

빅데이터 분석과정에서의 품질개선 시스템 제안

조찬영, 김재원, 이선영*
순천향대학교

{ccy0531, jwk0016, *sunlee}@sch.ac.kr

Quality Improvement System Method in Big Data Collecting Process

Chan Yeong Cho, Jae Won Kim, Sun-Young Lee*
Dept. of Information Security Soonchunhyang Univ.

요약

빅데이터를 활용하려는 수요가 증가함에 따라 데이터 품질 문제 또한 더욱 중요해지고 있다. 기존 DMQ 시스템의 한계점으로 인한 데이터 분석과정에서의 오류를 개선하기 위해 통합 DQM 시스템이 필요하다. 본 논문에서는 기존 DQM 시스템의 데이터 품질관리 과정에서 질적인 측면의 한계와 메타데이터 관리에서의 데이터 분석 과정 없이 저장만 한다는 문제를 개선하기 위해 비즈니스 프로세스 중의 하나인 ERP 와 DQM 시스템을 결합한 통합 DQM 시스템을 제안한다.

I. 서론

최근 서비스업, 공공기관 등 다양한 분야에서 빅데이터를 활용하는 사례가 증가하고 있다. 정부에서도 빅데이터 분석을 통한 청사 시설관리, 인공지능과 빅데이터를 기반으로 한 공무원 교육 등 다양한 분야에서 빅데이터가 활용되고 있다.

빅데이터를 활용할 때 고품질 데이터를 사용하는 것이 중요하다. 빅데이터를 활용할 때 저품질 데이터를 사용하면 서비스의 동작 오류가 발생할 수 있고, 의사결정 정확도가 하락하며 비즈니스 프로세스에 오류를 발생시킬 수 있다[1]. 이러한 문제를 해결하기 위해 정부에서는 빅데이터 플랫폼 구축을 통한 데이터 축적 및 유통을 활성화하려는 노력을 기울이고 있다[2].

본 논문에서는 기존의 DQM(Data Quality Management)시스템의 한계점 개선 방향을 제안하여 저품질 데이터로 인해 발생하는 데이터 분석과정에서의 오류를 해소하고자 한다.

II. 관련 연구

2.1 빅데이터

빅데이터(Big-Data)는 높은 통찰력, 의사결정, 프로세스 자동화를 위해 비용 효과가 높은 혁신적인 정보처리 과정이 필요하며 규모(Volume), 속도(Velocity), 다양성(Variety)을 특징으로 하는 데이터이다. 과거 빅데이터는 의미 있는 결과 도출이 가능한 수십~수천 TB 에 달하는 거대 데이터 집합을 의미했으나 관련도구, 플랫폼, 분석기법까지 포괄하는 용어로 변환되고 있다[3].

2.2 데이터 품질의 정의와 특성

데이터 품질은 데이터가 의도한 목적에 적합한 경우를 말한다. 저품질 데이터는 데이터의 결함으로 인한 잘못된 의사결정, 데이터 중복으로 인한 관리 비용 증가 등의 문제점을 유발한다.

데이터 품질 관리란 기관이나 조직 내외부의 정보시스템 및 DB 사용자의 기대를 만족시키기 위해 지속적으로 수행하는 데이터 관리 및 개선을 의미한다[4]. 데이터 품질 관리는 데이터 관리 생애주기를 거쳐서 기능을 수행한다. 데이터 관리 생애주기는 <그림 1>과 같이 크게 데이터 수집, 데이터 가공 및 분석, 데이터 활용의 3 단계로 나누어져 있다.

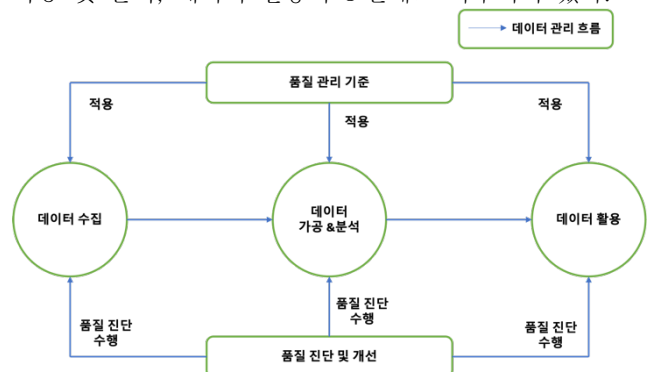


그림 1. 데이터 생애주기[1]

2.3 DQM 시스템

DQM(Data Quality Management)시스템은 데이터 품질 대상 선정 및 측정, 데이터 품질 지표 선정 및 데이터 품질 측정 규정 작성, 데이터 품질 절차 개선, 데이터 품질 개선 방법 및 프로세스 구현, 데이터 품질 모니터링의 5 단계로 이루어진다[1]. 또한, 데이터 품질 관리, 품질 측정 엔진, 외부 시스템 연계 등 3 가지 주요 기능과 DQM 데이터베이스로 구성되어 있다.

DQM 시스템은 데이터 품질평가, 데이터 정제, 메타데이터 관리의 세 가지 부분에서 한계점을 갖고 있다. 첫 번째, 비즈니스 관점 품질 정의 및 평가가 부족하다는 한계점이 있다. 두 번째, 핵심 정보의 선정과

같은 품질향상 추진이 부족하다는 한계점이 있다. 세 번째, 시스템 중심의 접근으로 단순히 데이터를 중앙 집중된 장소에 통합 관리하는 것에 그친다는 한계점이 있다.

2.4 ERP

ERP(Enterprise Resource Planning)란 전사적 자원 관리를 의미하며 기업에서 모든 정보를 관리하는 IT 시스템이다. ERP 시스템은 여러 비즈니스 프로세스를 한데 묶어 각 프로세스 간 데이터 흐름을 가능하게 해준다. ERP 시스템은 데이터 중복을 제거하고, 데이터 무결성을 제공한다. ERP 시스템의 구축은 기업의 경영환경 속에서 시스템을 효과적으로 사용할 수 있도록 하기 위해 필요한 모든 활동이라고 정의할 수 있다[5]. ERP는 기업 활동에 활용 가능한 모든 인적, 물적 자원을 효율적으로 계획, 관리하여 궁극적으로 기업의 경쟁력을 높이고 전사적 정보시스템을 구축하여 경영 자원의 최적화를 이루는 경영 혁신도구 중의 하나라고 할 수 있다[6].

III. 통합 DQM 시스템을 이용한 개선 방안

통합 DQM 시스템은 품질 규칙 설정, 품질 현황 모니터링, 품질 개선 및 조치결과를 통합 관리할 수 있는 통합 솔루션 기능을 제공한다. 또한, 기존 DQM 시스템의 비즈니스 관점 품질 정의 및 평가가 부족하다는 한계점을 개선하기 위해 비즈니스 프로세스와도 연동되어야 한다[1]. 본 논문에서는 DQM 시스템과 ERP를 함께 사용하는 통합 DQM 시스템을 제안한다.

통합 DQM 시스템은 <그림 2>와 같이 ERP와 개선 대상 업무 수행을 통한 품질 개선 활동을 수행한다. 품질 개선 활동 수행 시 관리 시스템과 상호간 품질관리 활동이 가능하다. 또한, 품질 개선 활동을 통해 두 가지 상시적 데이터 품질관리가 가능하다. 첫 번째, 개선 대상 업무를 통한 상시적 품질 개선활동은 오류 데이터를 조회하여 오류 발생 원인 파악, 데이터 오류조치 및 근본 해결을 위한 개선과제 수행, 품질개선 활동 현황 관리의 3 단계로 나뉘어진다. 두 번째, ERP를 활용한 상시적 품질 모니터링은 품질 모니터링 과정에서 점검 대상 데이터 추출 및 적재와 측정 결과 정량화를 가능하게 만든다. ERP를 결합시킨 DQM 시스템은 전사적 관점에서의 데이터 품질 관리 모니터링을 통해 고품질 데이터 유지를 가능하게 한다. 통합 DQM 시스템은 다양한 관점의 품질현황 모니터링 기능을 지원해야 한다. 여러 비즈니스 프로세스를 한데 묶어 프로세스 간 데이터 흐름을 가능하게 하는 ERP를 DQM 시스템과 결합함으로써 다양한 관점의 품질현황 모니터링을 가능하게 한다는 이점을 준다.

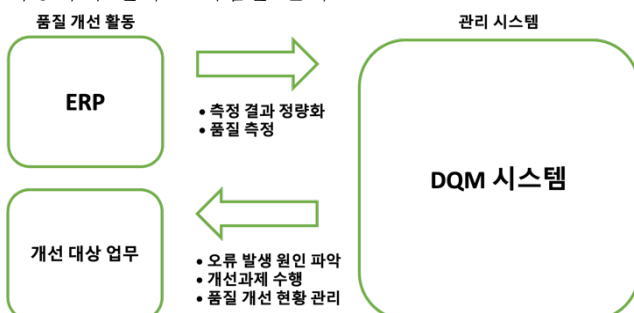


그림 2. 통합 DQM 시스템

IV. 결론

본 논문에서는 기존 DQM 시스템의 한계인 비즈니스 관점에서의 품질 정의 및 평가가 부족하다는 점을 개선하기 위해 ERP와 DQM 시스템을 결합한 통합 DQM 시스템을 제안한다. ERP는 다양한 관점의 품질현황 모니터링이 가능하고 여러 비즈니스 프로세스를 한데 묶어 각 프로세스 간 데이터 이동을 허용하기 때문에 통합 DQM 시스템에서 필수적인 다양한 품질현황 모니터링이 가능하다.

ACKNOWLEDGMENT

본 연구는 2018년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF2018R1D1A1B07047656).

참 고 문 헌

- [1]이주형, 이정원, 손영성.(2022). *2022 데이터산업 백서*. 한국데이터산업진흥원. pp. 226
- [2]문용식. (2021). *빅데이터 플랫폼 및 센터 데이터 품질관리 가이드*. 한국지능정보사회진흥원. pp. 07
- [3]김정숙. (2012). *빅 데이터 활용과 관련기술 고찰*. 한국콘텐츠학회지, 10(1), 34-40.
- [4]원영희. (2006). *데이터 품질관리 지침(ver 2.1)*. 한국데이터베이스진흥센터.
- [5]홍현기, 안중호. (1998). *프로세스 중심의 ERP 구축 방법론에 관한 연구 (A study on the Process-Oriented Methodology for ERP Implementation)*. 경영과학, 15(2), 33-44.
- [6]임창우, 이석희. (2007). *전사적 자원관리(ERP)의 도입효과에 관한 재무분석*. 정보시스템연구, 16(1), 91-109.