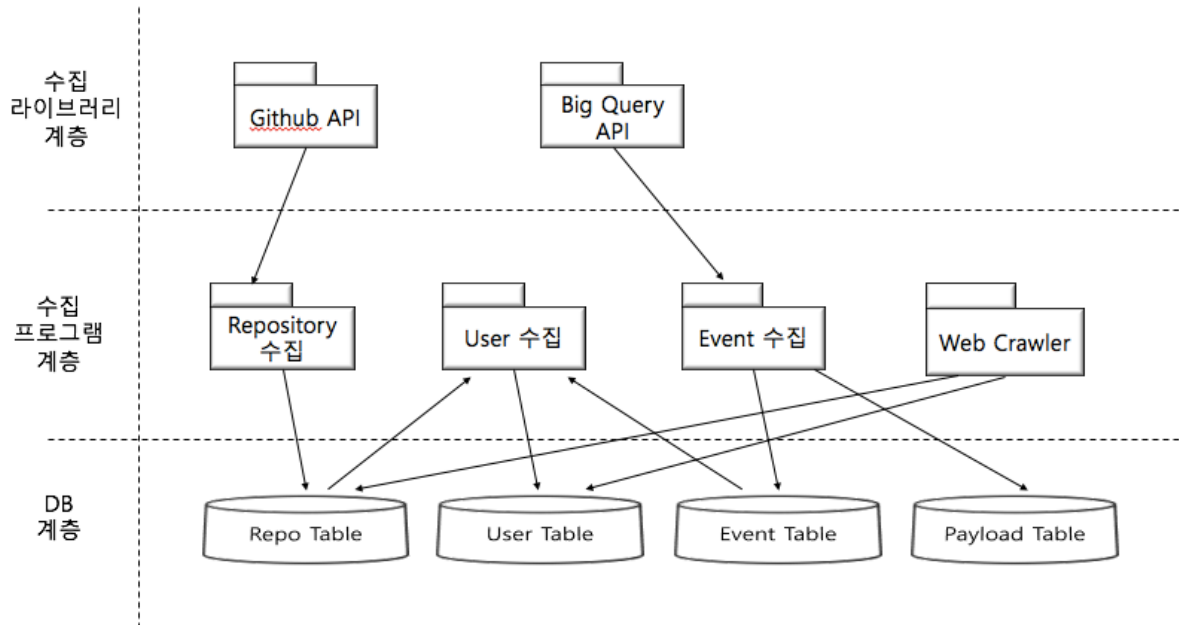


1. 수집 프로그램 아키텍처

1.1 클래스도



1.2 계층 정의

	계층명	개요
1	수집 라이브러리 계층	수집 라이브러리를 통해 데이터를 수집하는 계층
2	수집 SW 계층	수집 SW를 통해 데이터를 수집하는 계층
3	DB 계층	수집한 데이터를 DB에 저장하는 계층

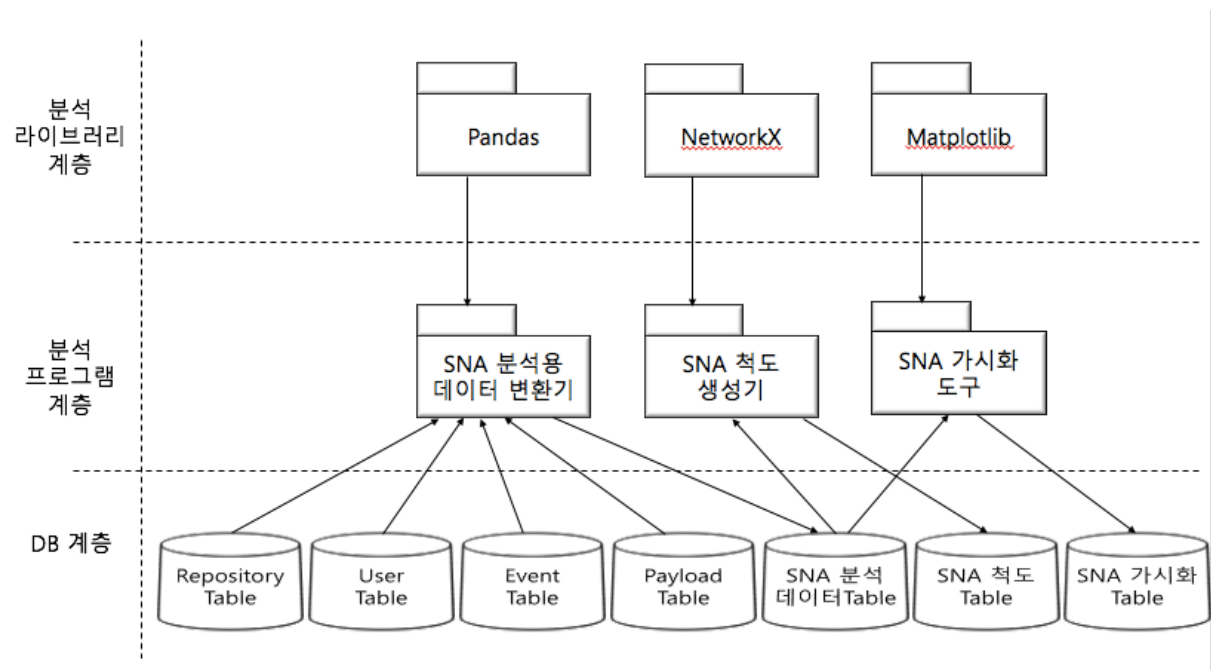
1.3 패키지 정의

	패키지명	개요	소속 클래스 목록
1	Github API	Github 내의 데이터를 호출하는 API	Github API, Httplib2, BeautifulSoup4
2	Big Query API	Big Query 내의 데이터를 호출하는 API	Big Query API, Httplib2, Apiclient, Oauth2client
3	Repository 수집	Github API를 통해 Repository 데이터를 수집하고 Repo_Table 에 저장하는 프로그램	OtherLanguageCollector, StarpReposCounter, RepoDataCollector
4	User 수집	Repository 데이터에서 Owner 데이터를 파싱해오고, Event 데이터에서 Actor 데이터를 파싱해서 User_Table에	OwnerDataParser, ActorDataParser

		저장하는 프로그램	
5	Event 수집	Big Query API를 통해 Event 데이터를 수집해서 Event_Table에 저장하고, 각 Event의 Payload 데이터를 파싱하여 Payload_Table에 저장하는 프로그램	EventDataCollector, PayloadDataParser
6	Web Crawler	Web Crawler를 통해 Repository의 Topic, License 데이터와 User의 추가 데이터 (Location 등)을 수집하여 Repo_Table과 User_Table에 저장하는 프로그램	RepoDetailCollector, UserDetailCollector
7	Repo_Table	Repository와 관련된 DB Table	Repo
8	User_Table	User와 관련된 DB Table	User
9	Event_Table	Event와 관련된 DB Table	Event
10	Payload_Table	Payload와 관련된 DB Table	Payload

2. 분석 프로그램 아키텍처

2.1 클래스도



2.2 계층 정의

	계층명	개요
--	-----	----

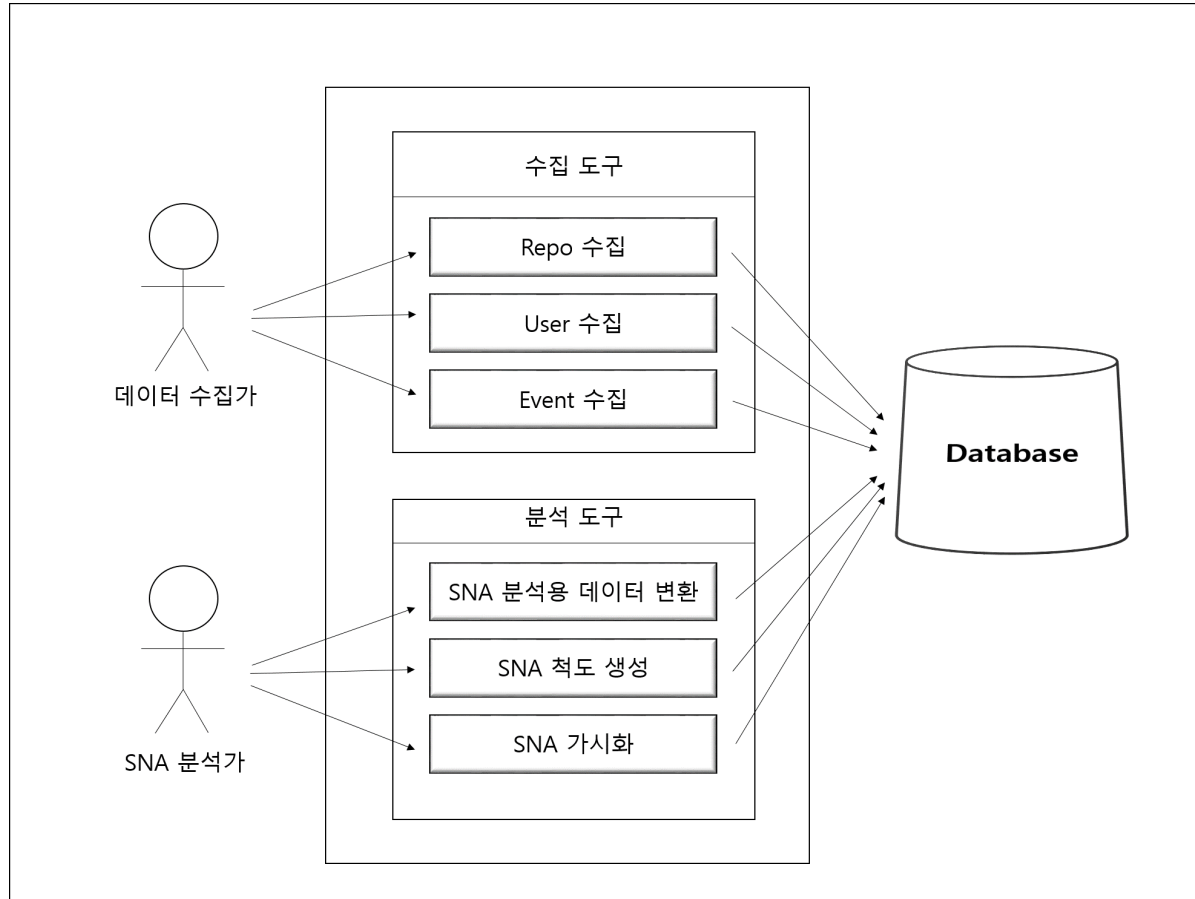
1	분석 라이브러리 계층	분석에 필요한 라이브러리에 대한 계층
2	분석 SW 계층	분석용 데이터 변환, 척도생성, 가시화 등 분석 작업에 대한 계층
3	DB 계층	분석에 관련해서 필요한 입력 데이터와 저장 데이터에 대한 계층

2.3 패키지 정의

	패키지명	개요	소속 클래스 목록
1	Pandas	데이터 구조화와 데이터 분석 라이브러리	Pandas
2	NetworkX	네트워크 생성 및 조작 라이브러리	NetworkX
3	Matplotlib	2D Plotting 라이브러리	Matplotlib
4	SNA 분석용 데이터 변환기	수집된 원시 데이터(Repository, User, Event, Payload Table)를 네트워크 분석에 사용되는 분석용 데이터로 변환하고 이를 SNA 분석 데이터 Table에 저장하는 프로그램	snaDataConverter
5	SNA 척도 생성기	분석용 데이터(SNA 분석 데이터 Table)의 값을 기반으로 네트워크 척도를 계산하고 그 값들을 SNA 척도 Table에 저장하는 프로그램	snaMeasureCreator
6	SNA 가시화 도구	분석용 데이터(SNA 분석 데이터 Table)의 값을 기반으로 가시화하여 이에 대한 이미지들을 생성하고 SNA 가시화 Table에 저장하는 프로그램	snaVisualizer
7	Repository Table	저장소에 관련된 DB Table	Repo
8	User Table	유저에 관련된 DB Table	User
9	Event Table	이벤트에 관련된 DB Table	Event
10	Payload Table	이벤트 Payload에 관련된 DB Table	Payload
11	SNA 분석 데이터 Table	SNA 분석 데이터에 관련된 DB Table	snaAnalysisData
12	SNA 척도 Table	SNA 척도에 관련된 DB Table	snaMeasure
13	SNA 가시화 Table	SNA 가시화에 관련된 DB Table	snaVisualize

3. Use Case

3.1 Use Case



3.2 패키지 정의

	패키지명	패키지 개요	소속 사용사례 목록
1	Repo 수집	저장소 데이터를 수집 및 저장	언어목록 수집
			스타수별 저장소 수 카운팅
			저장소 데이터 수집
			저장소 레이블 수집
			저장소 라이선스 수집
2	User 수집	유저 데이터를 수집 및 저장	Owner 데이터 수집
			Actor 데이터 수집
			유저 추가 데이터(Location 등) 수집
3	Event 수집	이벤트 데이터를 수집 및 저장	이벤트 데이터 수집
			이벤트 Payload 데이터 수집
4	SNA 분석용 데이터 변환	수집된 원시 데이터를 SNA 분석용 데이터로 정제하여	SNA 분석용 데이터 변환

		저장함	
5	SNA 척도 생성	SNA 분석용 데이터를 기반으로 네트워크 척도를 구하고, 이 값들을 저장함	SNA 척도 생성
6	SNA 가시화	SNA 분석용 데이터를 기반으로 네트워크들을 가시화하고 결과들을 저장함	SNA 가시화

3.3 사용사례

	사용사례명	사용사례 개요
1	언어 목록 수집	Github API 를 사용해, 수집 저장소에 대한 언어 목록과 각 언어별 검색 가능 최하위 스타수를 구함
2	스타 수 별 저장소 수 카운팅	Github API 를 사용해, 스타수별 저장소 수를 세어서 스타 수 별 검색 범위를 구함
3	저장소 데이터 수집	Gituhub API 를 사용해, 언어별, 스타 수 별 검색 범위로 데이터를 수집하고 Repo 전체 데이터가 모이면 이를 Repo 테이블에 저장함
4	저장소 레이블 수집	Github API 를 사용해, 저장소 별 Label 을 수집하고 이를 저장소 데이터에 추가함
5	저장소 라이선스 수집	Web Crawling 을 이용해, 저장소 별 라이선스를 수집하고 이를 저장소 데이터에 추가함
6	Owner 데이터 수집	저장소 데이터에서 Owner 데이터를 추출하여 User 테이블에 저장함
7	Actor 데이터 수집	이벤트 데이터의 Actor 를 User 테이블에 저장함
8	유저 추가 데이터 (Location 등) 수집	Web Crawling 을 이용해, 유저 추가 데이터를 수집하고 유저 데이터에 추가함
9	이벤트 데이터 수집	저장소 데이터와 매칭되는 이벤트 데이터를 BigQuery 에서 수집하고 이벤트 테이블에 저장함
10	이벤트 Payload 데이터 수집	이벤트 데이터에서 Payload 데이터를 따로 Payload 데이터에 저장함
11	SNA 분석용 데이터 변환	수집한 원시 데이터들을 분석용 데이터로 변환하여 저장함
12	SNA 척도 생성	SNA 분석용 데이터로 네트워크 척도를 생성하고 이 값들을 저장함

13	SNA 가시화	SNA 분석용 데이터로 값들을 가시화하고 이 결과들을 저장함
----	---------	--------------------------------------