



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

기술경영학 석사 학위 프로젝트 보고서

빅데이터 분석 플랫폼을 이용한 생산 프로세스 개선



2018년 8월

부경대학교 기술경영전문대학원

기술경영학과

원 석 래

기술경영학 석사학위 프로젝트 보고서

빅데이터 분석 플랫폼을 이용한 생산 프로세스 개선

지도 교수 김 민 수

석사학위 논문에 준하는 보고서로 제출함.

2018년 8월

부경대학교 기술경영전문대학원

기술경영학과

원 석 래

원석래 의 기술경영학 석사학위 프로젝트보고서를 인준함.

2018년 8월 24일

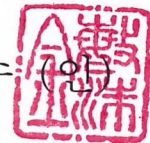
위 원 장 공학박사 신 승 준



위 원 기술경영학박사 천 동 필



위 원 공학박사 김 민 수



목 차

I. 서 론	1
1. 연구의 배경	1
2. 연구의 필요성	4
3. 연구의 목적	5
II. 선행연구	7
III. 빅 데이터 분석의 이론적 고찰	7
1. 데이터 마이닝	7
2. 프로세스 마이닝	8
3. 하둡	9
4. 데이터 처리 기술	10
IV. 분석 시스템 프레임워크	12
1. 프로세스 애널리틱스 플랫폼 아키텍처	12
2. 빅 데이터 분석 시스템 아키텍처	13
3. 시스템 기능	15
V. 분석 사례	21
VI. 결론	27
1. 연구 결과의 요약	27
2. 연구의 한계점과 향후 연구의 방향	27

참고 문헌	28
1. 국내 문헌	28
2. 해외 문헌	30
감사의 글	33



표 목 차

<표 1-1> Industry 4.0 Design Principles -----	2
<표 1-2> 프로세스 마이닝과 BPM 비교-----	5
<표 1-3> 프로세스 마이닝 분석도구 간 비교-----	6
<표 5-1> 분석 데이터 소개 -----	22
<표 5-2> 도입 전 후 분석 대상 데이터 정보-----	23



그 립 목 차

<그림 1-1> Industry 4.0 Design Principles-----	1
<그림 1-2> 스마트공장 시스템 구성도-----	3
<그림 1-3> Data Mining, BPM과 프로세스 마이닝의 관계-----	4
<그림 3-1> 데이터마이닝 단계-----	11
<그림 3-2> 프로세스 마이닝-----	12
<그림 3-3> 하둡 에코시스템-----	13
<그림 3-4> 오픈소스를 이용한 CEP 데이터 흐름 예시-----	14
<그림 4-1> 프로세스 애널리틱스 플랫폼-----	15
<그림 4-2> 프로세스 애널리틱스 플랫폼 데이터 흐름-----	16
<그림 4-3> 어플리케이션 구조-----	17
<그림 4-4> 어플리케이션 화면-----	18
<그림 4-5> 공정간 연관성 분석-----	19
<그림 4-6> 공정 흐름 분석 상세 화면-----	20
<그림 4-7> 공정 적합성 확인-----	20
<그림 5-1> 공정 흐름 도출-----	22
<그림 5-2> 기초 공정 흐름 분석(도입 전)-----	24
<그림 5-3> 기초 공정 흐름 분석(도입 후)-----	24
<그림 5-4> 평균 재공 시간(도입 전)-----	25
<그림 5-5> 평균 재공 시간(도입 후)-----	25
<그림 5-6> 실공정 데이터 유사도 및 공정 흐름 모델 적합도-----	26

The Improvement of Manufacturing Process Using Big Data Analytics Platform

Seok Rae Won

Graduate School of Management of Technology
Pukyong National University

Abstract

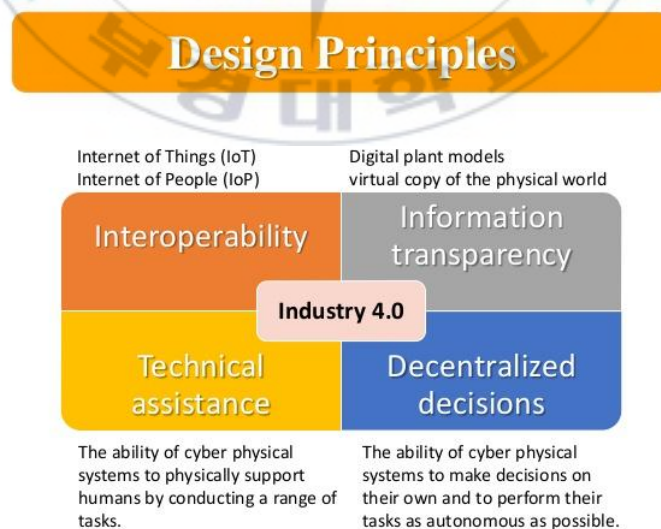
Since the rapid development of ICT technology after the third industrial Revolution, the information revolution, the growing interest in the convergence of traditional industries, such as manufacturing, has been actively conducting research on manufacturing and ICT convergence, beginning with Germany's industry 4.0. This lowers the manufacturing cost of traditional manufacturing and enables the expectation of quality improvement. The recent government Manufacturing 3.0 strategy has driven the high degree of manufacturing in the country, and in keeping with this trend, we are developing an analytical platform for big data on the production site of the manufacturing industry and we want to increase the competitiveness of manufacturing sites. The Big data analytics platform for improving production processes is designed to leverage the open source needed for big data analysis, develop on a scalable framework basis, and design a user interface that can run on a variety of devices. Distributed processing systems and distributed file systems were used to quickly process large amounts of data occurring in the manufacturing site. By processing real-time streamed data through process mining-based algorithms and analyzing plant status, we are able to identify problems in the field and support rapid decision making to improve the production process, which can be expected to increase the production efficiency of real enterprises.

I. 서론

1. 연구의 배경

정보화 혁명인 3차 산업혁명 이후 Information and Communications Technologies(ICT) 기술의 발달로 각종 정밀 센서를 생산 설비에 장착 가능하게 되고 Internet of Things(IoT) 기술을 통하여 센서들로부터 데이터를 수집, 전송이 가능하게 되었으며 다양한 형태 및 데이터의 양 또한 증가하게 되었으며 빅 데이터라는 용어가 생산 프로세스에도 등장하게 되었다. 과거 공장자동화를 통하여 각 공정에 대한 데이터 분석에서는 알 수 없었던 전 후 공정간 발생한 현상에 대하여 제조 공정 내 모든 설비가 연결되어 정보를 교환하고 생산된 빅 데이터를 기반으로 분석·판단하여 최적의 생산 프로세스를 만들 수 있는 환경이 만들어 지고 있다.

<그림 1-1> Industry 4.0 Design Principles



이러한 변화를 'Industry 4.0'이라는 키워드로 주도하고 있는 독일은 2003년부터 이에 대한 논의를 시작하였으며 이후 민간의 주도하에 다양한 분야의 단체들로부터 수많은 의견 수립·제안을 통하여 수정·보안되어 발표되었다. 이러한 Industry 4.0은 다음과 같은 네 가지 디자인 원칙을 가지고 있다(Mario Hermann, Tobias Pentek, Boris Otto, 2016).

<표 1-1> Industry 4.0 Design Principles

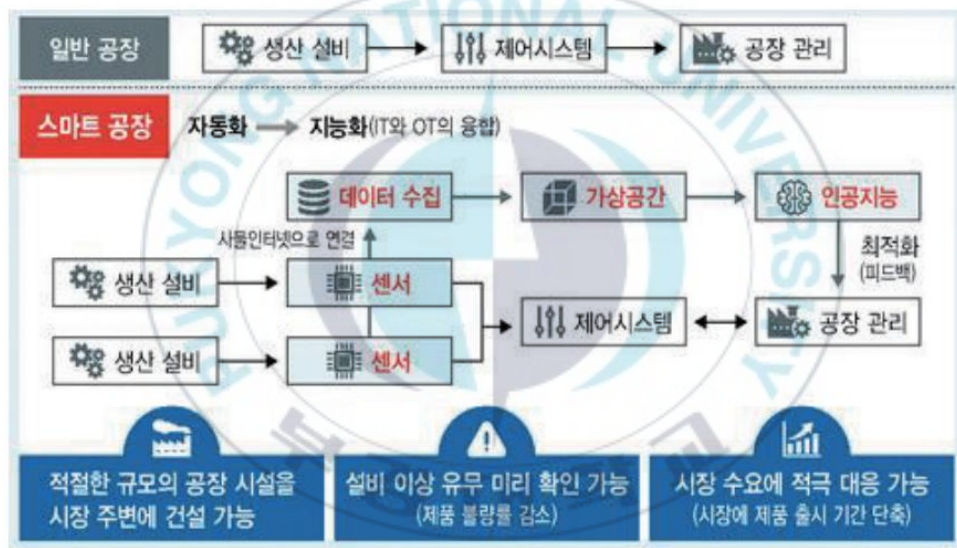
원칙	내용
1. 상호 운용성	IoT 또는 IoP를 통해 기계, 장치, 센서, 사람들을 연결하고 커뮤니케이션 할 수 있는 능력
2. 정보 투명성	센서 데이터들로부터 수집한 데이터를 기반으로 디지털 기반 모델을 정립하고 이를 이용하여 실제 물리 세계를 가상의 복제 본으로 창조할 수 있는 정보 시스템
3. 기술 지원	객관적인 정보 기반의 신속한 의사결정을 위한 데이터 수집 및 가공 시스템, 기피 업종이거나 위험한 일 등 인간의 노동을 물리적으로 지원하는 물리 시스템
4. 능동형 결정	물리 시스템이 스스로 결정을 내리고 수행

제조공정의 장비에 부착된 센서들을 통해 생산된 데이터들은 통신네트워크를 통하여 Database에 저장된다. 위에 언급된 Industry 4.0 디자인 원칙 중 정보 투명성을 만족하며 수집된 데이터들이 축적되어 대용량의 데이터가 된다. 이를 빅 데이터 기술로 분석하여 MES, PLM, POP등과 같은 생산제조시스템과 ERP와 같은 경영관리 플랫폼에 연동되어 객관적인 데이터를 기반으로 의사결정을 지원할 수 있다. 이를 통해 과거 인간의 경험적 판단에 의한 오차로 발생하는 문제점을 사전에 방지, 비용 낭비를 최소화

하고 변화에 훨씬 빠르게 대응하면서 품질을 획기적으로 높일 수 있고 중요한 경영 전략 결정에 까지 활용될 수 있다.

본 프로젝트 보고서는 최근 스마트공장 도입을 위하여 적용 되고 있는 센서 장비에서 발생되는 빅 데이터를 활용하여 공정의 흐름을 발견하고 나아가 설계된 공정 흐름과 실제 공정 흐름의 차이를 비교하여 최적의 생산 프로세스를 도출하여 개선 모델에 대한 의사 결정을 지원해 줄 수 있는 빅 데이터 분석 플랫폼을 도출하고자 한다.

<그림 1-2> 스마트공장 시스템 구성도



출처 : 대한민국 제조혁신 콘퍼런스(한국인더스트리 4.0협회, 2017)

근래 빅 데이터라는 용어의 등장으로 다양한 영역에서 그 데이터를 활용 및 분석하는 연구 및 적용이 활발히 진행되고 있다. 데이터 마이닝, 텍스트 마이닝, 데이터베이스, 머신러닝 등과 같이 데이터를 다루는 연구 분야이다. 이러한 연구는 축적된 데이터를 통하여 인자 간의 상관관계를 찾거나 각종 패턴을 발견하는 것들에 많이 활용되고 있다.

2. 연구의 필요성

최근 전 세계적으로 ICT 및 IoT 기술이 발전함에 따라 세계적으로 스마트공장의 기술이 주목 받고 있으며, 국내에서도 관심이 늘어나고 있는 상황이다. 하지만 각종 Factory Things(센서, 모바일, 디바이스, IoT 디바이스 등)에서 생성되는 수많은 데이터들이 ERP, MES, POP, MIS 등과 같은 시스템에 의하여 생성되고 축적되고 있지만 그 활용도는 저조하다.(John Gantz and David Reinsel, 2012).

대부분의 빅 데이터의 구조적 형태는 비정형 구조를 가지고 있으며 제조 공정에서 발생하는 비정형 데이터를 통하여 흐름을 분석하는데 사용할 수 있는 유용한 기법이 프로세스 마이닝이다(Wil M. P. van der Aalst, 2011). 프로세스 마이닝은 데이터 분석 기법(Data Mining)과 비즈니스 프로세스 관리(BPM)의 기능의 상호 보완적 역할을 위하여 개발된 기법으로 제조 공정에서 발생하는 빅 데이터를 통해 공정 간의 상관관계 또는 문제 발견을 통한 해결 방법을 도출 할 수 있도록 의사 결정 지원을 한다.

<그림 1-3> Data Mining, BPM과 프로세스 마이닝의 관계



<표 1-2> 프로세스 마이닝과 BPM 비교

구분	프로세스 마이닝	BPM
프로세스 모델 도출	과학적, 자동화 도출	경험적, 정성적(인터뷰, 자료 검토) 도출
분석 방식	데이터 중심	사람 중심
분석 기간	로그데이터 확보시 빠른 분석 가능	복잡한 프로세스로 이루어진 경우 3~4개월 소요
추가 분석	업무성과/조직모델/구성원들의 관계 분석, 업무 성과 특성 분석 지원	추가 분석 어려움
한계점	로그데이터가 없을 경우 분석 불가	정확한 현상 파악이 어려움 사실 왜곡의 발생

<표 1-2>에서와 같이 프로세스 마이닝과 BPM은 적용 영역에 대하여 많은 공통점을 가지고 있으며 상호 보완적인 역할을 할 수 있는 기능들을 가지고 있으며 생산 프로세스 개선을 위하여 BPM과 프로세스 마이닝이 융합된 기능 적용이 필요하다.

3. 연구의 목적

본 프로젝트 보고서의 결과를 통하여 국내 제조업의 제조 공정 고도화 기술로 활용된다면, 국내 제조업의 경쟁력 향상에 기여할 수 있을 것으로 기대된다. 본 프로젝트에 활용된 프로세스 마이닝 기법을 이용한 도구는 클라우드 및 빅 데이터 기반의 플랫폼에서 개발된 IOChord Process Analytics Platform 및 IOChord Process Re-Engineering 이다. 전 세계적으로 알려진 프로세스 마이닝 분석 도구는 오픈 소스 도구인

ProM(Process Mining Group Math&CS department, 2015)과 상용도구인 Disco(Disco Fluxicon, 2015)이다.

<표 1-3> 프로세스 마이닝 분석도구 간 비교

구분	ProM	Disco	Process Analytics
라이선스	GPL(오픈 소스)	상용 라이선스	상용 라이선스
플랫폼	데스크톱 설치	데스크톱 설치	클라우드 웹기반
로그저장관리	No(ProM 5.2) State(ProM 6.x)	Yes	Yes
데이터 연동	No	No	Yes
실시간 처리 여부	No	No	Yes
확장성	Plug-In	No	Service Plug-In
사용자환경	MDI	SDI	웹
기능	600++	6가지 주요기능	10가지 주요기능

ProM은 프로세스 마이닝 개념을 처음으로 만들어낸 네덜란드 아인트호벤 공대의 Aalst 교수를 중심으로 만들어 졌으며, 연구 결과를 자유롭게 Plug-In 형태로 공유할 수 있도록 구성되어 있다. 상용버전인 Disco는 ProM을 기반으로 한 상용화 제품으로 현재 프로세스 마이닝 도구로는 국내 및 해외에서 가장 잘 알려진 분석 도구 이다.

그러나 가장 잘 알려진 두 가지 분석도구는 분석자의 컴퓨터에 어플리케이션 형태로 설치되어 운영되므로 분석하고자 하는 데이터의 크기가 커지게 되면 분석 시간이 오래 걸리거나 또는 분석이 어려운 구조를 가지고 있으며 분석자의 컴퓨터에 설치되므로 분석자의 컴퓨팅 능력에 따라 많은 영향을 받게 된다. 또한 이벤트 데이터를 업로드 하여 분석하는 형태라 실제 시스템과 연동을 통하여 분석 결과를 도출하기 어렵다. 이러한 단점을 극복할 수 있는 방법으로 클라우드 시스템에서 운영되는 분산파일 및 분산처리 시스템을 통하여 그 단점을 극복하고자 하며 웹 어플리케이션 형태로

분석자에게 결과를 제공하여 분석자의 컴퓨팅 능력에 따라 소요되는 시간을 최소화 할 수 있는 방법으로 시스템을 설계하고 개발하게 되었다. 기존의 프로세스 마이닝 도구는 데이터로부터 프로세스 모델을 발견하는데 너무 치우쳐 있고, 연구를 위한 목적으로 그 기능이 충실히 구성되어 있어 실제 산업현장에 적용하기에는 어려움이 있으며 프로세스 마이닝 분석도구 간의 차이는 <표 1-3>과 같다.

본 프로젝트 보고서는 스마트 공장이라는 새로운 적용 환경에 적용하기 위하여 기본 개념을 충실히 적용하여 현장에 실제 적용할 수 있는 모델을 제시하고 BPM과 프로세스 마이닝이 조화를 이루어 생산 프로세스 개선을 위한 빅데이터 분석 플랫폼 적용 사례를 논하기로 한다.

Ⅱ. 선행연구

프로세스 마이닝은 정보시스템 등을 통하여 기록된 이벤트 로그에서 프로세스 지식을 도출함으로써 운영 프로세스에 대한 발견과 모니터링, 개선을 지원하는 연구 분야이다. 즉, 프로세스 실행 과정에서 발생한 기록을 기초로 프로세스를 도출하고, 개선방안을 수립하는데 필요한 유용한 지식과 정보를 추출하는 것이다. 정보시스템으로부터 작업의 실행 결과 및 이벤트 로그를 추출하여 기존의 통계 기법(Jansen-Vullers, M.H., W.M.P. van der Aalst, and M. Rosemann, 2006), 인공지능 기법(Rozinat, A., W.M.P. van der Aalst, 2006), (Van der Aalst, W.M.P., A.K. Alves de Medeiros and A.J.M.M. Weijters, 2007), 사회적 네트워크 기법(Van Der Aalst, W.M.P., H.A. Reijers, M.Song, 2005) 등을 이용하여 분석함으로써, 프로세스 모델을 유도하거나, 업무의 상관관계, 작업자의 업무 전달 관계 등 프로

세스 수행 상의 특징을 분석한다(Van der Aalst, W.M.P., T. Basten, 2002).

프로세스 마이닝의 기본적인 목적은 사람들이 머릿속에서 추정하고 있는 프로세스가 아닌 실제 현장의 업무 프로세스를 도출하고, 모니터링하며, 개선하는 것이다. 이를 수행하기 위해서는 프로세스 도출(process discovery), 적합성 검사(conformance check) 및 개선 및 향상 (enhancement)의 과정이 필요하다(Van der Aalst, W.M.P., A. Adriansyah, A.K. Alves de Medeiros, A. Karla Alves, F.Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van den Brand, R. Brandtjen, J. Buijs, 2012).

프로세스 마이닝 연구 초기에는 데이터 로그로부터 프로세스 모델을 추출하거나 데이터를 분석하는 기법에 관한 연구가 주를 이루었다면, 최근의 프로세스 마이닝의 연구는 실제 데이터를 이용하여 프로세스 마이닝을 수행하는 사례 연구 중심으로 진행되고 있다(이동하, 2014).

Alst et al.은 그들의 연구에서 공공 서비스 부분의 실제 데이터를 이용한 프로세스 마이닝을 시도하였으며(Van der Aalst, W.M.P., H.A. Reijers, A. J. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H. Verbeek, 2007), Medeiros et al.은 유전자 알고리즘을 적용하여 시행 로그 데이터의 예외 상황에 대해서 보다 안정적인 모델을 발견할 수 있도록 하였다(Van der Aalst, W.M.P., A.K. Alves de Medeiros and A.J.M.M. Weijters, 2007). Mans et al.은 의료 서비스 분야에서 환자가 의료서비스를 제공받는 흐름에 대한 연구를 수행하였고(Mans, R., M. Schonenberg, M. Song, W.M.P. van der Aalst., P.J. Bakker, 2009), 프로세스 수행 조직 관점에서 업무 프로세스를 분석하고, 작업간의 수행 관계를 분석한 사례도 있다(Song, M., W.M.P. van der Aalst, 2008). 제조업 분야에 적용한 사례는 Rozinat et al.과 Goedertier et al.의 연구가 있다. 위 연구들은 사례 연구

로서 기존의 프로세스 마이닝을 적용하고 평가한 연구이다. 발견된 모델의 결과가 복잡한 점을 지적하며 다른 프로세스 마이닝 기법의 필요성을 제시하였고(Rozinat, A., I.S. de Jong, C. Gunther, W.M.P. van der Aalst, 2009), Goedertier et al.은 통신 산업에 기존의 데이터마이닝 기법들을 적용하였다(Goedertier, S., J. D. Weerdt, D. Martens, J. Vanthienen, and B. Baesens, 2011). 금융 분야 사례연구에서도 프로세스 마이닝 기법들이 더욱 정교화 되고 성숙되어야 한다는 의견이 있었다(Jans, M., V.D. Werf, J. Martijn, N. Lybaert, and K. Vanhoof, 2011).

최근 국내에서도 다양한 사례연구가 진행되었다. 이동하는 조선 산업의 생산 공정 분석에 프로세스 마이닝 기법을 적용하여, 생산 공정상의 블록 이동과 블록 조립 프로세스를 분석하였다(이동하, 2014). 대학의 전자결재 프로세스를 프로세스 마이닝 기법을 이용하여 분석하였으며(강영식 외 5명, 2012), 감사 분야에 프로세스 마이닝 기법을 적용하여 유럽 은행의 실제데이터를 기반으로 분석한 사례도(강영식, 강창재, 이은상, 신철규, 2013) 있다. 그밖에 전시행사의 관람객 흐름과 관람 시간을 분석하였으며(손숙영, 문현실, 김재경, 송민석, 2012), 전자기기의 고장 수리 프로세스를 분석하였다(최상현, 한관희, 임진훈, 2013). 물류분야의 적용사례로는 Jeon and Bae의 연구(전대욱, 배혜림, 2013)와 Wang et al.(Wang, Y., F. Caron, J. Vanthienen, L. Huang, Y. Guo, 2014)의 연구가 있다.

본 프로젝트 보고서에서는 프로세스 마이닝 기법을 적용하여 냉연 철강 회사의 제조 프로세스를 분석하고자 한다. 냉연 철강 회사의 실제 작업 로그 데이터를 기반으로 다양한 프로세스 마이닝 기법들을 적용하여, 현실을 반영한 프로세스 도출하고 이를 분석하여 개선 방안을 수립하는 방법론을 제시할 것이다.

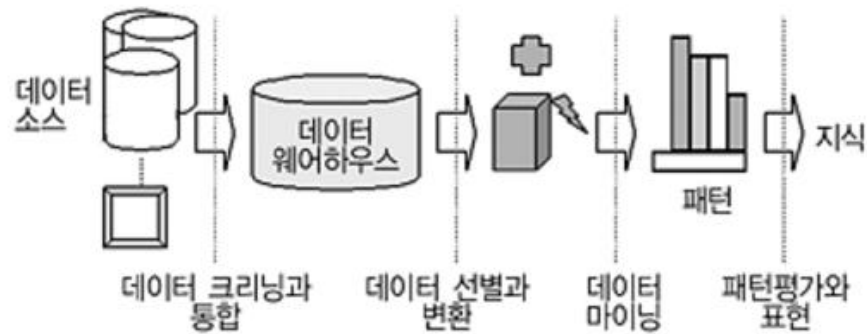
Ⅲ. 빅 데이터 분석의 이론적 고찰

1. 데이터 마이닝

데이터 마이닝은 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 기술로써 KDD(Knowledge-discovery in databases)라고 칭하기도 한다. 통계학에서부터 패턴 인식에 이르는 다양한 계량 기법을 사용하여 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견할 수 있다. 통계학 쪽에서 발전한 탐색적 자료 분석, 가설 검정, 다변량 분석, 시계열 분석, 일반선형모형 등의 방법론과 데이터베이스 쪽에서 발전한 OLAP(On-Line Analytic Processing), 인공지능 진영에서 발전한 SOM, 신경망, 전문가 시스템 등 기술적인 방법론이 쓰이며 이를 활용하여 신용평점 시스템(Credit Scoring System), 사기탐지시스템(Fraud Detection System), 장바구니 분석(Market Basket Analysis), 최적 포트폴리오 구축과 같이 다양한 산업분야에서 광범위하게 사용되고 있다. 대표적인 적용 방식으로는 다음과 같다.

- 1) 분류(Classification) : 일정한 집단에 대한 특정 정의를 통해 집단을 분류하고 그 구분의 인자를 추론한다.
- 2) 군집화(Clustering) : 수많은 데이터 속에서 구체적인 특성을 공유하는 군집(Cluster)를 찾는다.
- 3) 연관성(Association) : 동시에 발생한 사건간의 관계를 찾는다.
- 4) 연속성(Sequencing) : 특정 기간에 걸쳐 발생하는 관계를 규명한다.
- 5) 예측(Forecasting) : 대용량 데이터 집합 내의 패턴을 기반으로 발생할 데이터를 예측한다.

<그림 3-1> 데이터마이닝 단계

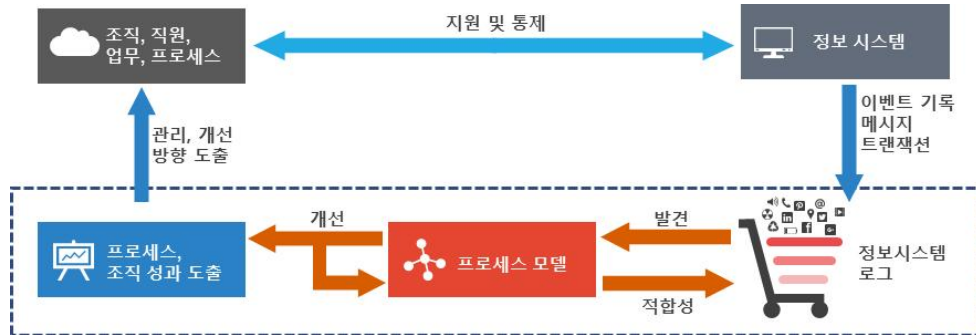


2. 프로세스 마이닝

프로세스 마이닝 기술은 정보 시스템에 기록이 되는 이벤트 로그를 분석하여 프로세스와 관련된 다양한 정보를 분석하는 기술로서 비교적 새로운 연구 분야로 데이터마이닝과 Business Process Management (BPM)의 연결고리 연결을 하며 데이터마이닝과 상호보완적인 관계에 있다.

데이터의 전체적인 분석에 대한 결과를 내는 데이터 마이닝에 비교하여 데이터와 데이터 사이의 관계를 모델화에 적합하며, BPM의 추상적인 모델 비교하여 프로세스 마이닝은 현실적인 모델의 시각화에 적합하다. 프로세스 마이닝의 기본적인 아이디어는 정보시스템에 기록되어 있는 이벤트 로그에서 지식을 추출함으로써, 사람들이 머릿속에서 추정하고 있는 프로세스가 아닌 실제 업무 프로세스를 도출하고 모니터링하며 개선한다.

<그림 3-2> 프로세스 마이닝

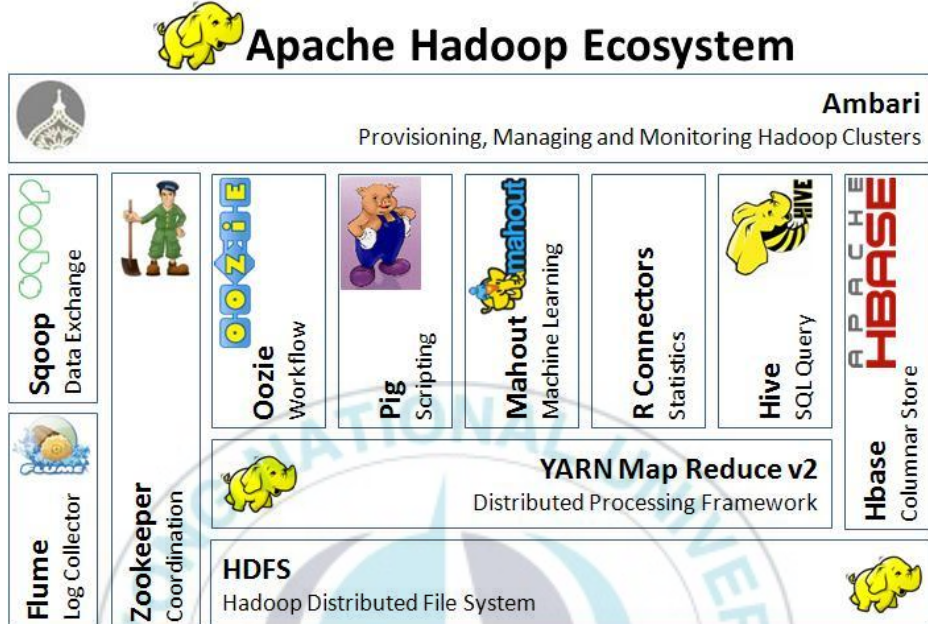


정보 시스템의 사용 기록에 대한 이벤트를 분석하여 기업의 프로세스를 발견, 병목지점 분석, 실적분석 및 상관관계 분석 등이 가능하다.

3. 하둡(Hadoop)

하둡은 대량의 자료를 처리하는 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프리웨어 자바 소프트웨어 프레임워크, 분산처리를 지원하기 위해 개발되어 하둡 분산 파일 시스템(HDFS: Hadoop Distributed File System)과 맵 리듀스를 구현, 아파치 루씬의 하부 프로젝트이다.

<그림 3-3> 하둡 에코시스템, Apache Hadoop



구글, 야후 그리고 페이스북등 글로벌 IT기업에서 핵심 기술로 사용되고 있으며, 그 외 많은 기업들이 자사 솔루션에 응용 및 활용하고 있다.

4. 데이터 처리 기술

가. 분산 데이터 처리(Distributed Data Processing)

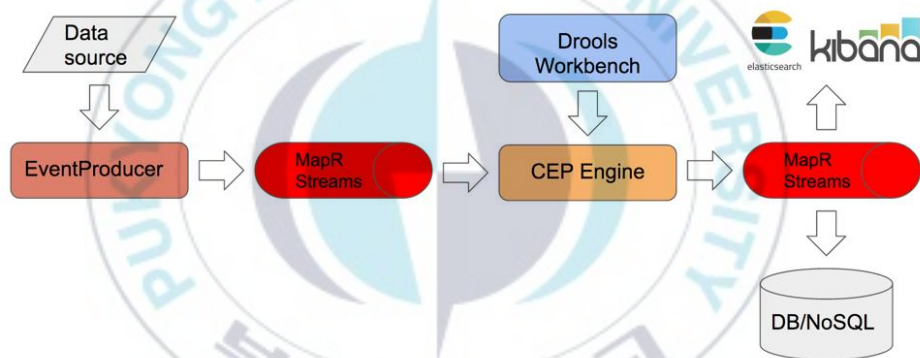
병렬 데이터 처리와 분산 데이터 처리는 데이터를 분할하여 프로세스 후 다시 결합하는 시스템데이터를 처리하는 측면에서는 유사한 처리 방식을 가지고 있다. 분산데이터처리는 클러스터로 네트워크에 결합되어 있는 물리적으로 분리된 컴퓨터에서 실행된다.

나. 병렬 데이터 처리(Parallel Data Processing)

대규모 작업을 작은 단위로 나누어 동시에 다수의 실행하는 방법으로 데이터를 처리하여 실행시간을 획기적으로 줄일 수 있다. 분산 처리와 달리 다수의 프로세스나 코어를 가지고 있는 단일 머신에서 데이터를 처리한다.

다. 복합 이벤트 처리 기술(CEP, Complex Event Processing)

<그림 3-4> 오픈소스를 이용한 CEP 데이터 흐름 예시



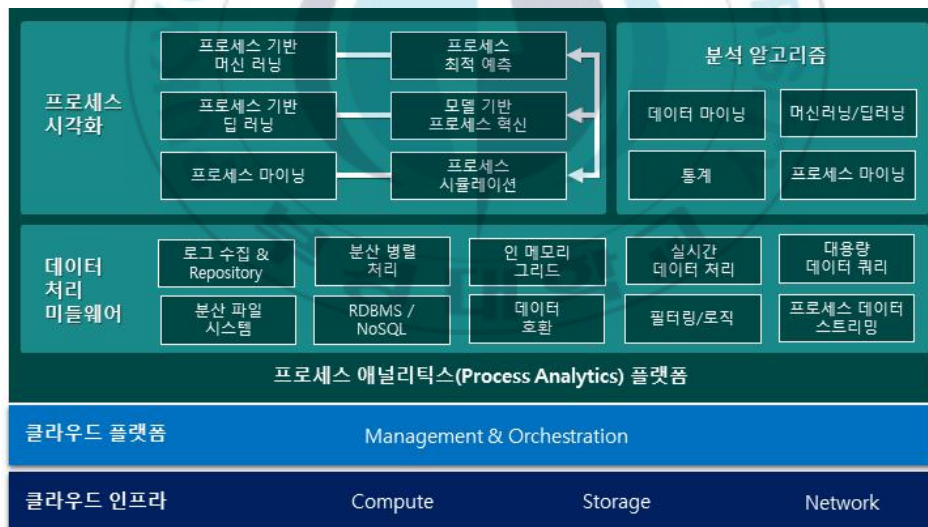
CEP는 여러 이벤트 소스로 부터 발생한 이벤트를 대상으로 실시간의 의미 있는 데이터를 추출하여 대응되는 액션을 수행하는 것을 말한다. 이때 이벤트 데이터는 스트림 데이터로써 대량의 지속으로 입력되는 데이터, 시간 순서가 필요한 데이터, 끝이 없는 데이터이다. 스트림 데이터는 전통적인 관계형 데이터베이스에서는 실시간 처리 분석이 불가능하다. CEP는 스트림 데이터를 실시간으로 분석하는 이벤트 데이터 처리 솔루션이다. 즉, 데이터베이스나 파일, 하둡에 저장하지 않고, 다양한 고속의 이벤트 스트림을 In-Memory 기반으로 초당 수백/수십만 건의 실시간 처리가 가능하다.

IV. 분석 시스템 프레임워크

1. 프로세스 애널리틱스 플랫폼 아키텍처

프로세스 애널리틱스 플랫폼 아키텍처는 <그림 4-1>과 같이 클라우드 인프라, 클라우드 플랫폼, 프로세스 애널리틱스 플랫폼으로 구성된다. 프로세스 애널리틱스 플랫폼에 사용되는 각각의 소프트웨어는 오픈소스 소프트웨어 기반에서 운영될 수 있도록 하였고 클라우드 인프라 및 플랫폼 영역은 활용 방향에 따라 상용제품 또는 오픈소스 제품이 병행되어 사용될 수 있다.

<그림 4-1> 프로세스 애널리틱스 플랫폼



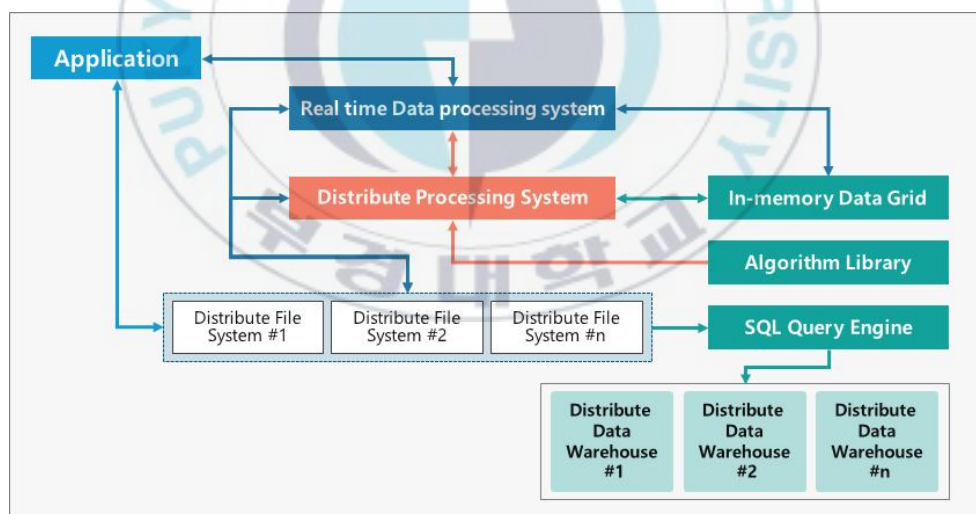
클라우드 인프라 영역은 서비스 리소스에 대한 가상화를 담당하고 클라우드 플랫폼 영역은 인프라 영역의 가상화 리소스를 관리할 수 있는 영역을 담당하며 시스템의 사용 상황에 따라 가상화 리소스의 확장성을 관리하

게 된다. 프로세스 애널리틱스 플랫폼 영역은 빅 데이터 분석에 필요한 데이터의 수집과 분석 영역을 담당하게 되며 분석 결과에 따른 시각화 영역과 분석 알고리즘 영역으로 나누어 구성되었다.

2. 빅 데이터 분석 시스템 아키텍처

빅 데이터 분석 시스템 아키텍처는 <그림 4-2>와 같이 실시간 데이터 분석 시스템, 분산 처리 시스템, 분산 파일 시스템, 분산 SQL Query Engine, In-Memory Data Grid 등 오픈 소스 소프트웨어 기반으로 구성하였으며 해당 시스템에 대하여 최적의 형태로 사용할 수 있도록 구성 하였다.

<그림 4-2> 프로세스 애널리틱스 플랫폼 데이터 흐름



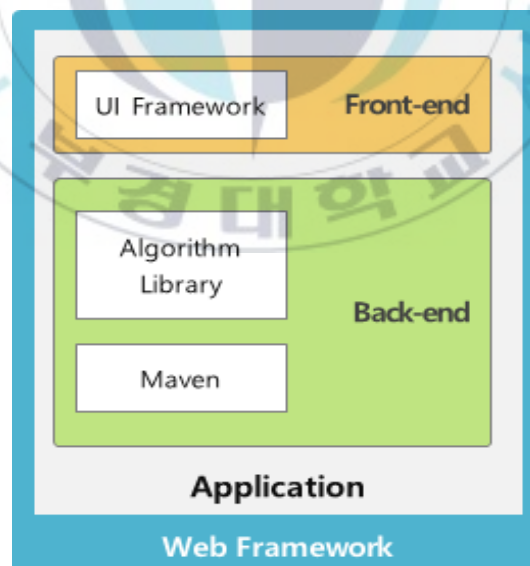
활용된 오픈소스 소프트웨어는 HDFS, Hadoop, Spark, Yarn, Sqoop, Presto 등 이미 글로벌 기업의 서비스 환경에서 운영 중인 소프트웨어를 사용하여 안정성에 대하여 보장할 수 있도록 하였다. 알고리즘 라이브러리

는 분산 환경에 적합하도록 개발 하였으며 운영 OS에 대하여 호환성 확보를 위하여 Java로 개발 하였다.

Back-end와 Front-end간의 통신은 RESTful API 방식을 활용하여 인증을 통한 타 어플리케이션에서도 활용 가능 할 수 있도록 설계 되었으며 알고리즘 호출과 반환은 JSON 데이터 형식을 사용하여 플랫폼에 독립적으로 개발 운영 될 수 있도록 하였다.

시스템 구성에서 가장 중요한 부분은 대용량 데이터를 처리할 수 있는 빅데이터 기술과 프로세스 마이닝이 상호 작용이 가능 하도록 개발 되었다. <그림 4-3>과 같이 어플리케이션은 프론트 엔드와 백 엔드로 구성되어 있으며, 백 엔드에서 처리 및 분석된 결과를 JSON 데이터 형태로 전달 받아 프론트 엔드의 유저 인터페이스를 통하여 시각화 시켜 사용자에게 결과를 전달하는 형태이다.

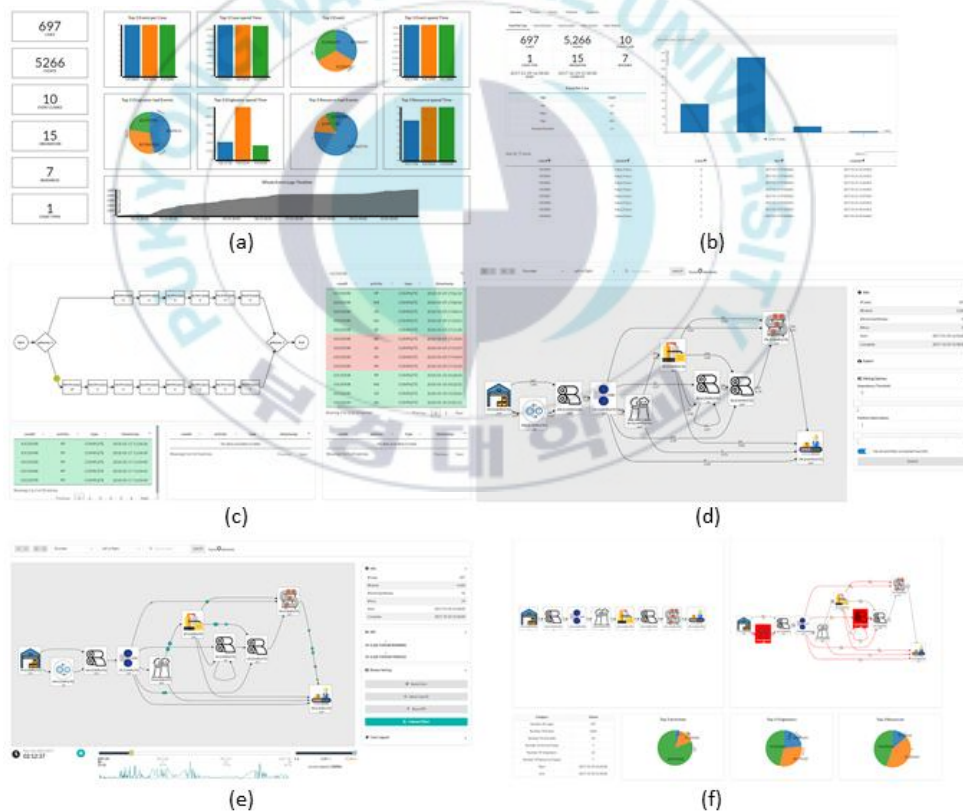
<그림 4-3> 어플리케이션 구조



3. 시스템 기능

시스템의 주요 기능은 <그림 4-4>와 같으며 요약화면(a)은 분석된 데이터에서 이벤트를 많이 발생 시키는 케이스와 오랜 시간이 소요되는 케이스, 가장 많이 발생하는 이벤트와 이벤트 별 소요시간 그리고 장비 및 사람에 대하여 요약 정보를 표출한다. 전체 데이터에 대한 대시보드(b)역할을 하는 화면에서는 분석에 대한 raw데이터를 확인 하고 조회 할 수 있다.

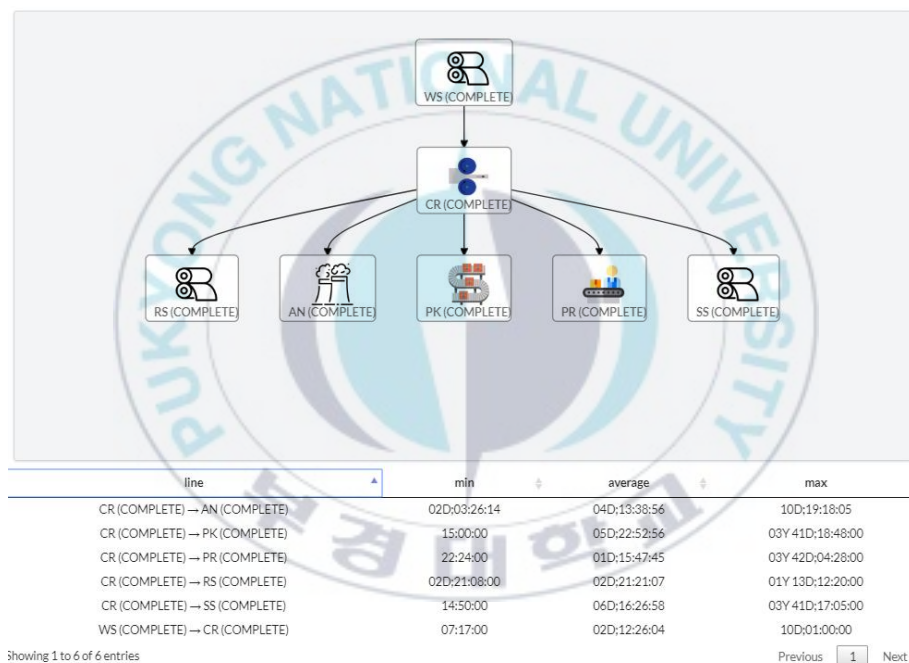
<그림 4-4> 어플리케이션 화면



실시간 모니터링(c) 화면에서는 설계된 프로세스 흐름에 대하여 생산되는

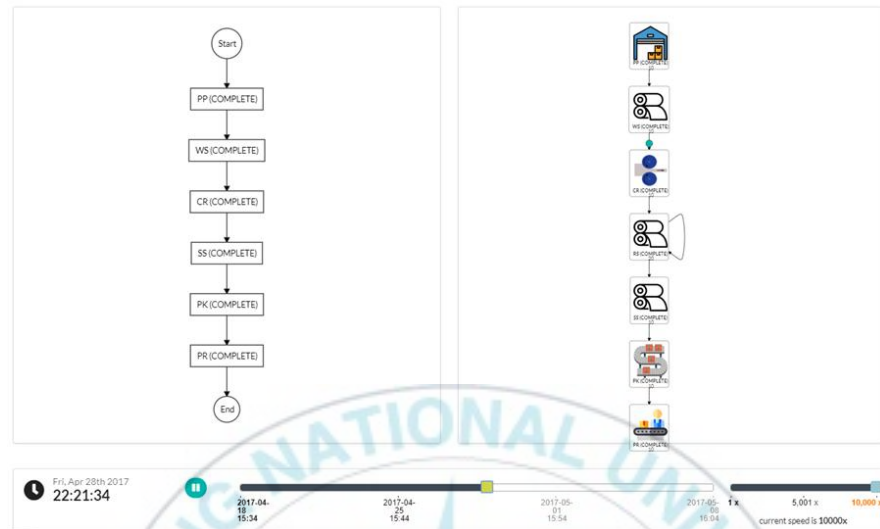
공정에서 수집된 데이터를 통하여 공정 흐름에 대한 이상 여부를 확인할 수 있다. 공정 흐름도(d)에서는 수집된 이벤트 데이터를 통하여 실제 공장의 공정 모델을 도출할 수 있으며 <그림 4-5>와 같이 특정 공정을 선택하여 전·후 공정에 대한 요약된 정보를 확인할 수 있으며 시간차 분석 알고리즘을 통하여 공정 간의 최소, 평균 및 최대 시간에 대한 정보를 확인할 수 있도록 한다.

<그림 4-5> 공정간 연관성 분석



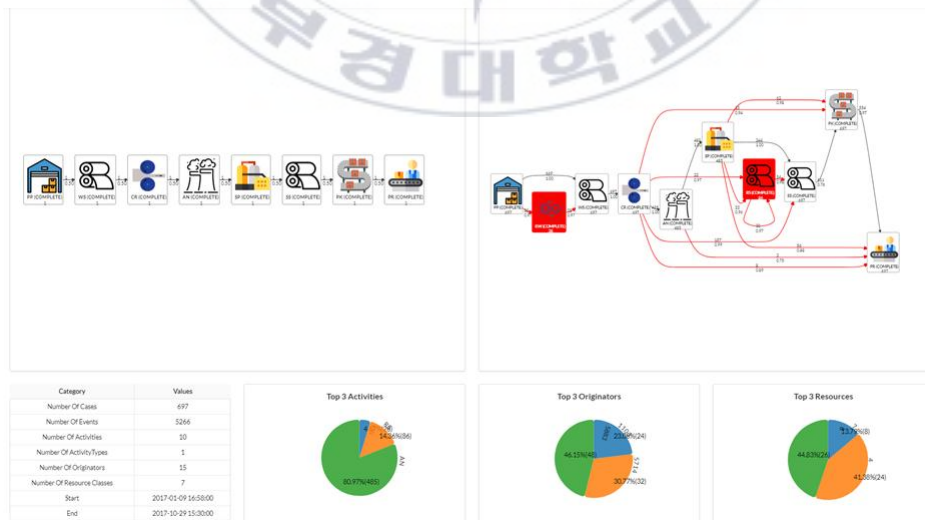
공정 애니메이션(e) 화면에서는 과거의 데이터를 통하여 공정개선을 위하여 공정 병목 등에 대한 이상 흐름을 발견할 수 있도록 지원하며, 공정 상에 진행 중인 토큰을 선택할 경우 해당 토큰에 대하여 <그림 4-6>과 같이 설계 공정과 이벤트에서 추출된 흐름을 비교하여 확인할 수 있다.

<그림 4-6> 공정 흐름 분석 상세 화면



공정의 적합성 검사(f)는 설계 공정과 이벤트에서 추출된 공정 모델간의 차이점을 분석하여 목표로 하는 공정 개선을 위한 의사 결정 지원을 할 수 있도록 구성되었다.

<그림 4-7> 공정 적합성 확인



<그림 4-7>과 같이 설계 공정과 이벤트에서 추출된 공정에서 공정 흐름에 대한 이상 흐름을 나타내고 이상 흐름에 대한 제품의 요약 정보를 제공한다. 또한 실제 공정에서 설계 공정 흐름의 공정을 거치지 않을 경우 설계 공정에 해당 공정을 표현하여 설계공정과 실제공정에 대한 비교를 직접적으로 할 수 있도록 지원한다.

공정 적합성 확인은 Business Process Modeling Notation(BPMN) 2.0 표준과 프로세스 마이닝을 통하여 도출된 프로세스 모델과의 상호 교차 비교가 가능하도록 설계하여 개발 되었으며 BPMN과 BPMN, BPMN과 프로세스 마이닝 모델 및 프로세스 마이닝 모델간의 비교가 가능하도록 구성되어 있다.

V. 분석 사례

개발된 본 제품을 통하여 실제 국내 냉연 철강 공장의 분석을 진행하였다. 해당 철강 공장의 공정 흐름은 일반적으로 코일 형태의 원재료를 가져와 두께와 폭, 철의 강도 등을 고객의 주문에 맞추어 가공하는 형태이다. 일반적인 공정의 흐름은 표면의 이물질 등을 제거하는 산세라인(PPL), 주문 형태에 맞게 폭을 절단 하는 전단라인(WSL, SSL), 코일의 두께를 줄이는 냉간압연(CRM) 및 강도 변경을 위한 소둔(ANN)공정으로 이루어지며, 가공 중 표면 및 형상 처리를 위한 스킨패스(SPL)공정을 거쳐 최종 고객에게 납품하기 위한 패키지(PRD, PKG) 단계를 거치게 된다.

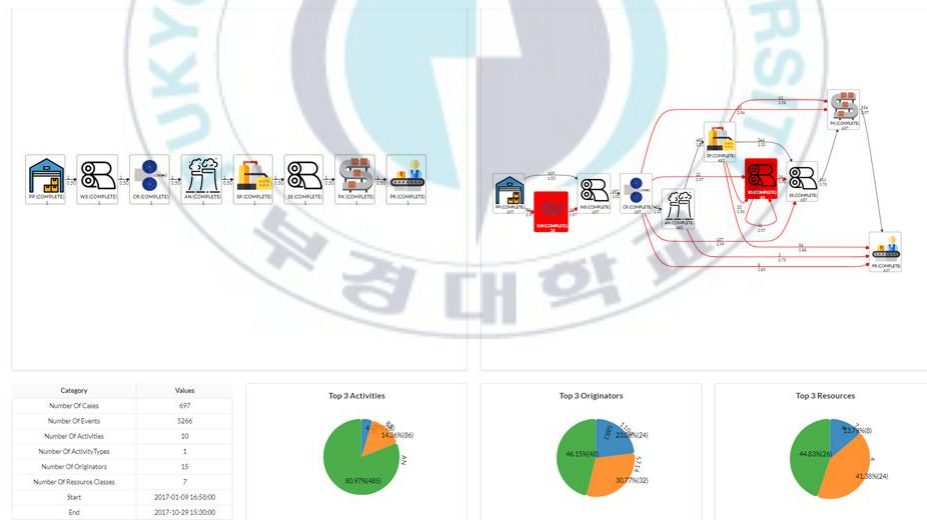
사례 분석은 두 가지 제조 제품을 선정하여 생산 공정에 대한 분석을 진행하였으며 분석한 데이터는 <표 5-1>과 같다.

<표 5-1> 분석 데이터 소개

구분	계
코일 수	697(개)
이벤트 수	5,266(개)
공정 수	10(개)
장비	7(개)
제품 수	2(개)
시작시간	2017/01/09
종료시간	2017/10/29

아래 <그림 5-1>은 <표 5-1>의 분석 데이터를 이용하여 생산 공정 흐름을 도출한 결과이다.

<그림 5-1> 공정 흐름 도출



좌측은 최초 제품의 생산 계획을 통하여 설계된 공정이고 우측은 수집된 이벤트 데이터를 통하여 도출한 공정 모델이다. 분석 데이터를 통하여 도출된 공정에서는 설계되지 않은 2개의 공정(RW, RS)가 발생하였으며 총

697개의 코일 중 114개의 코일(RW:28, RW:86)이 설계되지 않은 공정을 거친 것으로 나타났으며 붉게 표시 된 것과 같이 실제 공정 흐름에서도 정해진 순서가 아닌 다른 형태의 흐름을 보여 주고 있다. 이러한 공정 흐름의 분석을 통하여 공정의 재설계 또는 특정 공정에 집중되는 집중도 분석 등을 통하여 공정의 흐름을 개선할 수 있다.

시스템 도입 전과 도입 후의 실제 개선 수준을 비교하기 위하여 <표 5-2>의 데이터 정보를 사용하여 분석을 진행 하였다.

<표 5-2> 도입 전 후 분석 대상 데이터 정보

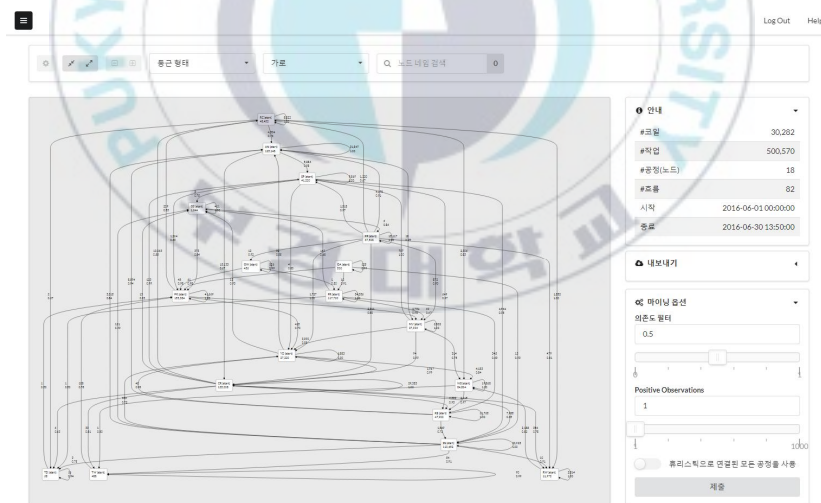
구분	도입전	도입후
코일 수	5,566(개)	30,282(개)
작업 수	84,596(회)	500,570(회)
시작시간	2014/11/24	2016/06/01
종료시간	2014/12/22	2016/06/30

기초 공정 흐름과 관련하여 도입 전 후를 비교하면 기본 발생 데이터가 5.98배가량 증가하였으며, 작업 코일의 양은 5.44배 증가하였다. 이는 도입 전 데이터가 A공장만 산출되었던 것에 비해 도입 이후 B공장이 시스템에 포함되며, 생산되는 제품의 종류와 공정이 증가하며 발생된 현상으로 분석되었음. 5.44배가량의 작업 코일의 증가에도 불구하고, 코일 당 작업 수는 1.08배 증가(15.19 -> 16.53)에 그쳐 공정 흐름을 상대적으로 개선한 것으로 판단된다.

<그림 5-2> 기초 공정 흐름 분석(도입 전)



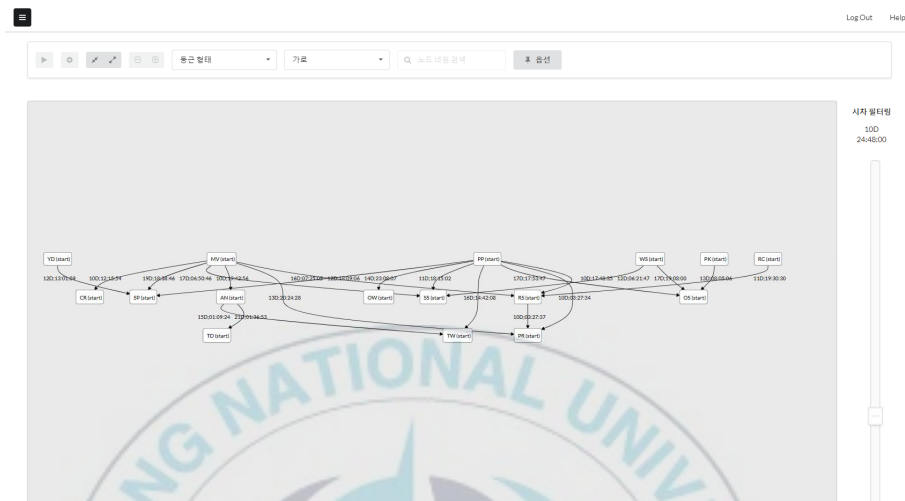
<그림 5-3> 기초 공정 흐름 분석(도입 후)



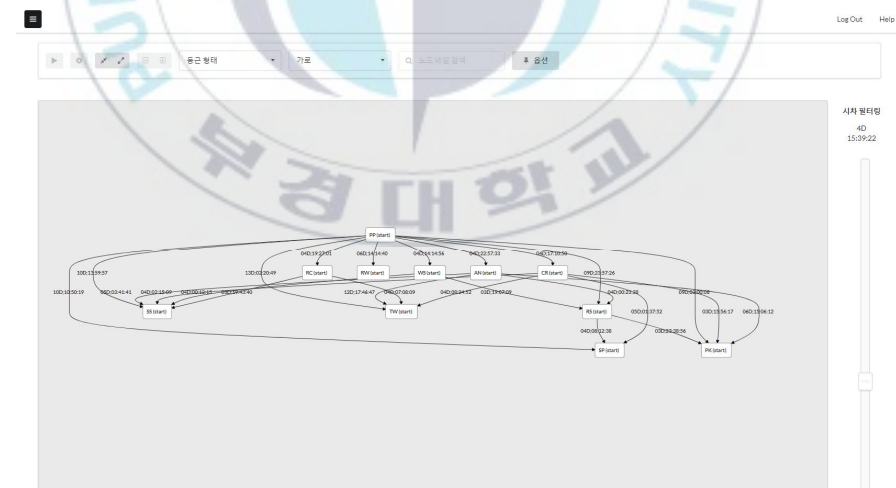
평균 재공 시간 감소에 대한 분석은 도입 전 평균 공정 재공 시간이 12.88일에서 도입 후 14.71일로 1.9일 가량 증가하였으나, 생산 프로세스 및 수집 데이터와 생산량 증가가 약 6개가량임을 감안 한다면 제품 당 평균

재공 시간의 변화는 개선된 것으로 분석된다.

<그림 5-4> 평균 재공 시간(도입 전)



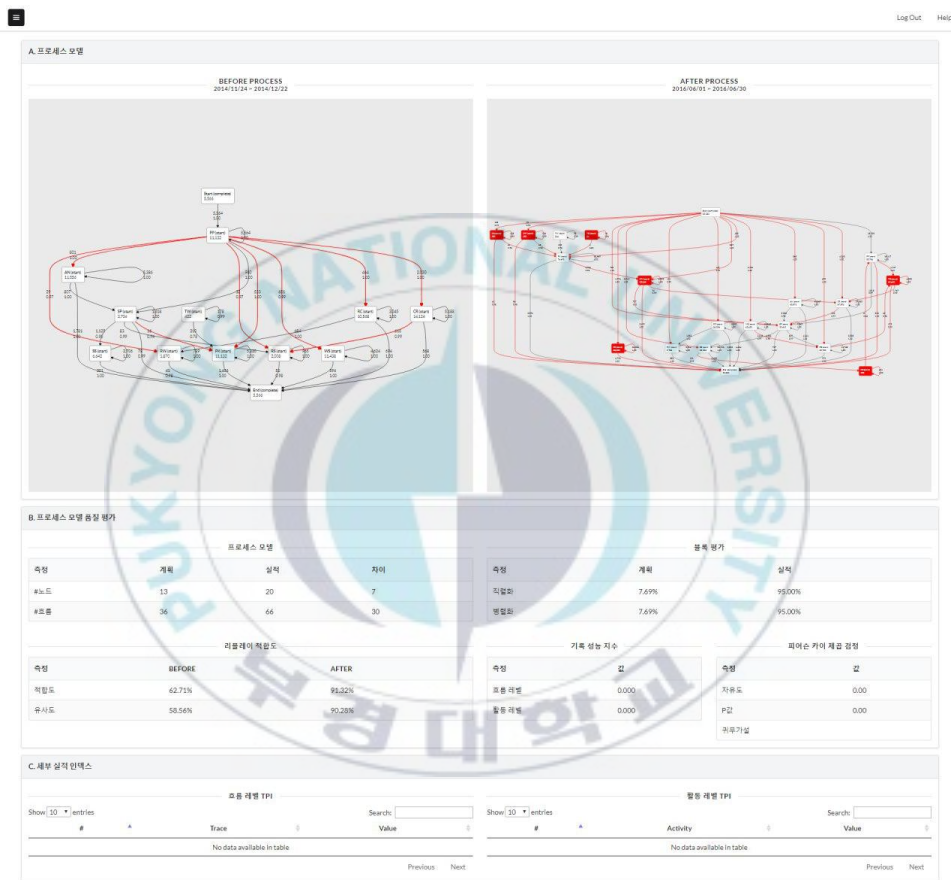
<그림 5-5> 평균 재공 시간(도입 후)



실 공정 데이터와의 유사도와 공정 흐름 모델의 적합도 에서 도입 전 적합도 62.7%, 유사도 58.56% 이었으며, 도입 후 적합도는 91.32%, 유사

도는 90.28%로 전후를 비교하였을 때, 적합도는 28.62%, 유사도는 31.72% 향상되었으며, 이는 도입 전 공정 생산 계획과 실적데이터가 상당 부분 일치 되면서 적합도와 유사도가 개선된 것으로 보인다.

<그림 5-6> 실공정 데이터 유사도 및 공정 흐름 모델 적합도



VI. 결론

1. 연구 결과의 요약

본 프로젝트 보고서의 핵심은 제조업의 제조 생산 과정에서 발생하는 빅 데이터를 효과적으로 분석하여 제조 공정의 문제를 파악하여 개선할 수 있는 의사결정을 지원 할 수 있는 시스템 개발에 초점을 두었다. 본 시스템을 통하여 기존 프로세스 마이닝 분석 도구들의 한계점인 타 시스템과의 연계 및 실시간 데이터를 통한 분석의 한계를 극복 할 수 있을 것이다.

시스템을 통한 사례 기업의 경우 기말 재공량을 현재 기말 재고량 대비 약 10% 정도 감소시킬 수 있으며 이에 따른 RUST발생에 따른 추가 공정 비용을 절감 할 수 있게 될 것이다. 평균 작업 시간 또한 현재 보다 줄일 수 있게 될 것으로 보이며 이러한 모든 것이 월 고정자산, 작업대기 및 지연시간, 추가 공정비용 절감 등을 통하여 기업의 매출 이익 증대에 많은 역할을 하게 될 것이다.

또한 다양한 Factory Thing으로부터 수집된 빅 데이터를 빠른 시간에 분석하고 실시간 공정 개선에 대한 분석이 가능한 시스템을 구현하는데 목표를 두었다. 공정 개선을 위한 근거가 정성적 판단에 기인한 것이 아닌 정량적 분석에 기인할 수 있도록 관련 기능을 지원하며 공정 개선을 위한 분석에 소요되는 시간 또한 각종 장비 및 시스템에서 실시간으로 수집하는 것을 통해 단축할 수 있다.

2. 연구의 한계점과 향후 연구의 방향

본 연구의 한계점과 향후 연구방향은 다음과 같다. 첫째, 서론에서 언급

한 수집 데이터의 투명성이 무엇보다 중요하다. 사례 분석에 활용된 데이터는 일부 Key-In을 통하여 수집된 데이터를 통하여 일부 오류를 발생시키고 있다. 이러한 문제의 해결은 발전하고 있는 각종 센서들을 통한 수집이 이루어 질 수 있도록 제조 현장의 변화가 필요하다. 둘째, 공정 개선 모델에 대한 적합도에 대한 다양한 정보를 제시하여 의사 결정 지원에 조금 더 편하게 접근할 수 있도록 시스템의 보완이 필요하다. 셋째, 실시간 공정 분석 시 시스템 리소스 사용에 대한 비용을 감소시킬 수 있는 In-Memory Data Grid 및 SQL Query Engine등의 아키텍처의 솔루션 변경 또는 보완이 필요하다. 마지막으로 실시간 스트리밍 데이터 수집 및 처리 등에 필요한 아키텍처의 보완이 필요하다.

솔루션의 활용도와 데이터의 신뢰성을 확보하여 생산 계획의 정확도가 높아진다면 생산 계획 예측의 성능을 높일 수 있는 머신러닝 / AI 기술을 현장에 폭넓게 적용하여 생산계획이나 설비 문제점 탐지 같은 공정 개선 부분에서 높은 정확도를 얻을 수 있을 것이다.

참고 문헌

1. 국내 문헌

과학기술일자리진흥원(2018), “스마트팩토리 산업 및 시장동향” S&T Market Report, 56.

이원창, 이이섭 (2016). “빅데이터 기반의 스마트 팩토리과 CEP를 결합한 실시간 분석시스템 개발.” 한국기술혁신학회 학술대회, 250-254.

배혜림 외 9명(2016), “스마트팩토리를 위한 운영빅데이터 분석 플랫폼,” 한국빅데이터학회지, 1(2), 9-19.

이동하 (2014), “프로세스 마이닝을 이용한 조선 산업의 생산 공정 분석”, 부산대학교 박사학위 논문.

전대욱, 배혜림 (2013), “프로세스 마이닝을 이용한 항만 터미널 환경에서의 위치 기반 컨테이너 이동에 관한 분석”, 대한산업공학회 추계학술대회 논문집, 869-883.

강영식, 강창재, 손동욱, 박형진, 허동한, 서필교(2012), “프로세스 마이닝 기반의 프로세스 혁신 지식 발견과 진단 : 전자결재 업무를 중심으로”, 산업혁신연구, 제28권, 제1호, 55-75.

강영식, 강창재, 이은상, 신철규(2013), “빅 데이터 환경에서 프로세스 마이닝의 감사 적용에 관한 연구”, 감사논집, 제20호, 5-35.

손숙영, 문현실, 김재경, 송민석(2012), “프로세스 마이닝을 이용한 전시행

사 분석”, 대한산업공학회 추계학술대회논문집, 131-137.

최상현, 한관희, 임건훈(2013), “프로세스 마이닝 기법을 활용한 고장 수리 프로세스 분석”, 한국콘텐츠학회논문지, 제13권, 제4호, 399-406.

2. 해외 문헌

Mario Hermann, Tobias Pentek, Boris Otto.(2016), “Design Principles for Industrie 4.0 Scenarios”, Hawaii International Conference on System Sciences, 49, 3928-3937.

John Gantz and David Reinsel.(2012) “THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East”, EMC Corporation.

Wil M. P. van der Aalst.(2011), “Process Mining: Discovery, Conformance and Enhancement of Business Processes”, Springer, 2011

Process Mining Group Math&CS department, Eindhoven University of Technology ProM, Process mining research tools application. <http://www.processmining.org/prom/start>, Accessed 14 May 2015.

Rozinat, A. and C.W. Gunther, Fluxicon Process Laboratories, Disco Fluxicon. <http://fluxicon.com/disco/>, Accessed 15 May 2015.

Jansen-Vullers, M.H., W.M.P. van der Aalst, and M. Rosemann(2006), “Mining configurable enterprise information systems, ”Data and Knowledge Engineering, Vol.56, No.3, 195-244.

Rozinat, A., W.M.P. van der Aalst(2006), Decision mining in ProM, Business Process Management, Springer Berlin Heidelberg, 420-425.

Van der Aalst, W.M.P., A.K. Alves de Medeiros and A.J.M.M. Weijters(2007), "Genetic process mining : an experimental evaluation," Data Mining and Knowledge Discovery, Vol.14, No.2, 245-304.

Van Der Aalst, W.M.P., H.A. Reijers, M.Song(2005), "Discovering social networks from event logs," Computer Supported Cooperative Work, Vol.14, No.6, 549-593.

Van der Aalst, W.M.P., T. Basten(2002), "Inheritance of workflows : an approach to tackling problems related to change," Theoretical Computer Science, Vol.270, No.1, 125-203.

Van der Aalst, W.M.P., A. Adriansyah, A.K. Alves de Medeiros, A. Karla Alves, F.Arcieri, T. Baier, T. Blickle, J.C. Bose, P. van den Brand, R. Brandtjen, J. Buijs(2012), "Process mining manifesto," Business process management workshops, Lecture Note on Business Information Processing , 169-194.

Wang, Y., F. Caron, J. Vanthienen, L. Huang, Y. Guo(2014), "Acquiring logistics process intelligence : Methodology and an application for a Chinese bulk port," Expert Systems with Applications, Vol.41, No.1, 195-209.

Van der Aalst, W.M.P., H.A. Reijers, A. J. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H. Verbeek(2007), "Business process

mining : An industrial application,” Information Systems, Vol.32, No.5, 713-732.

Mans, R. M. Schonenberg, M. Song, W.M.P. van der Aalst., P.J. Bakker(2009), “Application of process mining in healthcare.a case study in a dutch hospital,” Biomedical Engineering Systems and Technologies, 425-438

Song, M., W.M.P. van der Aalst(2008), “Towards comprehensive support for organizational mining,” Decision Support Systems, Vol.46, No.1, 300-317.

Rozinat, A., I.S. de Jong, C. Gunther, W.M.P. van der Aalst(2009), “Process mining applied to the test process of wafer scanners in ASML,” Systems, Man, and Cybernetics, Part C : Applications and Reviews, Institute of Electrical and Electronics Engineers Transactions, Vol.39, No.4, 474-479.

Goedertier, S., J. D. Weerdt, D. Martens, J. Vanthienen, and B. Baesens(2011), “Process discovery in event logs : An application in the telecom industry,” Applied Soft Computing, Vol.11, No.2, 1697-1710.

Jans, M., V.D. Werf, J. Martijn, N. Lybaert, and K. Vanhoof(2011), “A business process mining application for internal transaction fraud mitigation,” Expert Systems with Applications, Vol.38, No.10, 13351-13359.

감사의 글

이 석사 학위 논문을 작성하는데 주변의 많은 지원과 독려와 가르침이 있었습니다. 바쁘신 가운데에도 항상 배려와 격려로 지도하여 주신 김민수 지도교수님께 먼저 진심으로 감사를 드립니다.

학업과 업무를 병행하느라 중도에 힘든 순간도 많았으나, 항상 응원해준 가족들에게도 감사의 마음을 전합니다.

다섯 학기 동안 많은 새로운 과목을 접하며 새로운 지식을 습득하고 새로운 현장 경험들을 전달해 주신 기술경영전문대학원 모든 교수님께 감사드립니다.

마지막으로, 학업과 업무를 병행하는 데 많은 도움을 주신 직장 동료들에게 감사드리며, 감사의 글을 마칩니다.

2018년 6월

원 석 래