\* 빅데이터의 수집의 8가지 기업은 크롤링, 스크래핑, FTP, 오픈 API , RSS, 스트리밍, 로그, RDB Aggergator 이며 각각의 기본 정의는 아래와 같다

-크롤링기법 : SNS, 웹, 뉴스 등의 인터넷상에 제공하는 웹 문서 정보를 수집한다.

-스크래핑기법 : 인터넷 웹 사이트에 보이는 내용 중에서 특정 정보만을 추출하고 모든 동작을 자동으로 수행한다.

-FTP기법 : TCP/IP 프로토콜을 이용하여 인터넷 서버로부터 각종 파일을 송수신한다.

-오픈 API기법 : 서비스, 정보, 데이터 등의 개방된 정보로부터 API를 통해 데이터를 수집한다.

-RSS기법 : 웹상의 최신 정보를 공유하기 위한 XML 기반의 콘텐츠 배급 프로토콜 이다.

-스트리밍(Streaming)기법 : 인터넷에서 음성, 오디오, 비디오 등의 멀티미디어 데이터를 송수신하는 기술이다.

-로그 Aggregator기법 : 웹 서버 로그, 웹 로그, 트랜잭션 로그, 데이터베이스 로그 등의 각종 서비스 로그를 수집하는 오픈소스 기술.

-RDB Aggregator기법 : 관계형 데이터베이스에서 정형 데이터를 수집하여 하둡 분산 파일 시스템(HDFS)이나 HBase 등의 NoSQL에 저장하는 오픈소스 기술이다.

\* 실무에서 공공데이터 수집 이용하여 프로젝트를 진행할때는 대표적으로 스트리밍 기법과 로그 기법은 사용할 것 같다.

빅데이터의 주요 특성에는 대규모, 다양성, 속도가 있는데 최근에는 대규모 및 다양성이 많이 존제하는 스트리밍 기법이 가장 유효할 것으로 예상된다.

대표적으로 유뷰트 영상의 스트리밍을 이용하는 것이 대규모와 다양성에 모두 해당하며 로그 체크도 확인이 용의하여 프로젝트 진행 시 유용할 것으로 예상된다.