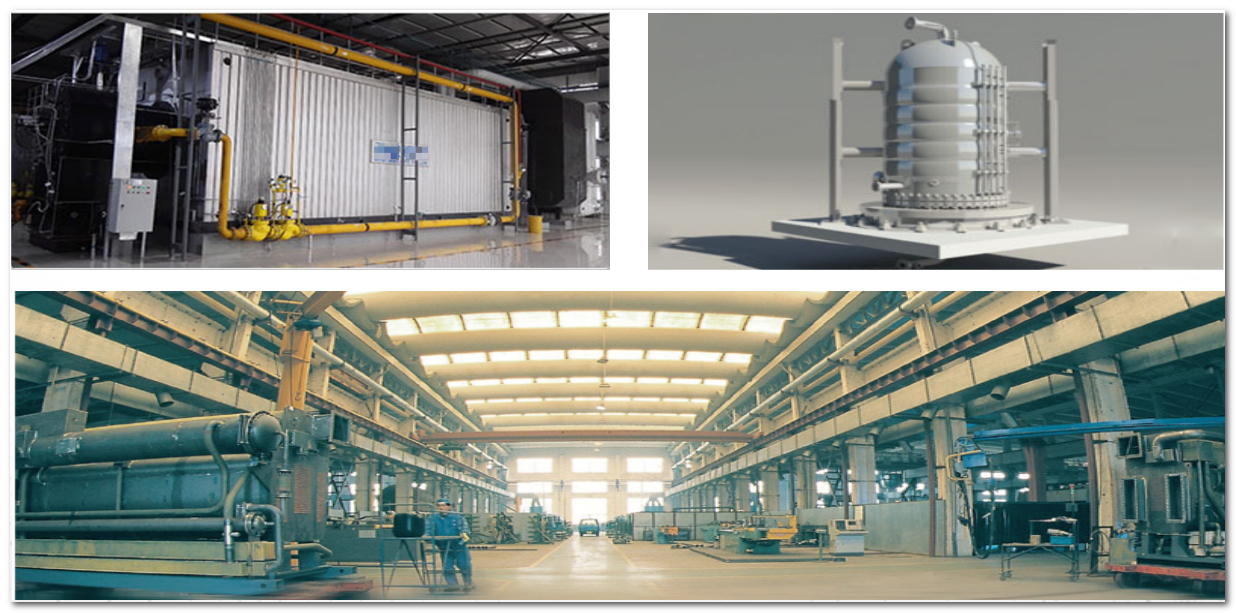
# 【工业数据可视化开发】 赛题1

## 1、项目背景

近年来，在国家“智能制造”、“数智化转型”战略的牵引和全球工业互联网产业创新发展的大背景下，国家要求大数据、人工智能、物联网等新兴技术，需要和工业应用场景进行深度融合应用。在企业生产制造、销售经营、设备运维、能源管理等领域，涌现了大量大数据技术+工业场景的分析应用需求。

某设备制造集团主要生产工业锅炉和工业空调（参见下图），名列中国机械工业500强，以节能环保高端装备制造产业为核心，为全球20000多家客户提供卓越的产品和服务，随着制造行业产能过剩、同行竞争激烈，卖设备利润急剧下跌，急需数智转型升级、服务化转型，从以前的卖设备转向为卖服务（设备安装服务、维修服务、保养服务、租赁服务等），最终实现商业模式的创新。



该集团设备运维事业部，针对客户分布广、运维服务成本高、管理难度大、客户投诉设备服务效率和质量低下等问题，希望采用大数据分析技术，对设备安装交付、故障报修、定期保养等业务流程和相关业务系统积攒的历史业务数据，进行大数据分析，开发：设备智慧运维监控大屏，从而辅助企业经营和管理决策，效果图参见下图。



## 2、任务要求

现要求根据提供的业务源数据，采用大数据分析流程，利用数据中台完成数据采集、数据处理开发、数据分析（数仓建模+业务指标分析）、离线开发、数据可视化等处理，实现对业务指标“故障类型分布Top5”及“故障原因及措施”进行分析并可视化展示，并实现点击故障查看对应原因和措施的交互。过程中所有任务及数据表命名时均需添加考号后缀（如：\_2030401），用以区分参赛人员数据。完成后效果类似下图：





### 2.1任务1：数据采集

1. 采集要求：请利用数据中台完成：从【设备后服务-业务源库】，采集如下数据表到【设备后服务-目标库】，并进行数据比对确保采集前后数据的一致性。

|  |  |
| --- | --- |
| 数据表名 | 数据表中文名 |
| kms\_categories | 故障分类表 |
| province | 省份表 |
| prod\_equipment\_temp | 设备表 |
| sms\_sos | 工单表 |
| sms\_so\_failure\_logs | 工单故障记录表 |
| kms\_causes | 故障原因表 |
| kms\_measures | 故障解决措施表 |

1. 命名规范要求：

（1）数据采集任务要求按如下规范命名：

数据采集\_故障类型分布Top5数据\_考号后缀

其中：考号后缀为自己的实际考号

（2）采集后的所有目标表添加自己考号后缀，例如：kms\_categories\_2030401；

### 2.2任务2：创建项目

1、制定项目的业务规划

按如下规划，分别创建：业务板块、数据域、业务过程三级分类目录，参考命名：

业务板块：设备后服务\_考号后缀 用于绑定---项目

-------数据域 ：设备大屏\_考号后缀 用于绑定---维度模型表

------业务过程 ：故障维修\_考号后缀 用于绑定----事实表 、原子指标

2、创建项目

* 中文名：设备后服务项目\_考号后缀
* 英文名：afterService\_考号后缀
* 数据源：设备后服务-目标库

### 2.3任务3：数据预处理（ETL）

1、清洗要求：

（1）待清洗表：目标库-设备表prod\_equipment\_temp\_考号后缀（对应自己后缀的表）

（2）清洗规则：

a、值替换：name字段：超高压锅炉AC791 替换为：超高压锅炉AC791-V2；

b、过滤：category\_name字段以“锅炉”结束并且create\_user 是null；

（3）清洗后存放到表：prod\_equipment\_etl\_temp\_考号后缀 （对应自己后缀的表）

（4）设置调度依赖：手动执行一次

### 2.4任务4：数仓建模指标分析

1、建维度模型：

要求创建三个维度模型表，并按月调度。

（1）创建设备维度：

参考命名：维度\_设备卡片\_考号后缀 dim\_equip\_考号后缀（换为自己后缀，下同）

对应原始表：prod\_equipment\_temp\_etl\_考号后缀（对应自己后缀的表）

主键：ID

调度：按月调度

选择列： DR、code、category\_id、category\_name、name

（2）创建故障类型维度：

参考命名：维度\_故障类型\_考号后缀 dim\_kms\_category\_考号后缀

对应原始表: kms\_category\_考号后缀

主键：ID

调度：按月调度

选择列：code、name、equipment\_type、product\_category\_id、product\_id

（3）创建省份维度：

参考命名：维度\_省份\_考号后缀 dim\_province\_考号后缀

对应原始表：province\_考号后缀

主键：id

选择列：全选

调度：按月调度

2.建事实模型：

要求创建一个事实模型表。

参考命名：事实\_工单故障\_考号后缀 sms\_sos\_failure\_考号后缀

主表：sms\_sos\_考号后缀

主键：ID

（1）表连接：

sms\_sos（关联主表字段id）、sms\_so\_failure\_logs（关联副表字段so\_id）

（2）选择物理表字段：id、category\_id、so\_id、equ\_id、province、ts

（3）自定义日期字段，并设计表达式：

* created\_date\_my 创建日期\_自定义
* yesterday\_my 昨天\_自定义

（4）事实表关联创建的三个维度，形成星型模型：

3、建原子指标：

（1）参考命名：原子指标\_故障维修数量\_考号后缀 breakdown\_categary\_num\_\_考号后缀

（2）统计日期：yesterday\_my 字段

（3）计算逻辑：排重后对id计数。

4、建派生指标：

1. 时间限定要求：截止至当日；
2. 业务限定：无；

（3）统计粒度要求：故障类型维度

（4）参考命名：

物理表英文名：sofar\_breakdown\_categary\_num\_考号后缀

物理表中文名：至今\_各故障类型个数\_考号后缀

### 2.5任务5：数据开发

通过SQL数据开发，对上述表进行处理，生成故障类型分布top5记录表和故障原因及措施表。

1、故障类型分布top5数据表

（1）创建开发任务：故障类型分布top5\_考号后缀

（2）创建故障类型分布Top5表，表名为：ads\_breakdown\_top5\_考号后缀：

表结构如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 字段名称 | 类型 | 字段长度 | 字段描述 |
| 1 | categoryCount | decimal | 41 | 故障发生次数 |
| 2 | categoryId | varchar | 64 | 故障类型id |
| 3 | name | varchar | 255 | 故障类型名称 |

（3）依据自己的派生指标汇总表和维度表数据生成故障类型分布Top5中表数据。

2、故障原因及措施数据表

（1）创建开发任务：故障原因及措施\_考号后缀

（2）创建故障原因及措施表，表名为：dws\_fault\_cause\_measure \_考号后缀：

表结构如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 字段名称 | 类型 | 字段长度 | 字段描述 |
| 1 | kms\_cause | varchar | 2000 | 故障原因 |
| 2 | kms\_measure | varchar | 2000 | 故障措施 |
| 3 | kms\_name | varchar | 255 | 故障名称 |

（3）依据故障类型分布Top5表、维度表、故障原因表及故障解决措施表数据生成故障原因及措施表数据。

### 2.6任务6：数据可视化

1、可视化整体要求：选用数字大屏，对上述数据开发处理的结果进行可视化展示，要求采用柱形图展示故障类型分布top5的数据，点击某故障柱形图，展示该故障的原因及相应措施。故障原因及措施采用表格展示；

2、样式要求：故障类型数值显示为整数、需设置图表标题。

3、页面布局合理，界面样式美观大方。