珠江数码大数据分析平台

需求规格说明书

项目编号：＿＿＿＿＿＿＿＿＿＿

项目名称：珠江数码大数据收视分析平台项目

填报单位：曙光信息产业（北京）有限公司

填报日期： 2015 年 11月 19日

验收单位：

验收日期：年月日（预）

年月日（终）

**文档修订历史**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 修订前版本 | 修订内容 | 完成日期 | 修订人 | 修订后版本 |
|  | 新建文档 | 2015.11.19 | *李晋钢* | 1.0.0 |
|  |  |  |  |  |
|  |  |  |  |  |

注：“草稿”状态的文档版本为0.Y.Z，Y≥0，Z>0，Y、Z的数值不断累加；

“正式发布”状态的文档版本为X.Y，X≥1，Y≥0，且X、Y值不断累加；

“正在修改”状态的文档指对“正式发布”后的文档进行修改，文档版本为X.Y.Z，其中X.Y同修改之前的文档版本号，Z>0，Z的数值不断累加

[1 项目概述 4](#_Toc461195576)

[1.1 项目定义 4](#_Toc461195577)

[1.1.1 项目名称 4](#_Toc461195578)

[1.1.2 项目背景 4](#_Toc461195579)

[2 需求概述 5](#_Toc461195580)

[2.1 子系统/功能模块划分 5](#_Toc461195581)

[2.2 需求索引 6](#_Toc461195582)

[3 数据接入模块 7](#_Toc461195583)

[3.1 数据接入调度框架 7](#_Toc461195584)

[3.2 收视维度表数据接入 8](#_Toc461195585)

[3.3 收视行为数据接入 8](#_Toc461195586)

[3.4 用户基础信息接入 8](#_Toc461195587)

[3.5 用户设备信息接入 9](#_Toc461195588)

[3.6 用户事件数据接入 9](#_Toc461195589)

[3.7 网络设备数据接入 9](#_Toc461195590)

[3.8 订购事件数据接入 10](#_Toc461195591)

[3.9 消费事件数据接入 10](#_Toc461195592)

[3.10 受理单数据接入 10](#_Toc461195593)

[3.11 支付事件数据接入 11](#_Toc461195594)

[4 数据分析服务模块 11](#_Toc461195595)

[4.1 订购产品分析 11](#_Toc461195596)

[4.2 订购事件分析 12](#_Toc461195597)

[4.3 消费金额分析 13](#_Toc461195598)

[4.4 消费ARPU分析 13](#_Toc461195599)

[4.5 支付金额分析 14](#_Toc461195600)

[4.6 支付人数分析 14](#_Toc461195601)

[4.7 支付次数分析 15](#_Toc461195602)

[4.8 受理单次数分析 15](#_Toc461195603)

[4.9 受理单人数分析 16](#_Toc461195604)

[4.10 收视时长分析 16](#_Toc461195605)

[4.11 人均收视时长分析 17](#_Toc461195606)

[4.12 观看人数分析 18](#_Toc461195607)

[4.13 观看次数分析 18](#_Toc461195608)

[4.14 到达人数分析 19](#_Toc461195609)

[4.15 忠诚度分析 20](#_Toc461195610)

[4.16 收视份额分析 21](#_Toc461195611)

[4.17 用户标签分析 21](#_Toc461195612)

[5 MQL数据分析模块 22](#_Toc461195613)

[5.1 过滤查询 22](#_Toc461195614)

[5.2 聚合分析 22](#_Toc461195615)

[5.3 SQL语法支持 23](#_Toc461195616)

# 项目概述

## 项目定义

## 项目名称

|  |  |
| --- | --- |
| **项目中文全称** | 珠江数码大数据分析平台 |
| **项目中文简称** | 大数据分析平台 |

## 项目背景

广州珠江数码集团有限公司（简称“珠江数码集团”）成立于1993年，是经广州市委、市政府授权，专门负责建设、维护和管理广州市行政区域内广播电视网络的主要运营机构。经过十多年的发展，珠江数码集团已成为华南地区最大规模的广播电视网络运营商之一，拥有有线电视用户超过260万户，宽带用户超过16万户，高清互动电视用户数正处于快速发展阶段。

珠江数码集团已建成完整覆盖广州各区（县级市）的有线传输与无线传输互为延伸、互为补充的广电宽带信息网络，实现了城区全程全网的双向覆盖，为广大市民提供有线数字电视、互联网接入服务（珠江宽频）、高清互动电视（Candytime甜果时光）、移动数字电视、CMMB手机电视、信息内容集成等多样化、跨平台的信息服务。

珠江数码集团正在从传统的有线电视运营商向“多媒体网络和信息服务运营商”作战略转型，集团下设综合服务、技术支持、市场、投资和分公司5大序列，共设15个部门、中心，7家分公司；旗下还拥有珠江在线多媒体信息有限公司和广州珠江数字电视及数字家庭技术应用研究院有限公司2家全资子公司，广州珠江数码番禺有限公司、花都有限公司、从化有限公司、增城有限公司4家控股子公司，珠江移动多媒体信息有限公司和南沙信息港有限公司2家参股子公司，已实现了集团化规模经营格局。

当前，珠江数码集团正紧紧把握“三网融合”所带来的发展机遇，加快高清交互数字电视的推广，将家庭电视机变成多媒体信息终端，为广大用户带来全新的数字媒体体验，并为广州信息化、数字化的发展提供新的途径，为政府部门架设通达市民的便捷通道。

由于众多新业务的推出和普及，需要建设一个大数据分析平台，通过这个平台可以整合集团各个大系统的所有和用户相关的数据，建立一套完整的数据分析及数据挖掘的系统平台，用于了解用户需求，促进各项业务的发展。近年来，随着国内媒介经营产业化、集团化的发展，收视率调查已越来越受到业界的重视。无论是电视机构运作，还是广告经营、节目评价，收视率都是备受关注的重要指标之一。为什么收视率会受到如此的重视呢？这应该从媒介经营本身说起。从表面看．媒介经营可以这样描述：一方面，媒介（电视台等）经营频道和节目受众收看电视节目；另一方面，媒介（电视台等）把时段卖给广告商，获取广告收入，广告商对广告产品进行广告创意和制作，把广告投放在购买的时段上，提供给广大受众，并从广告主身上获取利润。媒介经营的是电视频道和节目，与广告商／广告主交易的是播出时段（节目）。但仔细分析，播出时段（节目）并不能直接为媒介经营者和广告尚带来效益，媒介经营者最终效益的实现，是因为经过了“受众”这一环节。从本质上看，在媒介经营中，媒介与广告商之间交易的是受众，也就是附着在交易时段和节目上的“观众注意力”。

# 需求概述

## 子系统/功能模块划分

|  |  |  |
| --- | --- | --- |
| 子系统/模块名称 | 编号 | 主要功能 |
| 数据接入模块 | ZJSM-01 | 针对多源异构平台，实现跨平台的数据整合功能，支持平台包括Oracle、SqlServer、Excel、文本文件等。并实现对BOSS、新媒体、网维等数据的整合。 |
| 数据分析服务模块 | ZJSM-02 | 针对整合的多业务数据，提供聚合分析功能，并提供Restful数据访问接口。  分析的业务包括用户分析、消费分析、订购分析、收视分析、网维分析等。提供多维度全自由度组合分析，实现趋势、数值、排行等聚合分析。 |
| MQL数据分析模块 | ZJSM-03 | 实现MQL数据分析分服务，可通过编写简单的分析语句，进行业务分析扩展。提供Restful服务接口。 |
| 可视化分析模块 | ZJSM-04 | 可视化模块提供分析结果可视化服务功能，集成表格、曲线等多种展示方式。 |

## 需求索引

|  |  |  |
| --- | --- | --- |
| 所属子系统/模块 | 需求编号 | 需求名称 |
| 数据接入模块 | ZJSM-01-01 | 数据接入调度框架 |
| 数据接入模块 | ZJSM-01-02 | 收视维度表数据接入 |
| 数据接入模块 | ZJSM-01-03 | 收视行为数据接入 |
| 数据接入模块 | ZJSM-01-04 | 用户基础信息接入 |
| 数据接入模块 | ZJSM-01-05 | 用户设备信息接入 |
| 数据接入模块 | ZJSM-01-06 | 用户事件数据接入 |
| 数据接入模块 | ZJSM-01-07 | 网络设备数据接入 |
| 数据接入模块 | ZJSM-01-08 | 订购事件数据接入 |
| 数据接入模块 | ZJSM-01-09 | 消费事件数据接入 |
| 数据接入模块 | ZJSM-01-10 | 受理单数据接入 |
| 数据接入模块 | ZJSM-01-11 | 支付事件数据接入 |
| 数据分析服务模块 | ZJSM-02-01 | 订购产品分析 |
| 数据分析服务模块 | ZJSM-02-02 | 订购事件分析 |
| 数据分析服务模块 | ZJSM-02-03 | 消费金额分析 |
| 数据分析服务模块 | ZJSM-02-04 | 消费ARPU分析 |
| 数据分析服务模块 | ZJSM-02-05 | 支付金额分析 |
| 数据分析服务模块 | ZJSM-02-06 | 支付人数分析 |
| 数据分析服务模块 | ZJSM-02-07 | 支付次数分析 |
| 数据分析服务模块 | ZJSM-02-08 | 受理单次数分析 |
| 数据分析服务模块 | ZJSM-02-09 | 受理单人数分析 |
| 数据分析服务模块 | ZJSM-02-10 | 收视时长分析 |
| 数据分析服务模块 | ZJSM-02-11 | 人均收视时长分析 |
| 数据分析服务模块 | ZJSM-02-12 | 观看人数分析 |
| 数据分析服务模块 | ZJSM-02-13 | 观看次数分析 |
| 数据分析服务模块 | ZJSM-02-14 | 到达人数分析 |
| 数据分析服务模块 | ZJSM-02-15 | 忠诚度分析 |
| 数据分析服务模块 | ZJSM-02-16 | 收视份额分析 |
| MQL数据分析模块 | ZJSM-03-01 | 过滤查询 |
| MQL数据分析模块 | ZJSM-03-02 | 聚合分析 |
| MQL数据分析模块 | ZJSM-03-02 | SQL语法支持 |

# 数据接入模块

数据接入模块主要负责将数据定期导入到大数据平台，目前需要导入的数据源主要是存放在SQLServer上的收视数据，存放在Oracle上的BOSS数据和存放在MySQL上的网维数据等。按照导入的方式主要分数据的全量导入和数据增量的导入。

## 数据接入调度框架

|  |  |
| --- | --- |
| **需求名称** | 数据接入调度框架 |
| **需求编号** | ZJSM-01-01 |
| **优先级** | 1 |
| **功能描述** | 1. 定义数据接入流程，每一个数据源的接入定义为一个作业，可以通过配置文件定义多个作业。 2. 实现作业周期调度功能。 3. 支持异构数据源，实现对常用数据库、文本文件等的支持。 |
| **约束条件** |  |

## 收视维度表数据接入

|  |  |
| --- | --- |
| **需求名称** | 收视维度表数据接入 |
| **需求编号** | ZJSM-01-02 |
| **优先级** | 1 |
| **功能描述** | 1. 维度数据为全量导入 2. 配置维度数据全量导入作业，将SqlServer中的收视维度数据全量导出为本地文件 3. 将本地文件导入到Oracle数据库中，供收视数据事件数据关联处理使用 4. 每天通过自动任务，三次(6,15,23)导入全量数据表，每次导入作业均有日志记录 |
| **约束条件** | 源SQLServer中可以连接，且该表数据可访问 |

## 收视行为数据接入

|  |  |
| --- | --- |
| **需求名称** | 收视行为数据接入 |
| **需求编号** | ZJSM-01-03 |
| **优先级** | 1 |
| **功能描述** | 1. 配置收视行为数据作业，根据递增列，将该表的数据增量导出到本地文件 2. 将本地文件导入到Oracle数据库系统中，与维度数据和用户信息数据进行关联，关联结果写入大数据分析集群中 3. 每天通过自动任务，两次导入增量数据，每次导入作业均有日志记录 |
| **约束条件** | 源SQLServer中可以连接，该表数据可访问，且该表存在递增列 |

## 用户基础信息接入

|  |  |
| --- | --- |
| **需求名称** | 用户基础信息接入 |
| **需求编号** | ZJSM-01-04 |
| **优先级** | 1 |
| **功能描述** | 1. 配置用户基础信息接入作业 2. 从BOSS数据库中抽取用户基础信息，包括用户名、地址、所属分公司、用户状态等。将获取的数据同时写入Oracle数据库和大数据分析集群中。 3. 依据run\_time字段，增量抽取 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 用户设备信息接入

|  |  |
| --- | --- |
| **需求名称** | 用户基础信息接入 |
| **需求编号** | ZJSM-01-05 |
| **优先级** | 1 |
| **功能描述** | 1. 在BOSS数据库中抽取用户设备信息，抽取信息包括设备型号、序列号等内容。 2. 因每一个用户可以有多个设备，因此在大数据平台中采用嵌套数组的方式进行存储。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 用户事件数据接入

|  |  |
| --- | --- |
| **需求名称** | 用户事件数据接入 |
| **需求编号** | ZJSM-01-06 |
| **优先级** | 1 |
| **功能描述** | 1. 用户事件数据为用户注册后所有状态、品牌等变更的事件记录 2. 用户事件数据为全量同步数据。从BOSS用户历史表中抽取全量数据，构造用户状态时间区间，导入到大数据平台中。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 网络设备数据接入

|  |  |
| --- | --- |
| **需求名称** | 网络设备数据接入 |
| **需求编号** | ZJSM-01-07 |
| **优先级** | 1 |
| **功能描述** | 1. 网络设备数据包括网络的点评、流量、带宽等数据 2. 网络数据采用增量同步的方式 3. 每次同步新增网络设备数据，与用户设备信息和用户基础信息进行关联，导入到大数据平台中。 |
| **约束条件** |  |

## 订购事件数据接入

|  |  |
| --- | --- |
| **需求名称** | 订购事件数据接入 |
| **需求编号** | ZJSM-01-08 |
| **优先级** | 1 |
| **功能描述** | 1. 订购时间数据包括用户订购的时间、有效期、订购产品名称、类型、销售品等信息。 2. 订购数据采用增量同步的方式 3. 每次增量同步新增订购数据，与用户基础信息数据进行关联，导入到大数据平台中。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 消费事件数据接入

|  |  |
| --- | --- |
| **需求名称** | 消费事件数据接入 |
| **需求编号** | ZJSM-01-09 |
| **优先级** | 1 |
| **功能描述** | 1. 消费数据包括用户消费的时间、应收金额、账单类型、账单状态等信息。 2. 消费数据采用增量同步的方式 3. 每次增量同步新增消费数据，与用户基础信息数据进行关联，导入到大数据平台中。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 受理单数据接入

|  |  |
| --- | --- |
| **需求名称** | 受理单数据接入 |
| **需求编号** | ZJSM-01-10 |
| **优先级** | 1 |
| **功能描述** | 1. 受理单数据包括用户受理单的时间、受理单类型、受理单名称、备注等信息。 2. 受理单数据采用增量同步的方式 3. 每次增量同步新增受理单数据，与用户基础信息数据进行关联，导入到大数据平台中。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

## 支付事件数据接入

|  |  |
| --- | --- |
| **需求名称** | 支付事件数据接入 |
| **需求编号** | ZJSM-01-11 |
| **优先级** | 1 |
| **功能描述** | 1. 支付数据包括用户支付的时间、支付类型、支付名称、支付渠道等信息。 2. 支付数据采用增量同步的方式 3. 每次增量同步新增支付数据，与用户基础信息数据进行关联，导入到大数据平台中。 |
| **约束条件** | 源BOSS数据库可以连接，该表数据可访问 |

# 数据分析服务模块

数据分析服务模块接受数据分析服务请求，提取分析参数和指标信息，构造数据分析作业，在大数据平台中之行分析作业，返回分析结果。

本平台提供的所有分析流程均采用实时计算的方式，因此在数据同步到平台中后，即可获取分析结果。

## 订购产品分析

|  |  |
| --- | --- |
| **需求名称** | 订购产品分析 |
| **需求编号** | ZJSM-02-01 |
| **优先级** | 1 |
| **功能描述** | 将用户订购数据与用户基础信息数据关联，对给定时间区间内的有效订购产品进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、销售品名称、产品名称、片区、操作人员、分级订购目录等。  （2）趋势分析，统计指定时间范围下的有效订购产品总数量趋势。  （2）数值分析，统计指定时间范围下的有效订购产品总数量。  （3）排行分析，按指定字段分组，统计指定时间范围下的有效订购产品总数量排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、销售品名称、产品名称、片区、操作人员、分级订购目录等。 |
| **约束条件** | 消费数据正确存储在大数据分析平台中 |
| **备注** |  |

## 订购事件分析

|  |  |
| --- | --- |
| **需求名称** | 订购事件分析 |
| **需求编号** | ZJSM-02-02 |
| **优先级** | 1 |
| **功能描述** | 将用户订购数据与用户基础信息数据关联，对给定时间区间内的有效订购事件进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、销售品名称、产品名称、片区、操作人员、分级订购目录等。  （2）趋势分析，统计指定时间范围下的订购事件趋势。  （2）数值分析，统计指定时间范围下的订购事件总数量。  （3）排行分析，按指定字段分组，统计指定时间范围下的订购事件总数量排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、销售品名称、产品名称、片区、操作人员、分级订购目录等。 |
| **约束条件** | 消费数据正确存储在大数据分析平台中 |
| **备注** |  |

## 消费金额分析

|  |  |
| --- | --- |
| **需求名称** | 消费金额分析 |
| **需求编号** | ZJSM-02-03 |
| **优先级** | 1 |
| **功能描述** | 将用户消费数据与用户基础信息数据关联，对给定时间区间内的消费应收金额进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、消费项目、片区等。  （2）趋势分析，统计指定时间范围下的用户总消费金额趋势。  （2）数值分析，统计指定时间范围下的用户总消费金额。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户总消费金额排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、消费项目、片区。 |
| **约束条件** | 消费数据正确存储在大数据分析平台中 |
| **备注** |  |

## 消费ARPU分析

|  |  |
| --- | --- |
| **需求名称** | 消费ARPU分析 |
| **需求编号** | ZJSM-02-04 |
| **优先级** | 1 |
| **功能描述** | 将用户消费数据与用户基础信息数据关联，对给定时间区间内的消费应收金额进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、消费项目、片区等。  （2）趋势分析，统计指定时间范围下的用户平均消费金额趋势。  （2）数值分析，统计指定时间范围下的用户平均消费金额。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户平均消费金额排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、消费项目、片区。 |
| **约束条件** | 消费数据正确存储在大数据分析平台中 |
| **备注** |  |

## 支付金额分析

|  |  |
| --- | --- |
| **需求名称** | 支付金额分析 |
| **需求编号** | ZJSM-02-05 |
| **优先级** | 1 |
| **功能描述** | 将用户支付数据与用户基础信息数据关联，对给定时间区间内的支付金额进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。  （2）趋势分析，统计指定时间范围下的用户总支付金额趋势。  （2）数值分析，统计指定时间范围下的用户总支付金额。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户总支付金额排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。 |
| **约束条件** | 支付数据正确存储在大数据分析平台中 |
| **备注** |  |

## 支付人数分析

|  |  |
| --- | --- |
| **需求名称** | 支付人数分析 |
| **需求编号** | ZJSM-02-06 |
| **优先级** | 1 |
| **功能描述** | 将用户支付数据与用户基础信息数据关联，对给定时间区间内的支付人数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。  （2）趋势分析，统计指定时间范围下的用户总支付人数趋势。  （2）数值分析，统计指定时间范围下的用户总支付人数。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户总支付人数排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。 |
| **约束条件** | 支付数据正确存储在大数据分析平台中 |
| **备注** |  |

## 支付次数分析

|  |  |
| --- | --- |
| **需求名称** | 支付次数分析 |
| **需求编号** | ZJSM-02-07 |
| **优先级** | 1 |
| **功能描述** | 将用户支付数据与用户基础信息数据关联，对给定时间区间内的支付次数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。  （2）趋势分析，统计指定时间范围下的用户总支付次数趋势。  （2）数值分析，统计指定时间范围下的用户总支付次数。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户总支付次数排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、片区、操作人员、操作网点、缴费方式等。 |
| **约束条件** | 支付数据正确存储在大数据分析平台中 |
| **备注** |  |

## 受理单次数分析

|  |  |
| --- | --- |
| **需求名称** | 受理单次数分析 |
| **需求编号** | ZJSM-02-08 |
| **优先级** | 1 |
| **功能描述** | 将用户受理单数据与用户基础信息数据关联，对给定时间区间内的支付次数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、工单类型、工单名称、片区、操作人员、操作网点等。  （2）趋势分析，统计指定时间范围下的工单事件数量趋势。  （2）数值分析，统计指定时间范围下的工单事件数量。  （3）排行分析，按指定字段分组，统计指定时间范围下的工单事件数量排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司工单类型、工单名称、片区、操作人员、操作网点等。 |
| **约束条件** | 支付数据正确存储在大数据分析平台中 |
| **备注** |  |

## 受理单人数分析

|  |  |
| --- | --- |
| **需求名称** | 受理单人数分析 |
| **需求编号** | ZJSM-02-09 |
| **优先级** | 1 |
| **功能描述** | 将用户受理单数据与用户基础信息数据关联，对给定时间区间内的支付次数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、工单类型、工单名称、片区、操作人员、操作网点等。  （2）趋势分析，统计指定时间范围下的用户数量趋势。  （2）数值分析，统计指定时间范围下的用户数量。  （3）排行分析，按指定字段分组，统计指定时间范围下的用户数量排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司工单类型、工单名称、片区、操作人员、操作网点等。 |
| **约束条件** | 支付数据正确存储在大数据分析平台中 |
| **备注** |  |

## 收视时长分析

|  |  |
| --- | --- |
| **需求名称** | 收视时长分析 |
| **需求编号** | ZJSM-02-10 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的收视时长进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计用户在指定时间范围的总观看时长趋势。  （2）数值分析，统计用户在指定时间范围的总观看时长。  （3）排行分析，按指定字段分组，统计用户总观看时长的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）时间切面分析，统计用户在指定时间范围的总观看时长，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** |  |

## 人均收视时长分析

|  |  |
| --- | --- |
| **需求名称** | 人均收视时长分析 |
| **需求编号** | ZJSM-02-11 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的人均收视时长进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围的人均收视时长趋势。  （2）数值分析，统计指定时间范围的人均收视时长。  （3）排行分析，按指定字段分组，统计用户平均收视时长的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）时间切面分析，统计用户在指定时间范围的平均收视时长，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 观看人数分析

|  |  |
| --- | --- |
| **需求名称** | 观看人数分析 |
| **需求编号** | ZJSM-02-12 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的观看人数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围的总观看人数趋势。  （2）数值分析，统计指定时间范围的总观看人数。  （3）排行分析，按指定字段分组，统计总观看人数的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）时间切面分析，统计用户在指定时间范围的总观看人数，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 观看次数分析

|  |  |
| --- | --- |
| **需求名称** | 观看次数分析 |
| **需求编号** | ZJSM-02-13 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的观看次数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围的总观看次数趋势。  （2）数值分析，统计指定时间范围的总观看次数。  （3）排行分析，按指定字段分组，统计总观看次数的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）时间切面分析，统计用户在指定时间范围的总观看次数，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 到达人数分析

|  |  |
| --- | --- |
| **需求名称** | 到达人数分析 |
| **需求编号** | ZJSM-02-14 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的到达人数进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围的观看时间到达指定时间的总人数趋势。  （2）数值分析，统计指定时间范围的观看时间到达指定时间的总人数。  （3）排行分析，按指定字段分组，统计总观看人数的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）时间切面分析，统计用户在指定时间范围的总观看次数，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 忠诚度分析

|  |  |
| --- | --- |
| **需求名称** | 忠诚度分析 |
| **需求编号** | ZJSM-02-15 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的忠诚度进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围的用户忠诚度趋势。  （2）数值分析，统计指定时间范围的用户忠诚度。  （3）排行分析，按指定字段分组，统计用户忠诚度的排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）统计用户在指定时间范围的用户忠诚度，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 收视份额分析

|  |  |
| --- | --- |
| **需求名称** | 收视份额分析 |
| **需求编号** | ZJSM-02-16 |
| **优先级** | 1 |
| **功能描述** | 将用户收视数据与用户基础信息数据关联，对给定时间区间内的收视份额进行分析：  （1）分析维度，分析可组合维度包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （2）趋势分析，统计指定时间范围下特定用户收视占总用户收视的百分比趋势。  （2）数值分析，统计指定时间范围下特定用户收视占总用户收视的百分比。  （3）排行分析，按指定字段分组，统计指定时间范围下特定用户收视占总用户收视的百分比排名，可进行排行的字段包括：用户编号、客户ID、客户名称、客户分级地址、所属分公司、媒体节目类型、直播节目名称、直播频道名称、VOD节目名称、VOD节目提供商、VOD分级包信息、导演、演员、节目分类、节目系列剧、节目地区信息等。  （4）统计指定时间范围下特定用户收视占总用户收视的百分比，并按指定时间切面进行二维分布展示。 |
| **约束条件** | 媒资数据与收视数据正确存储在分析系统中 |
| **备注** | VOD分类依据：节目包、节目分类、节目系列、供应商、导演、演员 |

## 用户标签分析

|  |  |
| --- | --- |
| **需求名称** | 用户标签分析 |
| **需求编号** | ZJSM-02-17 |
| **优先级** | 1 |
| **功能描述** | 基于用户的历史收视行为、网络行为、工单行为、消费行为，计算用户的离网倾向、收视喜好、消费等级等信息 |
| **约束条件** |  |
| **备注** |  |

# MQL数据分析模块

MQL是基于Spark和ElasticSearch的数据分析语言，通过Restful接口对外提供数据服务，MQL使用管道式的数据传输方式，可实现数据的查询、分组和聚合操作。

## 过滤查询

|  |  |
| --- | --- |
| **需求名称** | 过滤查询 |
| **需求编号** | ZJSM-03-01 |
| **优先级** | 1 |
| **功能描述** | 对指定索引进行检索：索引名为CSV格式的一个或多个索引，支持通配符“\*”，fields为csv格式的字段名，可通过WHERE关键字指定过滤条件，可通过LIMIT关键字限制返回记录条数，offset指定查询偏移位置，可通过MAP关键字对结果集映射；  EXPORT和INSERT INTO可对返回结果导出到文件或索引。 |
| **约束条件** |  |
| **备注** |  |

## 聚合分析

|  |  |
| --- | --- |
| **需求名称** | 聚合分析 |
| **需求编号** | ZJSM-03-02 |
| **优先级** | 1 |
| **功能描述** | 执行聚合操作：聚合MQL语句操作性能优于SQL操作，每条聚合MQL语句可以包含一个或多个聚合元素，且必须衔接SEARCH的中间结果集，通过USING关键字可指定该MQL语句所衔接的中间结果集。 |
| **约束条件** |  |
| **备注** |  |

## SQL语法支持

|  |  |
| --- | --- |
| **需求名称** | 聚合分析 |
| **需求编号** | ZJSM-03-02 |
| **优先级** | 1 |
| **功能描述** | 执行SQL查询操作：支持SQL92标准常用SQL语句，可通过MAP关键字对SQL查询的结果集映射；  EXPORT和INSERT INTO可对返回结果导出到文件或索引 |
| **约束条件** |  |
| **备注** |  |