

backpacker 과제

사용 언어 및 기술 스택

- 언어: Scala 2.12
- 이유:
 - Spark는 Scala로 개발된 프레임워크로, 내부 API 및 문서가 가장 잘 정리되어 있음
 - Java에 비해 concise한 문법으로 배치 파이프라인 설계가 효율적임
 - Spark의 최신 기능 및 Dataset API와의 자연스러운 호환성을 고려하여 선택
- Spark Version: 3.5.5
- 라이브러리:
 - spark-core
 - spark-sql
 - spark-hive

WAU(Weekly Active Users) 계산

6-a. user_id 기준 WAU

```
SELECT
  DATE('$weekStartStr') AS week_start_date,
  COUNT(DISTINCT user_id) AS wau_user
FROM ecommerce_events
WHERE event_date BETWEEN DATE('$weekStartStr') AND DATE('$weekEndStr')
```

6-b. session_id 기준 WAU

```
SELECT
  DATE('$weekStartStr') AS week_start_date,
```

```

COUNT(DISTINCT session_id) AS wau_session
FROM ecommerce_events
WHERE event_date BETWEEN DATE('$weekStartStr') AND DATE('$weekEndStr')

```

결과

기준 주: 2019-10-07 ~ 2019-10-13

WAU (by user_id):

```

+-----+-----+
|week_start_date|wau_user|
+-----+-----+
|2019-10-07    |1049906 |
+-----+-----+

```

WAU (by session_id):

```

+-----+-----+
|week_start_date|wau_session|
+-----+-----+
|2019-10-07    |2107803  |
+-----+-----+

```