

# 실시간 데이터의 효율적인 저장을 위한 알고리즘 설계

이원조<sup>o</sup> 이단영 권순덕 고재진  
울산대학교 대학원 컴퓨터정보통신공학부  
wjlee@mail.ulsan-c.ac.kr<sup>o</sup>

## Design of Algorithm for Efficient Storage with Real-Time Data

Won-Jo Lee<sup>o</sup> Dan-Young Lee Soon-Duk Kwon Jae-Jin Koh  
School of Computer Engineering and Information Technology University of Ulsan

### 요 약

정보통신 기술의 급격한 발전으로 환경변수 데이터의 수집과 분석자료의 활용에 대한 요구가 증대되고 있다. 따라서 본 연구에서는 DAQ(Data Acquisition System)를 통해 수집된 실시간 데이터의 효율적인 데이터베이스 구축을 위한 저장 알고리즘의 설계를 제안하고, 온도 데이터 수집모델을 통하여 효과를 검증한다.

### 1. 서 론

최근 정보통신 기술의 발전으로 공정관리 데이터와 환경변수 데이터의 실시간 데이터베이스 구축과 활용에 대한 요구가 급격하게 증가하고 있는 추세에 따라 데이터 수집을 위해 다양한 DAQ 시스템(Data Acquisition System)과 응용프로그램들이 사용되고 있다. 그러나 기존의 시스템들은 DAQ에 의한 데이터의 수집과 분석에 한정되어 있어, 이들의 데이터베이스 구축에 대한 효율적인 알고리즘의 개발이 절실히 요구되고 있다. 따라서 본 연구에서는 기존의 DAQ와 응용프로그램을 이용하여 데이터를 수집하여 실시간 데이터의 효율적인 데이터베이스 구축을 위한 알고리즘의 설계를 통한 저장과 복원 기법 개발과 이를 기반으로 다양한 분석정보의 응용기법을 연구한다. 그러나 이번 연구에서는 선행단계로 DAQ를 통하여 수집된 실시간 데이터의 효율적인 데이터베이스 구축을 위한 저장과 복원 알고리즘을 제안하고, 온도 데이터 수집모델을 통하여 효과를 검증한다.

### 2. 본 론

#### 가. 관련연구

##### (1) 데이터 수집장치(Data Acquisition System)

물리량의 변화 값을 측정하는데, 사용되는 것이 데이터 수집장치(DAQ)이다. 예를 들어 금형의 온도가 제품의 불량에 미치는 영향이 큰 경우에 이 온도의 변화를 측정하는데, 근접 온도계를 사용한다면 사용자는 일정한 간격을 두고 온도를 눈으로 읽어서 체크시트에 기록 관리해야 한다. 그러나 측정시간이 짧고 온도의 측정빈도가 적다면 이러한 방법이 더 유용할 수도 있다. 하지만 연속적으로 데이터를 측정하고, 데이터를 보관하여 분석 자료로 활용해야 한다면 이러한 작업은 거의 불가

능하다. 따라서 이렇게 반복적이고 연속적인 작업을 가능하게 하는 것이 DAQ 시스템인데, 응용분야는 다음 표 1과 같다.[15][16][19][20]

분 야	내 용
실험실	전압측정/전류측정/온도측정/압력측정/길이측정, 기록, 분석
온 실	전압전류/온도변화 측정, 기록/광량측정/습도 측정
공 장	농도측정/온도측정/유량측정/유속측정/회전수 측정/전압, 전류, 전력량 측정/압력측정/하중측정
기계제 어	RPM 측정/Torque측정/유압측정/변위량(위치)측정/공구압력측정/가공정도 측정

표1. DAQ 시스템의 응용사례

##### (2) 실시간 데이터베이스 시스템

실시간 시스템에서 사용하는 데이터의 양이 증가함에 따라 데이터의 효율적인 저장과 조장이 요구되어 기존 데이터베이스 시스템과 결합된 것을 실시간 데이터베이스 시스템이라고 한다. 기존의 데이터베이스 시스템은 데이터의 무결성 유지, 데이터 공유, 비순차적 질의어 사용, 편리한 사용자 인터페이스 등의 장점이 있으나, 데이터가 디스크에 존재하므로 트랜잭션 시간의 대부분을 디스크 접근에 할당하기 때문에 결과 도출시간이 보장되지 않는 낮은 성능과 예견성 부재로 실시간 응용에 부적합하였다. 그러나 실시간 데이터베이스 시스템은 데이터의 일관성을 유지하면서 마감 시간을 만족할 수 있도록 트랜잭션을 스케줄링 하는데 있다.[3][4][5][6][7][8]

##### (3) 능동 데이터베이스(Active Database)

능동 데이터베이스는 특정 이벤트에 의하여 자동적으로 수행되는 동작을 명시하는 규칙들은 오랫동안 데이터베이스 시스템에 대한 중요한 기능이었는데, 능동 데이터베이스 규칙들을 명시하기 위해 사용된 모델을 Event-Condition-Action 또는 ECA모델이라고 부른다.

그 규칙은 다음 표2와 같이 세가지 요소를 갖는다.[9][10]

규 칙(Rules)	내 용
규칙을 수행하는 이벤트(Event)	이벤트 데이터베이스에 명시적으로 적용된 데이터베이스 갱신연산
동작을 수행할지 결정하는 조건(Condition)	이벤트가 발생시 조건에 참이라면 수행, 그렇지 않으면 다른 처리
수행할 동작(Action)	동작은 대개 일련의 SQL문장이다. 그러나, 특정한 트랜잭션이나 프로시저 또는 외부프로그램

표2. ECA(Event-Condition-Action) 모델

나. 제안 DAQ 모델

(1) 제안 DAQ 모델의 개요

실시간 데이터의 데이터베이스 구축을 위해서 이력 데이터베이스의 적용이 가능하나, 공정관리와 환경변수 등의 데이터 특성상 대량의 데이터 저장이 요구됨으로 적합하지 않아 다음과 같은 시스템을 제안한다. DAQ 시스템은 공정관리와 환경변수 등의 데이터를 실시간으로 수집하는 데이터 수집장치와 통신장비, 제어장치 및 모니터링 위한 응용 프로그램으로 구성되어 있다. 그러나 데이터의 측정빈도를 시간간격으로 설정되어 있어, 이들 데이터의 데이터베이스 구축 시 대량의 데이터로 자원의 낭비와 성능저하의 요인이 되고 있다. 이러한 문제점을 해소하기 위해서 이들 데이터의 특징을 조사해 보면 대부분 데이터 값의 변화 폭이 작기 때문에 시간가격 설정 기법이 아닌 데이터 변화량 측정기법으로 데이터베이스를 구축하면 데이터 량을 줄일 수 있게된다.[1][2][10]

(2) 제안 DAQ 모델의 구성

제안된 DAQ 모델은 다양한 센스(S1, S2, ..., Sn)로부터 DAQ로 수집되어 유.무선 랜을 통하여 실시간으로 데이터베이스에 저장되고, 응용프로그램에 의해서 분석과 모니터링 된다.

(3) 제안 DAQ 모델의 업무 흐름도

제안된 DAQ 모델의 업무 흐름도 그림1은 단일노드 데이터 처리기법을 보여주고 있다.

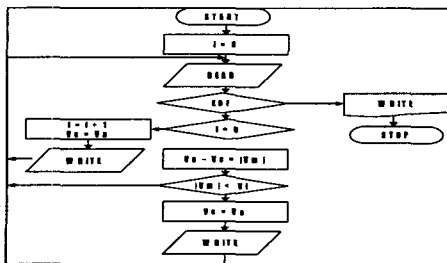


그림1. 제안 DAQ 모델의 업무 흐름도

- Temperature : 테이블 명
- Comparison Variable Value = Vc ; 비교 변수값
- Acquisition Value = Va ; 수집값
- Duration Time = Td ; 수집간격
- Limit Value = Vl ; 정밀도

- ① 최초 수집값 1 튜플을 [temperature] 테이블에 Insert 하고, 그 값을 비교 변수값(Vc)에 저장  
( $Vc \leftarrow Va$ )
- ② 새로 수집된 값과 비교 변수값(Vc)을 비교하여 정밀도의 범위내에 있으면 해당 튜플은 저장하지 않음  
( $Va - Vc = |Vm|, |Vm| < Vl$ )
- ③ 새로 수집된 값과 비교 변수값(Vc)을 비교하여 정밀도의 범위를 벗어나면 수집값은 비교 변수값(Vc)에 저장하고, 해당 튜플은 저장  
( $Vc \leftarrow Va, Va - Vc = |Vm|, |Vm| \geq Vl$ )
- ④ 수집 응용프로그램의 종료시 최종 튜플은 저장
- ⑤ 데이터베이스의 테이블에 저장된 전체 튜플은 정밀도의 범위 내에 해당되어 생략된 튜플 수 보다 작아진다. 값의 변동이 큰 경우에는 생략된 튜플의 수가 적으나, 값의 변동폭이 적은 경우에는 저장 튜플의 수를 현저히 감소시킬 수 있다.

다. 제안 DAQ 기법의 데이터 처리절차

제안 DAQ 기법의 데이터 처리절차는 단일 데이터 수집장치를 기준으로 설명한다. 다중노드 데이터 처리절차는 수집된 각각의 데이터는 각 테이블에 저장되므로 처리절차가 동일하다.[11][12][13][14]

(1) 데이터 처리절차

① 단일 노드에서 1초에 1건의 데이터를 수집한다(10건의 데이터 발생).

ID	Date	Time	Temp	Unit_Code	Limit
1001	02-08-11	12:00:00	130	U10	2
1002	02-08-11	12:00:01	131	U10	2
1003	02-08-11	12:00:02	131	U10	2
1004	02-08-11	12:00:03	132	U10	2
1005	02-08-11	12:00:04	133	U10	2
1006	02-08-11	12:00:05	134	U10	2
1007	02-08-11	12:00:06	134	U10	2
1008	02-08-11	12:00:07	134	U10	2
1009	02-08-11	12:00:08	132	U10	2
1010	02-08-11	12:00:09	132	U10	2

표3. 원본 데이터 현황

② 새로운 데이터 값(Temp)과 이전 데이터 값을 비교하여 정밀도(Limit)의 허용범위 값(Limit = 2)일 경우에는 저장하지 않는다(5건의 튜플을 데이터베이스에 저장).

ID	Date	Time	Temp	Unit_Code	Limit
1001	02-08-11	12:00:00	130	U10	2
1004	02-08-11	12:00:03	132	U10	2
1006	02-08-11	12:00:05	134	U10	2
1009	02-08-11	12:00:08	132	U10	2

표4. 저장 데이터 현황

③ 생략된 ID를 근거로 1초 단위의 생략된 튜플들을 복원한다. 복원된 값은 정밀도의 허용범위에 존재한다.(10건의 튜플을 복원하여 조회가 가능)

ID	Date	Time	Temp	Unit	Code	Limit
1001	02-08-11	12:00:00	130	U10		2
1002	02-08-11	12:00:01	130	U10		2
1003	02-08-11	12:00:02	130	U10		2
1004	02-08-11	12:00:03	132	U10		2
1005	02-08-11	12:00:04	132	U10		2
1006	02-08-11	12:00:05	134	U10		2
1007	02-08-11	12:00:06	134	U10		2
1008	02-08-11	12:00:07	134	U10		2
1009	02-08-11	12:00:08	132	U10		2
1010	02-08-11	12:00:09	132	U10		2

표5. 복원 데이터 현황

④ 그림2는 원본 데이터의 시간대별 온도 변화값을 그래프로 보여준다.

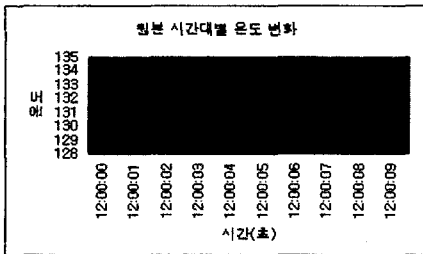


그림2. 원본 데이터의 온도변화 그래프

⑤ 그림3은 복원된 데이터의 시간대별 온도 변화값을 그래프로 보여주고 있다. 온도값의 변화는 정밀도 허용 오차범위 내(Limit = 2)에 존재한다.

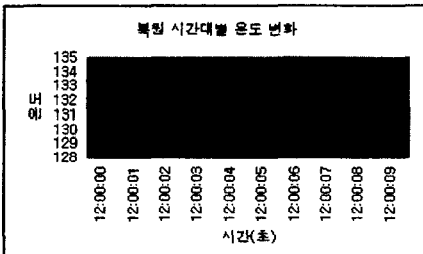


그림3. 복원 데이터의 온도변화 그래프

(2) 제안 DAQ 모델의 데이터 처리결과

제안된 단일 DAQ 모델에서는 1초당 1건 즉 10건의 데이터를 수집하여 이들 중 5건의 데이터가 데이터베이스에 저장되었다. 1일 86,400건의 데이터가 수집된다면 약50%의 저장효율이 발생하게 된다. 만약 다중노드 DAQ 시스템에서 수집값의 변화량이 적고 정밀도가 낮다면 효율이 매우 높아진다. 따라서 이 기법의 적용은 데이터의 변화량이 적은 환경변수의 데이터베이스 구축에 매우 효과적이라 평가된다.

### 3. 결 론

본 논문은 서론에서 언급된 바와 같이 DAQ로부터 수집된 데이터의 효율적인 실시간 데이터베이스 구축 연구를 위한 선행단계로 환경변수 데이터의 저장 알고리즘 설계 및 복원기법을 제안하였다. 따라서 기존의 시간간격 설정 기법의 단점을 보완하는 제안된 알고리즘은 환경변수와 같이 데이터의 변화량이 적고, 정밀도가 낮은 데이터의 관리에 매우 효율이 높다. 이 알고리즘의 적용 시 정보시스템 자원과 네트워크의 부하 감소를 통한 성능향상이 기대된다. 향후과제는 제안된 알고리즘을 수학적 이론으로 정립하고, 최신 정보통신 기술을 적용한 실무사례 모델을 구현하여 보다 경제적이고, 효율적인 공정관리 데이터와 환경변수 데이터의 실시간 데이터베이스 구축을 위한 지속적인 연구가 필요하다.

### 4. 참고문헌

- [ 1 ] A.J.H. Peddemors, A high performance distributed database system for enhanced Internet services, 1998
- [ 2 ] S. CASTTANO, Conceptual Schema Analysis : Technique and Applications, ACM, 1999
- [ 3 ] Lisa Cinginer DiPippo, Real-Time Databases, 1995
- [ 4 ] Saud A. Aldarmi, Real-Time Database Systems : Concepts and Design, 1998
- [ 5 ] Hermann Kopetz, Real Time and Dependability Concepts, 1999
- [ 6 ] Ben Kao, An Overview of Real-Time Database Systems, 1998
- [ 7 ] P.D.V. van der Stock, Modeling and Construction of Real-Time Database Schedulers, 1994
- [ 8 ] Bhaskar Purimetla, Real-Time Databases : Issues and Applications, 1994
- [ 9 ] NORMAN W. PATON, Active Database Systems, 1999
- [10] 황규영, 데이터베이스 시스템, 생능출판사, 2000
- [11] 홍준호, Oracle Bible, (주)영진출판사, 1999
- [12] 이상운, 이동 DBMS의 데이터 동기화 기술분석, 2000
- [13] 홍동권, 실시간 데이터베이스 시스템, 1999
- [14] 최용락, 데이터모델링을 이용한 데이터베이스시스템 성능 향상, 데이터베이스연구회지 13권4호
- [15] <http://www.ableduck.com/control/dcs03.htm>
- [16] <http://yfer.snu.ac.kr/juhwan>
- [17] <http://www.campbellsci.com>
- [18] <http://www.youngusa.com>
- [19] <http://www.bandp.co.kr/nong.htm>
- [20] <http://sunent.yeungnam-c.ac.kr/~hncho>