하둡 기반의 통합설비 모니터링시스템 설계 및 구현 사례 연구

김상락¹ · 장길상² · 조지운^{3*}

¹비케이앤씨 대표 / ²울산대학교 경영정보학과 / ³울산대학교 산업경영공학부

Case Study of Design and Implementation for Hadoop-Based Integrated Facility Monitoring System

Sangrak Kim¹ · Gilsang Jang² · Chiwoon Cho³

¹BK&C

²Department of Management Information System, University of Ulsan

³School of Industrial Engineering, University of Ulsan

SCADA and DCS that have performed automatic control and monitoring activities increase the productivity of enterprise in industries. In such systems, although their performance had been improved, there are still many deficiencies in predictive maintenance which can foresee the risk of any kinds of accidents. Because the data acquisition systems of main facilities are being distributed throughout the whole plant and therefore, integration of data obtained from the systems is very difficult. Accordingly, techniques that acquire meaningful information from the gathered data through realtime analysis still need to be improved. This paper introduces a developed facility monitoring system which can predict equipment failure and diagnose facility status through big data analysis to improve equipment efficiency and prevent safety accidents.

Keywords: Big Data, IoT, Hadoop, HDFS, Facility, Monitoring

1. 서 론

1.1 연구 배경

우리는 이미 제타바이트(약 1조 8000억 기가바이트)의 데이터 홍수 시대에 살고 있으며, 폭발적인 데이터 증가 추세는 스마트 단말기, 사물인터넷, 소셜 네트워크 등의 확대에 더욱 가속화되고 있다. IDC는 2011년도 전 세계의 디지털 정보량은 약 1.8 제타바이트이며 정형, 비정형데이터가 기하급수적으로 늘어나면서 2020년에는 디지털 정보의 양이 35.2제타바이트에 달할 것으로 전망했다(Yun, 2012).

스마트 시대의 소셜 네트워크, 사물인터넷, 라이프로그 데이터 등은 빅데이터 시대 진입에 중요한 요소들이다. 스마트 단말기는 수많은 데이터를 생산하고, 그 기기들로부터 생산되는

수많은 데이터들은 분산 파일 형태로 수집되어 중요한 정보로 가공된다. 즉, 기기나 설비에 설치된 센서로부터 수집한 데이터에서 현실의 문제점을 발견하고 해결책을 찾는 것이 가능하게 되었다. 빅데이터는 지금까지 이해할 수 없었던 정보를 이해하고, 분석할 수 없었던 대용량의 비정형 데이터를 처리하는 기술로서 데이터를 이용한 지능형 서비스 구현의 기반 기술이다. 이 기술은 가까운 미래에 사용자에게 제공하는 서비스의 질을 높이고 전문가의 역할과 가치를 바꿔놓을 것이다. 아울러 불가능할 것으로 생각되었던 음성 인식이나 자동 번역이 빅데이터를 기반으로 현실화되면서 IT 업계는 다시, 생각할 수 있는 기계를 꿈꾸고 있다(Son, 2012).

오늘날 수많은 국내 제조공장들은 설비노후화로 인해 각종 사고 발생 가능성이 잠재되어 있으며 근로자들의 안전 또한 위협받고 있다. 이러한 문제 해결을 위해 많은 기업들은 모니

본 논문은 2012년 울산대학교 연구비에 의하여 연구되었음.

* 연락저자 : 조지운 교수, 680-749 울산광역시 남구 대학로 93 울산대학교 산업경영공학부, Tel : 052-259-2287, Fax : 052-259-2180,

E-mail : chiwoon6@ulsan.ac.kr

2013년 11월 15일 접수; 2013년 12월 9일 수정본 접수; 2013년 12월 31일 게재 확정.

터링 시스템을 구축하고 있으나 대부분 문제가 발생했을 때 알람을 발생시키는 수준으로 이 정보로는 오직 사후 대응만 가능하다. 보다 높은 수준의 스마트 시그널 시스템을 도입한 업체들도 있으나 이것 역시 예지보전을 할 수 있을 만큼 신뢰성은 높지 않은 편이다. 최근 들어 많은 기업에서 배관부식, 설비진동 상태 파악을 위해 각종 운전 데이터, 검사 데이터를 수집하고 있으나 수집된 데이터 분석을 통한 정보로의 활용은 거의 하지 못하고 있는 실정이다. 또한 현재 구축되어 있는 설비관련 시스템들은 개별 데이터베이스를 사용하므로 장치별 모니터링 수준에 그치고 통합 관점에서의 모니터링은 이루어지지 않고 있다.

현실적으로 제조공장에 설치된 각종 설비들은 실시간으로 수많은 데이터를 생산하고 있다. 이러한 데이터들은 규모면에서 수 테라바이트 정도이며, 다양성 측면에서는 정형 데이터와 비정형 데이터로 구분할 수 있다. 또한 데이터 발생 속도 측면에서 초당 만건 이상의 데이터를 발생시키는 설비들도 있다. 이와 같은 데이터를 빅데이터라고 부르며, 이를 현재 제조공장에서 가장 많이 사용하고 있는 기존의 관계형 데이터베이스 기술을 활용하여 처리하면 인덱싱 작업에 과부하가 발생하여 고속 및 대용량 데이터 처리, 그리고 저장이 어렵다. 하지만 하둡 등 빅데이터 생태계 기술을 사용하면 대용량의 데이터 처리 및 저장이 가능하다. 자동차 생산업체 볼보는 빅데이터 기술을 활용해 커다란 성과를 거두었다. 고객의 자동차에 내장된 센서, 고객관계 관리 시스템, 딜러, 공장 등으로부터 수 테라바이트나 되는 데이터를 수집해서 자동차의 결함과 같은 문제의 사전 정보를 파악하였다(Hahm, 2012). GE관계센터에서는 비행 중인 비행기의 각종 설비들로부터 발생되어진 데이터를 위성을 통해 실시간으로 수집하여 비행기 고장을 사전에 예측하는 시스템을 운영하고 있다.

본 연구에서는 현장의 이러한 다양한 문제점들을 해결하기 위해 빅데이터 기술을 활용하여 각종 설비에서 발생하는 데이터를 하둡 플랫폼에 저장하고, 그 데이터를 배치 분석하여 설비상태를 진단하고, 이상 징후가 발견되면 즉시 담당자들에 통지하는 통합설비모니터링 시스템을 개발하였다. 본 통합설비모니터링 시스템 개발을 위해 울산 소재 5개 기업들의 실 상황을 조사하고 개선 방향을 모색하였고 개발 시 이를 반영하였다.

본 논문의 구성은 다음과 같다. 제 1장에서는 기존의 빅데이터 관련 기술을 정리하였다. 제 2장에서는 제조현장의 요구사항과 적용대상 업무의 정의, 이어 제 3장에서는 빅데이터 관련 기술을 활용한 통합설비모니터링 시스템에 대해 설명하였다. 제 4장에서는 연구결과에 따른 시사점 및 향후 연구방향 등을 제시하였다.

1.2 관련 기술

(1) 하둡(Hadoop)

하둡은 현재 빅데이터 처리에 가장 선호되고 있는 분산처리 기술을 핵심으로 하는 솔루션이다. 이것은 야후와 페이스북

등에서 핵심 기술로 사용되고 있으며, 많은 기업들이 자사 솔루션에 응용하여 사용하고 있다. 주요 구성은 하둡 분산 파일 시스템(HDFS, Hadoop Distributed File System), 에이치베이스(HBase), 맵리듀스(MapReduce) 등이다. 하둡 분산 파일 시스템과 에이치베이스는 각각 구글의 구글 파일 시스템과 빅테이블의 영향을 받았다(Vertica Systems, 2009). 기본적으로 비용 측면에서 효율적인 x86 서버로 가상화된 대형 스토리지를 구성하고, 하둡 분산 파일 시스템에 저장된 거대한 데이터셋을 간편하게 분산처리 할 수 있는 자바 기반의 맵리듀스 프레임워크를 제공한다. 하둡은 텍스트 검색 라이브러리로 폭넓게 사용되고 있는 아파치 루씬의 창시자인 더그 커팅에 의해 개발되었다. 하둡은 오픈소스 웹 검색엔진인 아파치 너치(Nutch)를 탑재했으며, 너치는 2002년에 처음 만들어진 실행 가능한 크롤러와 검색시스템이다(Kang, 2012; Kim, 2012).

(2) 맵리듀스(MapReduce)

맵리듀스는 데이터 처리를 위한 프로그래밍 모델이다. 자바, 루비(Ruby), 파이썬(Python), C++ 등 다양한 언어들로 쓰인 맵리듀스 프로그램들을 하둡으로 구동시킬 수 있다. 맵리듀스는 맵 단계와 리듀스 단계로 처리 과정을 나누어 작업한다. 각 단계는 입력과 출력으로써 키/값 쌍을 가지고 있고, 그 타입은 프로그래머가 선택 가능하다.

(3) 하이브(Hive)

하이브는 하둡 맵리듀스 프레임워크 상에서 구축되어진 오픈소스 데이터 웨어하우징 솔루션이다. 하이브는 데이터를 테이블로 표현하여 구조체와 HDFS에 저장된 데이터가 연결되는 수단을 제공한다. 테이블 구조 같은 메타데이터는 메타스토어라는 데이터베이스에 저장된다(Shim, 2011). 사용자의 데이터 접근을 위해 하이브는 SQL과 비슷한 쿼리 언어를 지원한다. 그것을 HiveQL이라고 한다. HiveQL에서 쿼리가 발행될 때, 그것은 파서에 의해 최적화된 쿼리 실행계획으로 변환되고 일련의 맵과 리듀스로 바뀐다. 이러한 작업은 출력 결과를 파일로 쓰기 위해 하둡 분산 파일 시스템에서 실행된다.

기본적으로 사용자는 사용자 화면으로 접속해서 HiveQL 명령어를 실행하여 그것을 드라이버로 보낸다. 그 드라이버는 세션을 만들고 컴파일러에 쿼리 문장을 전송하여 실행계획을 생성한다. 그때 쿼리 문장은 하이브 쿼리 최적화 구성요소에 의해 최적화되고 다수 맵과 리듀스 단계로 구성된 실행 쿼리 계획으로 변환된다. 그 계획은 한 개의 잡 트래커와 맵리듀스 단계 마다 몇 개의 태스크 트래커로 구성된 맵리듀스 실행 엔진에 의해 실행되어진다. 또한 하이브는 외부시스템의 인터페이스를 통해 데이터에 접근할 수도 있다. 대표적인 외부시스템은 PostgreSQL과 MySQL이다(Anja Gruenheid, 2011).

(4) 스콥(Sqoop)

관계형 데이터베이스에서 데이터를 추출해서 하둡으로 보

내 처리할 수 있도록 해주는 오픈 소스 도구이다. 스콕은 맵리듀스 작업을 실행하여 데이터베이스로부터 테이블을 임포트하며, 레코드를 하둡 분산 파일 시스템에 기록한다. 이것을 이용해서 데이터를 하둡 분산 파일 시스템에 저장할 수 있고, 관계형 데이터베이스의 데이터를 배치 프로그램을 이용해서 하둡 분산 파일시스템에 저장할 수도 있다(Shim, 2011).

(5) 포스코 IMC(Integrated Monitoring and Control) 센터

포항에 소재하고 있는 포스코 IMC 센터의 역할은 설비가동의 안정성 확보 및 고장예방이 주목적이다. 포스코 IMC 센터는 2010년에 세워진 통합 설비 모니터링 관제센터이다. 각종 설비로부터 취득한 데이터를 통합데이터베이스에 입력하고, 입력 데이터는 1분 주기로 SAS 분석 프로그램의 실행을 거쳐 분석 결과용 데이터베이스에 저장된다. 저장된 데이터는 모니터링시스템을 통해 현장 작업자, 관리자 등 관련 담당자들에게 다양한 전달 매체를 통해 전달된다. <Figure 1>은 IMC 센터의 운영 내용을 개념적으로 설명한다. 문제가 되는 설비 상태의 판단 기준은 전문가에 의해 설정된 설비상태 파라미터별 임계치 값이다. 특정 값이 임계치에 도달하지 못하거나 초과할 경우에 알람이 발생된다. 과거 데이터 추이 분석을 통한 결과 모델 데이터를 기준으로 설비상태를 파악할 수 있는 예지보전 시스템은 아직 개발되어 있지 않다. 이것은 향후 IMC 센터 발전 방안에는 포함되어 있다(POSCO, 2013).

(6) GE Proficy

GE Intelligent Platforms사에서 만든 Proficy 모니터링 & 분석 소프트웨어는 산업 현장의 빅데이터 관리와 분석을 위한 최초

의 통합 솔루션이다. Proficy는 모니터링 & 진단 센터에 설치되어 글로벌 제조공장들이 원격으로 모니터링 및 분석 작업을 가능하게 한다. 그것은 <Figure 2>와 같이 6개의 통합 제품들로 구성되어 있다. Proficy Historian은 데이터 수집 소프트웨어이다. Proficy Historian Analysis는 데이터 마이닝과 시각화 기능을 한다. Proficy SmartSignal은 상태기반 모니터링을 위한 예측 분석 소프트웨어이다. Proficy CSense는 프로세스 문제해결, 프로세스 상태 모니터링, 그리고 반복해서 프로세스 최적화 기능을 수행한다. Proficy Historian HD는 하둡 클러스터에 대규모 대용량의 데이터를 저장할 수 있게 하는 소프트웨어이다. Proficy Knowledge Center는 프로세스 가시화, 자산 상태 평가, 그리고 프로세스 최적화 등의 기능을 내장하고 있다(GE Intelligent Platforms, 2013).

Figure 2. GE Proficy

Figure 1. The concept of IMC center

2. 적용대상 업무 분석 및 정의

2.1 문제 정의

제조공장에 설치된 각종 설비들은 시시각각 실시간으로 수많은 데이터를 생산하고 있고, 이러한 데이터 속에는 설비의 상태를 파악할 수 있는 귀중한 정보가 숨겨져 있다. VMS, COMPASS, SYSTEM 1 등은 글로벌 선진기업들이 내놓은 설비상태를 감시하는 상용 소프트웨어들이다. 이들은 개별 데이터베이스를 내장하고 있어 데이터 통합을 위해서는 별도의 시스템 구축이 필요하다. 단일 설비에 대한 분석 및 진단에 대해서는 우수한 성능을 보이고 있으나 설비데이터 통합관점에서는 단점을 보이고 있다. 이러한 단점을 보완하기 위해 정형, 비정형 데이터를 신속하게 처리할 수 있는 하둡 기반 설비 데이터 통합 데이터베이스를 구축하고, 표준 데이터 수집체계를 수립한 후 데이터 분석, 시각화를 통해 설비상태를 보다 정확하게 진단 할 수 있는 시스템 설계가 필요하다.

실제 현장 상황을 보다 정확히 파악하기 위해 대기업 및 중견 제조업체 5개사를 대상으로 설비관리 업무의 고질적인 문제점과 개선 사항에 대해 팀장급 이상 관리자들과 인터뷰를 실시하였다. 그 결과 “제조현장에서 발생하고 있는 폭발사고 및 화학물질 누출로 인한 사고를 미연에 방지하기 위하여 각종 설비에서 발생하는 데이터를 실시간으로 수집 및 분석하여 사전 조치를 취할 수 있는 시스템이 필요하다”는 것이 공통적인 의견으로 나왔다.

2.2 현황 및 요구사항 조사

공장 내 설비관리 현황과 관리자들이 요구하는 기능에 대해서 조사한 결과는 <Table 1>과 같다. 현재 설비상태 관리를 위해 공장별로 DCS, PI System, 절연유 가스 분석시스템, SCADA

Table 1. The facility management issues for manufacturing

현황 및 요구사항
<ul style="list-style-type: none"> 설비로부터 현재 각종 운전 데이터 및 검사 데이터는 수집되고 있으나 그 데이터로부터 필요한 정보는 획득하지 못하고 있음. 일부는 스마트 시그널을 도입해서 초보적인 수준의 예지보전을 하고 있으나 신뢰성이 높지 않은 편임. 현재의 알람은 문제가 발생했을 때 정보를 제공하기 때문에 사후 대응만 가능함. 사전 대응이 가능한 고급 정보를 제공하여야 각종 사고를 미연에 방지할 수 있음. 통합 설비 모니터링시스템은 현장의 데이터를 단순 모니터링하는 수준에서 벗어나 즉각 실행에 옮길 수 있는 정보를 시스템으로부터 제공받을 수 있도록 설계되어야 함. 통합관점에서 데이터를 한 곳으로 모아 종합적인 설비상태 모니터링이 가능한 체계를 구축해야 함.

시스템, SYSTEM1, VM600 등 상용 시스템을 사용하여 데이터를 수집 및 분석하고 있다. 데이터 통합관점에서 이들은 개별적으로 데이터관리 저장소를 가지고 있어 데이터 통합을 위해서는 별도 프로그래밍 작업이 필요하다. 데이터 통합 작업 이후에 데이터 속에서 신뢰성 있는 정보를 얻기 위해서는 고도의 인공지능 기술(시계열 분석, 패턴인식, 기계학습, 클러스터링 등)을 사용한 분석 작업이 필요하다. 현재 많이 사용되고 있는 알람시스템은 문제가 발생했을 때 정보를 제공하므로 사후 대응만 가능하다. 사전 대응이 가능한 고급 정보를 제공하여야 각종 사고를 미연에 방지할 수 있다. 현장에서 필요한 통합 설비 모니터링시스템은 데이터를 단순히 모니터링 하는 수준에서 벗어나 사고발생 이전에 담당자들이 즉각 실행에 옮길 수 있는 정보를 시스템적으로 제공받을 수 있어야 한다.

3. 통합 설비 모니터링시스템 설계 및 구현

3.1 개념 설계

통합 설비 모니터링시스템의 기본 개념은 <Figure 3>과 같다. 현장의 각종 설비 장치들로부터 데이터를 수집하여 빅데이터 미들웨어에서 그것을 하둡 분산 파일 시스템으로 분산 처리 및 저장한다. 저장한 데이터는 데이터베이스, 웹 응용시스템, 모바일 디바이스 등으로 전송되어 업무와 관련한 담당자들이 설비관리 업무에 즉각 이용할 수 있도록 한다.

Figure 3. The concept of integrated facilities monitoring system

3.2 상세 설계

(1) 시스템 아키텍처

통합 설비 모니터링시스템의 아키텍처는 <Figure 4>와 같다. 아키텍처는 크게 데이터 수집, 데이터 저장 및 처리, 데이터 분석 및 시각화 영역으로 구성된다. 데이터 수집 영역은 각종 설비와 계측 장비들로 이루어져 있고, 데이터 저장 및 처리 영역은 빅데이터를 저장하고 처리하는 미들웨어 서버로 구성된다. 데이터 분석 및 시각화 영역에서는 설비관리와 관련한 업무용 시스템을 제공할 수 있는 웹 애플리케이션 서버로 구성된다.

Figure 4. System architecture

데이터 수집 서버는 각종 장치들에서 발생하는 데이터를 실시간으로 전송받아 처리하는 기능을 수행하고, 빅데이터 미들웨어는 하둡, 하이브, 스쿱, MySQL 등이 설치되어 데이터를 분산 저장 및 처리하며, 웹 애플리케이션 서버는 사용자가 필요한 정보를 보여주는 웹 응용시스템이 설치된다. 사용자는 웹 브라우저를 통해 웹 애플리케이션 서버로 데이터를 요청하면 서버는 요청에 대한 결과 데이터를 보내준다.

전체적인 흐름은 다음과 같다. 공장 내에 설치되어 있는 각종 장치, 배관, 회전기계, 전기, 계장 등으로부터 실시간으로 수집한 데이터를 오라클 데이터베이스로 저장한다. 이 데이터는 스쿱 데몬이 주기적으로 하둡 파일 시스템으로 저장한다. 하둡 파일 시스템에 저장된 데이터는 하이브 시스템 내의 HiveQL을 실행하여 데이터를 조회할 수 있다. 저장된 데이터는 분류 작업을 거쳐, 실시간 또는 배치 작업으로 관련 담당자들이 업무에 활용할 수 있도록 정보로 제공된다. 정보는 사용자 그룹의 업무 역할에 맞게 선택적으로 제공된다. 아울러 정보는 데이터 보다는 다양한 유형의 차트, 게이지, 기타 시각화 도구로 표현되며, 현장 작업자들의 작업 용이성을 위해 이동용 단말기를 통한 업무지원도 가능하다.

(2) 소프트웨어 아키텍처

대용량 데이터가 처리되는 다양한 단계에서의 문제점을 해

결하기 위해 분산 처리 소프트웨어 플랫폼인 하둡을 통합설비 모니터링시스템의 핵심 기술로 사용하도록 설계하였다. 하둡은 자바로 구현된 맵리듀스 프로그래밍을 이용하여 대용량 데이터를 클러스터 환경에서 병렬로 처리한다.

소프트웨어 아키텍처는 <Figure 5>에서 보는 바와 같이 다양한 오픈 소스들을 결합하여 설비 데이터를 수집, 저장, 분석, 그리고 시각화하도록 구성하였으며 다소 기능이 부족하거나 적용이 어려운 부분은 자체 개발하여 연계하였다. 또한 관계형 데이터베이스와 하둡 파일 시스템을 함께 사용하는 하이브리드 타입으로 설계하였다(Armbrust *et al.*, 2010; Dean *et al.*, 2004; Shvachko *et al.*, 2010).

3.3 시스템 구현 절차

실제 시스템 구현은 <Table 2>와 같이 진행되었다. 크게 3단계로 나누어 진행되었는데 1단계 사전준비 단계에서는 시스템 환경 정의, 개발·운영환경 구성, 파일럿시스템 설치 등과 같은 작업을 수행하였다. 2단계 요구사항 분석 단계에서는 요구사항 수집 준비, 요구사항 수집, 요구사항 정리 등의 작업을 실행하였다. 3단계 시스템 구축 단계에서는 시스템 구축 범위 정의, 시스템 구축 명세서 작성, 소프트웨어 설치 명세서 작성, 시스템 구현 등의 작업을 수행하였다. 이외에도 품질보증에

Figure 5. Software architecture

Table 2. System implementation steps

단계	활동	활동 설명
1. 사전준비	시스템 환경 정의	◦ 응용 시스템 구축, 데이터베이스 구축, 시스템 연계 등 시스템 설치 및 운영환경 구성을 위한 H/W, S/W 인프라 환경, 아키텍처 정의
	개발·운영환경 구성	◦ 시스템 구축 및 전환 후 운영에 필요한 기본환경 구성
	파일럿시스템 설치	◦ 파일럿 시스템 표준 및 절차에 따른 운영환경 내 시스템 설치
2. 요구사항 분석	요구사항 수집 준비	◦ 면담계획 수립 및 보유 업무영역 파일럿 시스템 점검
	요구사항 수집	◦ 사용자 파일럿 시스템의 구성 요소 및 기능 확인 후(보유 업무영역) 면담을 통한 요구사항 수집
	요구사항 정리	◦ 수집된 요구사항을 기초로 요구사항 기술서 작성 ◦ 구현기능과 요구사항간의 GAP 분석 및 협의·결정 ◦ 구축활용 기초 자료 준비
3. 시스템 구축	시스템 구축 범위 정의	◦ 구축에 대한 구성항목 정의
	시스템 구축 명세서 작성	◦ 구축 시스템별 하드웨어 구분 및 기능 정의
	소프트웨어 설치 명세서 작성	◦ 소프트웨어 논리적, 물리적 구조 정의
	시스템 구현	◦ 범위에 대한 시스템 기능 구현

대한 작업을 추가적으로 수행하였다.

3.4 하둡 미들웨어 설치

빅데이터 플랫폼에서 가장 중요한 단계는 미들웨어 설치이며 본 시스템 개발에서는 다음과 같은 내용으로 하둡을 설치하였다. 맵리듀스 기술을 기반으로 하는 하둡 소프트웨어 프레임워크를 설치하였고, 하둡 데이터 웨어하우징 솔루션인 하이브를 설치하였다. <Table 3>은 설치된 빅데이터 관련 소프트웨어 목록이다.

Table 3. The list of middleware software

장비명	IP 주소	설치 소프트웨어	버전
네임노드, 데이터노드 1	192.168.11.2	Cent OS	6.5
		JDK	1.6
		Apache Hadoop	1.0.3
		FLUME Agent	1.3.0
		HBase	0.94.1
		Zookeeper	3.3.5
		Sqoop	1.3.0
		Pig	0.9.2
		MySQL	5.5
데이터노드 2 데이터노드 3 데이터노드 4	192.168.11.3	Cent OS	6.5
	192.168.11.4	JDK	1.6
	192.168.11.5	Apache Hadoop	1.0.3
		FLUME Agent	1.3.0
		HBase	0.94.1
		Zookeeper	3.3.5

빅데이터 미들웨어는 리눅스 OS를 기본 운영체제로 사용하였다. 하둡을 설치하고 네임노드, 데이터노드를 설정한 다음 필요한 하둡 관련 소프트웨어들을 설치하였다. 하이브의 메타데이터 관리를 위해 MySQL 데이터베이스도 함께 설치하였다.

3.5 설비통합모니터링 시스템 구현 내용

(1) 데이터 수집 및 저장

데이터 수집 및 저장을 위해 두 가지 방법을 사용하였다. 첫 번째 방법은 마이크로프로세서와 각종 키트를 활용해서 온도, 조도, 습도 값을 취득하여 데이터를 수집하는 시스템을 개발하였다. 기본 펌웨어 소스는 HBE-MCU-Multi의 MCU가 연결된 포토센서 CDS와 온/습도 센서로부터 데이터 값을 주기적으로 수신하여 MCU와 연결된 TEXT LCD에 디스플레이하는 동시에, UART 통신으로 데이터 수집 서버에 데이터를 송신하는 동작을 수행하도록 구성하였다. <Figure 6>은 전체 시나리오에 따른 시스템 구성을 보여준다.

Figure 6. System configuration for collecting data from sensors

MCU I/O 기능은 ADC(Analog-to-Digital Converter) 기능, Timer and Counter 기능, UART(Universal Asynchronous Receiver/Transmitter) 기능 등이 있다. ADC 기능은 아날로그 신호를 디지털 신호로 바꾸는 기능이다. Timer and Counter 기능은 센서로부터 데이터 값을 주기적으로 읽어 들이기 위해 사용한다. UART 기능은 MCU가 센서로부터 취득한 데이터를 서버로 송신할 때 사용한다. TEXT LCD를 통해 온도, 조도, 습도 수집 값을 확인할 수 있다. 직렬 통신 프로그램을 통해 수집된 데이터들은 UART 통신을 활용하여 서버로 전달된다. UART를 통해 송신되는 데이터들은 총 6바이트이고, 이 데이터들은 직렬 통신 방식으로 MCU에 연결된 장비로 보내진다.

<Figure 7>은 데이터 수집용 장치 제작에 사용된 모듈들이다. 크게 마이크로프로세서인 ATmega128, 온도, 습도, 조도 센서, LCD, UART 통신모듈 등으로 구성되었다.

Figure 7. The Picture for MCU, Sensors, LCD, and UART

Figure 8. Data acquisition system diagram

두 번째 데이터 수집 및 저장 방법으로 제조현장 설비로부터 직접 데이터를 수신 받아 데이터베이스화 하였다. <Figure 8>

은 현장 설비와 각종 장비들의 전체적인 레이아웃이며 각 장치 및 개략적인 기능은 다음과 같다. ①번 장치는 데이터 수집 대상 설비이고 ②번 장비는 설비에 부착된 센서로부터 데이터를 측정하는 계측장비이다. ③번은 연소분석장치이고 ④번은 전력 부하를 조절하는 로드뱅크이다. ⑤번은 설비의 상태를 제어하는 PLC이고 ⑥번은 전력의 주요 값을 측정하는 파워미터기이다. ⑦번은 연소분석기 데이터 수집 서버이고 ⑧번은 통신장비인 허브이다. ⑨번은 설비로부터 수집한 상태 데이터를 데이터베이스로 저장하기 위한 모듈이 설치되어 있는 수집 PC이다. ⑩번은 설비 상태 값을 저장하기 위한 데이터베이스이다. 현장의 장치들로부터 수신 받은 데이터들은 원격지 데이터베이스에 저장된 후 스크에 의해 하이브 데이터 웨어하우스에 저장된다.

(2) 수집데이터의 하둡 파일시스템 저장

오라클 데이터베이스에서 하이브로 데이터를 옮기는 스크 실행 명령어는 다음과 같다. 이 스크 명령어를 리눅스 크론탭으로 설정하면 해당 명령이 주기적으로 실행된다.

```
./sqoop import --connect jdbc:mysql://XXX.XXX.XXX.XXX/test_db --table lu_customer --username testuser -password ***** --split-by customer_id --hive-import
```

(3) 모니터링 및 분석용 웹시스템 화면

모니터링 및 분석용 웹시스템은 스프링 프레임워크와 MyBatis 기술을 사용하였다. 웹시스템 개발 및 운영환경은 <Table 4>와 같다.

Table 4. Development and operating environments

운영체제	Microsoft Windows 7, CentOS 6.5
컴파일러	JDK 1.6.0
데이터베이스	Oracle 11g, MySQL 5.0
웹 서버	Apache
웹 컨테이너	Tomcat 6.0
데이터 서비스	MyBatis 3
개발언어	Java
개발도구	Eclipse Europa

<Figure 9>는 온도, 압력, 속도, 전력 항목들에 대한 실시간 측정값들을 종합적으로 모니터링할 수 있도록 구성한 화면이다. 온도, 압력, 속도, 전력 등의 값을 실시간으로 수집하여, 임계치를 벗어나는 값들에 대해서는 알람정보를 담당자들에게 제공하여 관련된 문제를 조치할 수 있도록 개발하였다.

<Figure 10>은 엔진 메인베어링 온도 측정값 및 속도와 전력 측정값들에 대한 조회 화면이다. <Figure 11>은 엔진 내부 배출 가스의 온도 측정값에 대한 조회 화면이다. 측정된 베어링별 온도 값, 전력 값, 그리고 내부 배출 가스의 온도 값이 임계치를 벗어나는 경우 알람을 발생하도록 되어 있다.

Figure 9. Integrated display

Figure 10. Display for temp., speed, and electricity

Figure 11. Display for temp. of gas

4. 맺음말

4.1 시사점 및 기대효과

대부분의 제조현장에서는 설비의 상태를 감시하기 위하여 관계형 데이터베이스를 사용하고 있다. 이러한 관계형 데이터베이스는 테이블에 데이터를 저장하는 데에 막대한 관리 비용이 소요될 뿐만 아니라 확장성 측면에서도 유연하지 못하다.

또한 기존의 관계형 데이터베이스는 데이터를 디스크로부터 읽거나 파싱(parsing)하거나 직렬화해야 하기 때문에 데이터의 적재 속도가 매우 느리다. 하지만 이 시스템에 사용된 하둡의 요소기술인 하이브(Hive)는 단지 파일 복사 또는 이동 연산만으로 실행되기 때문에 데이터의 적재 속도가 매우 빠르다.

하둡 기반으로 개발된 이 시스템은 제조현장에 있는 각종 설비들의 대용량 데이터에 대한 실시간 수집을 가능하게 하였다. 데이터노드 3대를 기반으로 본 연구의 결과를 검증하였고, 그 결과 실시간 데이터 분석에는 미흡한 부분(특히 수행속도 저하)이 확인되었다. 100만 건의 데이터에서 원하는 데이터를 조회하는데 10분 정도의 시간이 소요되었으며 이를 관계형 데이터베이스에서 실행했을 경우는 1초 이내에 결과가 조회되었다. 하지만 이는 한정된 데이터노드의 수에 따른 결과이며 데이터노드를 확장했을 경우는 조회 수행 속도가 개선될 수도 있다. 결과적으로 이와 같은 하둡 기술이 가지고 있는 실시간 분석의 문제점을 보완하기 위해서는 Splunk, Storm, S4 등의 기술이 추가적으로 필요하다. 본 연구 내용에 이러한 기술을 추가적으로 적용한다면 보다 개선된 실시간 기반의 설비상태 관리 시스템을 구현할 수 있을 것으로 기대한다.

개발되어진 하둡 기반의 통합 모니터링시스템을 통해 현장의 데이터를 단순히 모니터링하는 수준에서 벗어나 사고 방지를 위해 즉각 실행에 옮길 수 있는 정보를 제공할 수 있게 되었을 뿐만 아니라 통합관점에서 입체적으로 설비 상태를 관리할 수 있게 되었다.

4.2 향후 연구 방향

현재 빅데이터 기술과 사물인터넷 기술은 전 세계적으로 화두가 되고 있는 미래기술이다. 이 기술들은 사람의 느낌에 의한 판단이 아니라 데이터에 기반을 둔 판단을 하게 하여 제조 현장에서 발생하고 있는 각종 작업들을 지능화하는데 핵심적인 역할을 할 것이다. 본 논문에서는 하둡 플랫폼 사용시 실시간 데이터 분석 테스트에서 수행속도가 다소 떨어지는 것을 확인하였다. 빅데이터의 핵심은 분석이다. 수집한 대용량의 데이터로부터 의미있는 정보를 신속하게 제공할 수 있어야 한다. 하둡은 데이터 처리와 저장에 최적화되어 있어 제조 현장에서 실시간으로 발생되고 있는 수많은 데이터들을 실시간으로 분석하는 데는 한계가 있다. 이러한 스트리밍 데이터를 실시간으로 분석 처리하기 위해서는 Splunk, Strom, S4 등의 기술을 추가적으로 사용해야 한다. 향후 연구에서는 실시간으로 설비 상태 데이터를 조회하기 위해 이러한 기술을 추가적으로 사용하여 기존 시스템을 보완할 예정이다.

참고문헌

- Anja, G. *et al.* (2011), Query Optimization Using Column Statistics in Hive, IDEAS, Proceedings of the 15th Symposium on International Database Engineering and Applications.
- Armbrust, M. *et al.* (2010), A View of Cloud Computing, *Communications of the ACM*, **53**(4), 50-58.
- Dean, J. *et al.* (2004), MapReduce : Simplified Data Processing on Large Clusters, Sixth Symposium on Operating System Design and Implementation, 137-150.
- GE Intelligent Platforms (2013), Proficy Monitoring and Analysis Suite, <http://www.ge-ip.com/files/files/13513.pdf>.
- Hahm, Y. K. *et al.* (2012), Big Data changes business management, Samsung Economic Research Institute, Seoul, Korea.
- Kang, M. M. *et al.* (2012), Analytics and Utilization of Big Data, *Journal of The Korea Information Science Society*, **30**(6), 25-32.
- Kim, S. R. *et al.* (2012), The Future of Big Data, *Journal of The Korea Information Science Society*, **30**(6), 18-24.
- POSCO (2013), Introduction Document for POSCO IMC Center.
- Shvachko, K. *et al.* (2010), The Hadoop Distributed File System, In Proceedings of the 26th IEEE Transactions on Computing Symposium on Mass Storage Systems and Technologies, 1-10.
- Son, M. S. *et al.* (2012), To be a leader in the Big Data era, LG Economic Research Institute, LGERI Report, 2-6.
- Shim, T. G. *et al.* (2011), Hadoop Complete Guide, Hanbit Media, Seoul, Korea.
- Vertica Systems (2009), Managing Big Data with Hadoop and Vertica.
- Yun, M. R. (2012), Taking advantage and problem of Big Data, The Federation of Korean Information industries Issue Report, 10-13.