大数据架构剖析及数据安全融合技术

刘晓军，武娟，徐晓青
（中国电信股份有限公司研究院，广东 广州 510630）

摘　要：从剖析大数据通用技术架构入手，总结数据处理的相关流程；分析和对比数据中台、分级构建两种模式的业务场景、管理架构、建设难度、技术要求、数据处理等，总结其适用范围和优缺点；最后聚焦多级架构模式下数据安全融合技术问题，提出相应的解决办法，为行业大数据构建选型提供有益的借鉴和参考。

关键词：大数据；技术框架；数据安全融合
中图分类号：TP399
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2020052

Big data architecture analysis and data security fusion technology

LIU Xiaojun, WU Juan, XU Xiaoqing
Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

Abstract: First of all, the general technical architecture of big data was deeply analyzed, and the relevant processes of data processing were further summarized, Through the analysis and comparison of the two modes of data middle platform and hierarchical construction, the relevant business scenarios, management architecture, construction difficulty, technical requirements, data processing, etc. were summarized, and their adaptation scope, advantages and disadvantages were summarized. The multi-level architecture mode was focused on according to the security fusion technology problems. Then the corresponding solutions were put forward. Finally, for the industry big data construction mode, useful reference was provided.

Key words: big data, technical framework, data security fusion

1　引言

随着信息技术从 IT 向 DT 时代的转变，数据已成为企业的核心竞争力，伴随着大数据开源架构的逐步成熟和生态体系的完备，大数据已进入实体经济的各行各业。

传统的企业在构建大数据平台时，会遇到例如如何构建大数据框架、各分公司/分支机构数据如何有效融合等问题。同时传统行业对大数据基础技术积累不足，架构规划稍有偏差，将严重影响整体平台建设进度、数据流通质量等，可能导致各个系统、各分公司/分支机构的数据无法有效共享和融合。

2　大数据技术框架分析

Hadoop 架构技术的成熟和生态的完备使其

收稿日期：2019−11−01；修回日期：2020−02−20

已成为了大数据平台架构标准配置。通过不同组件的搭建，构建从底层数据源、数据接入、数据预处理、分布式数据存储、分布式资源管理、分布式计算、数据建模和共享分发等一套完善的大数据处理架构。

大数据平台因功能定位、业务特点不同，存在不同技术组件搭建/组合方式，但对整体大数据平台而言，技术框架具有较为广泛的通用性。Hadoop 技术架构如图 1 所示。

构建大数据平台的关键，是实现海量数据的接入、处理、存储、计算和共享/分发一整套数据链服务；同时辅助有关资源管理、配置管理等组件，形成融合离线数据和实时数据的整套解决方案。

（1）数据源

大数据平台必须支持海量数据的平滑接入，包括海量非结构化和结构化离线数据、各种实时流式数据等。

（2）数据接入

离线数据支持“SFTP+Sqoop”等接入方式按照一定时长等策略批量导入及 Flume 日志采集等；流式数据支持通过 Kafka 实时消息队列导入等。

（3）数据预处理

即 ETL 阶段，可基于 Kettle 开源工具或满足业务特点的自研工具，完成数据的校验、清洗、入库等工作，实现将原始数据归整和优化，之后进入存储阶段。

（）4 数据存储

ETL 处理以后的数据，可采用直接导入 HDFS 或 HBase 列存储；并在数据入库时完成元数据采集，一般存储到 MySQL 等数据库中。

（5）资源管理

Hadoop 系统具有原生的资源管理和调度 YARN；Spark 架构的 Mesos 模式，Spark 也支持基于 YARN 的运行。

（6）分布式计算

需要支持海量离线数据计算和实时流式数据计算，其中海量离线数据采用 MapReduce、Spark 等架构完成；实时流式计算，可采用 Structured Streaming 或 Flink 等框架完成。

（7）数据建模

数据建模是指针对已经完成数据初步归整和初步计算的结果数据，进行相关业务场景的数据

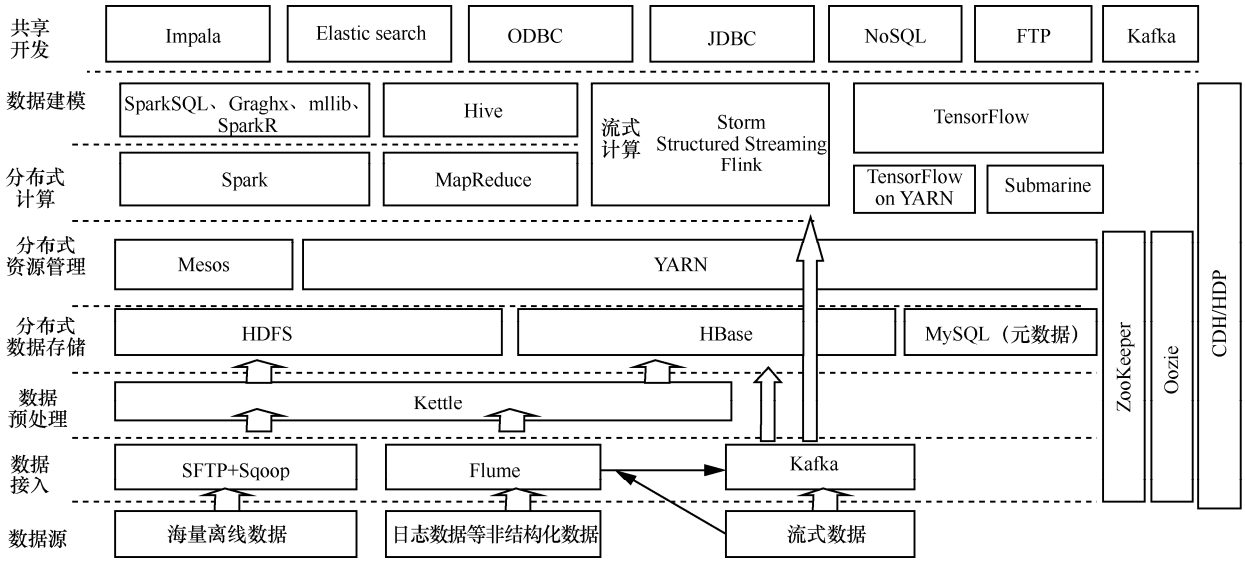

图 1　Hadoop 技术架构

统计、关联和价值挖掘等过程。主要包括离线计算、实时计算和 AI 3 种数据建模方式。离线计算建模，包括基于 Spark 的 SparkSQL、Graghx、mllib、SparkR 等建模方式以及基于 MapReduce 的 Hive 等汇总/关联建模方式；实时计算建模，包括 Storm、Structured Streaming 和 Flink 等方式实现实时汇总/关联；AI 建模方面，可采用 Submarine、TensorFlow on YARN（TonY）等工具，实现整合深度学习框架 TensorFlow/Caffe 等，基于 YARN 运行调用 Hadoop 分布式能力完成复杂的深度学习训练和预测。

（8）数据共享分发

面向平台数据用户，提供各种计算结果数据的检索、共享和分发功能，包括 Impala、Elastic search、NoSQL 等数据检索；ODBC/JDBC 等数据共享/分发接口；FTP、Kafka 等批量数据/实时数据分发工具等。

（9）资源管理配置

需要具备分布式计算任务协同管理和集群维护管理功能，包括 ZooKeeper、Oozie 等调度协同工具以及 CDH/HDP 等集群管理/配置/维护等框架。

大数据平台技术框架具有通用性，技术选型并不是构建大数据平台的关键所在，针对大中型企业如何有效构建大数据成为重点，尤其针对具有多种业务模式以及分支机构或分公司的集团级企业更加重要。

3 大数据构建模式

大数据构建过程是一家企业管理理念的集中展现，采用“大一统”集中模式还是“分而治之”的分级模式，各有优点。集中探讨大数据构建模式，重点针对“大一统”的数据中台和分级构建两种模式进行分析和对比。

3.1 数据中台模式

数据中台首先由阿里巴巴网络技术有限公司提出，并在其庞大的电商业务中获得了成功。数据中台是指通过数据技术，对海量数据进行采集、计算、存储、加工，采用统一标准和口径处理成标准数据，再存储为高价值的数据资产，最后具备集中面向客户提供高效数据服务一体化的大数据平台。

数据中台可以描述为将所有相关数据集中汇总，完成 ETL 处理后；按照统一标准提取元数据和完成规范化计算，存入 HDFS 或 HBase 中；通过汇总关联、机器学习建模和深度学习建模等方式，完成数据价值挖掘，再转存为高价值的数据资产；通过共享分发，实现针对数据资源或资产的查询/共享/分发等服务。

数据中台从其构建的模式来讲，具有数据集中、处理集中、计算集中、存储集中、挖掘集中和服务集中六大特征。数据中台示意图如图 2 所示。

（1）数据集中

即各条产品线、各个业务线和各个下属单位（分公司/分支机构）所有数据全部统一到一个平台，集中完成处理、计算、存储、挖掘和服务。

（2）处理集中

采用统一的数据处理规则，在统一数据质量控制下，将原始数据做标准化处理，并完成数据的统一规范化入库。

（3）计算集中

提供统一计算资源池（包括云资源），包括统一的离线和实时流式计算架构，并在统一的计算资源管理机制下进行各种计算任务的资源分配。

（4）存储集中

按照统一存储规则，完成数据存储，包括海量数据的分布式存储和备份、元数据存储和处理等；根据数据的冷热程度构建统一的数据仓库；在分布式计算或对外服务过程中，提供统一的缓存机制等。

（5）挖掘集中

基于统一数据建模管理和展示界面，利用统

图 2　数据中台示意图

一规范化的建模工具等，针对初步处理的标准数据进行多维的关联汇总分析、AI 预测分析等，实现对数据资源的资产化，完成数据价值的充分挖掘。

（6）服务集中

通过集中提供数据检索机制、数据开放接口，进行统一的资源共享和数据资产分发，并提供结果数据的可视化展示等。

数据中台遵循大中台、小前台的构建理念，一个综合平台将所有的数据接入、处理、存储、计算、挖掘等事项统统搞定；众多前台（各业务、各分支机构/分公司）直接在该大中台上获取数据资源和数据资产，不必承担任何存储、计算、挖掘等核心工作。

3.2　分级构建模式

分级构建的模式，不同于数据中台，一般按照业务或管理层级构建多个大数据平台，不同平台上分别具有数据接入、数据处理、数据计算、数据存储、数据挖掘和数据服务等功能，而且不同级别平台之间存在数据共享机制，但同一级平台间一般不进行数据共享/分发。上一级平台主要针对下属平台关联数据的汇总和转发，对更高层面数据的分析/汇总，一般不涉及原始数据初步处理和整理工作。

为了进一步阐述分级构建的特点，以业界比较常见的两级架构（集团、省分公司）来举例，直观阐述该种模式的特点，两级构建示意图如图 3 所示。

本模式属于业界较为常见的大数据构建模式，采用集团/省分公司两层格局，相互独立建设 Hadoop 集群，配备对应的数据接入、处理、计算、存储、挖掘和服务等功能。其特点如下。

（1）全国总平台/集团平台

- 数量特点：全国总平台处理的数据一般为各省分平台经过初步处理和价值挖掘的结果数据，各省相互关联而无法单独处理的少量原始数据。
- 主要任务：汇总各省上报的结果数据，分析各省相互关联的数据，挖掘全国性（或集团层面）数据价值，制订总体的数据接入/处理/关联规则等。
- 数据交互：作为全国或集团数据中转和管理中心，具有与各省分平台数据导入/下发/共享等功能，作为中介完成各省间平台数据的共享/分发。

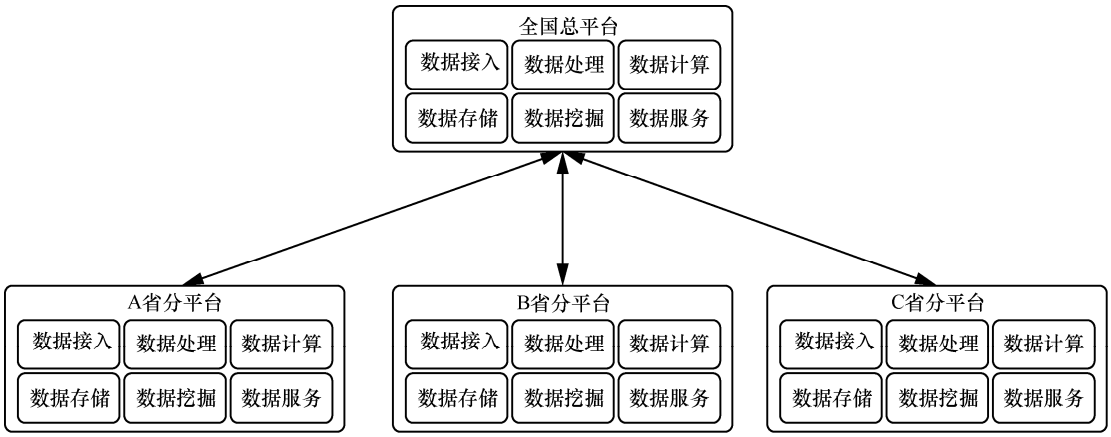

图 3　两级构建示意图

- 其他功能：除图 3 中的基本功能外，还需承担各省平台的对应管理、KPI 考核和各省分平台间数据共享/分发/融合等功能。

（2）各省分平台

- 主要功能：满足各省公司/分支机构自身的业务需求，集中独自处理各省数据，面向各省自有用户提供相关的数据服务。
- 数量特点：各省自有的原始数据及少量全国平台转发的相关省关联数据。
- 主要任务：汇总各省自有原始数据，分析并挖掘相关数据价值，面向各省用户提供数据服务，面向全国平台上报各省 KPI 等考核或全国关联的结果数据等。
- 数据交互：只与全国平台存在数据交互功能，各省平台间不存在数据交互关系。该平台作为一线支撑生产系统的服务平台，面向各种应用使用者具有大量数据分发/共享/检索/个性化建模、价值挖掘等数据服务功能。

大数据分级构建的体系，就是前端和平台一一对应，采用多个平台、多个相互独立的 Hadoop 集群的大数据分析体系。全国总平台/集团平台更多的是数据交互中介、规则管控和 KPI 考核等作用；与业务相关的海量计算和价值挖掘建模，更多体现在各省分平台上，各省分平台所承担的责任更大，在整个系统中作用更加突出。目前有些企业并没有建设全国或集团级平台，直接建设了多个隶属于省分公司或分支机构的省分平台，有关规则下发 KPI 考核等依靠原有文件传输方式来完成。

3.3　两种模式对比分析

数据中台模式具有便于数据统一管理、数据资产价值挖掘、避免重复建设等优点；分级构建模式，存在资源浪费和数据资产的严重分散，难以适应 DT 时代发展等问题。

随着 5G 环境下大带宽流式创新业务/IoT 海量链接，数据量呈指数级爆发，网络/终端测计算能力的提升及低时延的要求，边缘计算将在用户侧终端、设备接入网元、边缘计算节点等环节涌现，出现云边网端融合的新型数据处理架构。

数据中台、分级构建都具备融合边缘计算的能力，但从架构灵活性层面上，分层构建机制与云边端融合的处理机制具有较强契合性。二者的不同需从大数据本质综合衡量，通过两种构建方式适合的业务场景、服务灵活性、建设难度、技术条件和数据存储等特点的对比，实现在大数据部署过程中择优选用。

（1）业务场景

数据中台作为“大一统”的大数据处理平台，要具备贯穿整体业务和产品的数据主线 ID，例如

淘宝面向消费者方面的消费者 ID、面向商家的商家 ID 等，淘宝的业务基本是基于两大 ID 体系展开和延伸，所有数据集中能够确保在两大主线 ID 体系架构内有效地融合和价值提升。实体经济的很多大型企业，具有多种互不相关的业务体系，各业务间也没有直接关联的数据。例如航天、日用品零售、炼油等，数据之间关联基本没有，强制关联会出现大量没必要的解释、标注数据，严重影响后续数据查询效率和增加价值挖掘困难。在该种情况下，分级构建或分业务构建是更好的选择，不仅可以实现价值的共同提升，而且充分增加了业务灵活性。

数据中台架构体系，为了满足边缘计算需求，各边缘节点都需要符合数据主线 ID 的元数据体系或者按照统一的数据治理机制进行数据处理分析，但各个边缘节点能够接触的数据量和数量类型相对有限，统一的处理机制可能会限制边缘节点计算的灵活性和高效性。

分级构建模式，对不同的物联网接入设备以及不同别的数据类型，在云边网端融合体系中更适应边缘节点的个性化和灵活性，采用边缘节点根据自身业务场景和数据特点先行计算，再进行集中统计分析的方法。为了更好地适应边缘节点，分级构建模式需要改进固有的地域、层级限制，采用面向各个数据处理中心的业务场景融合性向更高的分级演进。

（2）管理架构

从本质上讲，大数据构建模式是企业管理理念的集中展现。互联网扁平化的管理模式，数据集中和统一挖掘的紧迫性和时效性更强。对于传统企业，为了便于本地化经营和不同业务线差异化管理，采用多级分层管理模式，普遍存在不同业务或不同区域之间不对应的情况，企业数据理解、数据规范和数据管理等方面难以短时间内达成统一。

对于存在多层管理架构，具有很多不同业务分公司的传统大型企业，采用分级构建模式，分业务/分公司自行搭建大数据平台相比，构建统一数据中台模式会更加适合。

同时针对 5G 时代业务的物联网场景，分领域构建大数据处理机制，可能更加适应业务的多样性和场景的多变性。

（3）建设难度

数据中台采用数据集中统一模式，必须构建超大型的 Hadoop 集群，要确保灵活资源调配和数据快速获取等能力，建设难度颇大。如果仅采用开源的 Hadoop 系统搭建，不采用任何优化处理，当整体集群服务器达到数千台规模，整体的计算性能将难以通过增加节点方式提升。

分级构建机制的数据相对较为分离，单个大数据平台规模相对数据中台具有明显的下降，一般不会达到 Hadoop 开源系统服务的规模边界，建设难度较小，也较易实现。

融合边缘节点计算方面，分级构建机制能够较为顺畅融合，更加能够突出边缘节点的个性化和灵活性，相对数据中台模式建设难度相对也较小。

（4）灵活服务

数据中台的数据统一管理和统一服务是核心，如单个业务方向或者领域需要修改元数据或更改数据表字段很难，甚至影响到整体的统一性。

分级构建模式采用较小规模大数据平台模式，具备业务强相关性，针对元数据的调整和字段的添加等，对整体数据影响不大，具备较为灵活的数据规则修改特性，能够有效地响应个性化的数据服务要求。

（5）技术条件

数据中台在技术或管理理念角度对行政管理领导提出考验，同时在整个数据中台构建过程中一旦确定前行则无法停止或后退，这就需要具备大数据集群设计、Hadoop 底层优化、各组件开发、数据治理、数据管理等全方位的大数据人才团队，

任意一个环节不能出现纰漏。

分级构建模式具备较为灵活的建设模式，可以采用“小步快跑”方式搭建，整体损失可控，技术要求相对较低，具有逐步分模块开发和完善的可能。

（6）数据处理

数据中台一般处理的数据为 PB 级别，不仅有巨量数据生命周期管理，还要针对不同数据冷热程度采用差异性存储策略，跨机房、跨区域、低时延调用和分布式计算，需完成较为复杂的技术处理，一般需采用难度较高的批流一体化的处理模式。

分级构建模式数据处理量相对数据中台具有数量级的降低，一般不需要跨机房或区域的数据存储、调用和计算等，通常的离线数据处理采用类 SQL 模式即可解决，相对数据处理技术要求较为简单。

针对数据中台和分级构建两种模式，通过上述 6 个方面的对比，可以大致得出的结论见表 1。

表 1　两种模式对比

对比项	数据中台	分级构建
业务场景	具备贯穿整体业务和产品的数据主线	各业务数据间缺乏直接或有效关联性
管理架构	互联网扁平化	多级分层管理架构
建设难度	难度较大、需优化 Hadoop 底层	难度较小、开源 Hadoop 生态可满足
灵活服务	需要为统一性做些牺牲	小平台、灵活性较高
技术条件	需要全方位的大数据顶级人才团队	边学习、边实践，逐步开发完善
数据处理	PB 级、难度颇大	TB 级、大多类 SQL 模式

从以上分析可知，数据中台和分级构建两种模式都有其适用范围。数据中台更加适合具备扁平化管理、贯彻业务/产品的数据主线、较高水平大数据人才团队和较强数据处理技术的公司；分级构建模式，较为适合多级分层管理、业务较为分散、技术水准有待提升和需要逐步学习/实践的公司。

对于分级构建模式，为了实现数据价值的最大化，需将部分全域关联数据进行有效融合，但其数据调度和融合的安全性就成为了难题。多方安全计算技术已成熟，但其作为处理多方非可信数据源的技术相对较为复杂，针对多级构建内部可信数据来说，通过较为简单的数据同步加密机制来实现，以便减少处理的复杂程度、提高计算效率。

4　分级架构数据融合技术

针对分级构建模式，各省分和全国/集团级平台均可为可信的数据源，可采用构建黑盒式的密码分发和共享机制，进行有效的数据融合，从根本上避免敏感数据泄露，有效促进数据融合和价值挖掘。

4.1　功能实现架构

这种技术是基于一套大数据平台体系，多方可信数据源统一加密入库、加密融合和面向管理人员屏蔽敏感信息；同时提供在应用调用过程中黑盒式解密，在敏感数据不曝光的情况下支撑相关应用的价值挖掘，既满足整体数据的安全性，又能有效支持挖掘数据价值。

在功能架构方面，采用数据安全统一管理、数据源同步加密和应用价值挖掘分离等机制实现。整体架构如图 4 所示，主要包括数据保护管理中心、数据加密前端和业务解密模块等。

（1）数据保护管理中心

位于全国或集团大数据平台，提供数据源加密/业务解密管理、前端监控、保护规则设置、数据使用轨迹管理等功能。

- 数据源管理：实现对各省分平台多方数据源接入验证，数据源对接接口、共享模式等进行管理和更新，确保实现数据源的统一安全保护。

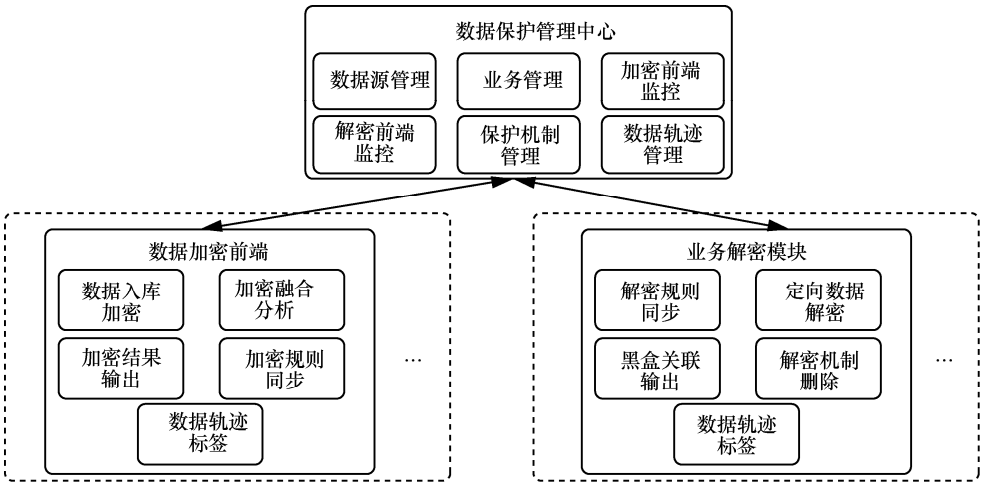

图 4　数据融合技术功能框架

- 业务管理：实现对数据融合业务应用的统一管理，包括业务登记、对接方式等。
- 加密前端监控：针对内置于各省分平台的数据源加密前端进行监控，包括版本更新、身份认证、异常监控等功能。
- 解密前端监控：针对内置于业务的解密模块进行监控，包括版本更新、身份认证和异常监控等功能。
- 保护机制管理：包括可灵活自动调整的加密算法、保护机制下发和同步、保护机制历史遍历等功能。
- 数据轨迹管理：通过获取各数据源和相关业务的使用情况，实现对行为轨迹记录，以便统一安全审计和追踪数据违规使用情况等。

（2）数据加密前端

位于各省分大数据平台，提供数据加密规则同步、入库、融合、输出、加密，数据使用轨迹标签等功能。

- 加密规则同步：实时获取保护管理中心制订的统一加密规则，并记录该规则开始使用的相关时间戳。
- 数据入库加密：所有原始数据都需要通过数据保护管理中心预定的统一规则，进行数据脱敏加密后方可入库，确保底层的 HDFS 文件和与 HBase 文件相关的敏感信息都是加密状态。
- 加密融合分析：针对完成加密入库数据，按照对应的计算规则进行内部的融合分析，实现加密数据无损融合分析。
- 加密结果输出：完成加密数据分析计算后，输出对应计算结果，其中敏感数据仍然保持加密状态，避免用户查询时有所泄露。同时该功能还包括按照全国/集团平台要求将对应加密全域关联数据上报等功能。
- 数据轨迹标签：针对敏感数据内部融合、计算、查询过程中的任何操作，都进行对应的轨迹标签，以便追踪数据处理过程。

（3）业务解密模块

位于全国或集团大数据平台，提供业务解密规则同步、定向融合数据解密、黑盒管理输出、白盒数据保护、数据轨迹标签等。

- 解密规则同步：按照业务需求驱动模式，当通过认证的业务具有解密需求时，按照所需数据生成入库时间，获得对应时间戳的解密规则。
- 定向数据解密：在业务解密计算过程中，通过对应的解密算法针对各省分平台上传加

密数据进行解密，业务或平台管理的相关人员不得人工干涉解密过程或人工查看。

- 黑盒关联输出：完成的解密文件，隐含在平台业务处理逻辑中，任何人都不得进行查询或删改，仅可直接关联相关业务使用场景，例如对接短信接口向特定用户发送短信，但用户电话号码一直处于加密状态等。
- 解密机制删除：当完成相关业务处理后，相关的解密机制对应删除，并确保不留任何痕迹。
- 数据轨迹标签：记录业务解密处理和使用过程中，相关的操作情况和使用轨迹等。

在上述安全性保障前提下，数据处理部分将采用边缘计算的理念将相关的本地化计算放到各省分平台，按照数据就近处理机制，完成本地化的数据加密入库/融合分析处理；跨各省分平台的数据融合部分，通过完成加密/解密机制放到全国或集团大数据平台来完成。

4.2 技术实现流程

本方法能够有效支撑采用多级构建的大数据平台，各省分平台间全域关联数据遵循统一的安全保护机制上报，并在全国大数据平台上基于业务驱动、完成黑盒解密价值挖掘。主要包括数据加密、数据解密等技术实现流程。

数据加密实施流程如图 5 所示，这一部分主要在各省分平台内部实现。

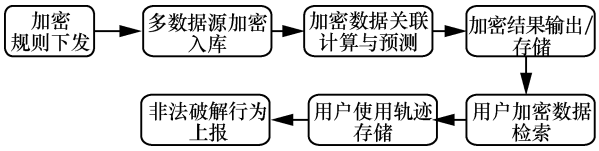

图 5 数据加密实施流程

（1）加密规则下发

全国/集团平台管理中心将定期更改的机密规则，可灵活采用 Base64 编码、MD5、SHA1、SHA256、SHA512 等多种加密算法以及对应时间戳同步下发到各省分平台等多方可信数据源。

（2）多数据源加密入库

按照统一的加密规则，各省分平台对多个数据源原始数据进行对应的加密处理后入库。

（3）加密数据关联计算与预测

完成原始数据加密入库后，各省分平台在加密状态下进行数据的关联计算与预测分析。

（4）加密结果输出/存储

完成关联计算的数据，以加密的状态存储、对外展示，并上报全国平台等。

（5）用户加密数据检索

各数据源的管理人员，针对数据进行检索时，相关的敏感数据以加密的方式展现。

（6）用户使用轨迹存储

用户检索、查找对应敏感数据使用轨迹，通过统一存储模式进行对应记录。

（7）非法破解行为上报

当数据源的加密前端，检测到非法入侵盗取数据或者非法破解时，及时上报到全国平台管理中心。

业务解密相关流程如图 6 所示，该部分主要在全国/集团平台内实现。

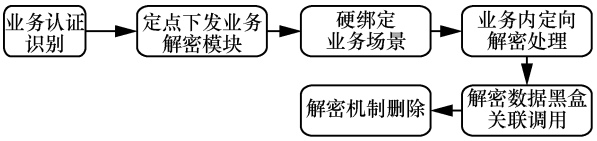

图 6 数据解密流程

（1）业务认证识别

需要调用各省分平台上报的存在全域关联数据的业务，首先完成在全国平台的管理中心识别和认证，方可进行相关各省分平台多方数据的解密处理和融合计算。

（2）定点下发业务解密模块

针对需要进行解密处理的业务，管理中心直接定位所需数据，包括时间段特征等，定点下发业务相关数据对应时间戳的解密模块。

（3）硬绑定业务场景

根据原来预设的相关接口与业务处理过程，

完成自动化对接，屏蔽人为干预。

（4）业务内定向解密处理

基于全国/集团平台，在业务处理过程内部，采用严格的进程权限管理，以禁止用户操作和查看，进行定向数据解密处理。

（5）解密数据黑盒关联调用

基于全国/集团平台，采取业务黑盒捆绑模式下的关联和调用解密。

（6）解密机制删除

相关业务完成对应的处理后，在全国/集团平台完成解密机制和解密后的数据全部定点自动清除。

在上述统一的安全保护机制下，能够基于密码同步共享、业务黑盒解密融合等方式，做到在分级构建下各省分平台具备关联的可信数据，安全上报到全国/集团平台，并实现在全国/集团平台基于业务捆绑的黑盒式解密融合计算，提供一体化的分级构建大数据系统相关数据的安全共享/融合/分析，而且相对常用的多方安全计算更加简便，并可实现与业务场景的有效融合。

5　结束语

本文通过多维度重点剖析数据中台和分级构建两种大数据建构模式，并尝试总结其适应范围；同时针对分级构建存在的数据安全问题，提出一套关联数据安全融合/分析等安全保护机制。随着 5G 等新技术的应用、边缘计算广泛应用，采用中心平台+多个边缘节点计算分级架构将成为大数据平台的主流，数据中台和分级构建两种模式也会在发展中不断借鉴和相互融合。

参考文献：

[1] 苏洋, 刘晓军, 唐勇. 游戏大数据平台研究与实践[J]. 电信科学, 2014, 30(10): 21-26.
SU Y, LIU X J, TANG Y. Research and practice of big data platform of games[J]. Telecommunications Science, 2014, 30(10): 21-26.

[2] 高强, 张凤荔, 王瑞锦. 轨迹大数据:数据处理关键技术研究综述[J]. 软件学报, 2017, 28(4): 959-992.
GAO Q, ZHANG F L, WANG R J. Big data: a review of key technologies in data processing[J]. Journal of Software, 2017, 28(4): 959-992.

[3] 曹珍富, 董晓蕾, 周俊, 等. 大数据安全与隐私保护研究进展[J]. 计算机研究与发展, 2016, 53(10): 2137-2151.
CAO Z F, DONG X L, ZHOU J, et al. Research advances on big data security and privacy preserving[J]. Journal of Computer Research and Development, 2016, 53(10): 2137-2151.

[4] MCAFEE A, BRYNJOLFSSON E. Big data: the management revolution[J]. Harv Bus Rev, 2012, 90(10): 60-66.

[5] LYNCH C. Big data: How do your data grow?[J]. Nature, 2008, 455(7209): 28-29.

[6] XUELONGL I, HAIGANG G. A survey on big data systems[J]. Scientia Sinica Informationis, 2015, 45(1): 35-38.

[7] 陈杰, 苏洋, 唐勇. 手机游戏大数据实时计算框架研究与实践[J]. 移动通信, 2016, 40(5): 79-84.
CHEN J, SU Y, TANG Y. Research and practice on real-time big data computing framework for mobile games[J]. Mobile Communications, 2016, 40(5): 79-84.

[8] 刘晓军, 武娟, 徐晓青. 大数据平台 AI 赋能技术方案探讨与验证[J]. 广东通信技术, 2019(6): 12-16.
LIU X J, WU J, XU X Q. Discussion and verification of AI enabling technology scheme for big data platform[J]. Guangdong Communication Technology, 2019(6): 12-16.

[作者简介]

刘晓军（1977− ），男，中国电信股份有限公司研究院工程师，主要研究方向为云计算、互动媒体技术、核心网络大数据分析与人工智能等。

武娟（1977− ），女，中国电信股份有限公司研究院高级工程师，主要研究方向为云计算、互联网及电信网络技术、大数据、人工智能等。

徐晓青（1990− ），男，博士，中国电信股份有限公司研究院工程师，主要研究方向为互联网及电信网络技术、人工智能和大数据等。