

试飞数据平台数据架构设计与应用

邓国宝, 查晓文, 冯 灿, 张逸飞, 薛博文

(中国商用飞机有限责任公司 民用飞机试飞中心, 上海 200232)

摘要: 试飞数据是民机飞行试验的重要产物, 具有测量参数数量大, 数据体量大, 飞行试验数据与试飞任务信息关联性强等特征, 支撑飞机型号取证与设计优化等任务; 试飞数据平台数据架构对多源异构数据集成接入、多类形态数据存储管理、多种层次数据处理分析等技术进行了研究, 采用湖仓一体的关键技术和方法打造试飞数据全集; 试飞数据接入采用流批一体的数据处理技术, 融合 Spark 和 Flink 主流数据处理引擎, 具备试飞数据快速入库能力; 提出按秒聚合方法, 具备 PB 级多维度试飞数据压缩存储功能, 存储性能提升近 10 倍; 采用以秒为索引条件支持快速检索, 强化数据湖查询能力; 研究数据仓库技术, 设计试飞数据多层数据模型, 具备多维信息精细查询, 多层数据灵活钻取, 多功能自定义函数集成等功能, 并成功应用在某型国产民机的飞行试验数据管理中, 服务于试飞数据用户, 提高了试飞数据管理效率与试飞数据应用价值。

关键词: 试飞数据; 数据架构; 湖仓一体; 流批一体; 按秒聚合; 数据仓库

Design and Application of Data Architecture of Flight Test Data Platform

DENG Guobao, ZHA Xiaowen, FENG Can, ZHANG Yifei, XUE Bowen

(Flight Test Center Flight-Test Instrumentation Dept., COMAC, Shanghai 200232, China)

Abstract: Flight test data is an important output of civil aircraft flight tests, it has the characteristics of many measurement parameters, large data volume, and strong correlation between flight test data and test flight mission information, supporting aircraft type certification and design optimization tasks. The data architecture of the flight test data platform integrates multiple heterogeneous data sources, manages various types of data in different formats, and provides multi-level data processing and analysis functions, creating a unified data lakehouse for flight test data. The data processing technology used for flight test data integration adopts a hybrid approach of stream-batch integration, incorporating the mainstream data processing engines such as Spark and Flink, with the ability to quickly ingest the flight test data into the data platform. The platform proposes a method of aggregation at the second level, with PB-level multidimensional flight test data compression and storage capabilities, the storage performance is improved by nearly 10 times, supporting fast retrieval based on second-level indexing conditions and enhancing data lake query capabilities. Data warehousing technology is also studied, a multi-layer data model for flight test data is designed, supporting the fine-grained queries for multidimensional information, flexible drilling down into multiple layers of data, and integration of custom functions, which is successfully applied in the management of flight test data for a certain type of domestic civil aircraft, serving flight test data users and improving the efficiency and value of flight test data management.

Keywords: flight data; data architecture; lake house; stream-batch integration; second-level aggregation; data warehouse

0 引言

飞行试验是在真实环境中对飞机性能和功能进行测试的过程, 用于验证飞机设计的合理性及其性能指标是否符合相关条款规定, 试飞数据是开展飞行试验获得的最重要的产物^[1-2]。试飞数据主要用户是飞机设计工程师、试飞课题工程师等人员, 为数据分析提供数据资源^[3-6]。

传统试飞数据管理模式存在文件便查询、数据关联使用困难等痛点, 不能支持数据用户挖掘使用试飞数据价值。试飞数据平台数据架构采用新一代信息技术, 设计基于 Flink 和 Spark 的试飞数据流批处理引擎, 打造试飞数据快

速处理工具; 研究试飞数据时序特征, 提出试飞数据秒内聚合存储方法, 创新数据存储方式与数据检索形式, 压缩数据存储空间, 提升数据查询效率; 引进数据仓库和数据湖技术, 设计试飞数据仓库数据模型, 实现了试飞数据 PB 级存储管理与应用功能^[7-8]。

1 系统结构及原理

如图 1 所示, 试飞数据平台按照三层架构的指导思想, 从数据流向及应用的角度, 整体系统分为数据层、服务层和应用层。数据层具备多源异构数据集成接入、多类形态数据存储管理、多种层次数据处理分析等功能, 打造湖仓

收稿日期: 2023-06-05; 修回日期: 2023-06-06。

作者简介: 邓国宝(1988-), 男, 硕士, 工程师。

通讯作者: 冯 灿(1984-), 男, 博士, 研究员级高级工程师。

引用格式: 邓国宝, 查晓文, 冯 灿, 等. 试飞数据平台数据架构设计与应用[J]. 计算机测量与控制, 2023, 31(12): 271-276.

一体的试飞数据全集^[9-12]。

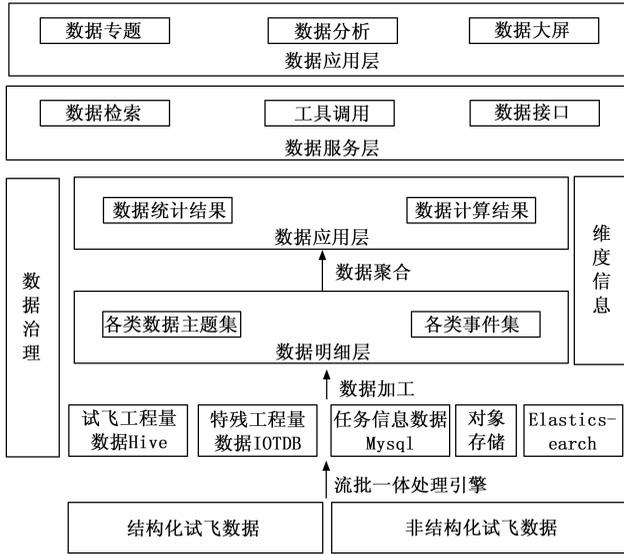


图 1 数据架构图

1.1 多源异构数据集成接入

多源异构数据集成接入功能通过数据同步软件、Socket、自定义接入方式(切片分发)将各种数据源数据同步到试飞数据处理与分析平台,数据源包含工程量数据、试飞原始数据、音视频数据、第三方系统同步数据、日志数据等^[13-14]。其中视频数据的同步、日志文件的采集、外部数据同步后的数据清洗的任务可通过数据集成工具提供支撑。针对不同的数据类型,接入方式也有所不同。工程量数据及原始数据作为系统中最重要业务数据,通过分片处理后传输到试飞数据平台后采用基于Flink流式处理引擎进行解析和处理并存储到数据湖中,存量历史数据采用基于Spark的批处理引擎实现数据解析并批量写入数据湖。入湖后数据通过基于Presto构建的试飞数据查询引擎提供数据的查询能力。主要功能如下:

1) 流处理:主要处理经分片后的原始数据,处理完成之后的数据写入到数据湖文件中,文件大小设定为200 M,文件滚动写完之后会自动新建下一个文件的写入,并将上一个写完的文件以load方式加载到数据湖中。

2) 批处理:主要处理数据湖中保存的原始数据文件。

3) 文件采集:针对接口采集到的半结构或结构化数据进行数据清洗等处理。采用Flume或Loader对分析中间结果文件、其他文件数据的处理^[15-16]。

1.2 多类形态数据存储管理

多类形态数据存储管理采用主流数据库组件,包含Hive、IoTDB、对象存储、MySQL、Elasticsearch等组件,存储结构化数据和非结构化数据,打造湖仓一体存储体系。

1) Hive:作为数据湖存储主要载体,用于存储经过流批一体处理引擎处理后的工程量数据。

2) IoTDB:存储分析过程中获取的时序数据段。

3) 对象存储:基于其海量、安全、高可靠的数据存储能力,存储非结构化数据(视频、文件)。

4) MySQL:存储结构化数据,包含从第三方系统同步的基础数据(试飞任务信息、试飞参数等),也用于存储试飞处理系统本身的基础数据(用户、权限、角色等)。

5) Elasticsearch:存储查询搜索关键信息,采用分词方式建立索引,支持智能检索快速查询功能。

1.3 多种层次数据处理分析

多种层次数据处理分析功能采用数据仓库技术,根据数据粒度设计数据贴源层、数据明细层和数据应用层,涵盖原始数据存储、分类数据管理、聚合数据应用等不同数据管理功能^[17-18]。按照数据治理方法,开展数据抽取、转换、格式统一、元数据管理、主数据管理、数据目录、数据地图、数据血缘等全生命周期操作,在数据的质量、数据标准、数据模型等维度都采用统一的管理。并发布数据服务,构建基于业务场景的数据集,提供各类主题明细、各类主题数据集市,支撑上层多样化的数据应用^[19]。

2 系统软件设计

2.1 软件设计思路和编程方法

试飞数据平台数据架构软件包括流批一体试飞数据处理引擎、试飞数据秒内聚合存储软件和试飞数据多层数据模型管理软件^[20]。

2.1.1 流批一体试飞数据处理引擎

流批一体试飞数据处理引擎基于大数据主流处理工具Spark批处理计算引擎和Flink流处理计算引擎,搭建数据处理服务器集群,立足试飞数据特征,开发试飞数据分布式处理软件。基于Spark的分布式批处理系统形成150 GB/h处理能力,满足大批量历史原始数据和工程量数据集中处理需求,具备系统拓展能力,能按照数据需求,实现有效扩容。基于Flink的分布式流处理系统具备1 GB/s处理能力,主要面对新增试飞任务产生的试飞数据,具备边同步边传输边处理边入库功能,也能够按照服务器数量扩容实现处理能力线性拓展。流批一体试飞数据处理引擎处理后的数据按照一定格式写入试飞数据湖,进行集中统一存储。

2.1.2 试飞数据秒内聚合存储软件

传统试飞数据是按照(K, V)方式存储,即(时间戳,数据值)存储试飞工程量数据。此种数据存储方式保留了数据原子粒度,能够支持精确查询到某个数据点。但是一个参数工程量数据存储规模根据采样率一般为百万级到亿级,存储内容较多,数据库设计时需要考虑各种分区设计,且数据查询时耗时较长,数据压缩程度也不高,为解决上述问题,本文提出了试飞数据秒内聚合存储方法,精简数据库存储记录数,建立时间索引为检索条件,提高数据压缩比,提升数据查询效率。

试飞数据秒内聚合存储方法:试飞工程量数据以秒为单位,将同一秒内数据合并为一条数据,该条数据内的每个数据点以逗号分隔并以字符串的形式进行存储。该条数

据只存储起始时间, 每个数据点根据采样率计算时间间隔后得出具体时间。

t_0 Value₁, Value₂,, Value_n

试飞数据湖内采用 hive 存储秒内聚合的试飞工程量数据, 按照试飞架次存储试飞工程量数据。数据存储方式为时间列和多个参数列, 时间间隔根据数据采样频率进行划分。建立数据存储表结构, 按单个 txt 文件为一张表的形式存储到 hive 中, 使用 orc 格式进行保存并使用 Snappy 方式进行压缩, 以减小数据的存储空间, 每个表中包括时间列和多个参数列。执行查询时使用 presto 进行基于内存的并行计算查询。

按照存储范式要求, 需要考虑特殊极限情况, 对于参数采样率过大较大或者参数数值长度较大的情况, 造成秒内数据点较多, 导致单条数据占用空间过大, 这会降低系统的存储和查询效率。针对此种特殊情况, 采用时序数据库方式进行数据存储。时序数据库可以提供数据收集、存储和分析等服务, 同时也支持跨设备和测量的时间序列数据的时间对齐、时间序列字段的计算(频域转换)和时间维度的丰富聚合函数支持。为了减少这部分的影响, 对于超过 8 192 频率的数据, 和数据类型为数据块的数据存储到 IOTDB 中, 此类型的数据时间精确到微秒值。

2.1.3 试飞数据多层数据模型

试飞数据平台根据参照数据仓库功能设计多层数据模型, 构建试飞数据仓库数据贴源层, 数据细节层、数据服务层等功能数据层, 并建立维度信息管理层, 统一存储数据仓库元数据信息, 描述各功能数据层库表元信息。

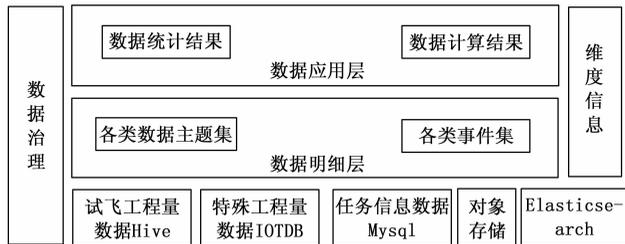


图 2 数据仓库

数据贴源层包含 Hive 数据库、时序数据库、关系型数据库、对象存储组件以及搜索数据库等, 主要存储试飞工程量数据、任务信息数据、文件数据以及检索关键信息。

数据明细层基于分析型数据库, 存储数据治理后的各类数据主题集和事件数据集, 提供高质量分类分主题中间数据。包含核心参数概览表、高频参数概览表、科目实验点明细表、架次飞行事件信息明细表、架次飞行时刻油量明细表、高频 FFT/小波等分析结果表、以及动力故障字专题架次结果明细表、CAS 告警专题架次分析结果明细表、飞行履历信息和容差架次统计结果明细表、刹车系统故障诊断架次分析结果明细表等数据表。

数据应用层存储经过数据治理后的聚合统计信息, 支持数据应用软件直接进行数据可视化展示等功能。包含跨

架次任务的动力故障字专题结果统计表、CAS 告警专题结果统计表、科目应用分析结果统计表、飞行履历信息和容差统计结果统计表、飞机交付关键状态参数预警结果统计表、大客刹车系统故障诊断结果统计表以及实验点信息统计表等数据表。

维度信息存储各类元数据, 包含试飞数据业务元数据、技术元数据等元数据信息, 支持试飞数据查询、影响分析和血缘分析等任务。业务元数据主要为参数信息映射表、合成参数映射表、核心参数映射表、概览参数信息映射表、试飞大纲信息基础表、架次任务信息表、架次飞行时间标签表、专题规则等数据表, 技术元数据主要为数据表描述信息。

2.2 软件实现流程

2.2.1 流批一体试飞数据处理引擎软件实现流程

如图 3 所示, 流批一体试飞数据处理引擎接收试飞原始数据和试飞工程量数据, 经过 Flink 进行流计算和 Spark 进行批计算后的结果数据写入试飞数据湖。

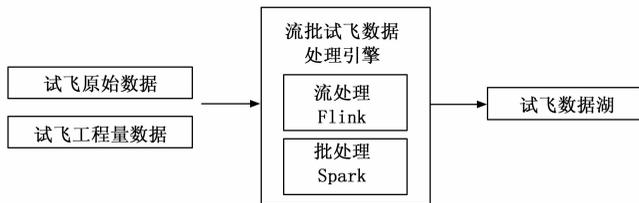


图 3 流批一体试飞数据处理引擎流程图

如图 4 所示, 流处理通过 Flink 接入通过 Kafka 传输的试飞原始数据流, 来完成原始数据到 HDFS (hadoop distributed file system, 分布式文件系统) 的存储、原始数据的解析和解析结果到数仓的存储。

1) 流处理任务启动时, 会首先读取启动参数, 获取当前流处理任务对应的数据处理任务信息(该信息会在创建数据处理任务写入到 MySQL), 包括数据总线类型、数据存储目录、合成配置文件路径等信息; 读取合成参数配置文件, 生成用于数据解析的参数配置信息;

2) 根据指定的任务运行环境和参数配置信息, 创建 ODS (Operational Data Store, 贴源数据存储) 表和参数映射信息;

3) 根据指定的任务运行环境、任务 ID 等创建对应的 Kafka 连接器, 接入对应的流数据;

4) 对接收到的数据进行解压(在同时对数据进行了压缩), 并将解压数据输出流, 到 HDFS 中;

5) 将解压的原始数据数拆分成一个一个的试飞数据包, 并对每个试飞数据包进行解析, 得到参数值 ParameterValue 的数据流;

6) 将 ParameterValue 的时间戳设置为水位线, 并根据前面生成的参数映射信息, 将 ParameterValue 按参数名所属的 ODS 表进行流分区 (KeyBy);

7) 设置 1 s 的窗口, 窗口聚合函数中对每个流分区的数

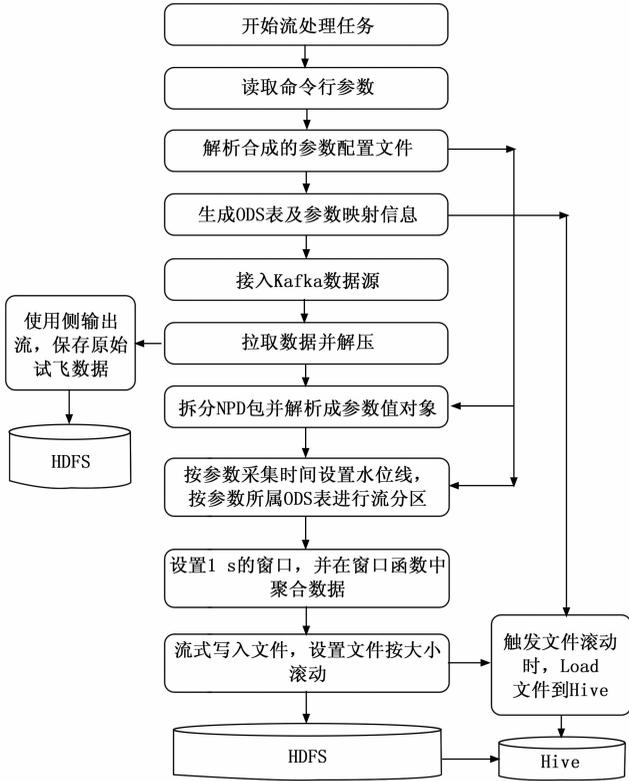


图 4 试飞数据流处理逻辑

据进行收集,并在窗口闭合操作函数中将这一秒的数据进行按秒聚合,这里会对不同频率的参数进行插值处理到同一频率;

8) 将窗口操作聚合的结果以流式写入文件,并设定文件超过 200 M 后进行滚动,创建新的文件写入,在触发文件滚动时,将前一个写满 200 M 数据的 HDFS 文件 Load 到对应的 Hive 表中。

批处理用于处理指定服务器同步过来的试飞数据处理并存储到数据湖 Hive 中,如图 5 所示。

1) 批处理任务启动时,会首先读取启动参数,获取当前流处理任务对应的数据处理任务信息(该信息会在创建数据处理任务写入到 MySQL),包括原始数据存储目录、合成配置文件路径等信息;读取合成参数配置文件,生成用于数据解析的参数配置信息;

2) 读取 ODS 表结构,删除后并重新创建和原来一样的 ODS 表,保留原来的参数映射信息;

3) 从 HDFS 中加载原始试飞数据文件流,将原始数据拆分成一个一个的试飞数据包,并对每个数据包进行解析,得到参数值的数据流;

4) 将参数值数据流按参数名进行分组,并在分组内将参数按时间排序,并按参数所属 ODS 表的频率进行对齐;

5) 将按参数名称分组处理的结果按相同的 Hive 表聚合在一起(再次按参数所属 ODS 表进行分组),在分组内进行按秒聚合;将分组内按秒聚合好的数据写入到对应的 Hive

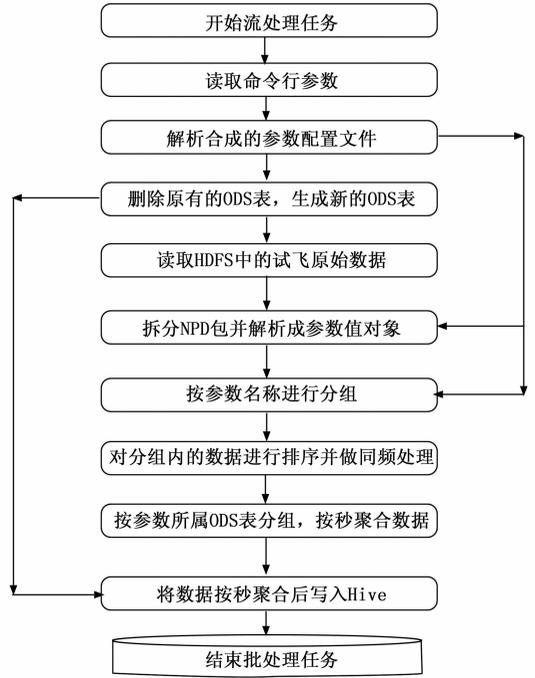


图 5 试飞数据批处理逻辑

表中。

2.2.2 试飞数据秒内聚合存储实现流程

如图 6 所示,试飞数据秒内聚合存储软件根据数据频率判断数据存储形式。当试飞参数采样率大于 8 192 Hz 时,数据采用顺序存到 IotDB 数据库中,否则数据存储到 Hive 数据库中,采用秒内聚合的方式和 ORC (Optimized Record Columnar, 优化后的列式记录) 格式。

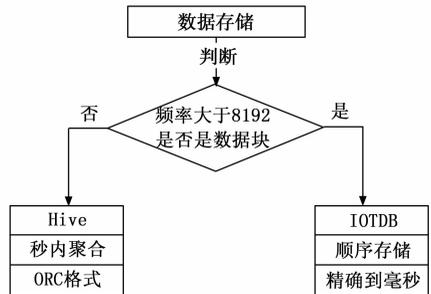


图 6 试飞数据存储流程

针对试飞数据平台存在的两种数据处理方式:流处理和批处理,在数据湖存储时的逻辑上有所不同。

流处理首先会根据试飞任务对应的参数信息生成多个不同参数采样率的 hive 表,这些表与该架次关联。流处理时,为了保证数据尽量实时地写入 Hive,同时避免频繁写入带来的小文件问题,系统采用 Flink 的流式写文件 Sink (Streaming File Sink) 先将数据写入 HDFS 中,并设定其滚动策略为最大 Part Size 200 M,也即每超过 200 M 后进行滚动产生一个新的文件进行写入。在每次发生文件滚动时系统会将写满 200 M 数据的文件 load 到对应的 Hive

表中。

批处理查询该架次在 ODS 表及参数映射信息, 然后根据试飞任务对应的参数信息生成多个不同参数采样率的 hive 表, 这些表与该架次关联。批处理时, 会将处理好的所有数据全部写入到 HDFS 中, 然后进行一次 load 操作将数据写入到 Hive 表中。

2.2.3 试飞数据多层数据治理流程

如图 7 所示, 试飞数据多层模型具备数据抽取、数据转换、格式统一、数据质量、数据标准以及数据资产等数据治理功能, 提供数据清洗等服务。

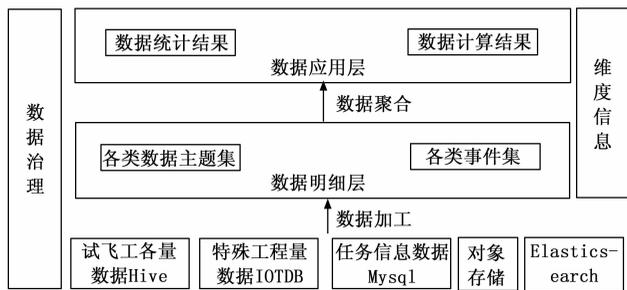


图 7 试飞数据多层数据治理流程图

1) 数据加工生成数据明细层。如图流程所示, 试飞数据平台存储在多类数据库中的多源数据使用离线数据处理工具定时执行任务, 将各类数据经过数据预处理、融合等数据加工方式生成数据主题数据表, 并结合维度信息表进行数据加工生成各类试飞时间表, 完成生成数据明细层, 实现业务过程建模进行驱动。根据业务特点构建不同业务明细表, 主要包含架次实验点按架次分析结果明细表、专题架次分析结果数据明细表、架次飞行事件信息明细表、高频参数 FFT/小波等分析结果明细表以及参数概览表等;

2) 数据聚合生成数据应用层。数据明细层数据通过数据聚合等方式, 进一步生成数据统计表和数据计算结果, 对外提供主题统计结果数据集, 主要包含专题数据结果集、科目应用分析结果集以及其他 BI 统计分析结果集等。

3) 维度信息支持数据建模, 丰富数据查询。维度信息表均统一放到数据库 schema 中, 维度信息包含解析参数信息映射表、合成参数信息映射表、核心参数信息映射表、概览参数信息映射表、试飞大纲信息基础表、架次任务信息基础表、系统专题规则及参数表、科目专题规则集参数表以及架次飞行时间标签表等数据表信息。

(1) 解析参数信息映射表是数据源层解析参数与 Hive 表字段映射关系, 该表每个试飞任务预计数据量在上万条信息, 多个试飞任务数据按试飞任务号进行分区存储。表主要字段有等试飞任务编号、参数名称、参数 uuid、值类型、采样率、hive 库名称、表名称、字段名称、参数编号、录入日期等信息。

(2) 合成参数信息映射表是已发布合成参数对应拆表存储的参数与 Hive 表字段映射关系, 该表为固定拆表, 每个机

型发布合成参数数量在千级左右, 因此每个机型表数据量在千级, 该表按机型、参数名称做为主键。表主要字段有等机型编号、参数名称、参数 uuid、值类型、采样率、hive 库名称、表名称、字段名称、参数编号、录入日期等信息。

(3) 核心参数信息映射表是已发布核心参数对应拆表存储的参数与 Hive 表字段映射关系, 该表为固定拆表, 每个机型发布核心参数数量在千级左右, 因此每个机型表数据量在千级, 该表按机型、参数名称做为主键。表主要字段有等机型编号、参数名称、参数 uuid、值类型、采样率、hive 库名称、表名称、字段名称、参数编号、录入日期等信息。

(4) 概览参数信息映射表是已发布核心参数、高频参数概览表对应拆表存储的参数与 sh 数据仓库表字段映射关系, 该表为固定拆表, 每个机型发布概览参数数量在千级左右, 因此每个机型表数据量在千级, 该表按机型、参数名称做为主键。表主要字段有等机型编号、参数名称、参数 uuid、值类型、采样率、表名称、字段名称、参数编号、录入日期等信息。

(5) 试飞大纲信息基础表是同步试飞大纲管理数据, 表主要字段有大纲编号、大纲名称、机型编号等信息。

(6) 架次任务信息基础表是同步试飞任务管理数据, 表主要字段有大纲编号、试飞任务号、机型编号、架机编号、试飞任务名称等信息。

(7) 系统专题规则及参数表是系统初始化录入动力故障字专题的参数和对应规则信息, 表主要字段有机型编号、专题名称、参数名称、参数 uuid、规则、起始值、结束值等信息。

(8) 科目专题规则及参数表主要字段有机型编号、科目名称、参数名称、参数 uuid 等信息。

(9) 架次飞行时间标签表是同步的时间标签管理已发布数据, 表主要字段有架次任务号、标签类型、标签名称、开始时间、结束时间。

多源异构数据经过数据加工生成数据明细层, 数据明细层经过数据聚合生成数据应用层, 实现数据质量提升, 满足面向数据应用的需求。

3 实验结果与分析

3.1 实验步骤和方法

试飞数据根据以下步骤开展数据处理。

1) 试飞数据经过流批处理引擎处理后生成中间数据存储到数据平台内存中;

2) 数据平台内存中中间数据按照秒内聚合存储方式生成若干条数据;

3) 按照秒内聚合的数据存储到湖仓一体的数据贴源层。

主要实验方法采用已有试飞数据进行实验验证, 分类分场景进行在数据存储和数据查询。数据存储方面按照高频参数、一般频率参数以及低频参数等各类情况进行存储性能验证。数据查询方面单表查询、多表查询以及不同频率下的 join 等场景开展验证。

3.2 实验数据

采用秒内聚合存储数据在空间占用方面,按秒聚合方式约占常规存储的 70%左右,存储性能优于常规存储。选择高频参数、一般频率参数以及低频参数开展测试,结果如表 1 所示。

表 1 数据存储对比表

序号	参数类型	传统存储	秒内聚合存储/GB	对比/%
1	高频	1.1 TB	378.5	33.6
2	中频	305 GB	101.7	32.5
3	低频	221 GB	68	29.6

在查询性能方面,分场景验证常规存储与秒内聚合存储的性能。

1) 在单表查询的场景下,按秒聚合比常规存储的查询性能更好;

2) 对于多表查询的场景,在相同频率下的 join 操作,按秒聚合查询性能更好,且随着查询数据的增大,按秒聚合相对常规存储的耗时更少;

3) 在不同频率下的 join 操作,常规存储无法进行查询。

表 2 数据查询验证

序号	查询条件	查询耗时/s		备注
		常规存储	按秒聚合	
1	单表单参数查询	35.94	9.82	单表 1 个参数,8 192 Hz, 30 min 数据
2	单表多参数查询	45.73	22.71	单表 3 个参数,8 192 Hz, 30 min
3	多表多参数查询	3.51	3.4	多表 6 个参数,64 Hz, 60 min
4	多表多参数查询	69.48	22.73	多表 20 个参数,1 024 Hz, 120 min 数据

3.3 实验结果分析

数据存储性能较常规存储有较好提升,后续在数据压缩方面可以继续优化。数据查询性能满足常用场景需求,在多表多参数大数据量情况下继续优化。

4 结束语

综合比较,按秒聚合查询性能优于常规存储,具备存储较省、查询较快的优势。试飞数据平台深度应用后,可以归集通用功能,整合为一款工业大数据领域通用工具,推广至试验数据等诸多领域,形成数据存储标准规范。

参考文献:

[1] AMADEO M, CAMPOLO C, MOLINARO A. Multi-source data retrieval in IoT via named data networking [C] // Proceedings of the 1st ACM Conference on Information-Centric Networking, 2014.

[2] AKERKAR R. Big data computing [M]. Crc Press, 2013.

[3] ROSSMANN M G, VAN BEEK C G. Data processing [J]. Acta Crystallographica Section D: Biological Crystallography,

1999, 55 (10): 1631-1640.

[4] BAKSHI K. Considerations for big data: architecture and approach [C] //2012 IEEE Aerospace Conference, IEEE, 2012: 1-7.

[5] YUAN B N, HUO Z, BAI X. Technology development and prospects of big data in flight test [J]. Computer Measurement & Control, 2015, 23 (6): 1844-1847.

[6] LIU Z Z. Model and simulation validation based on the data of the aero experimentation [J]. Journal of System Simulation, 2002, 14 (3): 281-284.

[7] TIAN F Z, HUANGPU D H, SHA Y Z. Summary of the development of a new generation of airborne data acquisition system [J]. Measurement and Control Technology, 2007, 26 (3): 16-18.

[8] 马卿云, 季航旭, 赵宇海, 等. 一种分布式异构带宽环境下的高效数据分区方法 [J]. 计算机研究与发展, 2020, 57 (12): 2683-2693.

[9] 丁岩, 杨万祥, 汪清, 等. 大数据统一 SQL 引擎研究与设计 [J]. 科技视界, 2019 (29): 1-4.

[10] ZHOU Z. Using heuristics and genetic algorithms for large-scale database query optimization [J]. Journal of Information and Computing Science, 2007, 2 (4): 261-280.

[11] GARCIA-GIL D, RAMIREZ G S, GARCIA S, et al. A comparison on scalability for batch big data processing on Apache Spark and Apache Flink [J]. BigData Analytics, 2017, 2 (1): 1.

[12] ARDAGNO C A, BELLANDI V, BEZZI M, et al. Model-based big data analytics-as-a-service: take big data to the next level [J]. IEEE Transaction Services Computing, 2021, 14 (2): 516-529.

[13] KUNE R. The anatomy of big data computing [J]. Software: Practice and Experience, 2016, 46 (1): 79-105.

[14] JI C. Big data processing in cloud computing environments [C] // 2012^{12th} International Symposium on Pervasive Systems, Algorithms and Networks, IEEE, 2012, 16 (2): 69-75.

[15] MORELLI E A, SMITH M S. Real-time dynamic modeling: data information requirements and flight-test results [J]. Journal of Aircraft, 2019, 46 (6): 1894-1905.

[16] KORSUN O N, POPLAVSKY B K, PRIHODKO S J. Intelligent support for aircraft flight test data processing in problem of engine thrust estimation [J]. Procedia Computer Science, 2017, 103 (1): 82-87.

[17] MUGTUSSIDS I B. Flight data processing techniques to identify unusual events [C] //Diss. Virginia Tech., 2000.

[18] LADWIG G, TRAN T. Linked data query processing strategies [C] // ISWC, 2010.

[19] REN Y. Data query mechanism based on hash computing power of blockchain in internet of things [J]. Sensors, 2019, 20 (1): 207.

[20] KOSSMANN D. The state of the art in distributed query processing [J]. ACM Computing Surveys, 2000, 32 (4): 422-469.