IITP DABT PreProcessing

배포 설치 가이드

문서 버전: 1.0.0
작성일: 2025-11-26

(주)스위트케이

문서 History

버전	일자	작성자	변경 내용
1.0.0	2025-11-26	(주)스위트케이	최초 작성

목차

1. 개요

1.1. 문서 목적

이 문서는 IITP DABT PreProcessing 시스템을 Ubuntu 서버 환경(Python 3.8 이상)에서 설치 및 실행하기 위한 절차를 정리한 가이드입니다. 본 문서의 지침을 순서대로 수행하면 신규 서버에 해당 시스템을 설치하고 정상적으로 실행할 수 있습니다.

1.2. 적용 범위 및 대상 계정

- **운영 체제**: Ubuntu 22.04 LTS 이상 (Python 3.8+ 기본 제공)
- **서비스 계정**: `iitp-app`
- **권장 권한**: sudo 권한 보유(설치 및 서비스 관리 목적)
- **적용 환경**: 개발/검증/운영 서버

2. 시스템 구성 및 요구사항

2.1. 서버/OS/네트워크 요구사항

항목	최소 사양	권장 사양
CPU	2 Core	4 Core 이상
RAM	4 GB	8 GB 이상
Disk	20 GB	50 GB 이상 (로그/데이터 포함)
OS	Ubuntu 22.04 LTS	Ubuntu 22.04 LTS
Python	3.8 이상	3.10
Network	KOSIS API 사이트 outbound 허용 (443)	동일

2.2. 필수 소프트웨어 및 패키지

- Python 3.8 이상
- Python venv 모듈(`python3-venv`)
- Git (소스 다운로드용)
- PostgreSQL 클라이언트 라이브러리 (psycopg2-binary 의존)
- unzip, curl 등 기본 도구

설치 예시 (Ubuntu 22.04 기준):

```
sudo apt update
sudo apt install -y python3.10 python3.10-venv python3-pip git curl
```

> Ubuntu 22.04는 python3.10이 기본 제공됩니다.

2.3. 계정 및 권한 정책

- 모든 설치 및 실행은 `iitp-app` 계정에서 수행합니다.
- `/home/iitp-app` 하위에 소스, 가상환경, 로그, 데이터 디렉토리를 생성합니다.

- 필요한 경우 다음 명령으로 계정 생성/권한 부여:

```
sudo adduser --disabled-password --gecos "" iitp-app
sudo usermod -aG sudo iitp-app   # 필요 시
```

3. 사전 준비

3.1. 소스 코드 확보

1. `iitp-app` 계정으로 로그인:

```
sudo su - iitp-app
```

2. 작업 디렉토리 이동:

```
cd ~
```

3. Git 리포지토리 클론:

```
git clone <REPO_URL> IITP-DABT-PreProcessing
cd ~/IITP-DABT-PreProcessing
```

> REPO_URL은 실제 저장소 주소로 변경합니다.

3.2. 환경 변수 및 DB 정보 수집

설치를 진행하기 전에 다음 정보를 확보해야 합니다.

항목	내용	비고
DB_URL	PostgreSQL 연결 문자열	예: `postgresql://user:pass@host:port/dbname`
LOG_LEVEL	실행 로그 레벨 (INFO 권장)	
DB_BATCH_SIZE	DB 삽입 배치 크기	기본 100
DATA_COLLECTION_SCOPE	ALL 또는 PART	PART 사용 시 대상 목록 필요
TARGET_SRC_TBL_ID_LIST	수집 대상 통계 ID 목록	`.env` 섹션 형태

3.3. 디렉토리 구조 및 권한 설정

프로젝트 실행에 필요한 기본 디렉토리:

```
~/IITP-DABT-PreProcessing/
├── logs/
├── kosis_data/
│   └── YYYYMMDD/
│       ├── data/
│       ├── meta/
│       └── latest/
```

권한 설정:

```
mkdir -p logs kosis_data
chmod 755 logs kosis_data
```

4. 설치 절차

4.1. Python 가상환경 구성

1. Python 버전 확인:

   ```
   python3 --version
   ```

   3.8 이상이어야 합니다. 여러 버전이 있을 경우 명시적으로 python3.10 등을 사용합니다.
2. 가상환경 생성:

   ```
   cd ~/IITP-DABT-PreProcessing
   python3 -m venv venv
   ```

3. 가상환경 활성화:

   ```
   source venv/bin/activate
   ```

   프롬프트에 `(venv)` 가 표시됩니다. 비활성화는 `deactivate` .

4.2. 필수 패키지 설치

가상환경 내에서:

```
pip install --upgrade pip
pip install -r requirements.txt
```

`requirements.txt` 에는 `requests` , `python-dotenv` , `sqlalchemy` , `psycopg2-binary` 등이 명시되어 있습니다.

4.3. 환경 설정 파일 구성

1. `.env` 파일 생성:

```
cat > .env <<'EOF'
DB_URL=postgresql://USER:PASSWORD@HOST:PORT/DBNAME
LOG_LEVEL=INFO
DB_BATCH_SIZE=200
DATA_COLLECTION_SCOPE=ALL
PARALLEL_WORKERS_FILE=4
PARALLEL_WORKERS_DB=2
# [TARGET_SRC_TBL_ID_LIST]
# STAT_TBL_ID_1,2019
# STAT_TBL_ID_2,2020
EOF
```

실제 값으로 수정합니다.

- `DATA_COLLECTION_SCOPE=ALL` 이면 모든 활성 통계를 수집합니다.
- `DATA_COLLECTION_SCOPE=PART` 로 설정하려면 반드시 `[TARGET_SRC_TBL_ID_LIST]` 섹션을 작성해 수집 대상 통계 ID와 (선택적으로) 시작 연도를 지정해야 합니다.
- 섹션 예시는 아래와 같습니다.

```
DATA_COLLECTION_SCOPE=PART
...
[TARGET_SRC_TBL_ID_LIST]
STAT_TBL_ID_1,2019
STAT_TBL_ID_2,2020
STAT_TBL_ID_3
```

연도가 생략되면 DB에 등록된 `collect_start_dt` 값을 사용합니다.

2. `config.py` 는 `.env` 를 로드하여 값을 사용하므로 별도 수정이 필요 없지만, 필요에 따라 기본값을 조정할 수 있습니다.

5. 실행 환경 구성

5.1. 로그 및 데이터 디렉토리 준비

프로그램 실행 시 자동으로 생성되지만, 권한 문제를 방지하기 위해 미리 생성합니다.

```
mkdir -p logs
mkdir -p kosis_data
chmod 755 logs kosis_data
```

5.2. 서비스 사용자 기준 실행 설정

- `iitp-app` 계정으로 로그인한 상태에서 가상환경을 활성화합니다.
- 필요 시 `.bashrc` 에 alias 추가:

  ```
  echo "alias activate_iitp='cd ~/IITP-DABT-PreProcessing && source venv/bin/activate'" >> ~/
  source ~/.bashrc
  ```

- 실행 전 항상 `source venv/bin/activate` , 작업 종료 후 `deactivate` .

6. 실행 절차

두 가지 실행 옵션을 제공합니다.

6.1. `--mode file` 실행

목적

KOSIS 데이터를 파일로만 저장.

명령

```
cd ~/IITP-DABT-PreProcessing
source venv/bin/activate
python main.py --mode file
```

처리 내용

- API로 데이터/메타/최신일자를 수집
- `kosis_data/YYYYMMDD/` 폴더에 파일 저장
- DB 삽입은 수행하지 않음

6.2. `--mode db` 실행

목적

파일 저장 후 DB에 삽입/갱신까지 수행.

명령

```
cd ~/IITP-DABT-PreProcessing
source venv/bin/activate
python main.py --mode db
```

처리 내용

- `--mode file` 수행 내용 포함
- 추가로 db_processing 모듈이 데이터 삽입, 통합 테이블 이관, 메타데이터 저장, 관리 테이블 업데이트, 과거 데이터 정리를 수행
- 성공 시 `logs/db_YYYYMMDD.log` 에 상세 기록

6.3. 상태 확인 및 종료

- 실행 결과는 표준 출력과 `logs/YYYYMMDD.log` 를 통해 확인
- 비정상 종료 시 `[ERROR]` 메시지 및 스택 트레이스 확인
- 가상환경 종료: `deactivate`

7. 운영 시나리오 및 스케줄링

7.1. 수동 실행 시나리오

1. `ssh iitp-app@<server>`
2. `source ~/IITP-DABT-PreProcessing/venv/bin/activate`
3. `python main.py --mode db`
4. 로그 및 결과 확인

7.2. 자동 실행 예시(CRON)

매일 새벽 3시에 DB 모드 실행:

```
crontab -e
```

추가:

```
0 3 * * * /bin/bash -c 'source /home/iitp-app/IITP-DABT-PreProcessing/venv/bin/activate && cd /h
```

> cron에서는 가상환경 경로와 프로젝트 경로를 절대경로로 작성해야 합니다.

8. 검증 및 문제 해결

8.1. 실행 결과 검증

- `logs/YYYYMMDD.log` 와 `logs/db_YYYYMMDD.log` 확인
- `kosis_data/YYYYMMDD/` 에 데이터/메타/최신 파일 생성 여부 확인
- DB 모드 시 PostgreSQL 테이블에 최신 데이터가 삽입되었는지 확인

8.2. 로그 분석 포인트

- `[INFO]` : 정상 처리 단계 기록
- `[WARNING]` : 데이터 분할 등 주의 메시지(에러는 아님)
- `[ERROR]` : 실행 중단. 스택 트레이스를 확인하여 원인 파악
- `db_YYYYMMDD.log` : DB 삽입 상세 내역

8.3. 문제 발생 시 체크리스트

증상	확인 항목	조치
DB 연결 실패	`.env` 의 DB_URL, 네트워크	접속 정보 수정, 방화벽 확인
API 호출 실패	KOSIS 인증키, 네트워크	인증키 유효성, outbound 443 허용
파일 권한 오류	logs, kosis_data 권한	`chmod 755` , 소유자 확인
cron 미실행	PATH, venv 경로	절대 경로 사용, 로그 확인

부록

부록 A. 주요 디렉토리/파일 요약

경로	설명
`main.py`	실행 진입점
`db_processing.py`	DB 삽입 로직
`kosis_api.py`	API 호출 모듈
`file_utils.py`	파일 저장 유틸리티
`config.py`	환경설정 로딩
`logs/`	실행 로그 디렉토리
`kosis_data/`	수집 데이터 저장

부록 B. 자주 묻는 질문(FAQ)

Q1. Python 버전은 반드시 3.8 이상이어야 하나요?
A1. 네. psycopg2-binary 등 패키지 호환성 때문에 3.8 이상을 권장합니다.

Q2. DB 모드 실행 시 기존 데이터는 어떻게 되나요?
A2. 동일한 날짜의 과거 데이터는 삭제 후 신규 데이터로 대체하며, 통합 테이블 및 메타데이터도 최신 상태로 유지됩니다.

Q3. PART 모드를 사용하려면?
A3. `.env` 에 `DATA_COLLECTION_SCOPE=PART` 와 `[TARGET_SRC_TBL_ID_LIST]` 섹션을 정의해야 합니다. 목록에 없는 통계 ID는 실행 시 오류가 발생합니다.