

대용량 화물 DTG 데이터 분석을 위한 빅데이터 플랫폼 설계 및 구현

김범수^o, 김태학*, 김진욱*
한국건설기술연구원 미래융합연구본부^o
한국건설기술연구원 미래융합연구본부*
e-mail: {bumsookim^o, kimth*, jukim*}@kict.re.kr

Design and Implementation of Big Data Platform for Analyzing Huge Cargo DTG Data

Bum-Soo Kim^o, Tae-Hak Kim*, Jin-Wook Kim*
Dept. of Future Technology and Convergence Research,
Korea Institute of Civil Engineering and Building Technology^o
Dept. of Future Technology and Convergence Research,
Korea Institute of Civil Engineering and Building Technology*

● 요약 ●

본 논문에서는 대용량 화물 DTG 데이터 분석을 위한 빅데이터 플랫폼을 설계 및 구현한다. DTG(digital tacho graph)는 차량운행기록을 실시간으로 저장하는 장치로서, 차량의 GPS, 속도, RPM, 제동유무, 이동거리 등 차량운행 관련 데이터가 1초 단위로 기록된다. 차량 운행 패턴 및 분석을 하기 위해서는 DTG 데이터의 빠른 처리가 필수적이며, 특히 대용량 DTG 데이터를 가공 및 변환하기 위해서는 별도의 플랫폼이 필요하다. 본 논문에서는 오픈소스 기반의 빅데이터 프레임워크인 스파크(Spark)를 이용하여 대용량 화물 DTG 데이터의 전처리 플랫폼을 구현하였다. 실제 대용량 화물 DTG 데이터를 대상으로 데이터를 변환 및 지도상에 표현해 보인다.

키워드: 빅 데이터(big data), 화물 DTG(cargo DTG), 스파크 프레임워크(Spark framework)

I. Introduction

빅 데이터[1]는 네트워크, 데이터 저장 및 데이터 수집 용량의 빠른 발전으로 물리, 생물 및 의학을 포함한 모든 과학 및 엔지니어링 분야에서 급속도로 확대되고 있다. 또한, 숨겨진 가치에 대한 심층적인 이해를 돕고 새로운 가치를 발견 할 수 있는 기회를 제공할 수 있도록 데이터셋을 효과적으로 구성하고 관리하는 과정을 내포하고 있다.

최근 건설 분야에서는 과적단속 예방을 위해 화물 DTG 데이터를 활용한 빅데이터 분석 연구가 진행되었다[2]. 이 연구[2]에서는 화물 DTG 데이터에서 화물차량의 시간대별, 요일별 이동 패턴을 분석하여 과적단속 적정지점을 과적단속 담당자에게 제공하여 효율적인 과적단속 업무를 지원할 수 있도록 하였다.

본 논문에서는 대용량 화물 DTG 데이터 분석을 위해 대표적인 오픈소스 빅 데이터 플랫폼인 스파크(Spark)[3]를 기반으로 하는 전처리 플랫폼을 설계 및 구현한다. 화물 DTG 데이터는 DTG 데이터의 특성상 사이즈가 워낙 방대하여 데이터를 변환 및 가공하는 것이 쉽지 않다. 따라서, 제한된 빅 데이터 플랫폼에서는 스파크를 이용하여

- 1) 한국 내 존재하는 데이터를 얻고, 2) 운행정보가 많은 차량(30분 이상 주행차량)만을 대상으로 삼은 다음, 3) 화물단지로부터 출발도착한 차량만을 골라 4) 1초 단위 데이터를 10초 단위 데이터로 변환하여
- 5) 수원 국토관리사무소 영역 반영하는 작업을 수행한다.

구현 결과, 약 1TB 정도하는 화물 DTG 데이터를 실시간으로 처리되는 것을 확인하였다. 이는 1초 단위의 위치 데이터로서 단순한 가공, 정제뿐만 아니라 공간 정보로서 실시간으로 이해할 수 있도록 직관적으로 표현하는데 그 의미가 있다. 따라서, 본 논문에서 제안한 시스템은 매우 실용적이고 사용자 친화적이라 확신한다.

II. The Proposed Big Data Platform

스파크는 하둡(Hadoop)[4]이 추구하는 맵 리듀스 방식을 벗어나 RDD(resilient distributed dataset)이라는 고유의 자료구조를 생성해 빅데이터 분석을 처리한다. 또한, 스파크는 자바, 스칼라, 파이썬

R 등 프로그래밍에 친화적인 API를 제공하며, SQL, 스트리밍 데이터, 머신러닝 및 그래프 처리를 지원한다. 본 연구에서는 빅데이터 처리 뿐만 아니라 통계분석까지 활용할 수 있는 SparkR API를 사용한다.

화물 DTG 데이터 분석을 위해 사용하는 DTG 데이터 요소들을 Table 1과 같다. DTG 데이터의 특징은 운행차량의 상황을 파악할 수 있도록 차량 상태정보와 위치 정보가 함께 있는 것이다. 본 논문에서는 공간 데이터 분석을 위해 스파크가 제공하는 지오스파크 (Geo-Spark)를 이용하여 화물차량의 GPS와 새주소 도로 형상을 적재 및 처리하고 WGS84 좌표계 기반에서 미터단위의 평면좌표계로 통일하여 분석한다.

Table 1. 화물 DTG 정보

설명	자료형	단위
정보발생일시	String	-
총 누적주행거리(km)	Unsigned integer	km
총 누적주행거리(m)	Unsigned integer	m
일일 주행거리	Unsigned integer	km
자동차 속도	Unsigned integer	km/h
분당엔진회전수	Unsigned integer	RPM
브레이크 신호	Unsigned integer	-
차량위치, GPS x	Integer	-
차량위치, GPS y	Integer	-
GPS 방위각	Unsigned integer	-
가속도 x	Integer	m/s ²
가속도 y	Integer	m/s ²

구현을 위해 본 논문에서는 2.1GHz Intel Xeon Silver 4110 2CPU, 288GB 램, 약 2TB SSD 하드 디스크 사양의 두 대의 서버로 클러스터를 구성하여 사용하였다. 운영체제는 Ubuntu 18.04.1 LTS이며, 데이터베이스로 Postgres와 PostGIS를 함께 연동하여 환경을 구축하였다.

Fig. 1은 구현한 빅 데이터 플랫폼을 이용하여 화물 DTG 데이터를 맵에 표현한 화면 예제이다. 그림을 보면 수원 국토관리사무소 영역을 지나는 운행정보가 많은 화물 차량 중 화물단지로부터 출발/도착한 차량만을 표현한 것이다. 따라서, 본 논문에서 제안한 시스템은 대용량 화물 DTG 데이터를 다루는데 매우 효과적이라 사료된다.

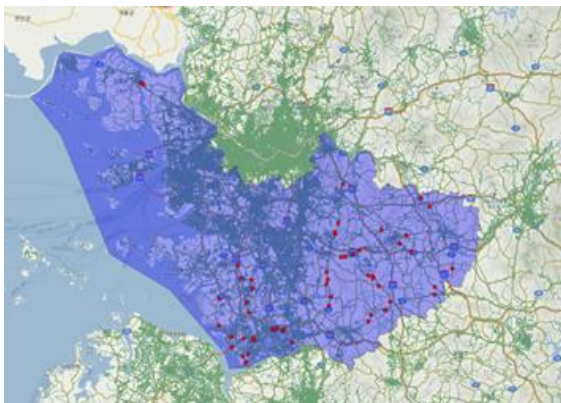


Fig. 1. 화물 DTG 데이터를 표현한 맵 화면 예제

III. Conclusions

본 논문에서는 대용량 화물 DTG 데이터 분석을 위해 빅 데이터 플랫폼을 설계 및 구현하였다. 대용량 DTG 데이터를 실시간으로 가공 및 변환하고 맵에 표현하기 위해 오픈소스 빅 데이터 플랫폼인 스파크를 활용하였다. 제안한 시스템은 DTG 데이터의 특성을 잘 반영하여 사용자가 쉽게 데이터를 이해하고 사용할 수 있도록 구현하였다. 향후 연구로는 본 연구를 기반으로 다른 시설물들의 데이터도 다룰 수 있는 IoT 분석 플랫폼을 설계와 구현할 예정이다.

ACKNOWLEDGEMENT

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 19SCIP-C146569-02).

REFERENCES

- [1] X. Wu, X. Zhu, G. Wu and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 1, pp. 97-107, Jan. 2014.
- [2] T.-H. Kim, B.-S. Kim, and J.-U. Kim, "A Study on the Analysis Method for the Optimal Location of Overstrength Restrictions Using Cargo DTG," In *Proc. of the KAIS Fall Conference*, pp. 1-3, Nov. 2018(In Korean).
- [3] Apache Spark, <https://spark.apache.org/>
- [4] Apache Hadoop, <https://hadoop.apache.org/>