문 서 명	기능 명세서	문서번호	FS-301	작성자	우건희
				작성일	2024-09-30
				버 전	1.0

요구사항 ID		FR-301
기능명		빅데이터를 통한 악성코드 블랙리스트 작성
기능 사항 내역	기능상세	Malwarebazzar 에서 가져온 2020 년도부터 보고된 악성코드 dataset.csv 를 통해 mongoDB vsapi DB 에 info collection 을 구축하는 기능입니다. Dataset.csv 에서는 md5 값으로 보고된 악성코드를 구분합니다. 기존 dataset.csv 에 있는 정보와의 virustotal api 를 사용해 나온 정보와의 불균형을 해결하기 위해서 dataset.csv 에서 악성코드 정보의 md5 를 모듈에서 읽어와 virustotal api 기능 중 hash 값으로 search 하는 기능을 통해 info collection 에 업데이트 합니다. process_hash.py 의 process_hash() 함수를 통해 실행합니다. 이 과정에서 처리된 해시 값은 processed_hashes.json 파일로 저장되고 해당 파일에서 같은 해시 값이 있으면 DB 에 해당 해시 데이터가 있는 것이므로 그 해시 값은 건너뜁니다. virustotal api 는 각 사용자마다 api key 로 query 할 수 있는 사용량이 제한되어있기 때문에 windows server 에 내장되어 있는 task schedular 를 사용하여 매달 3 일동안 사용량의 한계만큼 hash 를 검색해 데이터를 재가공 할 수 있게 trigger 를 걸어두었습니다.
	변경여부	
	변경내역	
	비고	