| **2022- 2 학기 Midterm Report** | | | | |
| --- | --- | --- | --- | --- |
| 교과목명 | 창직종합설계프로젝트1 PROJECT1 | | 교과목번호 | 101811 - 4 |
| 학점(설계)/시수 | 3 / Mon(10), Wed(10), Thu(10) | | | |
| 개설학과(학년) | 4 | | 담당교수 | Junyong Lee |
| 과제명 | recognize faces and recommend advertisements. | | | |
| 팀원 | 학번 | 이름 | | 연락처/e-mail |
| B711069 | 박선형 | | 010-5026-1897/  byjs0908@naver.com |
| B711134 | 이승환 | | 010-9135-4481/  devseung@g.hongik.ac.kr |
| B711211 | 허민욱 | | 010-8734-2429/  hmw1022@naver.com |
| 목 차  **Ⅰ. Introduction**  **Ⅱ. Review of 2022-1**  **Ⅲ. Progress.  Ⅰ. Making Kafka Cluster**  **Ⅱ. Making Producer Server    Ⅲ. Making Conumser Server**  **Ⅳ. Client**  **Ⅳ. What to do in Rest of 2022-2**  **V. Schedule of 2022-2** | | | | |

1. **Introduction**In the era of big data, the development of IoT devices and the development of cloud environments requires the ability to process vast amounts of real-time data generated at various endpoint stages.  
   To cope with this environment, we aim to develop a real-time distributed processing system that can provide customized information based on real-time user information. In particular, it aims to establish a pipeline for providing customized advertising data to individuals using individual facial data according to the value of future facial data utilization.  
   First, in order to extract the user's information, the user's face data is received and preprocessed, and then age and gender information is predicted through a trained AI model.  
   Thereafter, data is requested to the central messaging queue built through the Consumer WAS of the Pub/Sub structure based on the generated user prediction information. On the other hand, the information-generating endpoint client loads advertisements in the messaging queue regardless of the subject who consumes information through Producer WAS. After that, if matched information exists, it is transmitted to the Consumer and finally provided to the user.
2. **Review**The project is largely divided into AI parts and data pipeline parts. First of all, in the first semester, AI parts, that is, age and gender prediction models according to facial data, were designed and trained, and the user's facial data preprocessing process was conducted. User facial data collection utilized the OpenCv library and trained the model using the collected data and Kaggle's open-source facial data set. The model design is based on the Yolo model and utilizes Tensorflow.Keras.
3. **Progress**

**3.1 Kafka Cluster**A Kafka cluster is a server that controls and stores all data, consisting of a total of two instances. Each instance consists of a Kafka broker and a primary keeper who controls the cluster environment, and when data is generated through the cluster environment, each instance responds to server failures through data replication and provides high availability (HA).  
  
Until now, we have used two instances as demo versions, but if the domain scope expands and the topic volume increases later, we can add additional instances for scalability.  
  
Due to the nature of the demo version and advertising data, each partition in the Kafka cluster does not batch the data and provides a single data on a consumer request. It can provide a stream suitable for services that require real-time, although it requires a large amount of network costs compared to batch processing.

**3.2 Producer WAS**The producer application was implemented using Spring Boot's Spring-Kafka. Application configuration of MVC patterns allows data creators to request WAS in REST format to provide data to Kafka clusters. In this process, data providers can provide data without information about individual Consumer Clients.

**3.3 Cosummer WAS**The consumer application was also implemented using SpringBoot's spring-kafka. Consumer WAS stores subscription information for the user node according to the subscription information provided by the Face Recognition Client Application and filters the data based on it. In this process, it performs the function of filtering only customized information through age and gender information and providing it to the client node.

**4. What to do in Rest of 2022-2**

* **Connecting Python moudle & Kafka Server.**
* **Simulation & Test**
* **Refactoring Code.**
* **Release**

**5. Schedule of 2022-2**

|  |  |  |  |
| --- | --- | --- | --- |
| 달  설계요소 | 10월 | 11월 | 12월 |
| **Connecting Python moudle & Kafka Server.** | **o** | **o** |  |
| **Simulation & Test** | **o** | **o** |  |
| **Refactoring Code.** |  | **o** | **o** |
| **Release** |  |  | **o** |