

## 1. 学习配置和使用 Nginx - 了解如何设置虚拟 IP 登录并生成访问日志。

- Nginx 生成 access.log: [后端实践: Nginx 日志配置 \(超详细\) - 知乎 \(zhihu.com\)](#)
- access.log: 访问用户的 IP 地址、浏览器、请求处理时间等信息

```
access_log path [format [buffer=size] [gzip[=level]] [flush=time]
```

```
[if=condition]]; # 设置访问日志
```

```
access_log off; # 关闭访问日志
```

**学习进度:** 由于以前接触过 nginx，具体来说，自主配置过 nginx 工程，已经了解并熟悉 nginx 工程结构，各模块作用，以及简单的配置流程。包括设置接口、协议、查看 access 和 error 信息等。还需进一步了解设置虚拟 IP 登录和 nginx 连接 kafka 的技术流程。

## 2. 学习安装和配置 Kafka - 创建一个 Kafka 主题，并设置 Kafka 生产者将 Nginx 的访问日志发送到该主题。

- Kafka 是一个分布式消息队列系统，被广泛用于处理高吞吐量的实时数据流。它可以帮助收集、传输和存储数据 - 设置 Kafka 集群、创建 topics、发送和接收消息。
- Kafka 原理解析: [\(73 条消息\) Kafka 原理解析 \(一\) -- Kafka 主要流程概览 kafka 消息主键\\_知也无涯\\_Wileystar 的博客-CSDN 博客](#)
- Kafka3 教程: [06 尚硅谷 Kafka 入门 安装 Kafka 哔哩哔哩 bilibili](#)
- Kafka 官网&文档: [Apache Kafka](#)
- 按照教程在 linux ubuntu 上安装 kafka，linux 虚拟机环境:

```
architecture: ('64bit', 'ELF')
```

```
machine: x86_64
```

```
node: chriswang-virtual-machine
```

```
platform: Linux-5.15.0-56-generic-x86_64-with-glibc2.35
```

```
processor: x86_64
```

```
python_build: ('main', 'Nov 14 2022 16:10:14')
```

```
python_compiler: GCC 11.3.0
```

```
python_version: 3.10.6
```

```
release: 5.15.0-56-generic
```

```
system: Linux
```

```
version: #62-Ubuntu SMP Tue Nov 22 19:54:14 UTC 2022
```

- kafka 基本操作:

kafka-server-start.sh # 启动脚本

kafka-server-stop # 关闭脚本

# 其他重要对应脚本:

kafka-console-consumer.sh # 消费者

kafka-console-producer.sh # 生产者

kafka-topics.sh # 主题

**学习进度:** 对 kafka 的使用是本次项目的核心。在本地配置了 kafka 环境, 了解了 kafka 工程结构, 包括主要的命令和配置信息位置。了解了 kafka 技术的作用和意义, 初步了解了 kafka 的运作原理。由于不熟悉 java, 计划在使用 kafka 过程中同时了解需要用到的 java 的知识。初步熟悉了 java 的基本内容, 包括语言特性和简单语法。

### 3. 学习使用 Flink 进行流式计算 - 配置 Flink 消费 Kafka 主题中的数据, 并进行格式化和处理 - 在处理完成后, 将处理后的数据再次发送到另一个 Kafka 主题。

- Flink 是一个开源的流式计算框架, 它可以实现实时数据流的处理和分析。其特点是批流统一, 高性能, 大计算规模。
- Flink 系列教程: [05 Flink 快速上手 创建 Maven 工程&导入依赖 哔哩哔哩 bilibili](#)

**学习进度:** Flink 与 Kafka 的连接需要在掌握 kafka 的基础上进行学习。需要学习的内容包括如何使用 Flink 进行数据流的处理、转换、聚合和计算, 以及如何与 Kafka 集成以实现流式计算。此前接触的 ETL 的工具是基于 JavaScript 的 Kettle, 其图形化工具是 Spoon。在 Windows 环境下使用 IntelliJ Idea 初步熟悉了 Flink 工程的创建等。已经了解了 Flink 的基础知识和原理、部署提交、运行架构和基本 API, 需要继续学习其时间语义、窗口、处理函数、状态、容错, 以及 SQL-Client 连接。

### 4. 学习如何安装和配置 ELK 套件 - 使用 Logstash 将第二个 Kafka 主题中的数据导入到 Elasticsearch 进行存储 - 一旦数据存储在 Elasticsearch 中, 使用 Kibana 进行数据可视化和分析, 以展示统计结果。

- [【IT 老齐 203】20 分钟上手 ELK 日志监控系统 哔哩哔哩 bilibili](#)
- [【尚硅谷】ElasticSearch 教程入门到精通 \(基于 ELK 技术栈 elasticsearch 7.x+8.x 新特性\) 哔哩哔哩 bilibili](#)

**学习进度:** 了解了 elasticsearch, Kibana 的功能。曾有过使用 PowerBI 进行数据可视化和分析的经验。需要学习 ELK 日志监控系统的技术细节, Kibana 数据可视化等。