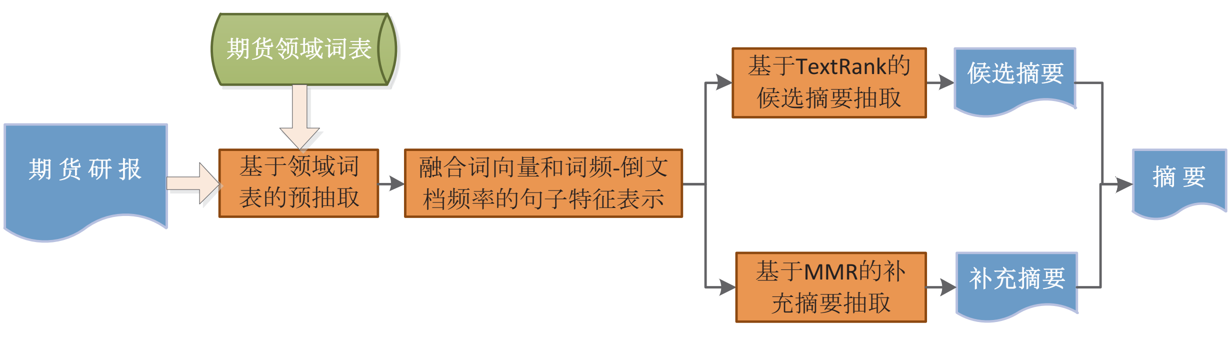


图 2：基于 TextRank 的期货研报自动摘要方法流程图

包含原文中心内容的摘要 Y={y1，y2，…，ym}，且满足 m<<n 且 $yi(1 \leqslant i \leqslant m) \in D$ [8]。图 1 描述了期货研报自动摘要问题，依靠自动化手段，从期货研报原文中抽取出含有关键信息的句子（红色字体所示），重组成研报摘要。解决该问题能够提升期货市场监管人员获取研报关键信息的效率，减轻人工工作量。

通过对期货研报调研发现：

1）研报文本长，且包括大量图表；

2）研报内容、组织结构与撰写风格多种多样，涉及到基本面、技术面等多种关键信息。关键句分布无规律，给研报的自动摘要带来极大困难；

3）因缺乏期货研报摘要的训练语料，机器学习和深度学习中的有监督方法均不适用；

4）某些期货研报在开头或结尾部分自带摘要，但从质量上看，存在不够全面、缺少交易所监管人员关注的关键信息等问题。

3 方法

根据期货研报的上述特点，提出一种期货研报自动抽取的方法，其流程图如图 2 所示。对获取到的期货研报，首先进行格式转换和数据清洗等预处理工作，然后，基于领域词表进行预抽取获得候选关键句集合，并对候选关键句采用融合 Word2Vec [9] 和 TF-IDF 的句子特征表示。第三步，采用 TextRank 算法自动抽取候选摘要，同时，对于一些非常重要但势单力孤的关键句，采用基于 MMR 算法进行补充抽取，获得补充摘要。最后，将候选摘要与补充摘要合并获得摘要。

3.1 基于领域词表的预抽取

从网上收集到的期货研报，因格式问题，不能直接用于自动摘要，需先将其通过 OCR 技术转成纯文本格式，研报中的表格、图片在研报正文中均有详细描述，故将其舍弃。此外，对研报中含有的如免责声明、联系方式和数据来源等干扰信息进行自动识别并删除。

经过上述处理后的期货研报文本还存在如下的问题：篇幅较长、信息量较大、关键句分布无规律。因此，直接对其进行基于 TextRank 的关键句抽取效果并不好。针对这个问题，采用基于期货领域的词表对研报进行预抽取，即根据调研阶段人工构建的期货领域关键词表对期货研报含有关键词的句子进行初步抽取，形成研报候选关键句集合，为后续精准抽取摘要提供数据基础。期货领域关键词表样例如表 1 所示。

3.2 融合词向量和词频－倒文档频率的句子特征表示

对于预抽取得到的关键句集合，需要对其进行句子特征表示，以便后续自动摘要抽取时使用。句子特征表示的目的是将不可计算和非结构化的研报文本，转变成可计算的、结构

表 1：期货领域关键词表（部分）

关键要素	关键词
供需	供需、供应、供给、需求
库存	库存、抛储、轮入、轮出
产业链	产业链、上游、下游、原料
天气与种植	天气、种植、成本
风险点	风险点、重要监测点、风险因子、风险因素

化的向量，同时减少语义信息的损失。因期货研报风格、用词等具有多样性，使用单一的词向量模型表征效果不佳。为兼顾词向量及词频对句子语义信息的影响，本文采用 Word2Vec 和 TF-IDF 相结合的方法表征研报关键句，Word2Vec 能够将单个词语映射为一个词向量，TF-IDF 根据研报文本中单个词语出现的频率计算出词语的重要程度。对每个句子中的非停用词，取其 TF-IDF 值作为权重，与经 Word2Vec 转换后的词向量相乘，得到句子的加权平均值，其计算如公式（1）所示，其中，Vi 表示第 i 个句子的句向量，wordij 表示第 i 个句子的第 j 个词，Word2Vec(wordij)、TF-IDF(wordij) 分别表示第 i 个句子的第 j 个词的词向量和 TF-IDF 值，Mi 表示第 i 个句子中词的个数。

$$V_i = \frac{\sum_{j=1}^{M_i} Word2Vec(word_{ij}) \times TF-IDF(word_{ij})}{M_i} \tag{1}$$

3.3 基于 TextRank 的候选摘要抽取

传统 TextRank 算法是一种基于图排序的文本摘要算法，将原文本中的句子作为图的结点，句子间通过边连接，边上的权重代表了句子间的相似度，得到相似度矩阵。通过相似度矩阵计算每个句子与其他句子的关联度，如果关联度越大，其重要程度也越大，最终选出最重要的若干个句子作为摘要。

传统 TextRank 算法仅利用了文本信息，忽略了文本所包含的语义信息。针对这个问题，对关键句采用融合 Word2Vec 和 TF-IDF 的句向量表示，通过句向量间相似度计算，得到关键句的相似度矩阵，其中，句子 i 与句子 j 的相似度计算方式如公式(2)所示。W_{ij} 为两句向量的相似度，V_{ik}、V_{jk} 分别表示句向量 V_i、V_j 的第 k 维度的值。

$$W_{ij} = \frac{\sum_{k=1}^{n} V_{ik} \times V_{jk}}{\sqrt{\sum_{k=1}^{n} (V_{ik})^2} \times \sqrt{\sum_{k=1}^{n} (V_{jk})^2}} \tag{2}$$

通过句向量相似度矩阵，迭代计算出各句子的评分（即句子的重要程度），计算公式见公式（3）。

$$WS(S_i) = (1 - d) + d \times \sum_{j=1}^{n} \frac{W_{ji}}{\sum_{k=1}^{n} W_{jk}} WS(S_j) \tag{3}$$

其中，S_i、S_j 表示句子 i、j，$WS(S_i)$ 和 $WS(S_j)$ 表示句子 i，j 的权重，该权重在计算过程中不断迭代，其最终结果作为判断句子 S_i 是否作为摘要中关键句的依据。参数 d 为阻尼系数，其取值范围为 0~1，一般为设定为 0.85，表示无向图中某句子指向其他句子的概率。

3.4 基于 MMR 的补充摘要抽取

在期货研报自动摘要过程中，TextRank 算法挑选候选摘要句子时会侧重于挑选那些与很多句子相似度高的句子，遗漏一些重要却“势单力薄”的句子。MMR 算法能够平衡摘要中句子与研报原文的相关性和冗余度，从而更好地完成文本摘要任务。基于 MMR 进行补充摘要抽取的过程如下：首先，计算某个句子与研报的相似度，代表

表 1：基于 TextRank 的候选摘要抽取算法

算法 1　基于 TextRank 的候选摘要抽取算法
输入：经预抽取后的研报候选关键句向量集合，所需抽取的摘要句子数 n
输出：研报的候选摘要
开始
①　初始化相似度矩阵
②　对研报句向量使用余弦相似度计算构建关键句子的相似度矩阵
③　初始化研报句子的得分集合（句子得分越高则表明该句子与其他句子关联度越大，即该句子越重要）
④　利用公式（3）迭代计算关键句的得分
⑤　判断研报句子得分集合是否收敛，若未收敛则重复④
⑥　对句子评分集合进行排序
⑦　选出评分最高的 n 个句子，构成摘要
结束

该句子与研报的相关性，再计算该句子与当前摘要的相似度，代表摘要的冗余度，利用超参数 λ 平衡相关性和冗余度，并采用贪心方法优化句子得分，经排序选取句子组成补充摘要。平衡相关性和冗余度的目标函数如公式（4）所示。

$$MMR(Q,C,R)=\arg\max_{d_i\in C}^{k}[\lambda\, sim(Q,d_i)-(1-\lambda)\, sim(R,d_i)]\tag{4}$$

其中，Q 表示需要摘要的研报文本，C 表示研报文本分句后形成的句子集合，R 表示当前已抽取的摘要。$sim(Q, d_i)$ 表示研报文本中的某个句子 d_i 与 Q 的相似度，即句子 d_i 与研报文本的相关性；$sim(R,di)$ 表示研报中句子 d_i 与 R 的相似度，代表着摘要的冗余度；λ 用来平衡摘要的相关性和冗余度。

以下通过一个样例展示 MMR 算法在保持摘要句子相关性和简洁性方面的作用。

在该样例中，展示了研报部分原文、基于 TextRank 的候选摘要和基于 MMR 的补充摘要，其中，原文中标红的部分是参考摘要。可以看到研报中的关键句——“另一方面，17/18 年度国内棉花丰产，产量的大幅提升也在一定程度上保障了棉花供应”——与上文所述的“国内棉花进入去库存周期”的相似度较低，因而被基于 TextRank 的抽取算法所遗漏。而 MMR 算法因其关注了句子与研报的相关性，故抽取到了被遗漏的但在参考摘要的关键句子。

4 实验及结果分析

4.1 实验数据

经过调研发现，文本摘要领域尚未有公开的期货研报人工摘要数据集。针对这个问题，精心收集由不同的期货分析师撰写，涵盖了周报、月报、季报、半年报、年报，涉及棉花、白糖、花生、玻璃、纯碱、油脂等热门期货品种，且均为研究单一品种的研报 15 篇，由四位领域专家分别人工构建了上述研报的摘要，作为标准答案来评价方法的有效性。

4.2 评价指标

文本摘要常用的自动评价指标是 ROUGE。ROUGE 是一种面向 n 元词召回率的评价方法。其基本思想是将生成的自动摘要和参考摘要进行对比，通过统计它们之间重叠的基本单元数目来评价摘要的质量，例如 ROUGE-N、ROUGE-L 等。ROUGE-N 主要统计 N-gram 上的召回率，其计算公式如公式（5）所示。

$$ROUGE-N=\frac{\sum_{S\in(reference-summaries)}\sum_{gram_N\in S}Count_{match}(gram_N)}{\sum_{S\in(reference-summaries)}\sum_{gram_N\in S}Count(gram_N)}\tag{5}$$

其中，N 代表 N-gram 的长度（实验中取 1 和 2），$Count_{match}(gram_N)$ 即自动摘要和参考摘要中相同的 N-gram 的数量，$Count_{match}(gram_N)$ 表示参

原文: 2.1.2.短期库存高位，但现货资源依旧偏紧 国内棉花从 2016 年开始进入去库存周期，棉花产量的大幅缩小，使得近几年产消缺口持续维持在 300 万吨左右，除了配额限制下的进口，国储棉轮出成为了保障市场棉花供应的主要途径。但是就近两年的轮出数据来看，15/16 年国内棉花缺口在 167 万吨，16/17 年度缺口为 236 万吨，而这两年的国储棉轮出量分别为 266 万吨、322 万吨，轮储量大幅高于市场缺口量也就意味着有相当一部分的国储库存转化为社会库存。另一方面，17/18 年度国内棉花丰产，产量的大幅提升也在一定程度上保障了棉花供应。因此，短期棉花供应充足。
基于 TextRank 的候选摘要：国内棉花从 2016 年开始进入去库存周期，棉花产量的大幅缩小，使得近几年产消缺口持续维持在 300 万吨左右，除了配额限制下的进口，国储棉轮出成为了保障市场棉花供应的主要途径。
基于 MMR 的补充摘要：国内棉花从 2016 年开始进入去库存周期，棉花产量的大幅缩小，使得近几年产消缺口持续维持在 300 万吨左右，除了配额限制下的进口，国储棉轮出成为了保障市场棉花供应的主要途径。另一方面，17/18 年度国内棉花丰产，产量的大幅提升也在一定程度上保障了棉花供应。

考摘要中 N-gram 的数量。

ROUGE-L 主要计算最长公共子序列的重合率，其计算如公式（6）所示。

$$ROUGE-L=\frac{Count(LCS)_{max}}{Count(words)} \quad (6)$$

其中，$Count(LCS)_{max}$ 是自动摘要和参考摘要最长公共子序列的单词个数，Count(words) 表示自动摘要中的单词个数。

4.3 实验设计

为验证期货研报摘要自动抽取方法各个处理环节的有效性，分别从预抽取模块有效性、句子表征有效性和方法整体有效性三个方面进行实验。实验以 TextRank 为基线，各处理环节的算法及功能介绍如表 2 所示。

4.4 实验结果与分析

为验证预抽取模块的有效性，对未使用和使用预抽取模块的 TR、WTI-TR 算法进行对比实验，实验结果如表 3 所示。

从表 3 可以看出，未使用预抽取模块的 TR、WTI-TR 在 ROUGE 评分上均低于使用预抽取模块的评分，说明对研报文本使用预抽取模块，能够有效地缩小文本候选摘要的搜索范围，进一步提升后续各抽取环节的抽取质量。

为验证本文提出的句子表征方法的有效性，分别对 TR、仅采用 Word2Vec 的句子表示的候选摘要抽取算法 W-TR、融合 Word2Vec 和 TF-IDF 的句子表征方法的 WTI-TR 进行对比实验，实验结果如表 4 所示。

从表 4 可以看出，TR、W-TR 在 ROUGE 评

表 2：算法名称及功能介绍

算法	功能
TR	TextRank 算法
W-TR	基于词向量句子表征的 TextRank 候选摘要抽取算法
WTI-TR	融合了词向量及词频-倒文档频率的句子特征的 TextRank 候选摘要抽取算法
WTI-TR-MMR	WTI-TR 融合基于 MMR 的补充摘要抽取算法

表 3：有无预抽取模块的实验结果

预抽取	算法	ROUGE-1	ROUGE-2	ROUGE-L
无	TR	0.69	0.50	0.62
	WTI-TR	0.70	0.55	0.62
有	**TR**	**0.74**	**0.59**	**0.68**
	WTI-TR	**0.74**	**0.61**	**0.70**

表 4：融合 Word2Vec 和 TF-IDF 的句子特征表示的摘要自动抽取消融实验结果

算法	ROUGE-1	ROUGE-2	ROUGE-L
WTI-TR	**0.74**	**0.61**	**0.70**
W-TR	0.71	0.56	0.65
TR	0.74	0.59	0.68

表 5：最终实验结果

算法	ROUGE-1	ROUGE-2	ROUGE-L
TR	0.74	0.59	0.68
WTI-TR	0.74	0.61	0.70
WTI-TR-MMR	**0.76**	**0.65**	**0.71**

分上均低于 WTI-TR，说明融合词向量和词频－倒文档频率的句子特征表示能够更充分的表征句子的语义特征，从而进一步提升后续各抽取环节的抽取质量。

综合上述对比实验结果，本文提出的方法及各个处理环节的最终实验结果如表 5 所示。

从表 5 中可以看出，WTI-TR 较 TR 在 ROUGE-2、ROUGE-L 上分别提升了 2 个百分点，说明融合词向量和词频－倒文档频率的句子特征的抽取算法在研报文本摘要中有效。WTI-TR-MMR 算法在 ROUGE-1、ROUGE-2、ROUGE-L 上均表现更好，表明基于 MMR 算法的补充摘要抽取为 WTI-TR 的摘要结果提供了有效补充，进一步说明本文提出的期货研报自动摘要方法的有效性。

5 结论

为提高期货市场监管人员的获取研报关键信息的效率、降低人工工作量，本文提出了一种期货研报摘要自动抽取方法。采用 TextRank 与 MMR 算法相结合，融入领域词表、词频－倒文档频率、词向量表示等多种特征，以保证摘要与原文的相关性和简洁性。实验结果表明，该方法能够有效地利用多种特征，快速、高效、准确地从期货研报中抽取出期货市场监管人员关注的关键信息构成研报摘要，且不需要大规模人工标注的参考摘要和依赖于人工标注的参考摘要的训练过程。针对该方法，下一步的工作包括通过研报文本表示和多种抽取算法融合上进一步优化并提升方法的有效性。

参考文献：

[1] 百度百科 .
https://baike.baidu.com/item/%E8%84%B1%E6%B0%B4%E7%A0%94%E6%8A%A5/22673277?fr=aladdin.

[2] 邱震宇、彭南博、朱阿柯、陈雨辉、肖驰 . 华泰资管研报智能分析技术探索和实践 . 交易技术前沿 ,2022:4–11.

[3] Luhn H. The Automatic Creation of Literature Abstract[J]. IBM Journal, 1958. 2(2):159–165.

[4] Aizawa A. An information–theoretic perspective of tf–idf measures[J]. Information Processing & Management, 2003, 39(1):45–65.

[5] Mihalcea R, Tarau P. TextRank: Bringing Order into Texts[J]. 2004:404–411.

[6] 程琨 , 李传艺 , 贾欣欣 , 葛季栋 , 骆斌 . 基于改进的 MMR 算法的新闻文本抽取式摘要方法 [J]. 应用科学学报 ,2021, 39(03):443–455.

[7] Lin C Y. Rouge: A package for automatic evaluation of summaries[C]//Text summarization branches out. 2004: 74–81.

[8] 李金鹏 , 张闯 , 陈小军 , 等 . 自动文本摘要研究综述 [J]. 计算机研究与发展 , 2021, 39(1):1–21.

[9] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv e–prints, 2013.

I 信息资讯采撷

Information

监管科技全球追踪

监管科技全球追踪

8 月 3 日，中国互联网金融协会与中国人民保险集团股份有限公司在北广金融科技中心签署合作备忘录并揭牌“中互金 - 人保金融科技联合实验室”。

8 月 3 日，英国政府设立了欺诈管理局。欺诈管理局将由反欺诈专家和数据专家组成，使用一流工具和高级分析技术来帮助政府和公共机构保护公共资金。

8 月 9 日，国家工业信息安全发展研究中心发布《2022 年上半年我国网络安全产融合作发展报告》，报告阐述了 2022 年上半年我国网络安全领域非上市投融资总体情况，分析当前产业发展优势和存在的问题挑战，预判未来发展趋势，并提出相关对策建议。

8 月 15 日，美国能源部发布人工智能风险管理手册，用于识别人工智能（AI）风险的综合参考指南，包含风险缓解的建议措施，以支持负责可信 AI 的使用与开发。

8 月 17 日，工信部、财政部联合发布《关于开展财政支持中小企业数字化转型试点工作的通知》。《通知》明确，发展数字经济是把握新一轮科技革命和产业变革新机遇的战略选择。

8 月 17 日，数字金融市场基础设施联盟宣布，计划在英国启动数字货币试点，并采用数字英镑作为稳定币资产。

8 月 18 日，根据《数字经济伙伴关系协定》（DEPA）联合委员会的决定，中国加入 DEPA 工作组正式成立，全面推进中国加入 DEPA 的谈判。

8 月 25 日，由公安部第一研究所牵头的“基于法定证件的数字身份区块链技术研究与应用”正式启动并进入实施阶段。

9 月 2 日，国家主席习近平签署第一一九号主席令，《中华人民共和国反电信网络诈骗法》已由中华人民共和国第十三届全国人民代表大会常务委员会第三十六次会议通过，自 2022 年 12 月 1 日起施行。

9 月 5 日，国家计算机病毒应急处理中心（CVERC）针对西北工业大学遭受美国网络安全局（NSA）网络攻击事件发布调查报告。报告全面还原了相关攻击事件的总体概貌、技术特征、攻击武器、攻击路径和攻击源头，初步判明相关攻击活动源自 NSA“特定入侵行动办公室”（TAO）。

9 月 8 日，人民银行副行长范一飞出席第二届中国（北京）数字金融论坛并讲话。范一飞表示，将构建开放型数字人民币生态，加快推进数

字人民币体系的标准化建设。

9月9日，为加强宏观审慎管理，强化系统重要性银行监管，根据《系统重要性银行评估办法》，近期人民银行、银保监开展了2022年度我国系统重要性银行评估，认定19家国内系统重要性银行，其中国有商业银行6家，股份制商业银行9家，城市商业银行4家。

9月9日至10日，第28次亚太经合组织中小企业部长会议在泰国以线上线下结合的方式举行，加强数字包容性和绿色经济发展。

9月16日，美国科技政策办公室（OSTP）发布《美国中央银行数字货币系统的技术评估》（Technical Evaluation for a U.S. Central Bank Digital Currency System）专题报告。

9月19日，产业链供应链韧性与稳定国际论坛在浙江省杭州市开幕。

9月26日，国际清算银行（BIS）发布报告《金融科技贷对于美国小型企业信贷准入的影响》。

9月28日，人民银行联合多部门印发《陕西省铜川市普惠金融改革试验区总体方案》《浙江省丽水市普惠金融服务乡村振兴改革试验区总体方案》，及《四川省成都市普惠金融服务乡村振兴改革试验区总体方案》。

10月8日，人民银行副行长范一飞在《中国金融》2022年第19期中发表文章指出，实施乡村振兴战略是决胜全面建成小康社会、全面建设社会主义现代化国家的重大历史任务。

10月11日，美国拜登政府为积极防护国家网络安全，提出了十一项安全防护措施以“锁好国家数字大门”。

10月14日，全国信息安全标准化技术委员会归口的14项涉网络支付服务数据安全要求的网络安全国标获批发布。

10月18日，国际清算银行美洲代表处发布报告《数据共享API标准（账户聚合平台）》。调研结果表明，尽管数据共享对于提升央行效率与促进合作有普遍好处，但一直以来都存在着机构间协调、标准化及技术基础设施建设等问题。

10月26日，香港金融管理局联合国际清算银行香港创新中心、泰国中央银行、人民银行数字货币研究所及阿拉伯联合酋长国中央银行发布报告《多边央行数字货币桥：以CBDC联通经济》，阐述多边央行数字货币桥项目近期试行成果。

10月26日，英格兰银行（BoE）和英国金融行为监管局（FCA）宣布将与产业界合作，通过联合转型计划改变金融部门的数据收集方式，从而以更低成本获取所需数据。

10月28日，为规范网络产品安全漏洞收集平台备案管理，工信部印发《网络产品安全漏洞收集平台备案管理办法》。

2022年四季度《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 年四季度征稿主题如下：

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能提升的作用等。

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），请通过投稿信箱 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 12 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。

《交易技术前沿》编辑部

证券信息技术研究发展中心（上海）

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）

上海证券交易所
SHANGHAI STOCK EXCHANGE

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2023年1月，印刷单位为上海长鹰印刷厂。