CEDC报表平台设计说明书

版本信息

|  |  |  |  |
| --- | --- | --- | --- |
| **修改日期** | **版本** | **描述** | **作者** |
| **2023-06-10** | **V0.1** | **初稿** | **梁健** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

目录

[1 综述 1](#_Toc137804955)

[1.1 编写目的 1](#_Toc137804956)

[1.2 预期读者 1](#_Toc137804957)

[1.3 术语 1](#_Toc137804958)

[2 报表平台架构 2](#_Toc137804959)

[2.1 报表平台的定位 2](#_Toc137804960)

[2.2 平台总体功能图 2](#_Toc137804961)

[2.3 平台总体技术架构 3](#_Toc137804962)

[2.3.1 数据集市数据处理策略 3](#_Toc137804963)

[2.3.2 任务调度描述 5](#_Toc137804964)

[3 Database开发规范 6](#_Toc137804965)

[3.1 SQL脚本规范(SP,View,DDL以外) 6](#_Toc137804966)

[3.2 DB Procedure规范 6](#_Toc137804967)

[3.3 DB View规范 6](#_Toc137804968)

[4 ETL开发规范 7](#_Toc137804969)

# 综述

## 编写目的

本文档是要定义与描述报表平台的功能，主要包括数据接入和报表平台功能架构两个方面。

在数据接入方面，本文会通过分层说明应用各个层次功能及组织结构，以确保将来的应用程序符合业务需求。

在报表平台功能方面，本文会说明平台中所提供的数据服务。

## 预期读者

本文档仅适用于对CEDC所有项目组成员可利用本文档快速了解报表平台的整体功能情况。

## 术语

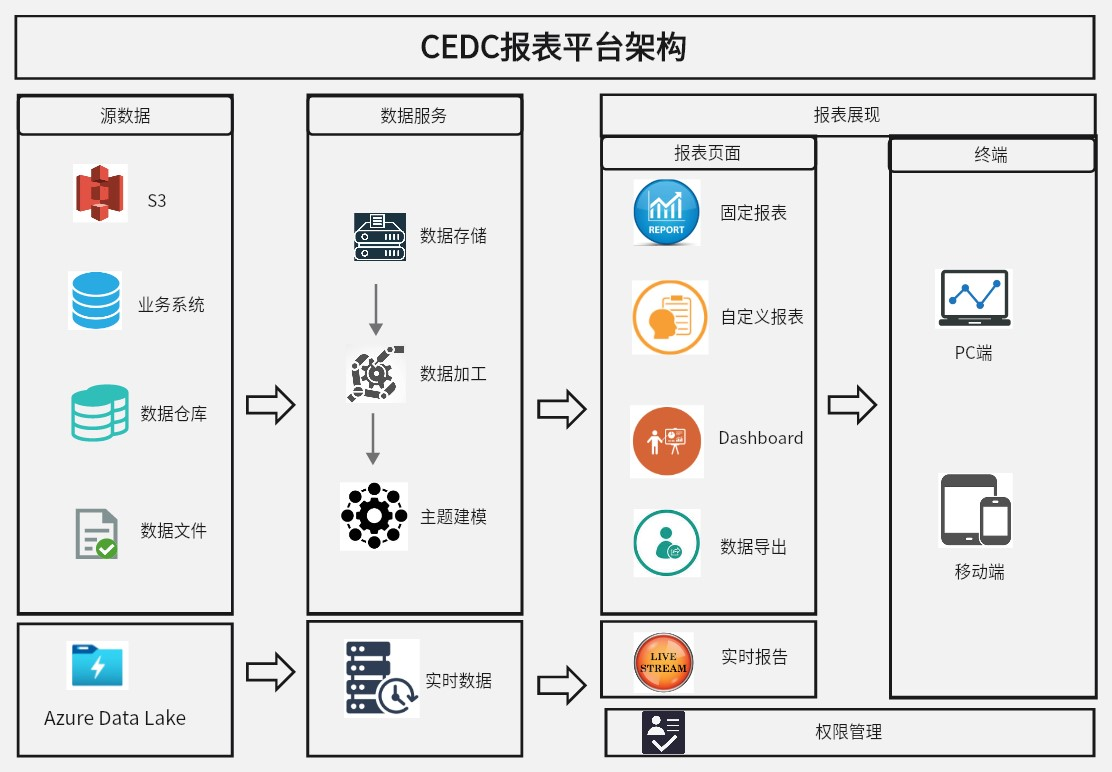
|  |  |
| --- | --- |
| 名称 | 描述 |
|  |  |

# 报表平台架构

## 报表平台的定位

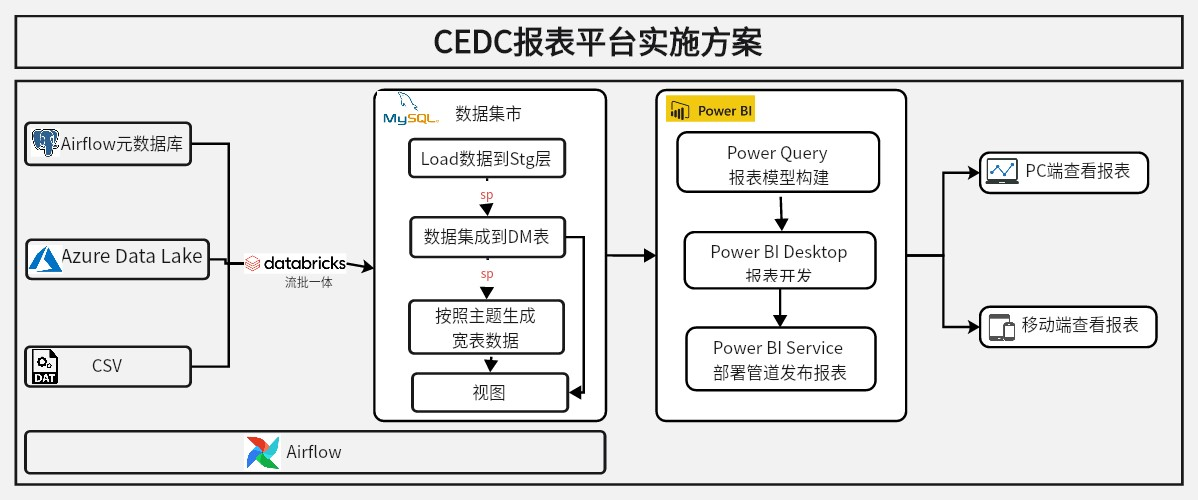
系统是一个独立的报表展现平台，会有独立的数据库作为支撑来提供数据的服务，平台可以承接各种渠道的数据并对数据进行加工整合，最终通过PC和移动终端呈现数据。

## 平台总体功能图



系统总体架构图旨在让用户直观地了解应用系统架构，通过每一层级的上下关联描述让用户清楚整体架构的分类以及各层级之间的协同关系。系统平台分为三个部分，分为源数据、数据服务和报表展现。源数据层是各种业务数据的来源；数据服务层是对加载到数据集市中的数据进行清洗、整合并按照主题域将数据进行分类处理；报表展现层主要负责页面展示。

## 平台总体技术架构



### 数据集市数据处理策略

#### 概要描述

数据集市是一个独立的系统，会接入来自不同渠道的数据，通过统一的接口提供报表展现，因此会数据集市中会将来自不同系统的数据进行整合并根据主题对数据进行数据。

#### 数据集市分层

|  |  |  |
| --- | --- | --- |
| 层次 | 功能 | 描述 |
| Stg层 | 存放本期源数据 | Databricks将源系统的数据从不同的渠道抽到DM的Stg表中，需要支持重跑 |
| DM层 | 存放所有支持报表的基础数据 | 承接各种渠道的数据，并对数据进行整合，对于同样的维度信息整合到一起，保证数据集市的统一性，同时避免数据库表的无限放大，同时需要支持重跑。 |
| 主题域宽表 | 根据业务主题区分成不同的数据集合 | 将数据按照不同领域和主题设置成能够服务于大多数同主题的报表的宽表,本次主要分成marketing和sales两个主题域 |
| 视图 | 将数据集市的物理表与报表隔离，报表都是从视图获取数据 | 将数据集市的物理表与报表解耦合，降低运维成本 |

#### 报表展现

平台中报表展现使用的是PowerBI,而且是多账号同时开发，开发人员需要使用统一的报表开发规范和模型，创建报表模型尽可能使用星形模型，从而达到报表页面的互动，增强用户体验。

用户可以在PC端登陆报表平台查看报表，也可以使用手机或pad查看报表，同时支持用户导出数据，对于不同的用户只会看到自己对应的数据。

### 任务调度描述

数据集市中的业务数据处理是使用Mysql的存储过程来处理，会将所有SP存放在一个配置表中，通过一个main接口调用所有的SP方式来实现数据的刷新；

SP的调用以及源数据到数据集市的Stg表的过程会使用到airflow通过APi调用databricks的notebook。

# Database开发规范



## SQL脚本规范(SP,View,DDL以外)

格式： ${SQLDescription}.sql

说明： ${SQLDescription}描述sql脚本功能或含义

举例： 如EXPORT\_DIM\_ORG.sql是导出表dim\_org的sql 处理

功能： 支持数据重跑

## DB Procedure规范

格式： SP\_${TargetTableName}.sql

说明： SP\_${TargetTableName}是存储过程名称。

举例： 如SP\_DIM\_ORG.sql进行生成表dim\_org的存储过程。

功能： 支持数据重跑，SP中脚本模板如下

## DB View规范

格式： ${使用的场景}\_VIEW\_${描述}.sql

说明： ${使用的场景}是指哪里会用到，比如报表、提供给下游数据等，暂时有两

种：RPT:给报表使用；PUB：给下游数据使用

${描述}：根据实际内容定义

举例： 如RPT\_VIEW\_DIM\_ORG.sql将给报表使用的org数据

PUB\_VIEW\_DIM\_ORG.sql指视图是给下游提供org数据

功能： 支持数据重跑

# ETL开发规范

等待后期datbricks 规范再更新