**数据中心开发规范**

**V1.0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **版本号：** | V1.0 | **修改人：** | 数据技术部 | **日期：** | 2016/7/1 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

目录

[一、 概述 1](#_Toc455151254)

[二、 基础运营开发规范 1](#_Toc455151255)

[1. Tableau基础运营报表 1](#_Toc455151256)

[2. BA基础运营报表 2](#_Toc455151257)

[三、 生态报表开发规范 3](#_Toc455151258)

[四、 资源工时评估规范 3](#_Toc455151259)

# 概述

本文档主要说明数据中心开发流程中需要满足的规范，从而使各类数据需求在开发过程中可管理、可追踪、可回溯，同时减少开发过程中的人力成本、沟通成本、资源成本。

数据中心开发规范主要有三个方面：

1. 基础运营开发规范

介绍基础运营报表的主要组成部分（tableau报表和ba报表），罗列各项报表样例的开发条件、所需工时、占用资源。游戏运营方可以通过此部分内容了解基础运营报表涵盖的内容，以及可以获取的数据信息。

1. 生态报表开发规范

生态报表是游戏研发和运营根据游戏后台数据结构和特殊的生态需要而进行开发的报表。由于此类报表无法抽象通用，因此需要游戏研发和运营明确数据源、数据逻辑、报表样式等。数据中心将依据具体的需求进行工时评估和资源评估，评估可行之后再进行开发。

1. 资源需求评估规范

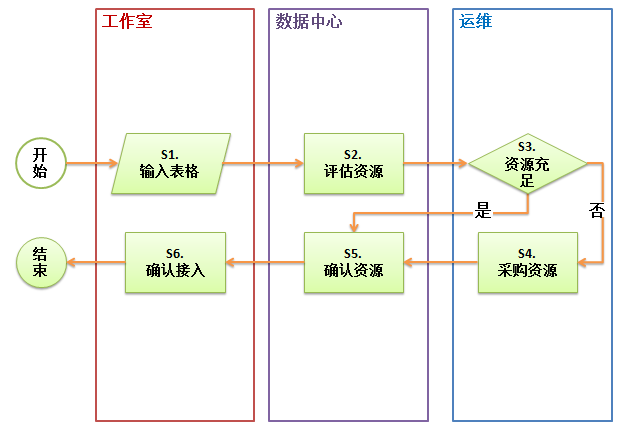
资源需求评估分为两类，基础运营资源评估和生态报表资源评估。基础运营资源评估基于游戏方对每日流量的预估，例如：每日登陆玩家约有a人、充值次数约有b次。

生态报表资源预估基于接入的游戏库原始数据和报表开发所需中间表和结果表。游戏研发和运营需要明确生态报表所需每张源表的数据量、更新策略、清理策略，并依此告知数据中心生态需求所需的空间。

# 操作流程

1. 数据接入申请流程图：

在OA上提交“数据中心游戏数据开发流程”：



1. 标准操作流程具体工作说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **工作编号** | **责任人** | **通知人** | **工作名称** | **工作描述** |
| S1 | 工作室游戏运营 | 无 | 输入表格 | 工作室按实际开发要求，输入《数据中心生态需求文档》及《数据中心生态需求工时及资源评估文档》，如仅需开发tableau以及BA的基础报表，则在申请中说明。 |
| S2 | 数据中心分析组 | 数据技术部负责人 | 评估资源 | 数据中心根据申请文档做资源消耗以及工期评估。 |
| S3 | 运维 | 无 | 资源充足 | 运维检查集群等资源准备情况。 |
| S4 | 运维 | 运维技术部负责人 | 采购资源 | 如果资源准备不充足，进行资源采购。 |
| S5 | 数据中心分析组 | 数据技术部负责人 | 确认资源 | 数据中心确认资源准备满足数据接入需求。 |
| S6 | 工作室游戏运营 | 工作室负责人 | 确认接入 | 工作室负责人确认资源消耗等数据以及开发排期。 |

# 基础运营开发规范

## Tableau基础运营报表

Tableau报表基于数据中心sdk开发，能够提供游戏的日活、充值、留存等信息。主要包含以下报表。

1. 集成数字
2. 单日分服数据
3. 单日渠道数据
4. 活跃、新增用户分析
5. 在线分析
6. 累计付费分析
7. 充值分析
8. 总/新充值分析
9. 运营商付费
10. top10付费渠道登录及转化分析

具体的报表样例可见以下附件《基础运营报表样例》

Tableau基础运营工时预估（单位：人日）。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **基础运营报表需求清单** | | **数据接入** | **脚本开发** | **数据验证** | **tableau报表开发** | **汇总** |
| 1 | 集成数字 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 2 | 单日分服数据 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 3 | 单日渠道数据 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 4 | 活跃、新增用户分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 5 | 在线分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 6 | 累计付费分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 7 | 充值分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 8 | 总/新充值分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 9 | 运营商付费 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |
| 10 | top10付费渠道登录及转化分析 | 0.1 | 0.3 | 0.2 | 0.1 | 0.7 |

**总计：7人日（若需要进行历史补数，额外工时视所需补数时长而定）**

## BA基础运营报表

BA系统包含活跃分析、付费分析、玩家留存、生命价值、运营商分析等模块。用户可以查看游戏整体、单服、多服、单运营商、多运营商的各项指标，并且提供了日、周、月不同的时间窗口。

活跃分析：日活、周活、月活、新老用户活跃构成、PCU、ACU、DT

付费分析：充值金额、充值人数、充值次数、付费率、ARPU、ARPPU、新老充值构成、充值档次构成。

玩家留存和生命价值：留存数据、ltv数据、首周内7留/2留、3登/2登、4登/3登。

运营商分析：通过supersdk数据计算，提供各运行商的登录、付费、转化、留存信息。

BA基础运营工时预估如下（单位：人天）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **BA基础运营报表模块** | | **数据接入** | **脚本开发** | **数据验证** | **BA系统开发** | **汇总** |
| 1 | 活跃分析 | 0.1 | 0.5 | 0.3 | 0.1 | 1 |
| 2 | 充值分析 | 0.1 | 0.5 | 0.3 | 0.1 | 1 |
| 3 | 玩家留存 | 0.1 | 0.5 | 0.3 | 0.1 | 1 |
| 4 | 生命价值 | 0.1 | 0.5 | 0.3 | 0.1 | 1 |
| 5 | 运营商分析 | 0.1 | 0.5 | 0.3 | 0.1 | 1 |

**总计：5人日（若需要进行历史补数，额外工时视所需补数时长而定）**

# 生态报表开发规范

生态报表是对游戏内生态体系的汇总和分析，因此绝大部分需求需要接入游戏后台数据。受此影响，生态需求对数据中心来说存在着数据源不明确、逻辑不明确、资源需求不明确的因素。为了避免或减少这些因素对开发的影响，生态报表开发需要遵循以下步骤。

1. 游戏研发方和运营方需要提供明确的数据源清单以及报表清单，并说明数据源表结构和报表逻辑。此步骤确定该需求的业务可行性。
2. 游戏研发方和运营方需要对源数据表、中间表和报表的结果表进行数据量预估。此步骤确定该生态的技术可行性。具体细节请见第四章：资源需求评估规范
3. 在业务可行性和技术可行性确定后，数据中心将按照运营提供的逻辑文档进行开发。

**注意：**数据中心的开发将严格遵循逻辑文档上的内容，如果研发需要调整逻辑，请提供新的文档，数据中心将更具改动的大小调整工时。

生态需求逻辑文档需要提供以下内容：

1. 数据源表表清单，各表的表结构和字段说明。
2. 生态报表清单，各报表的样例，报表各维度和指标的取值逻辑。
3. 如果发生需求变化，需要记录文档变更信息：变更时间、变更内容

具体请见附件《数据中心生态需求文档样例》

# 资源工时评估规范

资源工时评估规范包含两部分，基础运营和生态报表。

基础运营的工时预估前文已经罗列：

Tableau报表需要7个人日，BA报表需要5个人日。

基础运营资源预估如下：

请游戏研发和运营对上线后游戏的运营情况进行评估，资源占用预估如下。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 日活量玩家 | Hadoop集群每日 | 集群资源年预估 | 数据库（mysql） | 数据库年预估 |
| Tableau | 每10w | 2G | 0.8T | 10M | 3.6g |
| BA | 每10w | 1G | 0.4T | 10M | 3.6g |

生态需求的工时和资源预估主要根据需要接入的源表数量和需要开发的报表数量进行预估。

工时预估：

对于每张需要接入的源表，数据中心工作分为：数据接入、集群建表、接入验证。

为了清洗规范数据，对于每张原始表将建立中间表。需要进行建表、脚本开发和数据验证。

对于每张报表，数据中心需要进行脚本开发、报表绘制以及结果验证。

资源预估：

资源预估分为源表预估、中间表预估、结果表预估。

源表预估需要游戏研发方填写，说明每张表的数据量、更新策略、清理策略。

例如：

Tablea 每天新增1g数据，只需要保留一个月

tableb 每日更新，数据不新增只更新，数据量为5g，无需清理

那么请填写如下信息

|  |  |  |  |
| --- | --- | --- | --- |
| 源表 | 数据量 | 更新策略 | 清理策略 |
| tableA | 1G | 每日新增 | 保留一个月 |
| tableB | 5G | 每日更新 | 无 |

中间表预估基于源表预估生成，每张源表对应一张中间表，中间表的更新策略为每日新增（或每周新增、每月新增，随源表变化，但是不会为每日更新），清理策略为不清理。中间表的数据量预估为源表的50分之一。

结果表预估为每张报表对应一张结果表，结果表的数据量、更新策略、清理策略请游戏研发方填写。需要注意的是结果表所占资源为hadoop集群资源和数据库mysql资源。

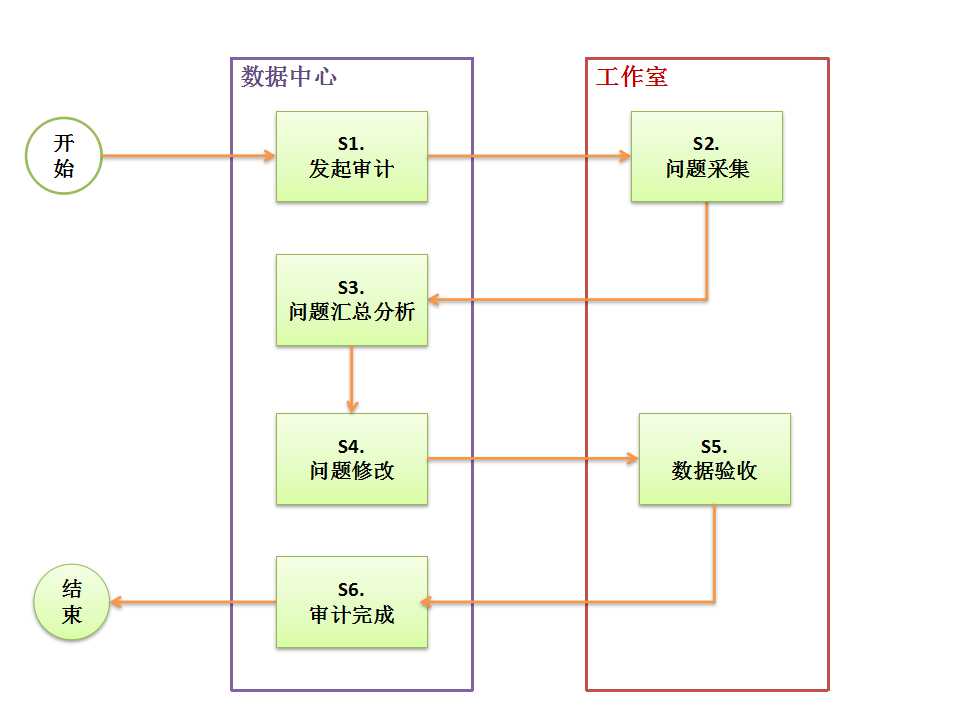
详见附件《数据中心生态需求工时及资源预估样例》

# 数据审计工作规范

数据开发流程完成后，为了进一步提升数据准确性，数据技术部会发起数据审计流程，以一段时间的游戏数据为对象，请工作室运营相关负责人进行数据检查及审计。

1. 数据接入申请流程图：

在OA上提交“数据中心数据审计流程”：



1. 标准操作流程具体工作说明

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **工作编号** | **责任人** | **通知人** | **工作名称** | **工作描述** |
| S1 | 数据中心分析组 | 数据技术部负责人 | 发起审计 | 以游戏的历史数据为目标，发起数据审计工作 |
| S2 | 工作室运营负责人 | 工作室负责人 | 问题采集 | 填写《数据审计问题统计》文档，文档详见附件，采集问题 |
| S3 | 数据中心分析组 | 无 | 问题汇总分析 | 对工作室提出的问题进行汇总分析 |
| S4 | 数据中心分析组 | 数据技术部负责人 | 问题修改 | 对工作室提出的问题按优先级排期修改 |
| S5 | 工作室运营负责人 | 无 | 数据验收 | 数据中心完成数据修改后，由工作室验收数据 |
| S6 | 数据中心分析组 | 数据技术部负责人、工作室负责人 | 审计完成 | 对审计完成的数据在系统中标记，表示这些数据已经准确。 |

# 附件

## 基础运营报表样例



## 数据中心生态需求文档样例



## 数据中心需求工时及资源评估样例



## 数据审计问题统计

