

(51) 국제특허분류:

G06F 17/30 (2006.01)

(21) 국제출원번호:

PCT/KR2017/013633

(22) 국제출원일:

2017년 11월 28일 (28.11.2017)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2016-0177345 2016년 12월 23일 (23.12.2016) KR

(71) 출원인: 부산대학교 산학협력단 (PUSAN NATIONAL UNIVERSITY INDUSTRY-UNIVERSITY COOPERATION FOUNDATION) [KR/KR]; 46241 부산시 금정구 부산대학교로63번길 2 (강전동, 부산대학교), Busan (KR).

(72) 발명자: 푸트라아흐마드 디줄피카 아디 (PUTRA, Ahmad Dzulfikar Adi); 48738 부산시 동구 범일로 120, 212호 (범일동, 현대고시원), Busan (KR). 배혜림 (BAE, Hye Rim); 46241 부산시 금정구 부산대학교로63번길 2, 특성화공학관 623호 (강전동, 부산대학교), Busan (KR). 풀샤시아이큐 레비세이 (PULSHASHI, Iq Reviessey); 46241 부산시 금정구 부산대학교로63번길 2, 공과대학 산업공학과 (강전동, 부산대학교), Busan (KR). 수트리스노와티

리스카아스리아나 (SUTRISNOWATI, Riska Asriana); 46241 부산시 금정구 부산대학교로63번길 2, 공과대학 산업공학과 (강전동, 부산대학교), Busan (KR). 앤디프라스티아부디 와후 (ANDY, Prastyabudi Wahyu); 48738 부산시 동구 범일로 120, 212호 (범일동, 현대고시원), Busan (KR). 최유림 (CHOI, Yu Lim); 48113 부산시 해운대구 좌동순환로311번길 8, 106동 801호 (좌동, 해운대영남아파트), Busan (KR).

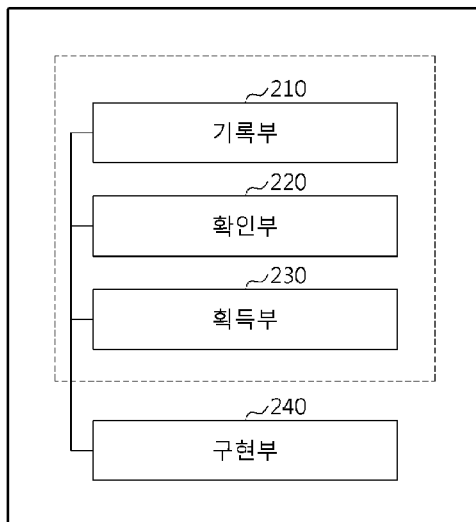
(74) 대리인: 특허법인 무한 (MUHANN PATENT & LAW FIRM); 06044 서울시 강남구 학동로 3길 9, 5층 (논현동, 명림빌딩), Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: HIGH-CAPACITY EVENT LOG REPLAY METHOD AND HIGH-CAPACITY EVENT LOG REPLAY SYSTEM

(54) 발명의 명칭: 대용량 이벤트 로그 재생 방법 및 대용량 이벤트 로그 재생 시스템

200



210 ... Recording unit

220 ... Checking unit

230 ... Acquisition unit

240 ... Implementation unit

(57) Abstract: A high-capacity event log replay method and a high-capacity event log replay system are disclosed. A large capacity event logs replay method according to an embodiment of the present invention comprises: a step of counting event logs generated between a process and a next process performed subsequent to the process and maintaining the counted event logs in a storage system as log files; a step of checking a log file corresponding to a process included in a connection command with respect to a log replay page in the storage system in accordance with generation of the connection command; and a step of dividing the checked log file into a predetermined size, and generating and acquiring a plurality of divided log files from the storage system.

(57) 요약서: 대용량 이벤트 로그 재생 방법 및 대용량 이벤트 로그 재생 시스템이 개시된다. 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 방법은, 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지하는 단계와, 로그 재생 페이지에 대한 접속 명령이 발생함에 따라, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인하는 단계, 및 상기 확인된 로그파일을 정해진 크기로 분할하여, 복수의 분할 로그파일을 생성하여, 상기 스토리지 시스템으로부터 획득하는 단계를 포함한다.



(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 대용량 이벤트 로그 재생 방법 및 대용량 이벤트 로그 재생 시스템

기술분야

- [1] 본 발명은 프로세스 마이닝(Process Mining)에 사용되는 이벤트 로그(Event Log)를 재생하는 로그 재생(Log Replay) 과정에서, 이벤트 로그의 처리 시간을 줄이는 기술에 관한 것이다.

배경기술

- [2] 로그 재생이란, 이번의 로그기록을 프로세스 모델을 통해 시각화하여 재생하는 기술로서, 프로세스 수행에 따라 로그파일에 기록되는 이벤트 로그로부터 지식을 추출하는 프로세스 마이닝 분야에서 사용되고 있다.
- [3] 기존의 로그 재생 기술에 따르면, 프로세스 수행에 따라 대량의 이벤트 로그가 발생되면, 이벤트 로그를 처리하여 토큰 애니메이션을 생성하고 real KPI를 계산하는 과정에서 병목과 편차가 발생하는 문제점이 생길 수 있다.
- [4] 이에 따라, 로그 재생 과정에서 발생하는 병목과 편차 문제를 해결하고, 연산 처리 속도를 향상시킬 수 있는 기술이 요구되고 있다.

발명의 상세한 설명

기술적 과제

- [5] 본 발명의 실시예는 맵 리듀스(MapReduce) 알고리즘을 이용하여, 분산 API 웹 서비스를 기반으로 대용량 이벤트 로그를 작은 파티션으로 분할 함으로써, 이벤트 로그를 재생하는 처리 속도를 향상시켜, 병목과 편차 문제를 해결하는 것을 목적으로 한다.

과제 해결 수단

- [6] 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 방법은, 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지하는 단계와, 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인하는 단계, 및 상기 확인된 로그파일을 정해진 크기로 분할하여, 복수의 분할 로그파일을 생성하여, 상기 스토리지 시스템으로부터 획득하는 단계를 포함한다.
- [7] 또한, 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템은, 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지하는 기록부와, 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인하는 확인부, 및 상기 확인된 로그파일을 정해진 크기로 분할하여, 복수의 분할 로그파일을 생성하여,

상기 스토리지 시스템으로부터 획득하는 획득부를 포함한다.

발명의 효과

- [8] 본 발명의 일실시예에 따르면, 이벤트 로그를 재생하는 로그 재생 과정에서, 맵리듀스(MapReduce) 알고리즘을 이용하여, 분산 API 웹 서비스를 기반으로 대용량 이벤트 로그를 작은 파티션으로 분할 함으로써, 이벤트 로그를 재생하는 처리 속도를 향상시켜, 병목과 편차 문제를 해결할 수 있다.
- [9] 또한, 본 발명의 일실시예에 따르면, 대용량 이벤트 로그를 분할하여 분산 API 웹 서비스 형태로 애니메이션을 생성해 Real KPI를 계산하고, 분할된 이벤트 로그를 순차적으로 클라이언트 웹 어플리케이션에 재생 함으로써, 대용량 이벤트 로그의 분할 및 분산 처리를 통해 로그 재생 처리 성능을 향상시킬 수 있다.

도면의 간단한 설명

- [10] 도 1은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템의 구성을 도시한 도면이다.
- [11] 도 2는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템의 내부 구성을 도시한 블록도이다.
- [12] 도 3은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 이벤트 로그 및 로그파일의 일례를 도시한 도면이다.
- [13] 도 4는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 이벤트 로그로부터 생성되는 프로세스 모델의 일례를 도시한 도면이다.
- [14] 도 5a 및 5b는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 로그 재생 페이지를 제공하는 일련의 과정을 도시한 흐름도이다.
- [15] 도 6은, 도 5b에 도시한 단계(517)를 세부적으로 나타낸 상세흐름도이다.
- [16] 도 7은, 도 5b에 도시한 단계(524)를 세부적으로 나타낸 상세흐름도이다.
- [17] 도 8은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 로그 재생 페이지를 시각화 하는 일례를 도시한 도면이다.
- [18] 도 9는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 방법의 순서를 도시한 흐름도이다.

발명의 실시를 위한 최선의 형태

- [19] 이하, 첨부된 도면들을 참조하여 본 발명의 일실시예에 따른 응용프로그램 업데이트 장치 및 방법에 대해 상세히 설명한다. 그러나, 본 발명이 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [20] 도 1은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템의 구성을 도시한 도면이다.
- [21] 도 1을 참조하면, 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템(100)은, 웹 브라우저(110), HDFS(120) 및 SPARK(130)를 포함하여 구성될

수 있다.

- [22] 웹 브라우저(110)는 관리자 단말에서 실행되는 웹 어플리케이션 중 하나로서, 일례로, 구글 사의 '크롬(Chrome)'을 예로 들 수 있다.
- [23] 웹 브라우저(110)는 관리자 단말에서 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정(프로세스)에 대응되는 로그파일을, HDFS(120)에서 확인하고, 확인된 로그파일을 정해진 크기로 분할한 복수의 분할 로그파일을, HDFS(120)으로부터 순차적으로 획득할 수 있다.
- [24] 웹 브라우저(110)는 REST_API 기반으로 상기 복수의 분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그를 처리하여, 공정 별 이벤트 발생 정도를 나타내는 로그 재생 페이지를 관리자 단말에 구현할 수 있다.
- [25] 즉, 웹 브라우저(110)는 REST_API 기반으로 복수의 분할 로그파일 각각을 처리하여, 상기 로그 재생 페이지를 단계적으로 시각화하여 구현할 수 있다.
- [26] 실시예에 따라, 웹 브라우저(110)는 HDFS(120)에 상기 접속 명령과 관련된 로그파일이 유지되지 않는 경우, SPARK(130)에서의 공정 수행에 따른 이벤트 로그의 발생을 대기할 수 있다.
- [27] 여기서, SPARK(130)는 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 HDFS(120)에 유지할 수 있다.
- [28] SPARK(130)는 범용 고성능 분할 처리 플랫폼의 일례로서, 맵리듀스(MapReduce) 함수에 따라, 메모리 기반으로 프로세스를 분할 수행하여 데이터를 분산 처리하는 기능을 할 수 있다.
- [29] 여기서, 맵리듀스(MapReduce)는 대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 제작되었으며, 2004년 발표한 소프트웨어 프레임워크로서 함수형 프로그래밍에서 일반적으로 사용되는 Map과 Reduce라는 함수로 구성될 수 있다.
- [30] 이에 따라, 본 발명의 대용량 이벤트 로그 재생 시스템(100)은 이벤트 로그를 재생하는 로그 재생 과정에서, 대용량 이벤트 로그를 작은 파티션으로 분할 함으로써, 이벤트 로그를 재생하는 처리 속도를 향상시킬 수 있다.
- [31] 여기서, REST_API(Representational state transfer API)는, 인터넷 상의 컴퓨터 시스템들 사이에서 일방향의 정보 처리 상호 운용(interoperability)을 제공하며, 시스템들에 접근 요청을 허용하거나, 웹 리소스의 텍스트 표시를 제어하는 웹 서비스 API의 하나를 지칭할 수 있다.
- [32] HDFS(Hadoop File System)(120)는, 대량의 자료를 분산 처리하는 스토리지 시스템의 하나로서, 대용량 파일을 나눠서 저장하여 데이터 처리를 용이하게 할 수 있다.
- [33] 도 2는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템의 내부 구성을 도시한 블록도이다.

- [34] 도 2를 참조하면, 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템(200)은, 기록부(210), 확인부(220), 획득부(230) 및 구현부(240)를 포함하여 구성할 수 있다.
- [35] 기록부(210)는 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지한다.
- [36] 즉, 기록부(210)는 분할 처리 플랫폼(예, 'SPARK')에 의한 프로세스 수행 과정에서 다양한 작업행위에 대해 발생하는 이벤트 로그를 발생순서에 따라 로그파일로서 기록할 수 있다.
- [37] 기록부(210)는 로그파일을, 대량의 자료를 분산 처리하는 스토리지 시스템(예, 'HDFS')에 분산하여 저장 함으로써, 로그 재생 시 대용량 이벤트 로그의 처리가 보다 용이해지도록 할 수 있다.
- [38] 예를 들어, 도 5a 및 5b를 참조하면, 기록부(210)는 로그 재생 알고리즘에 따라 프로세스를 처리하고, 처리한 결과를, HDFS에 유지하는 SPARK에 의해 구현될 수 있다.
- [39] 여기서, 이벤트 로그는, 프로세스(공정)를 구성하는 단위 작업인 이벤트(예컨대, 작업의 시작, 종료, 취소 등)들의 집합으로서, 하나의 케이스와 관련을 가지며, 누가 언제 어디서 어떤 작업을 수행하였는지에 대한 정보를 나타낼 수 있도록, 다양한 부가적인 정보를 포함하여 로그파일로서 기록될 수 있다.
- [40] 일례로, 이벤트 로그는 작업을 시작하거나 수행하는 리소스(예, 작업수행자, 시스템, 장비 등), 이벤트의 타임스탬프(timestamp, 이벤트 발생시간) 및 이벤트와 연관된 데이터(예컨대, '주문의 규모'), 이벤트의 액티비티(Activity) 중 적어도 하나의 여러 가지 부가적인 정보를 포함할 수 있다. 여기서, 액티비티는 프로세스를 구성하는 기본 단위로서 작업의 수행 등을 나타내는 정보일 수 있다.
- [41] 확인부(220)는 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인한다.
- [42] 확인부(220)는, 상기 접속 명령을 발생하는 관리자 단말에서 실행되는 웹 브라우저(예컨대, 구글 사의 '크롬(Chrome)')에 의해 구현될 수 있다.
- [43] 확인부(220)는 상기 스토리지 시스템(예, 'HDFS')에 상기 접속 명령과 관련된 로그파일이 유지되지 않는 경우, 분할 처리 플랫폼(예, 'SPARK')에서의 공정 수행에 따른 이벤트 로그의 발생을 대기할 수 있다.
- [44] 구체적으로, 획득부(230)는 상기 스토리지 시스템에 상기 접속 명령과 관련된 로그파일이 유지되지 않는 경우, 상기 분할 처리 플랫폼에서, 프로세스가 처리됨에 따라 이벤트 로그가 생성되어, 상기 이벤트 로그를 기록한 로그파일, 상기 스토리지 시스템 내에 유지되면, 상기 스토리지 시스템으로부터, 상기 생성된 로그파일을 획득할 수 있다.
- [45] 예를 들어, 도 5b의 단계(514 내지 518)를 참조하면, HDFS에 유효한 데이터가

존재하지 않는 것으로 확인되면, SPARK는, 로그 재생 알고리즘에 따라 프로세스를 처리한 결과(로그파일)를 HDFS에 저장하고, HDFS는 저장된 HDFS 파일을 앱 서버 & REST API로 반환하여, 웹 브라우저에 JASON으로 응답할 수 있다.

- [46] 보다 구체적으로, 도 6을 참조하면, 상기 분할 처리 플랫폼(SPARK)은, 이벤트 로그의 시퀀스로 구성된 케이스에 대한 리스트가 입력되면, 상기 각 케이스 내 모든 이벤트 로그를 맵(Map) 함수를 적용하여 이벤트 트랜지션으로 매핑하고, 상기 이벤트 트랜지션에 리듀스(Reduce) 함수를 적용하고, 소트(Sort) 함수를 이용하여 상기 이벤트 트랜지션을 생성시점에 따라 정렬하고, 파티션을 초기화하고, 인서트(Insert) 함수를 이용하여, 상기 파티션 각각에 일정 개수(예를 들어, '5,000개')의 이벤트 트랜지션을 삽입하고, 상기 각 파티션을 상기 스토리지 시스템에 유지하고, 상기 각 파티션에 대한 요약정보를 상기 스토리지 시스템에 유지할 수 있다.
- [47] 획득부(230)는 상기 확인된 로그파일을 정해진 크기(파티션)로 분할하여, 복수의 분할 로그파일을 생성하여, 상기 스토리지 시스템(예, 'HDFS')으로부터 획득한다.
- [48] 즉, 획득부(230)는 상기 스토리지 시스템으로 분할 명령을 전송하고, 상기 분할 명령에 따라, 상기 스토리지 시스템에서, 상기 로그파일을 파티션 단위로 분할하여 상기 복수의 분할 로그파일을 생성하도록 할 수 있다.
- [49] 이때, 획득부(230)는 상기 공정의 종류 또는 상기 로그파일의 용량에 따라, 분할 로그파일이 갖는 크기를 정하고, 정해진 상기 크기에 비례하여, 상기 스토리지 시스템으로부터의, 상기 분할 로그파일 각각의 획득 간격을 조정할 수 있다.
- [50] 상기 스토리지 시스템은, 상기 분할 명령에 의해 지정되는, 파티션 당 이벤트 로그의 개수 및 파티션의 개수 중 적어도 하나를 고려하여, 상기 로그파일을 분할할 수 있다.
- [51] 일례로, 획득부(230)는 상기 로그파일 내에 위치하는 순서에 상응하여, 상기 복수의 분할 로그파일 각각을 순차적으로 획득할 수 있다. 예를 들어, 획득부(230)는 HDFS로부터 확인한 로그파일을 파티션 단위로 분할한 복수의 분할 로그파일을, HDFS로부터 순차적으로 획득할 수 있다.
- [52] 또한, 상기 이벤트 로그는, 발생시점에 대응하여 상기 로그파일에 기록되고, 획득부(230)는 상기 복수의 분할 로그파일 중 제1 분할 로그파일에 포함되는 일정 개수의 이벤트 로그가 처리되는 동안, 상기 로그파일 내, 상기 제1 분할 로그파일 다음에 위치하는 제2 분할 로그파일을, 상기 스토리지 시스템으로부터 획득할 수도 있다.
- [53] 또한, 획득부(230)는 상기 스토리지 시스템으로 분할 명령을 전송하고, 상기 분할 명령에 포함되는 토큰에 의해 식별되는 특정의 분할 로그파일을 상기 스토리지 시스템으로부터 획득할 수도 있다.
- [54] 구현부(240)는 분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그에

대한 렌더링을 통해, 공정 별 이벤트 발생 정도를 나타내는 로그 재생 페이지를 구현한다.

- [55] 즉, 구현부(240)는 REST_API 기반의 웹 어플리케이션('웹 브라우저')을 통해, 상기 복수의 분할 로그파일을 처리하여, 상기 로그 재생 페이지를 시각화 할 수 있다.
- [56] 일례로, 구현부(240)는 스토리지 시스템(일례로, 하둡 파일 시스템 'HDFS')으로부터, 순차적으로 획득한 복수의 분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그에 대한 렌더링을 통해, 로그 재생 페이지를 도 8과 같이 관리자 단말에 구현할 수 있다.
- [57] 도 8을 참조하면, 구현부(240)는 이벤트 로그에 포함된 액티비티('QW', 'PP', 'RC', 'SP', 'AN', 'PR', 'RW' 및 'WS')를 각각 노드로서 표시하고, 각 노드를 아크로 연결하고, 노드 사이에서 아크를 따라 움직이는 원을 통해 프로세스의 흐름을 표현하고, 원의 크기를 통해 프로세스의 흐름의 양을 나타냄으로써, 관리자 단말에서 로그 재생 페이지를 시각화 할 수 있다.
- [58] 이와 같이, 본 발명의 일실시예에 따르면, 이벤트 로그를 재생하는 로그 재생 과정에서, 맵 리듀스(MapReduce) 알고리즘을 이용하여, 분산 API 웹 서비스를 기반으로 대용량 이벤트 로그를 작은 파티션으로 분할 함으로써, 이벤트 로그를 재생하는 처리 속도를 향상시켜, 병목과 편차 문제를 해결할 수 있다. 또한, 본 발명의 일실시예에 따르면, 대용량 이벤트 로그를 분할하여 분산 API 웹 서비스 형태로 애니메이션을 생성해 Real KPI를 계산하고, 분할된 이벤트 로그를 순차적으로 클라이언트 웹 어플리케이션에 재생 함으로써, 대용량 이벤트 로그의 분할 및 분산 처리를 통해 로그 재생 처리 성능을 향상시킬 수 있다.
- [59] 도 3은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 이벤트 로그의 구성 및 로그파일의 일례를 도시한 도면이다.
- [60] 본 발명에서, 이벤트 로그(Event Log)는, 프로세스 마이닝(Process Mining)을 위한 투입으로 사용되는 이벤트들의 집합을 나타내고, 이벤트(Event)는 프로세스(공정)를 구성하는 단위 작업으로서, 작업의 시작, 종료, 취소 등과 같이 로그파일에 기록되는 행위를 나타낼 수 있다.
- [61] 참고로, 프로세스 마이닝은 기계나 정보 관련 시스템에서 생성되는 이벤트 로그(transaction dataset)를 분석하여 지식을 추출하는 기법에 관한 것이다.
- [62] 이벤트(이벤트 로그)는 하나의 케이스와 관련을 가지며, 케이스(Case)는 이벤트들의 시퀀스로서, 분석할 프로세스에 의해 처리된 개체들인 프로세스 인스턴스(Process Instance)를 지칭할 수 있다.
- [63] 이벤트 로그는, 각 이벤트에 대해, 프로세스 마이닝 기법에 활용 가능한 다양한 부가적인 정보를 포함하여 구성될 수 있다.
- [64] 일례로, 이벤트 로그는, 누가 언제 어디서 어떤 작업을 수행하였는지에 대한 정보를 나타낼 수 있도록, 작업을 시작하거나 수행하는 리소스(예, 작업수행자, 시스템, 장비 등), 이벤트의 타임스탬프(timestamp, 이벤트 발생시간) 및 이벤트와

연관된 데이터(예컨대, '주문의 규모'), 이벤트의 액티비티(Activity) 중 적어도 하나의 여러 가지 부가적인 정보를 포함할 수 있다. 여기서, 액티비티는 프로세스를 구성하는 기본 단위로서 작업의 수행 등을 나타내는 정보일 수 있다.

[65] 도 3의 (i)은 이벤트 로그('W[4]')의 구성의 일례를 나타내고 있다.

[66] 도 3의 (i)을 참조하면, 이벤트 로그('W[4]')는, 이벤트('e')의 집합('E')과, 이벤트 타입('et')의 집합('ET')과, 액티비티('a')의 집합('A')과, 리소스('r')의 집합('R')과, 케이스('c')의 집합('C')과, 각 이벤트에 타임스탬프를 할당하는 함수('t:E→iR₀⁺')와, 각 이벤트에 이벤트 타입을 할당하는 함수('et:E→ET')와, 각 이벤트를 액티비티에 연관시키는 함수('a:E→EA')와, 각 이벤트를 리소스에 연관시키는 함수('r:E→R∪{⊥}'), 및 각 이벤트를 케이스에 연관시키는 함수('c:E→C')로 구성될 수 있다.

[67] 이벤트 로그는, 프로세스 수행 과정에서 다양한 작업행위에 대해 발생되며, 발생순서에 따라 로그파일에 기록될 수 있다.

[68] 도 3의 (ii)에 도시한 테이블은 이벤트 로그를 기록한 로그파일의 일례를 나타내고 있다.

[69] 도 3의 (ii)를 참조하면, 이벤트 로그는, 각 이벤트를, 이벤트 식별정보(ID)와, 타임스탬프(이벤트 발생시간), 액티비티(이벤트의 작업), 리소스(예, 작업수행자, 시스템, 장비 등) 및 코스트 중 적어도 하나의 부가정보에 연관시켜 로그파일에 기록할 수 있다.

[70] 이때, 이벤트 로그(이벤트들)은 하나의 로그파일이 아닌, 서로 다른 데이터베이스에 흩어져 있는 로그파일에 기록될 수도 있다.

[71] 도 4는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 이벤트 로그로부터 생성되는 프로세스 모델의 일례를 도시한 도면이다.

[72] 본 발명의 대용량 이벤트 로그 재생 시스템은, 프로세스 마이닝 기법에 따라 이벤트 로그에 대한 분석을 통해 프로세스 모델(Process Model)을 생성할 수 있다.

[73] 일례로, 본 발명의 대용량 이벤트 로그 재생 시스템은, 작업의 순서와 같은 프로세스의 흐름에 중점을 두고, 모든 가능한 경로의 특징을 파악 함으로써, 예를 들어 Petri net 모델이나, EPC, BPMN, UML activity diagram과 같은 프로세스 모델을 도출할 수 있다.

[74] 또한, 본 발명의 대용량 이벤트 로그 재생 시스템은, 이벤트 로그에 숨겨진 리소스에 대한 정보에 초점을 맞추어, 어떤 행위자(예, 작업자, 시스템, 역할, 부서)가 업무 수행에 관여하고, 그들이 어떻게 연결되어 있는지 파악하여, 역할이나 기능에 따라 사람들을 분류하여 조직 체계를 만들거나 사람들 사이의 소셜 네트워크를 보여주는 프로세스 모델을 도출할 수 있다.

[75] 또한, 본 발명의 대용량 이벤트 로그 재생 시스템은, 케이스의 특징에 초점을 맞추어, 프로세스 모델을 도출할 수 있다. 즉, 본 발명의 대용량 이벤트 로그 재생 시스템은, 프로세스 내에서의 경로 또는 케이스에 참여하는 작업자를 통해

케이스의 특성을 분석하거나, 케이스와 연관된 데이터의 값에 따라 케이스의 특성을 분석할 수도 있다. 예를 들어, 보충 주문을 나타내는 케이스가 있을 경우, 주문된 상품의 수량과 공급자는 유용한 정보가 될 수 있다.

- [76] 또한, 본 발명의 대용량 이벤트 로그 재생 시스템은, 이벤트 로그에 포함되는 타임스탬프를 통해 이벤트의 시간과 빈도를 분석하여, 프로세스 모델을 도출할 수도 있다. 일례로, 본 발명의 대용량 이벤트 로그 재생 시스템은, 병목점 도출, 서비스의 레벨 측정, 리소스 활용도 모니터링 및 수행 중인 케이스의 잔여 시간 예측 등을 수행할 수 있다.
- [77] 도 4의 (i)은 프로세스 모델('G')의 구성의 일례를 나타내고 있다.
- [78] 도 4의 (i)을 참조하면, 프로세스 모델('G')은 노드(node)의 집합('N')과, 노드들을 연결하는 아크(arc)의 집합('E $\subseteq$ N $\times$ N') 및 하나의 노드를 액티비티의 집합에 연관시키는 함수($na:N \rightarrow P(A)$)로 구성될 수 있다.
- [79] 도 4의 (ii)는, 도 3의 (ii)에 도시한 로그파일에 근거하여 도출한 프로세스 모델의 그래프를 나타내고 있다.
- [80] 도 4의 (ii)를 참조하면, 대용량 이벤트 로그 재생 시스템은, 도 3의 (ii)에 도시한 액티비티(예컨대, 'register request', 'examine thoroughly', 'check ticket', 'decide', 'reject request', 'examine casually', 'pay compensation')를 각각 노드로서 생성하고, 상기 각 노드를 아크로 연결하여, 도시한 것과 같은 프로세스 모델의 그래프를 도출할 수 있다.
- [81] 도 5a 및 5b는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 로그 재생 페이지를 제공하는 일련의 과정을 도시한 흐름도이다.
- [82] 도 5a를 참조하면, 단계(501 내지 502)에서, 웹 브라우저는, 관리자 단말로부터 로그 재생 페이지에 대한 오픈(open) 명령이 수신되면, 로그 재생 페이지를 앱 서버 & REST API로 요청한다.
- [83] 단계(503 내지 505)에서, 웹 브라우저는, 앱 서버 & REST API로부터 HTML로 응답을 수신하면, 휴리스틱 마이너 API(heuristic miner API)를 호출하여, HDFS 파일(로그파일)을 요청한다.
- [84] 단계(506)에서 HDFS는, HDFS 파일 요청이 수신되면, HDFS에 유효한 데이터가 존재하는지를 확인한다.
- [85] 단계(506)에서의 확인 결과, HDFS에 유효한 데이터가 존재하는 경우, 단계(507)에서, HDFS는 HDFS 파일을 앱 서버 & REST API로 반환하고, 단계(508)에서 앱 서버 & REST API는, 웹 브라우저에 JASON으로 응답한다.
- [86] 단계(506)에서의 확인 결과, HDFS에 유효한 데이터가 존재하지 않는 경우, 단계(509 내지 510)에서, SPARK는, 휴리스틱 마이너 알고리즘에 따라 프로세스를 처리하고, 처리한 결과(로그파일)를 HDFS에 저장한다. 단계(507)에서, HDFS는 저장된 HDFS 파일을 앱 서버 & REST API로 반환하고, 단계(508)에서, 앱 서버 & REST API는, 웹 브라우저에 JASON으로 응답한다.
- [87] 여기서, JASON(JavaScript Object Notation)은 경량 데이터 교환 포맷으로서,

사람에게 있어서 읽고 쓰기가 쉽고, 기계에 있어서 파싱과 생성이 용이하며, 자바스크립트 프로그래밍 언어의 서브셋에 기반을 두고 있다.

[88] 단계(511)에서, 웹 브라우저는, 프로세스 모델을 렌더링하고, 사용자 인터랙션 함수(User Interaction function)를 초기화 한다.

[89] 도 5b를 참조하면, 단계(512 내지 513)에서, 웹 브라우저는 로그 재생 페이지를 호출하여 HDFS 파일을 요청한다.

[90] 단계(514)에서 HDFS는, HDFS 파일 요청이 수신되면, HDFS에 유효한 데이터가 존재하는지를 확인한다.

[91] 단계(514)에서의 확인 결과, HDFS에 유효한 데이터가 존재하는 경우, 단계(515)에서, HDFS는 HDFS 파일을 앱 서버 & REST API로 반환하고, 단계(516)에서 앱 서버 & REST API는, 웹 브라우저에 JASON으로 응답한다.

[92] 단계(514)에서의 확인 결과, HDFS에 유효한 데이터가 존재하지 않는 경우, 단계(517 내지 518)에서, SPARK는, 로그 재생 알고리즘에 따라 프로세스를 처리하고, 처리한 결과(로그파일)를 HDFS에 저장한다. 단계(515)에서, HDFS는 저장된 HDFS 파일을 앱 서버 & REST API로 반환하고, 단계(516)에서, 앱 서버 & REST API는, 웹 브라우저에 JASON으로 응답한다.

[93] 단계(519)에서, 웹 브라우저는, 애니메이션을 초기화 하고, 모든 파티션을 리스팅 한다.

[94] 구체적으로, 웹 브라우저는, 애니메이션 핸들링을 위해, 모 타임라인(parent timeline, 'tmp')을 초기화 한다.

[95] 단계(520 내지 521)에서, 웹 브라우저는, 파티션 API를 호출하여 파티션 P_i 의 토큰 데이터를 획득하고, 단계(522)에서, HDFS는 상기 토큰 데이터를 이용해 HDFS 파일을 앱 서버 & REST API로 반환하고, 단계(523)에서, 앱 서버 & REST API는, 웹 브라우저에 JASON으로 응답한다.

[96] 단계(524 내지 526)에서, 웹 브라우저는, 파티션(P_i)의 토큰에 대한 애니메이션을 초기화 하고, 애니메이션을 재생한 후, 차기 파티션(P_{i+1})을 획득한다.

[97] 도 6은, 도 5b에 도시한 단계(517)를 세부적으로 나타낸 상세흐름도이다.

[98] 도 6에는, SPARK에서 로그 재생 알고리즘에 따라 프로세스를 처리하고, 처리한 결과를 HDFS에 저장하는 상세한 과정이 도시되어 있다.

[99] 도 6을 참조하면, 단계(601)에서, SPARK는 케이스('C')의 리스트를 시스템으로부터 검색한다.

[100] SPARK는 이벤트 로그('E')에서 이벤트 트랜지션('k')의 리스트를 매핑하는 것으로 시작한다. 케이스('C')의 리스트는 시스템으로부터 검색되고, 각 케이스('C_i')는 이벤트('e')의 시퀀스를 포함한다.

[101] 단계(602)에서, SPARK는 차기 케이스('C_{i+1}')가 존재하는지 확인한다.

[102] 상기 단계(602)에서의 확인 결과, 차기 케이스('C_{i+1}')가 존재하는 경우, 단계(603)에서, SPARK는, 케이스 내에 이벤트('e')를 리스팅 한다.

- [103] 본 단계(602 내지 603)에서, SPARK는 이벤트('e')의 시퀀스를 검사하기 위해 각 케이스('C_i')를 반복한다.
- [104] 단계(604)에서, SPARK는, 차기 이벤트('e_{i+1}')가 존재하는지 확인한다.
- [105] 상기 단계(604)에서의 확인 결과, 차기 이벤트('e_{i+1}')가 존재하는 경우, 단계(605)에서, SPARK는, MAP 함수를 적용하여 이벤트 트랜지션(K_{i+1}) 속성 값을 설정한다. 여기서, 이벤트 트랜지션(Event Transition)은, 연속된 2개의 이벤트 간 변화를 나타낼 수 있다.
- [106] 본 단계(604 내지 605)에서, SPARK는 이벤트 트랜지션('K_j') 속성을 설정하기 위해 케이스('C_i')에 이벤트('e_j')의 시퀀스를 반복한다.
- [107] 상기 단계(604)에서의 확인 결과, 차기 이벤트('e_{i+1}')가 존재하지 않는 경우, SPARK는 단계(602)로 이동하여, 차기 케이스('C_{i+1}')가 존재하는지 재확인한다.
- [108] 상기 단계(602)에서의 확인 결과, 차기 케이스('C_{i+1}')가 존재하지 않는 경우, 단계(606)에서, SPARK는, 이벤트 트랜지션('K')에 REDUCE 함수를 적용한다.
- [109] 본 단계(606)에서, SPARK는 맵 함수를 완료한 후, 리듀스 함수를 실행하여, 전체 이벤트 트랜지션 데이터 K를 분산 시스템으로부터 수집한다.
- [110] 단계(607)에서, SPARK는, 소트(SORT) 함수를 적용하여 이벤트 트랜지션('K')을 시작시간(start time)에 따라 정렬한다. 여기서, 결과값은 이벤트 트랜지션('K')의 시퀀스로서 형성되며, 이벤트 트랜지션('K') 데이터의 리스트는 시작시간에 따라 정렬된다.
- [111] 단계(608)에서, SPARK는, 파티션('P')의 리스트를 초기화 하고, 단계(609)에서, SPARK는 차기 파티션('P_{i+1}')이 존재하는지 확인한다.
- [112] 여기서, SPARK는 파티션('P') 당 제한되는 빈 슬롯 수를 포함하여 파티션('P')의 리스트를 반복적으로 초기화 한다.
- [113] 상기 단계(609)에서의 확인 결과, 차기 파티션('P_{i+1}')이 존재하는 경우, 단계(610)에서, SPARK는 파티션('P_i')를 초기화 하고, 단계(611)에서, 빈 슬롯이 존재하는지 확인한다.
- [114] 상기 단계(611)에서의 확인 결과, 빈 슬롯이 존재하는 경우, 단계(612)에서, SPARK는 빈 슬롯에 이벤트 트랜지션('K')을 입력한 후, 단계(611)로 이동하여 빈 슬롯이 존재하는지 재확인한다.
- [115] 본 단계(611 내지 612)에서, SPARK는 각 파티션은 빈 슬롯을 포함하기 때문에, 이벤트 트랜지션('K') 속성을 빈 슬롯에 반복적으로 입력한다.
- [116] 상기 단계(611)에서의 확인 결과, 빈 슬롯이 존재하지 않는 경우, 단계(613 내지 614)에서, SPARK는 파티션('P_i') 속성 값을 설정하고, HDFS로 파티션('P_i') 데이터를 저장한다.
- [117] 여기서, SPARK는 차기 파티션('P_{i+1}')을 반복하기 전에, 파티션('P_i') 속성 값을 설정한 후 파티션 데이터를 HDFS에 저장할 필요가 있다.
- [118] 상기 단계(609)에서의 확인 결과, 차기 파티션('P_{i+1}')이 존재하지 않는 경우, 단계(615 내지 616)에서, SPARK는 파티션 요약정보(Partitions Summary)('Ps')

- 속성 값을 설정하고, 파티션 요약정보('Ps') 데이터를 HDFS에 저장한다.
- [119] 여기서, SPARK는 마지막 파티션에서 이벤트 트랜지션의 처리를 완료한 후에, 파티션 요약정보('Ps') 속성 값을 설정한 뒤, 그 데이터를 HDFS에 저장한다.
- [120] 도 7은, 도 5b에 도시한 단계(524)를 세부적으로 나타낸 상세흐름도이다.
- [121] 도 7에는 웹 브라우저에서 파티션('P_i')의 토큰에 대한 애니메이션을 초기화 하는 과정이 도시되어 있다.
- [122] 도 7을 참조하면, 단계(701 내지 702)에서, 웹 브라우저는 모 타임라인('tmp')을 중지하고, 파티션('P_i')을 위한 애니메이션 핸들링을 위해, 자 타임라인(child time, 'tmc_i')을 초기화 한다.
- [123] 단계(703 내지 704)에서, 웹 브라우저는 파티션('P_i')에서 토큰(K)을 리스팅하고, 같은 소스(source), 타겟, 개시 시간 및 완료 시간에 따라, 토큰(K)을 그룹핑 한다.
- [124] 단계(705 내지 707)에서, 웹 브라우저는 파티션('P_i')에 그룹핑된 토큰(Kg)을 리스팅하고, 그룹핑된 토큰(Kg)을 위한 SVG를 생성하고, 생성한 KgSVG를 타임라인(tmc_i)에 추가한다.
- [125] 여기서, SVG(Scalable Vector Graphics)는 상호성(interactivity)과 애니메이션(animation)을 지원하는 2차원 그래픽 XML 기반의 벡터 이미지 포맷으로서, W3C(World Wide Web Consortium)에 의해 개발된 공개 표준으로서, 동적이고 상호적이라는 특징을 가진다.
- [126] 단계(708)에서, 웹 브라우저는 차기 그룹핑된 토큰(Kg_{i+1})이 존재하는지 확인하고, 차기 그룹핑된 토큰(Kg_{i+1})이 존재하는 것으로 확인되면, 단계(706)으로 이동하여 차기 그룹핑된 토큰(Kg)을 위한 SVG를 생성한다.
- [127] 차기 그룹핑된 토큰(Kg_{i+1})이 존재하지 않는 것으로 확인되면, 단계(709)에서, 웹 브라우저는 모 타임라인('tmp')을 재개한다.
- [128] 도 8은 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 시스템에서, 로그 재생 페이지를 시각화 하는 일례를 도시한 도면이다.
- [129] 도 8을 참조하면, 본 발명의 실시예들에 따른 대용량 이벤트 로그 재생 시스템은, 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정(Process)에 대응되는 로그파일을 정해진 크기로 분할하여, 복수의 분할 로그파일을 생성할 수 있다.
- [130] 대용량 이벤트 로그 재생 시스템은, 복수의 분할 로그파일 각각을 스토리지 시스템(일례로, 하둡 파일 시스템 'HDFS')으로부터 순차적으로 획득하고, 상기 복수의 분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그에 대한 렌더링을 통해, 로그 재생 페이지를 도 8과 같이 관리자 단말에 구현할 수 있다.
- [131] 도 8을 참조하면, 대용량 이벤트 로그 재생 시스템은, 이벤트 로그에 포함된 액티비티('QW', 'PP', 'RC', 'SP', 'AN', 'PR', 'RW' 및 'WS')를 각각 노드로서 표시하고, 각 노드를 아크로 연결하고, 노드 사이에서 아크를 따라 움직이는 원을 통해 프로세스의 흐름을 표현하고, 원의 크기를 통해 프로세스의 흐름의 양을 나타냄으로써, 관리자 단말에서 로그 재생 페이지를 시각화 할 수 있다.

- [132] 이하, 도 9에서는 본 발명의 실시예들에 따른 대용량 이벤트 로그 재생 시스템(200)의 작업 흐름을 상세히 설명한다.
- [133] 도 9는 본 발명의 일실시예에 따른 대용량 이벤트 로그 재생 방법의 순서를 도시한 흐름도이다.
- [134] 본 실시예에 따른 대용량 이벤트 로그 재생 방법은 상술한 대용량 이벤트 로그 재생 시스템(200)에 의해 수행될 수 있다.
- [135] 도 9를 참조하면, 단계(910)에서, 대용량 이벤트 로그 재생 시스템(200)은, 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지한다.
- [136] 대용량 이벤트 로그 재생 시스템(200)은 분할 처리 플랫폼(예, 'SPARK')에 의한 프로세스 수행 과정에서 다양한 작업행위에 대해 발생하는 이벤트 로그를 발생순서에 따라 로그파일로서 기록하고, 로그파일을, 대량의 자료를 분산 처리하는 스토리지 시스템(예, 'HDFS')에 분산하여 저장 함으로써, 로그 재생 시 대용량 이벤트 로그의 처리가 보다 용이해지도록 할 수 있다.
- [137] 단계(920)에서, 대용량 이벤트 로그 재생 시스템(200)은, 관리자 단말에서 로그 재생 페이지에 대한 접속 명령이 발생하는지 판단한다.
- [138] 단계(920)에서의 판단 결과, 로그 재생 페이지에 대한 접속 명령이 발생하는 경우, 단계(930)에서, 대용량 이벤트 로그 재생 시스템(200)은, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인한다.
- [139] 단계(940)에서, 대용량 이벤트 로그 재생 시스템(200)은, 확인된 로그파일을 정해진 크기(파티션)로 분할하여, 복수의 분할 로그파일을 생성하여, 상기 스토리지 시스템(예, 'HDFS')으로부터 획득한다.
- [140] 일례로, 대용량 이벤트 로그 재생 시스템(200)은 상기 스토리지 시스템으로 분할 명령을 전송하고, 상기 분할 명령에 따라, 상기 스토리지 시스템에서, 상기 로그파일을 파티션 단위로 분할하여 상기 복수의 분할 로그파일을 생성하도록 할 수 있다.
- [141] 이때, 대용량 이벤트 로그 재생 시스템(200)은, 상기 공정의 종류 또는 상기 로그파일의 용량에 따라, 분할 로그파일이 갖는 크기를 정하고, 정해진 상기 크기에 비례하여, 상기 스토리지 시스템으로부터의, 상기 분할 로그파일 각각의 획득 간격을 조정할 수 있다.
- [142] 대용량 이벤트 로그 재생 시스템(200)은, 복수의 분할 로그파일을 HDFS로부터 순차적으로 획득할 수 있다. 즉, 대용량 이벤트 로그 재생 시스템(200)은, 상기 복수의 분할 로그파일 중 제1 분할 로그파일에 포함되는 일정 개수의 이벤트 로그가 처리되는 동안, 상기 로그파일 내, 상기 제1 분할 로그파일 다음에 위치하는 제2 분할 로그파일을, 상기 스토리지 시스템으로부터 획득할 수 있다.
- [143] 단계(950)에서, 대용량 이벤트 로그 재생 시스템(200)은, 분할 로그파일 각각에 대한 처리를 통해, 로그 재생 페이지를 웹 기반으로 시각화 한다.
- [144] 즉, 대용량 이벤트 로그 재생 시스템(200)은, REST_API 기반의 웹

어플리케이션('웹 브라우저')을 통해, 상기 복수의 분할 로그파일을 처리하여, 공정 별 이벤트 발생 정도를 나타내는 로그 재생 페이지를 시각화 할 수 있다.

- [145] 실시예에 따라, 대용량 이벤트 로그 재생 시스템(200)은, 상기 스토리지 시스템(예, 'HDFS')에 상기 접속 명령과 관련된 로그파일이 유지되지 않는 것으로 확인되면, 분할 처리 플랫폼(예, 'SPARK')에서의 공정 수행에 따른 이벤트 로그의 발생을 대기할 수 있다.
- [146] 구체적으로, 대용량 이벤트 로그 재생 시스템(200)은, 상기 스토리지 시스템에 상기 접속 명령과 관련된 로그파일이 유지되지 않는 경우, 상기 분할 처리 플랫폼에서, 프로세스가 처리 됨에 따라 이벤트 로그가 생성되어, 상기 이벤트 로그를 기록한 로그파일이, 상기 스토리지 시스템 내에 유지되면, 상기 스토리지 시스템으로부터, 상기 생성된 로그파일을 획득할 수 있다.
- [147] 보다 구체적으로, 도 6을 참조하면, 상기 분할 처리 플랫폼(SPARK)은, 이벤트 로그의 시퀀스로 구성된 케이스에 대한 리스트가 입력되면, 상기 각 케이스 내 모든 이벤트 로그를 맵(Map) 함수를 적용하여 이벤트 트랜지션으로 매핑하고, 상기 이벤트 트랜지션에 리듀스(Reduce) 함수를 적용하고, 소트(Sort) 함수를 이용하여 상기 이벤트 트랜지션을 생성시점에 따라 정렬하고, 파티션을 초기화하고, 인서트(Insert) 함수를 이용하여, 상기 파티션 각각에 일정 개수(예를 들어, '5,000개')의 이벤트 트랜지션을 삽입하고, 상기 각 파티션을 상기 스토리지 시스템에 유지하고, 상기 각 파티션에 대한 요약정보를 상기 스토리지 시스템에 유지할 수 있다.
- [148] 본 발명의 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [149] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로

수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

- [150] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

- [청구항 1] 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트 로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지하는 단계; 로그 재생 페이지에 대한 접속 명령이 발생 함에 따라, 상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지 시스템에서 확인하는 단계; 및 상기 확인된 로그파일을 정해진 크기로 분할하여, 복수의 분할 로그파일을 생성하여, 상기 스토리지 시스템으로부터 획득하는 단계를 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 2] 제1항에 있어서, 상기 복수의 분할 로그파일을 이용하여, 공정 별 이벤트 발생 정도를 나타내는 상기 로그 재생 페이지를 구현하는 단계로서, 분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그에 대한 렌더링을 통해, 상기 로그 재생 페이지를 구현하는 단계를 더 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 3] 제2항에 있어서, 상기 로그 재생 페이지를 구현하는 단계는, REST_API 기반의 웹 어플리케이션을 통해, 상기 복수의 분할 로그파일을 처리하여, 상기 로그 재생 페이지를 시각화 하는 단계를 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 4] 제1항에 있어서, 상기 공정의 종류 또는 상기 로그파일의 용량에 따라, 분할 로그파일이 갖는 크기를 정하는 단계; 및 정해진 상기 크기에 비례하여, 상기 스토리지 시스템으로부터의, 상기 분할 로그파일 각각의 획득 간격을 조정하는 단계를 더 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 5] 제1항에 있어서, 상기 스토리지 시스템으로 분할 명령을 전송하는 단계; 및 상기 분할 명령에 따라, 상기 스토리지 시스템에서, 상기 로그파일을 파티션 단위로 분할하여 상기 복수의 분할 로그파일을 생성하는 단계를 더 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 6] 제5항에 있어서, 상기 복수의 분할 로그파일을 생성하는 단계는, 상기 분할 명령에 의해 지정되는, 파티션 당 이벤트 로그의 개수 및 파티션의 개수 중 적어도 하나를 고려하여, 상기 로그파일을 분할하는 단계를 더 포함하는 대용량 이벤트 로그 재생 방법.

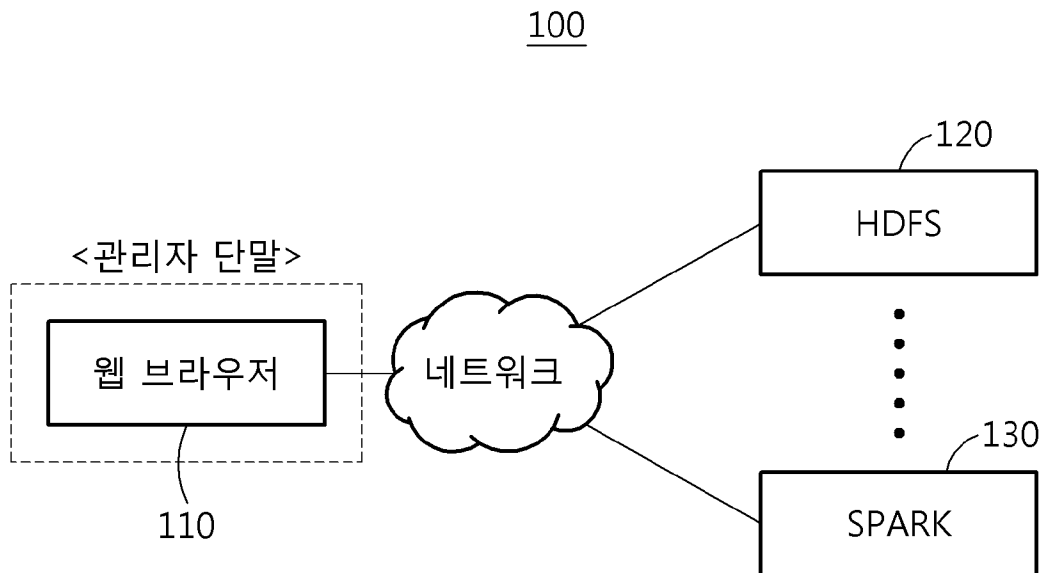
- [청구항 7] 제1항에 있어서,
상기 스토리지 시스템으로 분할 명령을 전송하는 단계; 및
상기 분할 명령에 포함되는 토큰에 의해 식별되는 특정의 분할
로그파일을 상기 스토리지 시스템으로부터 획득하는 단계
를 더 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 8] 제1항에 있어서,
상기 획득하는 단계는,
상기 로그파일 내에 위치하는 순서에 상응하여, 상기 복수의 분할
로그파일 각각을 순차적으로 획득하는 단계
를 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 9] 제1항에 있어서,
상기 이벤트 로그는, 발생시점에 대응하여 상기 로그파일에 기록되고,
상기 획득하는 단계는,
상기 복수의 분할 로그파일 중 제1 분할 로그파일에 포함되는 일정
개수의 이벤트 로그가 처리되는 동안, 상기 로그파일 내, 상기 제1 분할
로그파일 다음에 위치하는 제2 분할 로그파일을, 상기 스토리지
시스템으로부터 획득하는 단계
를 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 10] 제1항에 있어서,
상기 스토리지 시스템에 상기 접속 명령과 관련된 로그파일이 유지되지
않는 경우,
분할 처리 플랫폼에서의 공정 수행에 따른 이벤트 로그의 발생을
대기하는 단계
를 더 포함하는 대용량 이벤트 로그 재생 방법.
- [청구항 11] 공정과, 상기 공정에 연이어 수행되는 차기 공정 전에 발생하는 이벤트
로그를 카운트하여, 로그파일로서 스토리지 시스템에 유지하는 기록부;
로그 재생 페이지에 대한 접속 명령이 발생 함에 따라,
상기 접속 명령에 포함되는 공정에 대응되는 로그파일을, 상기 스토리지
시스템에서 확인하는 확인부; 및
상기 확인된 로그파일을 정해진 크기로 분할하여, 복수의 분할
로그파일을 생성하여, 상기 스토리지 시스템으로부터 획득하는 획득부
를 포함하는 대용량 이벤트 로그 재생 시스템.
- [청구항 12] 제11항에 있어서,
분할 로그파일 각각에 포함되는 일정 개수의 이벤트 로그에 대한
렌더링을 통해, 공정 별 이벤트 발생 정도를 나타내는 로그 재생 페이지를
구현하는 구현부
를 더 포함하는 대용량 이벤트 로그 재생 시스템.
- [청구항 13] 제11항에 있어서,

상기 획득부는,
상기 공정의 종류 또는 상기 로그파일의 용량에 따라, 분할 로그파일이
갖는 크기를 정하고, 정해진 상기 크기에 비례하여, 상기 스토리지
시스템으로부터, 상기 분할 로그파일 각각의 획득 간격을 조정하는
대용량 이벤트 로그 재생 시스템.

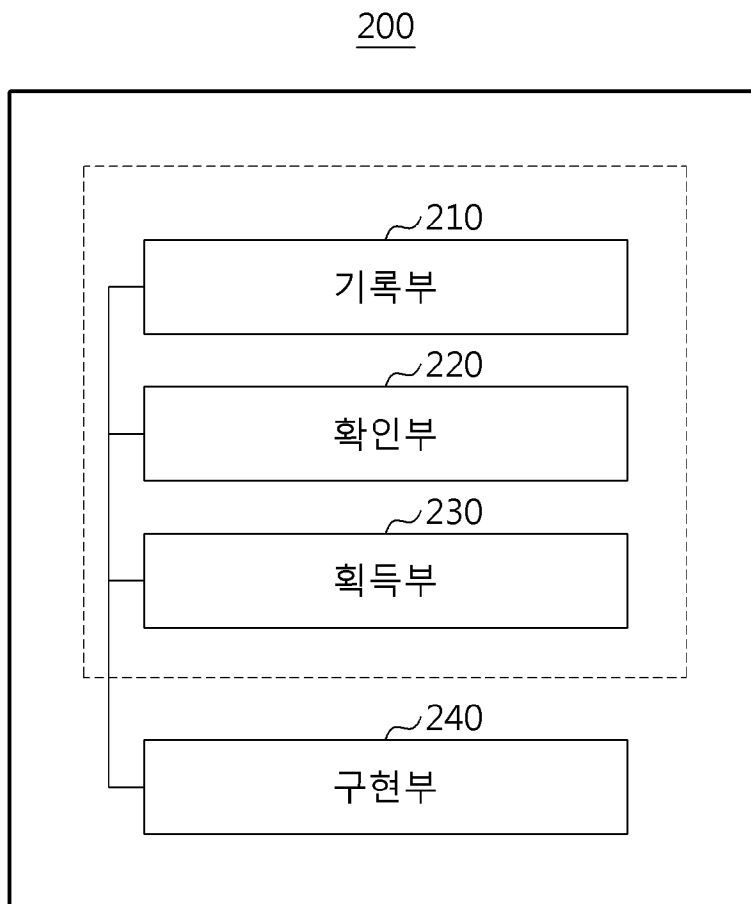
[청구항 14] 제11항에 있어서,
상기 획득부는,
상기 스토리지 시스템으로 분할 명령을 전송하고,
상기 분할 명령에 따라, 상기 스토리지 시스템에서, 상기 로그파일을
파티션 단위로 분할하여 상기 복수의 분할 로그파일을 생성하도록 하는
대용량 이벤트 로그 재생 시스템.

[청구항 15] 제11항에 있어서,
상기 이벤트 로그는, 발생시점에 대응하여 상기 로그파일에 기록되고,
상기 획득부는,
상기 복수의 분할 로그파일 중 제1 분할 로그파일에 포함되는 일정
개수의 이벤트 로그가 처리되는 동안, 상기 로그파일 내, 상기 제1 분할
로그파일 다음에 위치하는 제2 분할 로그파일을, 상기 스토리지
시스템으로부터 획득하는
대용량 이벤트 로그 재생 시스템.

[도1]



[도2]



[도3]

(i) 이벤트 로그 W[4]

W = (E, ET, A, R, C, t, et, a, r, c)

(ii)

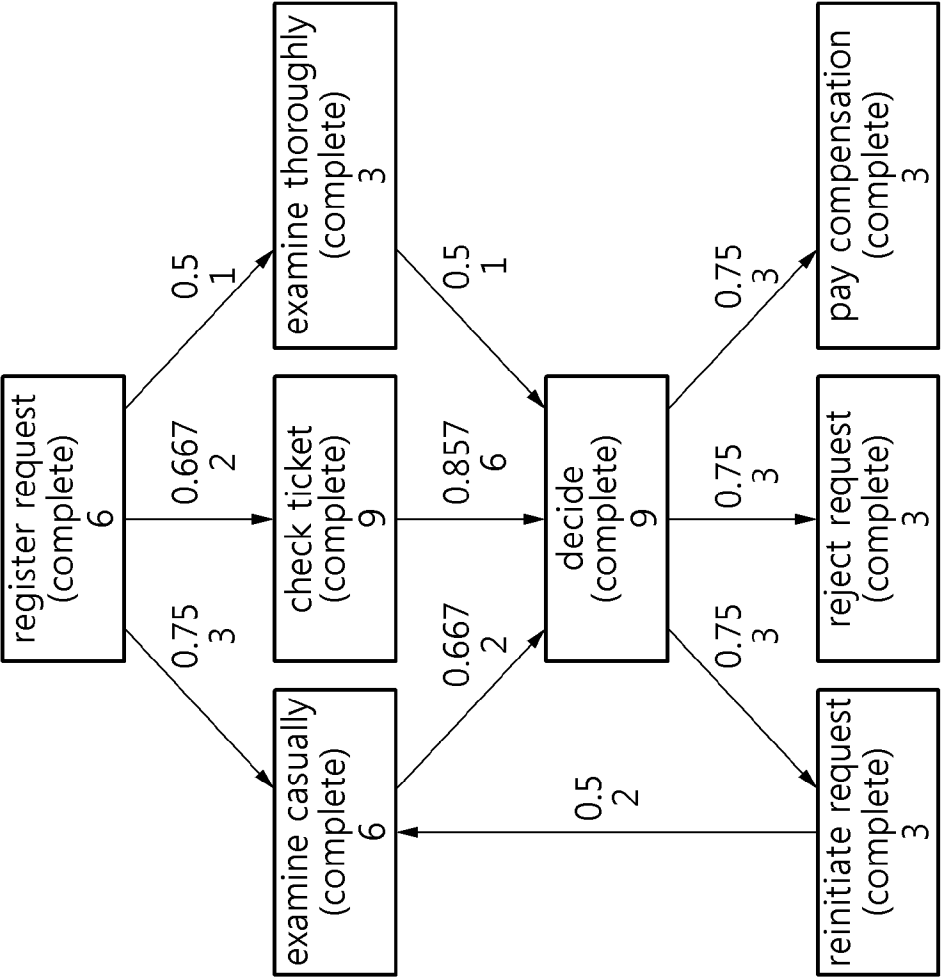
CID	Event ID	Timestamp	Activity	Resource	Costs
1	35654423	30-12-2010:11.02	register request	Pete	50
1	35654424	31-12-2010:10.06	examine thoroughly	Sue	400
1	35654425	05-01-2011:15.12	check ticket	Mike	100
1	35654426	06-01-2011:11.18	decide	Sara	200
1	35654427	07-01-2011:14.24	reject request	Pete	200
2	35654483	30-12-2010:11.32	register request	Mike	50
2	35654485	30-12-2010:12.12	check ticket	Mike	100
2	35654487	30-12-2010:14.16	examine casually	Sean	400
2	35654488	05-01-2011:11.22	decide	Sara	200
2	35654489	08-01-2011:12.05	pay compensation	Ellen	200
...	...	...	...	...	...
6	35654871	06-01-2011:15.02	register request	Mike	50

[도4]

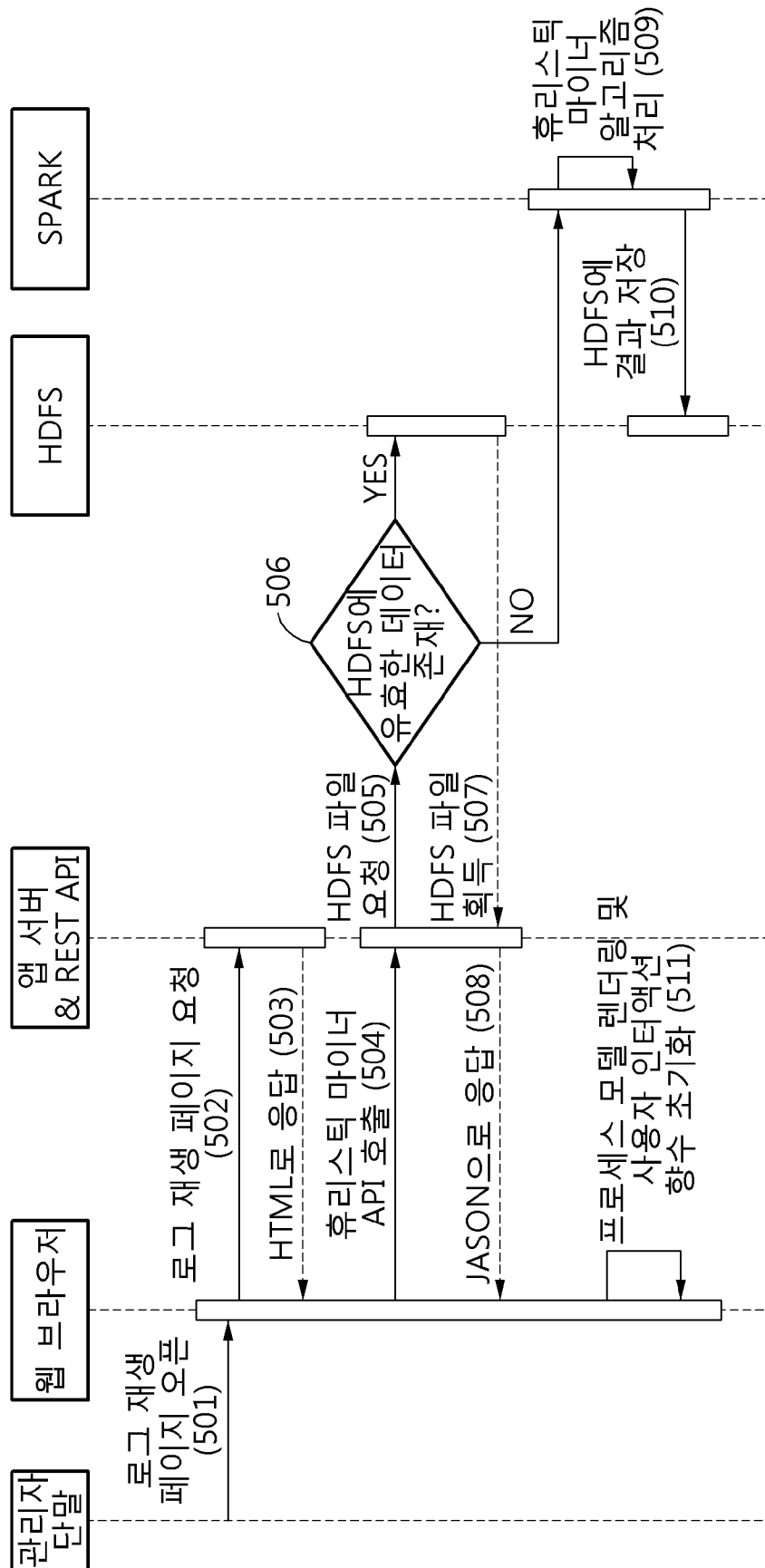
(ii)

(i) 프로세스 모델 G

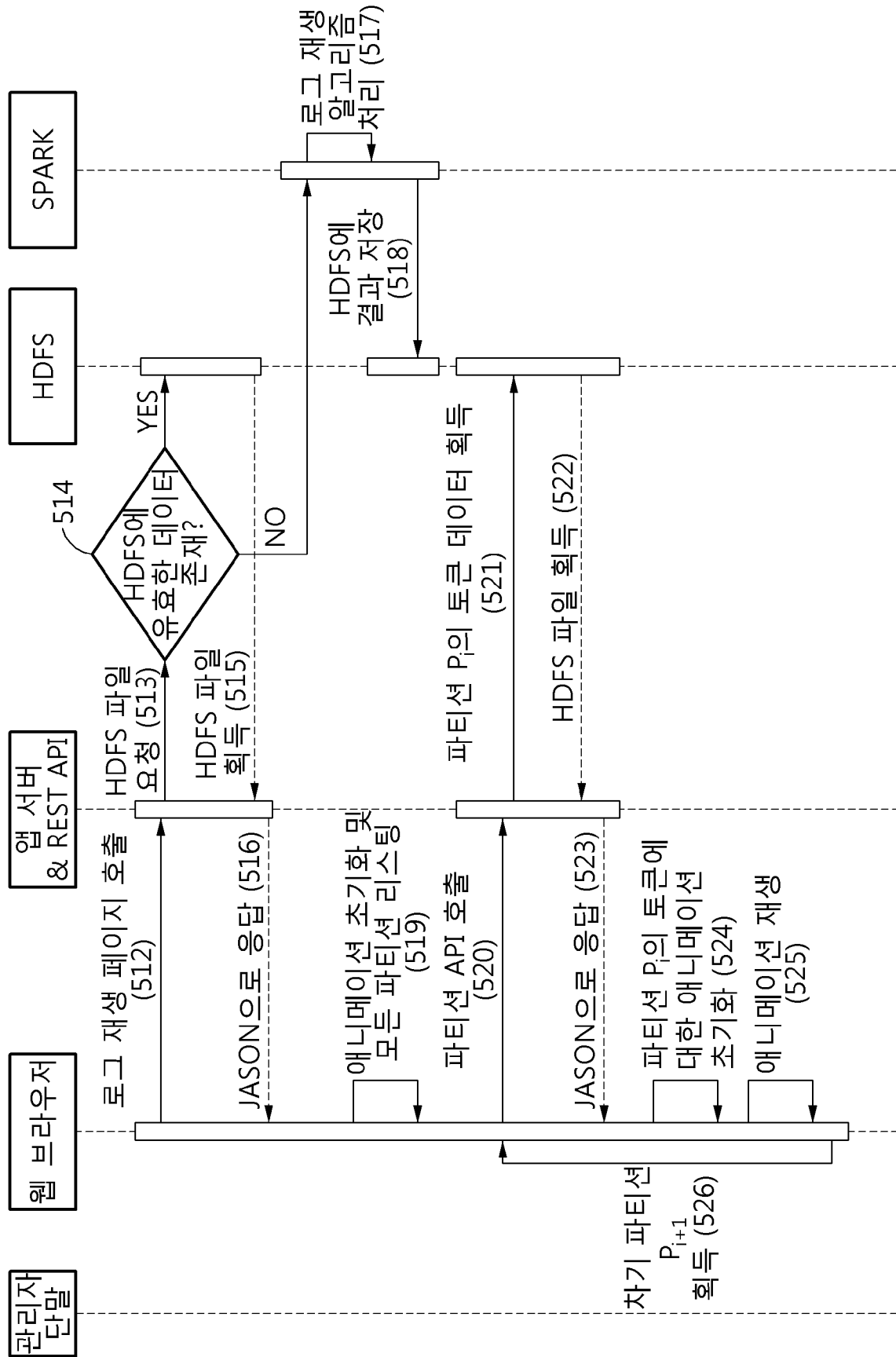
$G = (N, E, na)$



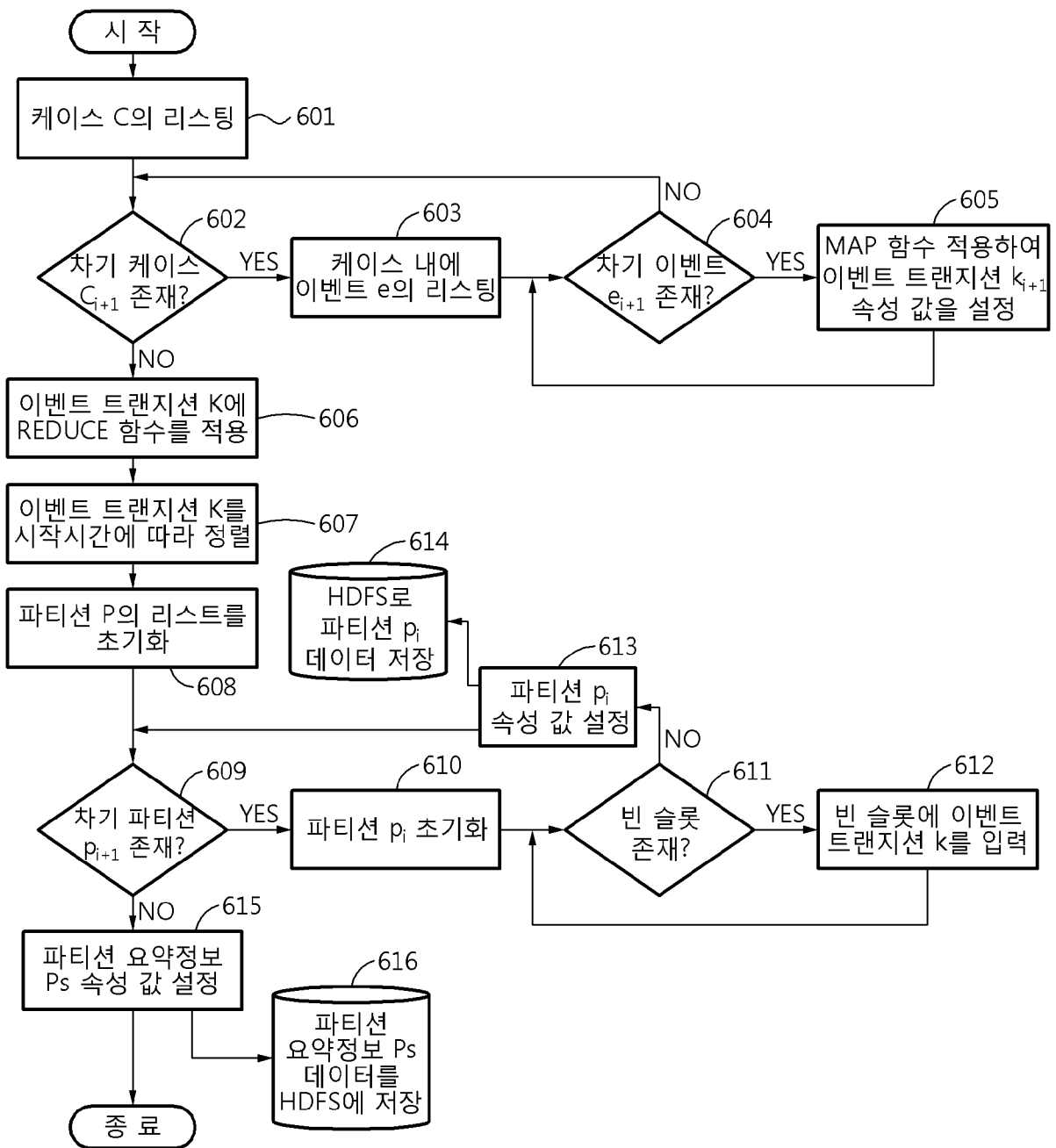
[도5a]



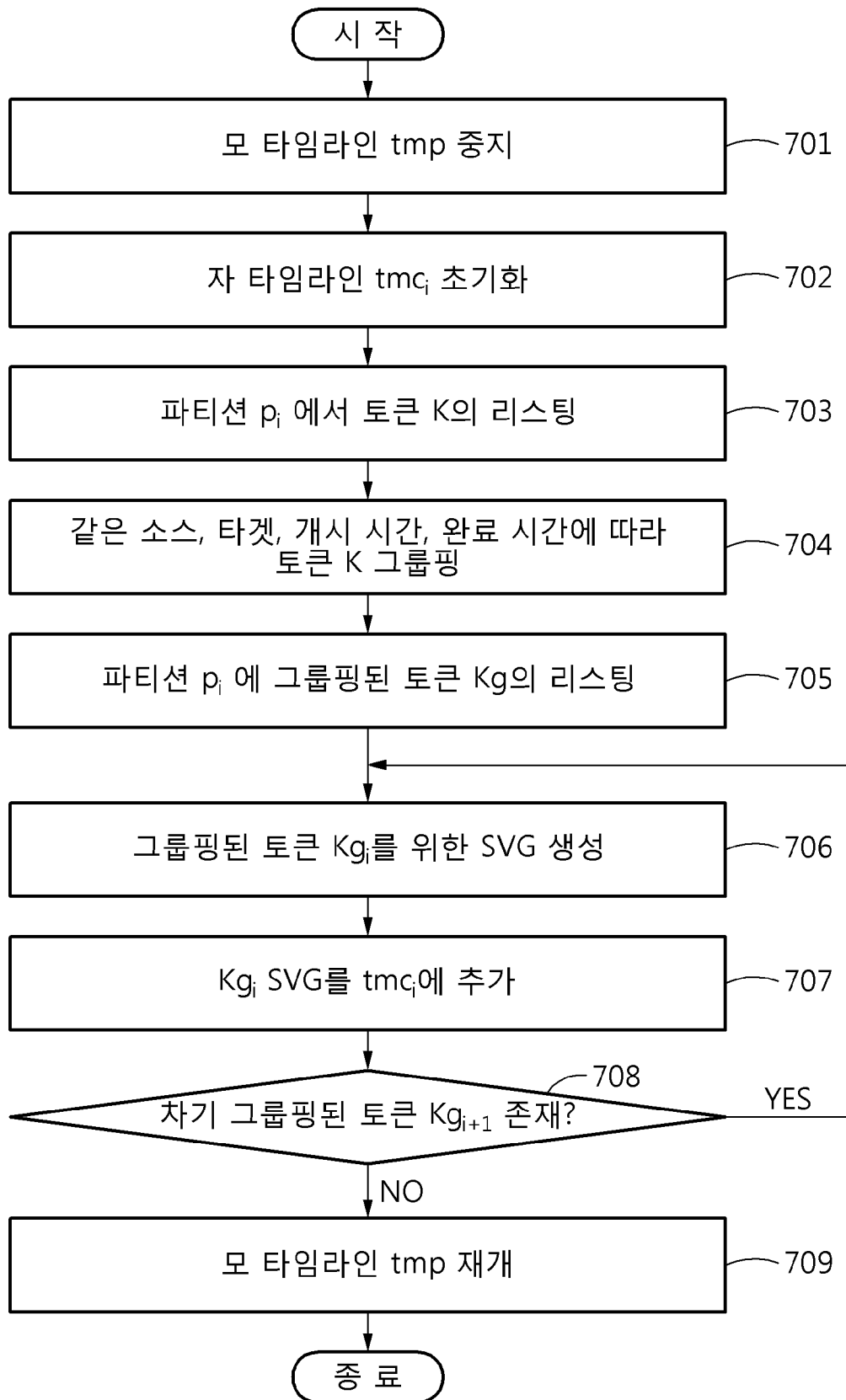
[도5b]



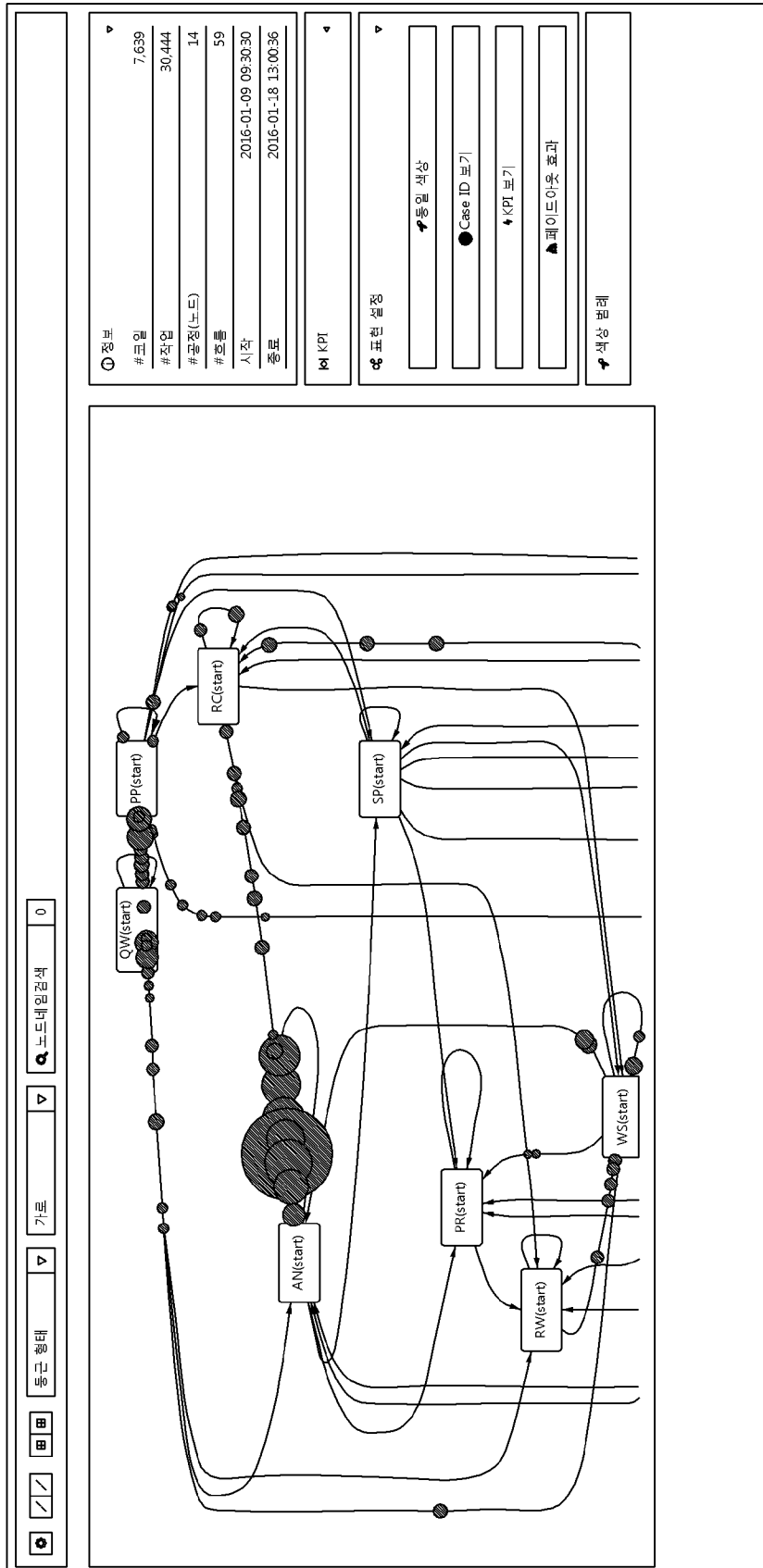
[도6]



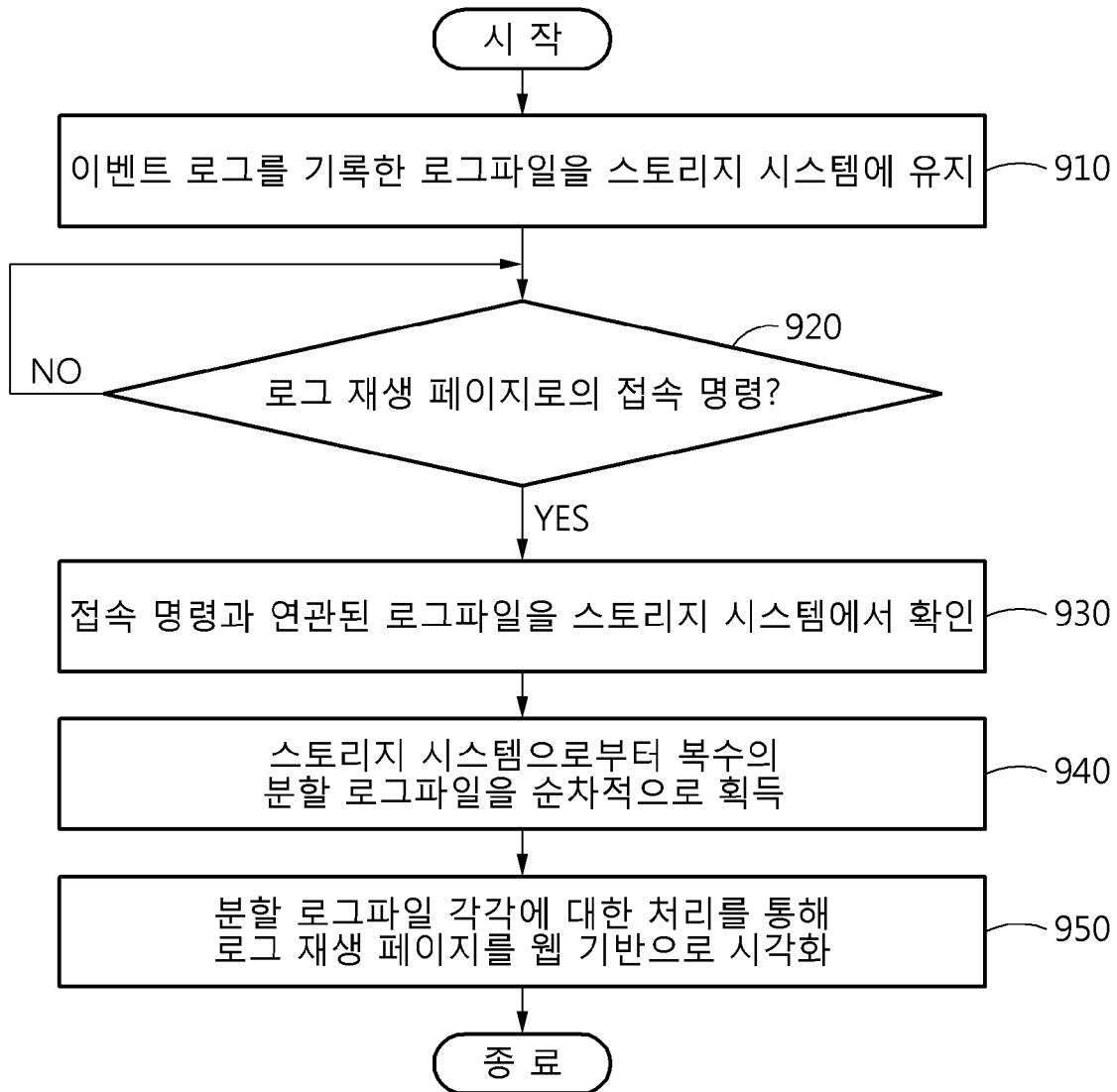
[도7]



[도8]



[도9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/013633

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/30; G06F 15/16; G06F 15/00; G06F 17/00; B29C 45/76; G06F 17/40; G06F 21/00; G06F 9/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: process, event, log, storage, partition

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2003-033958 A (SUMITOMO HEAVY IND. LTD.) 04 February 2003 See paragraphs [0016]-[0022]; claim 1; and figures 1-2.	1-15
Y	KR 10-2015-0063233 A (KONKUK UNIVERSITY INDUSTRIAL COOPERATION CORP.) 09 June 2015 See paragraphs [0041]-[0049]; claim 1; and figure 1.	1-15
A	KR 10-2007-0015871 A (SONY CORPORATION) 06 February 2007 See paragraphs [0144]-[0148]; claim 1; and figures 12-14.	1-15
A	KR 10-2010-0027836 A (THE INDUSTRY & ACADEMIC COOPERATION IN CHUNGNAM NATIONAL UNIVERSITY et al.) 11 March 2010 See paragraphs [0050]-[0083]; claim 1; and figure 6.	1-15
A	KR 10-1556541 B1 (KOREA INSTITUTE OF SCIENCE & TECHNOLOGY INFORMATION) 02 October 2015 See paragraphs [0028]-[0051]; claim 1; and figure 1.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

05 MARCH 2018 (05.03.2018)

Date of mailing of the international search report

05 MARCH 2018 (05.03.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Sconsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2017/013633

Patent document cited in search report	Publication date	Patent family member	Publication date
JP 2003-033958 A	04/02/2003	JP 3616997 B2	02/02/2005
KR 10-2015-0063233 A	09/06/2015	KR 10-1559206 B1	13/10/2015
		US 2015-0154288 A1	04/06/2015
KR 10-2007-0015871 A	06/02/2007	CN 1959673 A	09/05/2007
		CN 1959673 C	09/05/2007
		EP 1770558 A1	04/04/2007
		JP 2007-041722 A	15/02/2007
		US 2007-0025701 A1	01/02/2007
		US 8526795 B2	03/09/2013
KR 10-2010-0027836 A	11/03/2010	KR 10-1005866 B1	05/01/2011
KR 10-1556541 B1	02/10/2015	JP 2016-081495 A	16/05/2016
		US 2016-0110234 A1	21/04/2016

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 17/30(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 17/30; G06F 15/16; G06F 15/00; G06F 17/00; B29C 45/76; G06F 17/40; G06F 21/00; G06F 9/50

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 프로세스, 이벤트, 로그, 저장, 분할

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	JP 2003-033958 A (SUMITOMO HEAVY IND. LTD.) 2003.02.04 단락 [0016]-[0022]; 청구항 1; 및 도면 1-2 참조.	1-15
Y	KR 10-2015-0063233 A (건국대학교 산학협력단) 2015.06.09 단락 [0041]-[0049]; 청구항 1; 및 도면 1 참조.	1-15
A	KR 10-2007-0015871 A (소니 가부시끼 가이샤) 2007.02.06 단락 [0144]-[0148]; 청구항 1; 및 도면 12-14 참조.	1-15
A	KR 10-2010-0027836 A (충남대학교산학협력단 등) 2010.03.11 단락 [0050]-[0083]; 청구항 1; 및 도면 6 참조.	1-15
A	KR 10-1556541 B1 (한국과학기술정보연구원) 2015.10.02 단락 [0028]-[0051]; 청구항 1; 및 도면 1 참조.	1-15

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2018년 03월 05일 (05.03.2018)

국제조사보고서 발송일

2018년 03월 05일 (05.03.2018)

ISA/KR의 명칭 및 우편주소



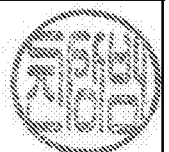
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

진상범

전화번호 +82-42-481-8398



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
JP 2003-033958 A	2003/02/04	JP 3616997 B2	2005/02/02
KR 10-2015-0063233 A	2015/06/09	KR 10-1559206 B1	2015/10/13
		US 2015-0154288 A1	2015/06/04
KR 10-2007-0015871 A	2007/02/06	CN 1959673 A	2007/05/09
		CN 1959673 C	2007/05/09
		EP 1770558 A1	2007/04/04
		JP 2007-041722 A	2007/02/15
		US 2007-0025701 A1	2007/02/01
		US 8526795 B2	2013/09/03
KR 10-2010-0027836 A	2010/03/11	KR 10-1005866 B1	2011/01/05
KR 10-1556541 B1	2015/10/02	JP 2016-081495 A	2016/05/16
		US 2016-0110234 A1	2016/04/21