

实训任务八：信息技术学院教学云平台运营分析

一、目的

- 1、分析信息技术学院教学云平台运营数据；
- 2、掌握使用 python 进行数据清洗的方法；
- 3、掌握信息技术学院教学云平台运营分析。

二、学时

6 学时

三、类型

复合型

四、需求

1、硬件

每人配备计算机 1 台，不低于双核 CPU、8G 内存、500GB 硬盘。
每组配备服务器 1 台。

2、软件

Windows 操作系统，安装 puTTY 管理终端软件，安装 VMware ESXi 控制台软件。
服务器安装 VMware ESXi 7.0。

3、网络

计算机、服务器、虚拟主机使用固定 IP 地址接入局域网，并支持对互联网的访问。

4、工具

无。

五、任务要求

- 1、完成信息技术学院教学云平台运营数据的预处理；
- 2、完成信息技术学院教学云平台运营数据分析；
- 3、完成不少于 4 个分析主题，每个分析主题不少于 15 个分析模型。

六、考核要求

- 1、提交《信息技术学院教学云平台运营分析报告》；
- 2、提交信息技术学院教学云平台运营分析可视化分析成果截图/演示视频。



扫码看任务步骤

七、任务步骤

任务 1：获取信息技术学院教学云平台运营数据

教学云平台运营数据可从系统运维实训平台（<http://ywsx.xg.hactcm.edu.cn>）下载，其数据如表 8-1 所示：

表 8-1 教学云平台数据分类

数据类型	数据来源	数据描述	格式	数据量 (条)	文件名称
日志数据	教学云平台身份认证	接口操作日志	Access	905432	QS_UMS_ApiLogInfo
		机构信息	Access	43	QS_UMS_OutfitInfo
		教师信息	Access	67	QS_UMS_TeacherInfo
		学生信息	Access	1719	QS_UMS_StudentInfo
		用户登录日志	Access	32923	QS_UMS_UserLogInfo
	教学云平台日志信息	业务访问日志	JSON	8216378	xg_it_yunwei_log
业务数据	教学云平台网站管理	网站信息	Access	33	QS_WCM_System_WebSite
		发文信息	Access	3127	QS_WCM_Site_Article
	教学云平台资源管理	资源文件信息	Access	4942	QS_DMS_Site_ResourcefileInfo
		资源信息	Access	559	QS_DMS_Site_Resourceinfo
		资源网站信息	Access	5	QS_DMS_System_Website

任务 2：解析信息技术学院教学云平台运营数据字段含义

信息技术学院教学云平台运营数据包含 11 种数据，在进行分析前需要先解析每个键值的含义并找出数据间的联系。

步骤 01：解析字段含义

教学云平台运营数据字段含义详见《运营数据字段含义》。

步骤 02：确定数据关系

依据解析的字段含义找出字段之间的联系，例如 `qs_dms_site_resourcefileInfo` 表中的 `FilePath` 字段和业务访问日志中的 `uri` 字段可相互对应，以此确定 `qs_dms_site_resourcefileInfo` 表和业务访问日志的数据关系，从而进行数据聚合，聚合后的数据信息如表《数据关系解析表》。

任务 3：使用 Python 预处理信息技术学院教学云平台运营数据

数据预处理在数据分析的过程中至关重要，它关系着最后数据呈现出来的效果，也关系着数据分析的精准和科学性，本任务使用 Python 语言进行数据处理，并加处理后的数据输出为文本文件，方便导入 Elasticsearch 中。

步骤 01：清洗数据

数据分析前需要先清洗格式和数值异常的数据，以及和分析无关的字段，例如在日志数

据中 `http_cookie` 字段就暂时使用不到，可使用 Python 将其清洗。

步骤 02： 根据数据关系进行聚合

根据《数据关系解析表》文件中确定的数据关系，使用 Python 进行数据聚合。

步骤 03： 输出聚合后数据

将数据聚合结果输出为 JSON 格式数据。

任务 4：使用 Elasticsearch 存储信息技术学院教学云平台运营数据

步骤 01： 使用 Filebeat 采集聚合后数据

安装 Filebeat 并使用 Filebeat 将预处理的数据上传至 elasticsearch 中。

配置文件：

```
filebeat.inputs:  
- type: log  
  enabled: true  
  paths:  
    - E:\user_log.txt  
  json.key_under_root: true  
  json.overwrite_keys: true  
  json.message_key: log  
  tags: ["access"]
```

```
output.elasticsearch:  
  hosts: ["10.10.2.152:9200"]
```

参考命令：

```
filebeat.exe -e -c filebeat.yml
```

步骤 02： 创建索引，并查看运营数据

登录 Kibana 平台，创建索引，查看聚合后数据是否正确。

任务 5：设计信息技术学院教学云平台运营数据分析模型

参照《运营数据分析模型》完成数据分析模型设计。

任务 6：信息技术学院教学云平台运营分析

根据《运营数据分析模型》依次创建数据分析报表。

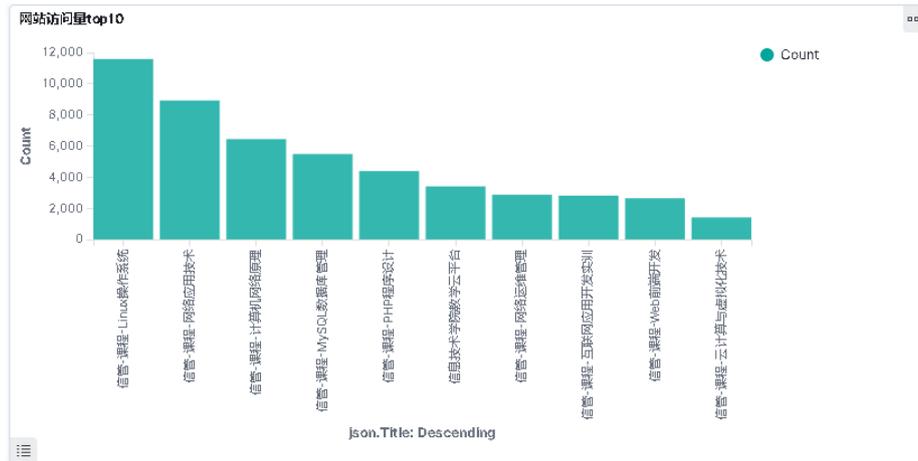


图 8-1 网站访问量

步骤 02: 创建“信息技术学院教学云平台运营分析”仪表盘

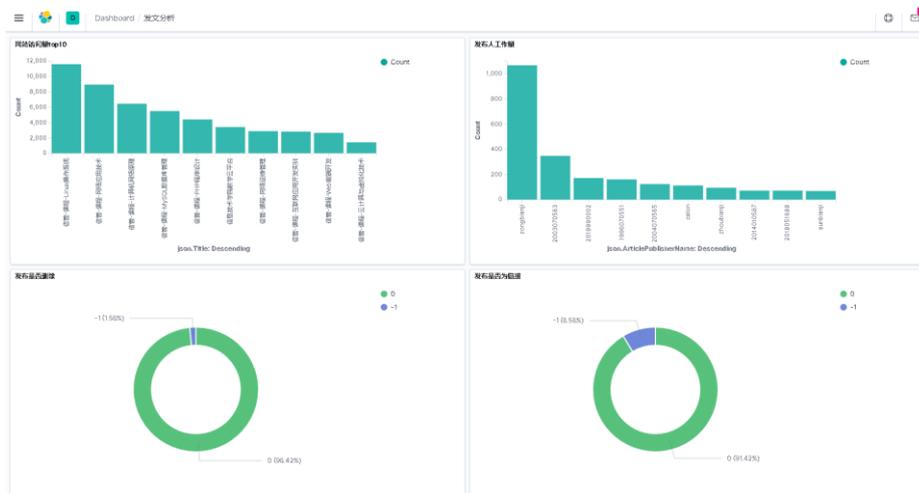


图 8-2 信息技术学院教学云平台运营分析

八、实验思考

1、深入理解大数据分析。

- (1) 大数据分析的流程是什么??
- (2) 大数据分析时需要考虑的问题有哪些?

2、了解日志数据与业务数据。

- (1) 日志数据和业务数据的区别是什么?
- (2) 日志数据和业务数据的联系是什么?

3、如何做好数据可视化?

- (1) 数据可视化的类型有哪些?
- (2) 如何凸显数据的价值?