

## 원전 상태 감시 및 조기 경보용 빅데이터 시범 플랫폼의 설계를 위한 시스템 엔지니어링 방법론 적용 연구

차재민<sup>1)\*</sup> 손충연<sup>2)</sup> 황동식<sup>2)</sup> 신중욱<sup>1)</sup> 염충섭<sup>1)</sup>

1) 고등기술연구원, 2) ㈜엠엔디

### A Study on the Application of SE Approach to the Design of Health Monitoring Pilot Platform utilizing Big Data in the Nuclear Power Plant (NPP)

Jae-Min Cha<sup>1)\*</sup>, Junguk Shin<sup>1)</sup>, Choong-Yeon Son<sup>2)</sup>, Dong-Sik Hwang<sup>2)</sup>, Choong Sub Yeom<sup>1)</sup>

1) Institute for Advanced Engineering (IAE)

2) MND

**Abstract** : With the era of big data, the big data has been expected to have a large impact in the NPP safety areas. Although high interests of the big data for the NPP safety, only a limited researches concerning this issue are revealed. Especially, researches on the logical/physical structure and systematic design methods for the big data platform for the NPP safety were not dealt with. In this research, we design a new big data pilot platform for the NPP safety especially focusing on health monitoring and early warning services. For this, we propose a tailored design process based on SE approaches to manage inherent high complexities of the platform design. The proposed design process is consist of several steps from elicitate stakeholders to integration test via define operational concept and scenarios, and system requirements, design a conceptual functional architecture, select alternative physical modules for the derived functions and assess the applicability of the alternative modules, design a conceptual physical architecture, implement and integrate the

---

\* 교신저자 : Jae-Min Cha, [jmcha@iae.re.kr](mailto:jmcha@iae.re.kr)

\* This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

physical modules. From the design process, this paper covers until the conceptual physical architecture design. In the following paper, the rest of the design process and results of the field test will be shown.

**Key Words** : Nuclear Power Plant, Big Data, Early Warning, Platform, Systems Engineering

## 1. 서론

원자력 발전소(이하 원전)는 석탄 화력 발전에 비하여 운전 시 오염물질 배출이 적고 발전 단가가 저렴하여 2013년 기준 국내 총 설비용량의 25.3%를 차지하는 중요한 설비이다(Korea Atomic Industrial Forum, 2014). 원전은 이러한 경제성에도 불구하고, 체르노빌 및 후쿠시마 사고에서 알 수 있듯이 사고 발생 시 인간과 환경에 치명적인 피해를 입힐 수 있기 때문에 안전성이 매우 중요하다. 원전 산업은 원전 안전성을 확보하기 위해 전통적으로 계획 예방정비(Preventative Maintenance)를 수행해 왔으며, 최근에는 원전에 설치된 현장데이터 수집망으로서 PI 시스템\*의 도입 확대에 따라 현장 데이터 확보가 가능해지면서 온라인 모니터링을 통한 상태 감시와 조기 경보 등을 수행해 오고 있다.

한편, 최근 Information Technology(IT) 기술의 발달과 함께 데이터의 생산이 폭발적으로 증가하면서, 데이터로부터 가치 있는 정보를 도출하는 빅데이터(Big data)가 세계적으로 주목을 받으며(Sagiroglu&Sinanc, 2013), 원전으로의 적용에 대한 관심 역시 높아지고 있다. 빅데이터는 초기에 주로 유통, 금융, 통신, 의료, 제조 산업 분야로 적용되고 있었지만, 최근에는 석유/정유, 철강 등 플랜트 산업에서도 빅데이터를 적용하고자 하는 시도가 보고되고 있다(Alam et al., 2014; Baaziz&Quoniam, 2014; Brul, 2013; Cameron, 2014; Feblowitz, 2012; Hems et al., 2013; Seshadri, 2013). 하지만, 이러한 플랜트 산업에서의 빅데이터에 대한 관심 확대에도 불구하고, 아직까지는 원전에서 빅데이터 기술을 적용한 시도는 거의 찾아보기 어려운 실정이다.

구체적으로, 원전에 적용되고 있는 상태 감시 및

조기 경보 시스템 측면에서 보면 SignalMate, PHI, EPI\*Center, SureSense 등 국내외에서 개발된 시스템이 적용되고 있으나, 대용량(Volume), 빠른 속도(Velocity), 다양성(Variety) 등의 특성을 갖는 빅데이터(Gantz&Reinsel, 2011)를 수집, 저장, 처리, 분석, 시각화 등의 전주기 활동을 지원하기 위한 빅데이터 기술을 적용한 시스템은 찾아보기 어렵다(Kang, 2015). Table 1은 국내외 상태 감시 및 조기 경보 시스템 현황을 보인다. 또한, 빅데이터의 전주기 활동을 지원하는 세부 기능 지원 SW 모듈들의 통합 시스템인 빅데이터 플랫폼 측면에서 보면, 다양한 산업 분야에 범용적으로 적용할 수 있는 일반적인 빅데이터 플랫폼은 여러 기업과 연구자들에 의해 보고되고 있으나(Cloudera, 2015; Hortonworks, 2015; MapR, 2015; NIST, 2014), 원전의 상태 감시 및 조기 경보 분야의 특성에 적합하도록 개발된 빅데이터 플랫폼은 아직까지 찾아보기 어려운 상황이다.

본 연구에서는 높은 관심에도 불구하고 아직까지 국내에서 개발된 적이 없는 원전 상태 감시 및 조기 경보를 위한 원전 빅데이터 시범 플랫폼(이하 시범 플랫폼)을 개발한다. 구체적으로, 시범 플랫폼은 최근 빠르게 개발되고 있는 빅데이터 관련 기술들을 원전의 상태 감시 및 조기 경보 분야에 적용하는 것이며, 이를 통해 기존 원전 내 상태 감시 및 조기 경보 시스템이 처리하기 힘들었던 실시간/대용량/다양성을 가진 원전 빅데이터로부터 보다 신속하고 정확하게 위험 상황을 경고함으로써, 원전 고장과 사고를 예방하는 것이다.

본 연구의 대상인 시범 플랫폼은 원전 상태 감시 및 조기 경보 분야와 빅데이터 기술 분야에 대한 통

\* 오피레이션 인텔리전스의 글로벌 리더인 OSIsoft의 데이터 수집 및 분배, 분석 제품

<Table 1> Current status of domestic/international condition monitoring and early warning system (Kang, 2015)

제품명 기준	SignalMate	PHI	EPI*Center /Proficy	SureSense
제작사	한전전력연구원	BNF테크놀로지	SmartSignal/GE	Expert MicroSystem
국가	대한민국	대한민국	미국	미국
적용 기술	Data-driven	Data-driven	Data-driven	Data-driven
	신호분석기법	통계적기법	통계적기법	통계적기법
	상관계수 가중 AAKR	AAKR	MSET, SPRT, SBM	EESE 등 다수
상태감시	○	○	○	○
조기경보	○	○	○	○
빅데이터	X	X	○	X

**AAKR:** Auto-Associative Kernel Regression; **MSET:** Multivariate State Estimation Technique; **SPRT:** Sequential Probability Ratio Test; **SBM:** Similarity Based Modeling; **EESE:** Expert State Estimation Engine

합적인 지식이 요구될 뿐만 아니라, 개발에 참고할 수 있는 기존 시스템이나 참고 자료를 찾아보기 어렵기 때문에, 개발의 복잡성이 높은 시스템으로 간주될 수 있다. 따라서 이러한 복잡성을 가지는 시범 플랫폼의 개발을 체계적으로 지원하기 위한 접근 방법이 요구되며, 시스템 엔지니어링(Systems Engineering; SE)은 이러한 한계를 극복하기 위한 하나의 좋은 대안적인 접근방법이 될 수 있다. SE는 비행기, 자동차, 기차 등 복잡한 시스템을 개발함에 있어 고객의 요구를 만족시키는 시스템의 솔루션을 도출하고 검증하기 위한 체계적이고 종합적인 엔지니어링 접근 방법이다(INCOSE, 2015). 특히, 시스템 엔지니어링은 새로운 시스템 혹은 이전 개발 경험이 부족한 시스템이나, 다분야 기술 지식이 필요한 시스템의 개발에도 적합한 특성을 가진다고 알려져 있으며, 이러한 특성은 시범 플랫폼을 체계적으로 개발하는데 매우 적합하다. 따라서 본 연구에서는 시범 플랫폼의 체계적인 개발을 위해 SE 접근 방법에 기반하였다.

이어지는 본 논문의 구조는 다음과 같다. 제 2장에서는 빅데이터의 정의와 개념을 소개하고, 제 3장에서는 SE 접근 방법에 기반한 시범 플랫폼 설계 절차 및 결과물에 대하여 소개하고, 마지막으로 제 4장에서는 결론 및 향후 연구 계획을 보인다.

## 2. 빅데이터의 정의와 개념

빅데이터라는 용어는 지난 수년간 전 세계적으로 사용되어 왔지만, 빅데이터의 정의에 대해서는 제안자의 관점에 따라 서로 다르게 정의되고 있으며, 이는 빅데이터 개념의 모호성을 증가시키는 주요한 원인 중 하나로 간주되고 있다(Ward&Baker, 2013). 따라서 본 연구의 대상인 시범 플랫폼을 설계하기 위해서는 관련된 용어부터 명확하게 정의할 필요가 있다. 이에 본 장에서는 기존 문헌들이 제시하고 있는 빅데이터에 대한 개념으로부터 보다 구체적인 빅데이터 관련 용어를 정의하고, 이를 통해 빅데이터 플랫폼의 개념을 구체화 한다.

### 2.1 기존 문헌의 빅데이터의 정의

전술된 바와 같이 빅데이터에 대한 정의는 제안자의 관점에 따라 상이하다. 본 절에서는 이 중 여러 문헌들로부터 주로 인용되고 있는 Gartner (Beyer&Laney, 2012), McKinsey (Manyika et al., 2011), Oracle (Dijcks, 2012), IDC (Gantz&Reinsel, 2011)가 제시한 빅데이터의 정의를 보인다.

- 강화된 의사 결정, 통찰의 발견, 프로세스 최적화를 가능하도록 새로운 형태의 처리를 요구하는 대용량, 빠른 속도, 그리고(또는) 다양성을

갖는 정보 자산(Beyer&Laney, 2012)

- 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합(Manyika et al, 2011)
- 전통적인 관계형 데이터베이스 중심의 비즈니스 의사 결정과 함께 새로운 형태의 비구조화된 데이터로부터 가치를 도출하는 것(Dijcks, 2012)
- 매우 빠른 속도의 수집, 발견 그리고/혹은 발견을 가능하게 함으로써 다양한 형태의 매우 큰 용량의 데이터로부터 효율적으로 가치 있는 정보를 추출하기 위해 설계된 차세대 기술 및 아키텍처(Gantz&Reinsel, 2011)

전술된 정의를 살펴보면 Beyer&Laney(2012)와 Manyika et al.(2011)은 빅데이터를 데이터 관점으로 정의하고 있으며, Dijcks(2012)와 Gantz&Reinsel(2011)는 각각 분석 과정 및 기술 관점으로 정의하고 있음을 알 수 있다. 이런 다양한 관점의 정의는 사람들이 빅데이터를 이해하는데 있어서 모호성을 증가시키는 요인이 된다. 본 연구에서는 이러한 모호성을 줄이기 위해서, 빅데이터의 정의를 빅데이터, 빅데이터 분석, 빅데이터 기술로 구분하여 정의하였으며, 더하여 본 연구의 대상이 되는 빅데이터 플랫폼의 정의를 도출하였다. 이어지는 절에서는 이를 보인다.

## 2.2 제안된 빅데이터 관련 용어의 정의

빅데이터 정의의 모호성을 줄이기 위해 본 연구에서 3가지로 구분한 빅데이터 정의는 다음과 같다.

- **빅데이터**: 기존 저장, 분석, 처리 기술 수준으로는 처리하기 힘들었던 대용량, 고속 및 다양함의 특성을 동시에 혹은 일부 갖는 정보 자산
- **빅데이터 분석**: 기존 저장, 분석, 처리 기술 수준으로는 처리하기 힘들었던 대용량, 고속 및 다양성의 특성을 동시에 혹은 일부 갖는 빅데이터로부터 의미있는 가치를 도출하는 것
- **빅데이터 요소 기술**: 기존 저장, 분석, 처리 기

술 수준으로는 처리하기 힘들었던 대용량, 고속 및 다양성의 특성을 동시에 혹은 일부 갖는 빅데이터로부터 의미있는 가치를 도출하기 위하여 데이터의 수집에서부터 저장, 처리, 분석, 시각화, 관리 등 빅데이터 분석을 가능하게 하는 기술들로 통계적/비통계적 방법, 알고리즘, SW/HW 기술 등을 포함

- **빅데이터 플랫폼**: 기존 저장, 분석, 처리 기술 수준으로는 처리하기 힘들었던 대용량, 고속 및 다양성의 특성을 동시에 혹은 일부를 갖는 빅데이터의 수집, 저장, 처리, 관리, 분석, 시각화, 관리 등 빅데이터 분석을 지원하는 빅데이터 요소 기술들을 포함하는 상위 개념의 통합 시스템

기존 문헌들이 데이터/분석 과정/기술 등 각각의 관점에서 빅데이터를 정의한 것 대비하여 빅데이터 관련 정의를 구분함으로써 모호성을 줄일 수 있도록 하였으며, 빅데이터 관련 과제를 추진 중인 관련 전문가들로부터 빅데이터 개념에 대한 이해가 명료하여졌음을 확인할 수 있었다.

## 3. SE 접근 방법 기반 시범 플랫폼 설계

### 3.1 SE의 필요성

1장에서 시범 플랫폼 개발의 어려움에 대하여 간단히 서술하였지만, 본 절에서는 이를 좀 더 구체화한다. 시범 플랫폼 개발이 가지는 주요 어려움은 다음과 같다.

- **신규성**: 참고할 수 있는 기존 시스템이나 참고 자료를 찾아보기 어려운 새로운 시스템임
- **개발 방법의 부재**: 현재 빅데이터 플랫폼의 설계 및 개발은 주로 선진 빅데이터 플랫폼의 구조를 바탕으로 일부 기능을 추가 및 삭제하고 있으며, 이에 대한 체계적인 설계 및 개발 방법을 찾아보기 어려움
- **다학제성**: 시범 플랫폼은 원전 상태 감시 및 조기 경보 분야와 빅데이터 기술 분야에 대한 통합적인 지식이 요구됨

이러한 특징은 시범 플랫폼의 개발을 복잡하게 만드는 요인이 되며, 뿐만 아니라 시범 플랫폼이 성공적으로 개발되기 위해서는 개발 초기에서부터 플랫폼에 관련된 이해당사자로부터 관련 요구사항을 수집 및 정의하여 개발 후방 단계에서 발생할 수 있는 설계 변경을 최소화 할 필요가 있다. 따라서 복잡한 시범 플랫폼의 개발을 성공적으로 수행하기 위한 접근 방법이 요구되며, 본 연구에서는 이를 위해 복잡한 시스템을 개발함에 있어 고객의 요구를 만족시키는 시스템의 솔루션을 도출하고 검증하기 위한 체계적이고 종합적인 엔지니어링 접근 방법인 SE를 활용하였다.

### 3.2 시범 플랫폼 설계를 위해 적용된 SE 접근 방법

SE는 시스템 개발을 체계적으로 수행하는 체계화 된 접근 방법이자 프로세스이며, 이미 ANSI/EIA-632(2003), IEEE 1220(2005), ISO/IEC 15288(2015)와 같은 국제표준 등을 통해서 이를 구체화하고 있다. 하지만, 이러한 국제표준에서 제시하고 있는 프로세스와 방법은 시범 플랫폼의 개발에 적용하기에는 투입되는 시간과 비용대비 효과성이 낮을 수 있다. 특히, 시범 플랫폼은 상용화된 제품으로 원전 현장에 적용하기에 앞서, 적용 가능성을 확인하고자 하는 특성이 강하므로, 기본적인 SE 접근 방법의 특성을 반영하면서 동시에 간소화 될 필요가 있다.

이에 우리는 원전 분야, 빅데이터 분야, SE 분야의 전문가들과 함께 시범 플랫폼 개발을 위한 테일러링(Tailoring) 된 SE 접근 방법 기반의 개발 프로세스를 정의하였다. Figure 1은 제안된 개발 프로세스를 보인다.

Figure 1에서 보이는 바와 같이 이해당사자의 도출부터 요구사항 기능 및 물리 아키텍처 설계, 구현 통합 및 테스트에 이르는 기본적인 SE 접근 방법을 적용하였으며, 세부 프로세스 별 수행 방법 역시 SE에서 사용되는 다양한 방법 등을 적용하였다. 참고로, 빨간 점선으로 표시된 영역은 본 논문에서 다



[Figure 1] SE based design process for pilot platform

루고 있는 범위이며, 나머지 하단 부분은 차후 논문에서 다룰 예정이다.

### 3.3 SE 접근 방법에 기반한 시범 플랫폼 설계

#### 3.3.1 이해당사자 도출 및 정의

시범 플랫폼이 어떠한 모습이 되어야 할지에 대한 요구사항을 도출하기 위해, 우선 시범 플랫폼과 이해관계를 가지고 있는 핵심 주요 이해당사자를 도출하였다. 구체적으로 시범 플랫폼과 관련된 다양한 이해당사자를 고려하기 위하여 설계/개발, 운영, 유지보수 단계에 이르는 시스템 수명주기를 구분하고 각 단계별로 관련된 주요 이해당사자 리스트를 도출하였다. Table 2는 도출된 주요 이해당사자의 예를 나타내었다.

#### 3.3.2 운영 개념 정의

흔히 시스템의 개발에 있어 시스템이 모두 개발된 이후 혹은 개발 도중에 주요 이해당사자로부터 당초에 원하는 시스템과는 다른 모습이라는 피드백을 받고, 개발 사양이 변경됨으로 인해 시간과 비용이 지연되는 일이 발생한다. 뿐만 아니라, 시스템을 실제 설계 및 개발하는 설계자/개발자들 역시 시스템이 개발된 이후 어떻게 운영될 것인지에 대한 공동의 비전이 부재하여 개발 과정상에 있어서 효율

<Table 2> Derived stakeholders

시스템 수명주기	주요 이해당사자	이해관계
플랫폼 설계/개발 단계	플랫폼 사업자 CEO	플랫폼의 설계/개발에 관련된 기능과 성능을 결정하는 최종 의사결정권자
	플랫폼 설계/개발자	플랫폼의 설계/개발을 실무적으로 담당하고 실행하는 자
플랫폼 운영 단계	플랫폼 설치자	플랫폼을 원전에 설치하여 약속된 기능과 성능으로 작동하게 하는 자
	플랫폼 설치 검수자	플랫폼을 원전에 설치할 때 약속된 기능과 성능으로 작동하는지 최종 확인하는 자
	플랫폼 사용자 (원전 운전자)	플랫폼을 원전에 도입 및 운영하여 조기 이상상태 진단 서비스를 받는 자
플랫폼 유지보수 단계	플랫폼 유지보수자	플랫폼의 기능과 성능이 공급 계약 조건을 만족할 수 있도록 정기적/비정기적으로 유지 보수하는 자

성이 떨어지고는 한다. 이러한 문제를 예방하기 위해서는 시스템의 설계에 들어가기에 앞서 시스템이 완성된 다음 어떻게 운영될 지에 대한 공통의 개념인 운영 개념을 정의하는 것이 필요하다.

본 연구에서는 현재 원전 내 시범 플랫폼과 유사 기능을 수행하고 있는 시스템(As-Is system)은 어떻게 운영되고 있고, 어떠한 한계점이 있는지 파악하기 위하여, 도출된 주요 이해당사자를 대상으로 현재 운영 개념(As-Is Operational Concept)과 향후 개발이 완료된 후 어떻게 플랫폼이 운영될 것인지에 대하여 표현한 미래 운영 개념(To-Be Operational Concept)을 도출하였다. 더하여, 일관성 있는 운영 개념의 도출을 위해 미국 운영개념서 가이드(BSR/AIAA G-043A-201, 1992)에서 제시된 5W1H(Who, What, Where, When, Why, and How) 질문법을 통해 운영 개념을 도출하였다. 5W1H는 6가지 관점의 핵심적인 질문을 통해 운영 개념을 도출하는 직관적인 방법이다. Figure 2는 운영개념 도출을 위한 5W1H를 보인다.

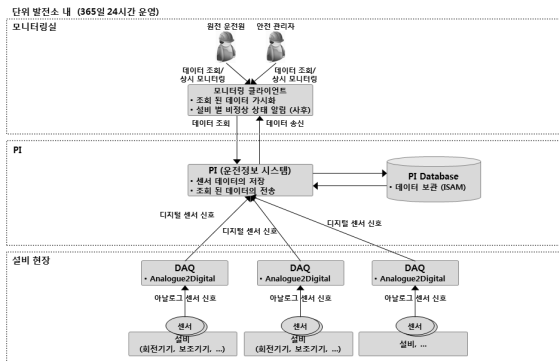
원전 현장 유지보수 경험이 있는 전문가를 대상으로 현재 원전 내 설비의 상태 모니터링 및 이상상

**Who:** Who are the stakeholders involved with the system?  
**What:** What are the elements and the high-level capabilities of the system?  
**Where:** What is geographic and physical extent of the system?  
**When:** What is the sequence of activities that will be performed?  
**Why:** What is the problem or opportunity addressed by the system?  
**How:** How will the system be developed, operated, and maintained?

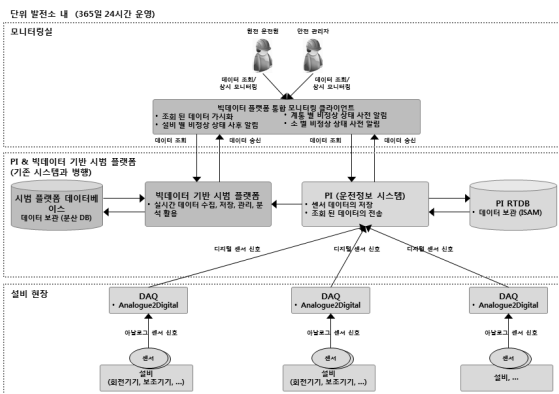
[Figure 2] 5W1H for operational concept derivation

태 진단에 관한 현황을 도출하였다. 설비에 달려 있는 각종 센서들로부터 DAQs(Data Acquisition System)를 거쳐 PI 시스템으로 정보가 수집되고, 수집된 데이터는 대용량의 데이터베이스에 저장된다. 원전 운전원, 안전 관리자 등은 PI 시스템에 원격으로 접속되어 있는 모니터링 클라이언트를 통해 PI 시스템에 접속하여 필요한 데이터를 조회하고 상시 모니터링하며, 설비 별로 이상상태가 발생하는 것을 파악할 수 있다.

앞서 도출된 현재 운영 개념을 바탕으로, 현재 시스템이 장기 데이터 분석 및 처리, 계통 별/소 별 종합적 데이터 분석 및 처리, 실시간 데이터 분석 및 처리, 반정형/비정형 데이터 등 다양한 데이터 분석 및 처리 등의 측면에서 한계가 존재하며, 개선의 여지가 있음을 확인하였다. 이를 바탕으로, 향후 시범 플랫폼이 적용되었을 때 어떻게 운영될지에 대한 미래 운영 개념(To-Be Operational Concept)을 도출하였다. 미래 운영 개념서는 현재 운영되고 있는 시스템에 빅데이터 시범 플랫폼을 연계함으로써, 현재 운영 개념에서 한계로 파악되는 대용량 데이터 분석 및 처리, 실시간 데이터 분석 및 처리, 다양한 데이터 분석 및 처리 역량을 높이고, 현재 운영되고 있는 시스템 대비 설비 및 시스템의 이상 상태를 보다 조기에 감지할 수 있게 한다. Figure 3과 4는 도출된 현재 및 미래 운영 개념을 보인다.



[Figure 3] As-Is Operational concept



[Figure 4] To-Be Operational concept

### 3.3.3 시스템 요구사항 정의

도출 된 미래 운영 개념을 실현하기 위한 플랫폼을 성공적으로 개발 및 구현하기 위해서는 이해당사자로부터 시스템이 가져야 하는 기능, 성능 및 제약에 관련된 시스템 요구사항을 도출 및 정의할 필요가 있다. 이를 위하여, 시범 플랫폼의 설계 및 개발에 참여하고 있는 본 연구의 참여기관 전문가들과 원전 운전 및 유지보수 경험을 보유하고 있는 전문가들을 대상으로 시스템 요구사항을 도출하고자 하였다. 하지만, 시범 플랫폼의 경우 아직까지 전례를 찾아보기가 쉽지 않은 새로운 시스템으로서 이해당사자가 시범 플랫폼에 대한 이해가 어려워 명확한 요구사항을 도출하는데 어려움을 겪었다. 본 문제를 해결하기 위하여, 본 연구에서는 시범 플랫폼이 가질 수 있는 시스템 수준의 세부 기능을 정의하고, 이를 바탕으로 전문가와 함께 운영 개념을 상세화하는 운영 시나리오를 도출하였으며, 시나리오

도출 과정에서 식별된 요구사항을 바탕으로 시스템 요구사항을 정의하였다.

#### 3.3.3.1 시스템 수준 공통 기능 도출

일반적으로 빅데이터 플랫폼 시장은 Cloudera, Hortonworks, MapR 등과 같은 특정 플랫폼들을 기반으로 개발되고 있다. 하지만, 이러한 빅데이터 플랫폼들은 주로 IT, 통신, 금융, 물류 산업 등을 중심으로 한 아키텍처에 초점이 맞추어져 있을 뿐만 아니라, 이러한 특정 플랫폼에 한정될 경우 플랫폼의 설계 변경에 대한 유연성과 추후 기능 추가를 위한 확장성에 제약이 생긴다. 따라서, 본 연구에서는 특정 빅데이터 플랫폼 솔루션에 국한되지 않으면서, 차후 다양한 분야로의 확장 적용이 가능한 플랫폼의 설계를 위하여, 현재 개발되었거나 개발 진행 중에 있는 빅데이터 플랫폼 관련 국내외의 주요 표준/문헌을 분석하여 공통적인 플랫폼 기능을 도출하였다.

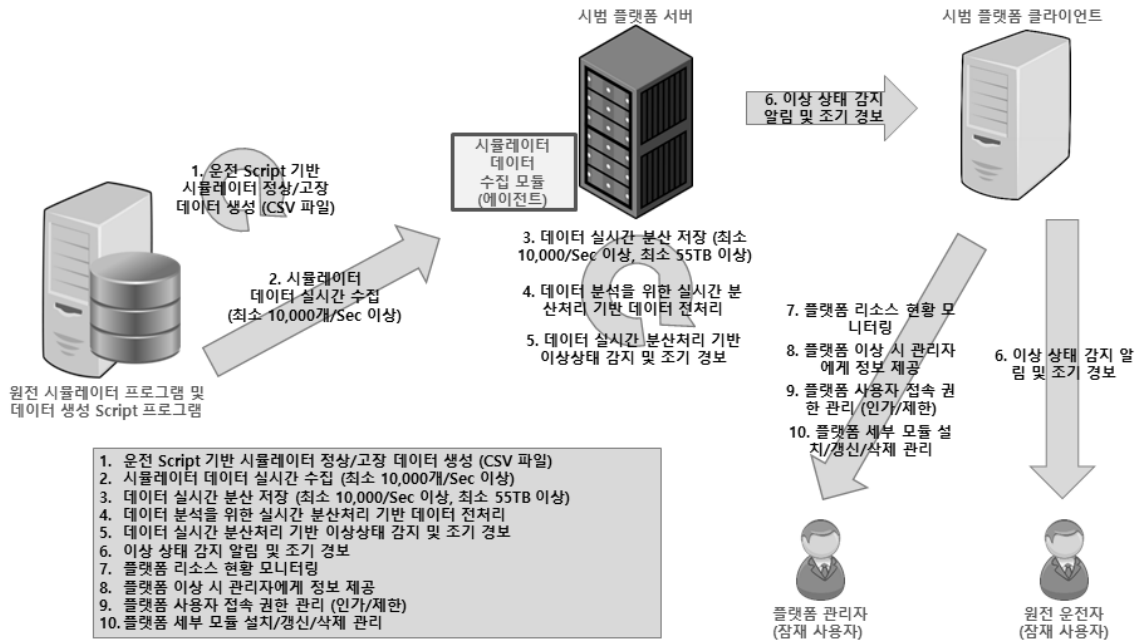
본 연구에서는 한국정보화진흥원(NIA), 한국정보통신기술협회(TTA), 미국립표준기술연구소(NIST)에서 보고되고 있는 6건의 공개된 주요 자료를 참고 자료로 선정하였으며(NIA, 2013; NIA, 2014; NIST, 2014; TTA.KO-10.0700, 2013; TTA.KO-10.0705, 2013; TTA.KO-10.0778, 2014), 각 자료의 분석을 통해 각 자료에서 제안하고 있는 빅데이터 플랫폼이 가지는 기능을 일관된 기준으로 정리하고, 이를 같은 기능별로 정렬함으로써, 공통 기능을 도출하였다. 최종적으로 도출 된 6개의 주요 공통 기능은 데이터 수집, 데이터 저장/추출, 데이터 처리, 데이터 분석, 시각화, 플랫폼 관리이며, Figure 5와 같다.

#### 3.3.3.2 운영 시나리오 정의

도출된 시범 플랫폼의 6개 주요 공통 기능을 바탕으로 다양한 전문가들과 함께 운영 개념을 보다 상세화하는 시범 플랫폼 운영 시나리오를 도출하였다. 도출된 시나리오는 원전 시뮬레이터로부터 실시간으로 생성되는 빅데이터를 실시간으로 수집, 저장, 처리, 분석 과정을 통해 잠재 사용자인 원전 운

NIA (2013)	빅데이터 수집	빅데이터 저장	빅데이터 처리	빅데이터 분석	빅데이터 시각화		
TTAK.KO-10.0700 (2013)	빅데이터 수집	빅데이터 저장·관리	빅데이터 처리	빅데이터 분석		빅데이터 시스템 관리	
TTAK.KO-10.0705 (2013)	데이터 수집	데이터 저장 및 관리	데이터 처리	데이터 분석 및 시각화		시스템 관리	빌링/서비스 품질 관리
TTAK.KO-10.0778 (2014)	데이터 수집	데이터 처리 및 저장 관리		데이터 분석	데이터 응용/시각화		
NIA (2014)	데이터 수집	데이터 저장·관리 (데이터 처리/저장/보안관리/품질관리)		데이터 분석			데이터 제공·이용
NIST (2014)	데이터 수집		데이터 준비	데이터 분석	데이터 시각화		데이터 접속
	데이터 수집	데이터 저장/추출	데이터 처리	데이터 분석	시각화	플랫폼 관리	

[Figure 5] Derivation of common platform system functions of pilot platform



[Figure 6] Operation Scenario

전자와 플랫폼 관리자에게 이상 상태 감지 알림 및 조기 경보와 플랫폼 관리 기능을 제공하는 세부 활동들로 구성된다. 참고로, 원전 현장에서의 데이터 수집의 보안성이 높아 확보가 어렵기 때문에, 본 연구에서는 원전 현장의 데이터와 유사한 수준의 실시간 데이터를 확보할 수 있는 Western Services Corporation (WSC)사의 원전 시뮬레이터를 사용하였다. Figure 6는 시범 플랫폼 운영 시나리오의 단계별 순서와 이를 도식화한 것이다.

### 3.3.3.3 시스템 요구사항 정의

전문가들과의 시범 플랫폼 운영 시나리오의 도출 과정에서 플랫폼이 가져야 하는 기능과 성능, 개발 및 운영에 관련 된 제약사항 등 다양한 요구사항이 식별되었다. 도출된 요구사항들을 종합 및 정리하고, 전문가들과 함께 재검토함으로써 플랫폼의 시스템 요구사항을 정의하였다. 정의된 시스템 요구사항은 도출된 주요 공통 기능에 따라 데이터 수집 요구사항, 데이터 저장/추출 요구사항, 데이터 처리 요구

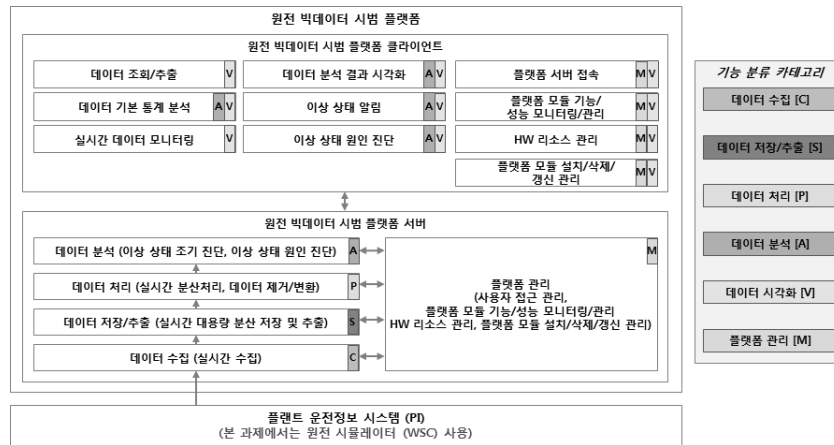


<Table 3> Some examples of system requirements for pilot platform

분류	요구사항
데이터 수집 요구사항	플랫폼은 원전 내의 다양한 센서들 (온도, 습도, 가속도, 유량, 변위, 전류, 전압, 회전, 토크 등)로부터 방대한 양의 데이터를 딜레이 없이 실시간 (최소 1초 이하)으로 수집할 수 있어야 한다.
	플랫폼은 원전 내의 많은 센서들로부터 발생하는 방대한 양의 데이터를 동시에 (최소 10,000개 이상) 수집할 수 있어야 한다.
데이터 저장/추출 요구사항	플랫폼은 다양한 센서들로부터 수집 된 센서 로그 데이터를 딜레이 없이 실시간 (최소 초당 10,000개 이상)으로 저장할 수 있어야 한다. (원전 단일 발전소 평균 약 10,000개의 센서가 있으며, 이로부터 1초당 한 번씩 데이터 수집 전제)
	플랫폼은 많은 센서들로부터 수집 된 대용량의 센서 로그 데이터를 (최소 55TB 이상) 충분히 저장할 수 있어야 한다. (센서 데이터 10,000개 x 데이터 당 크기 8byte x 60초 x 60분 x 24시간 x 365일 x 24개호기 = 약 55.07TB)
	플랫폼은 데이터 분석 시 필요한 데이터를 데이터 저장공간으로부터 딜레이 없이 실시간으로 추출할 수 있어야 한다.
데이터 처리 요구사항	플랫폼은 대용량의 데이터를 빠르게 분석할 수 있도록 여러 대의 서버를 통해 분산 처리 할 수 있어야 한다.
	플랫폼은 데이터 저장공간으로부터 추출 된 데이터를 분석할 수 있는 형태로 처리 할 수 있어야 한다.
데이터 분석 요구사항	플랫폼은 원전 운전자에게 원전 내의 다양한 센서들의 센서 로그 데이터를 실시간으로 조회할 수 있는 기능을 제공해야 한다.
	플랫폼은 원전 이상 상황 발생 전에 이상 상황을 진단하고 원전 운전자에게 사전 경고를 줄 수 있어야 한다.
	플랫폼은 원전 이상 상황 발생 시 원전 운전자에게 경고를 줄 수 있어야 한다.
	플랫폼은 원전 이상 상황 발생 시 원전 빅데이터 분석자에게 이상 상황의 잠재 원인을 제공할 수 있어야 한다.
데이터 시각화 요구사항	플랫폼은 원전 빅데이터 분석자에게 센서 로그 데이터를 딜레이 없이 실시간으로 모니터링 할 수 있는 기능을 제공해야 한다.
	플랫폼은 원전 빅데이터 분석자에게 데이터의 다양한 시각화 기능을 제공할 수 있어야 한다.
	플랫폼은 원전 빅데이터 분석자에게 빅데이터 분석 결과를 다양한 시각화 된 표현 형태로 제공할 수 있어야 한다.
플랫폼 관리 요구사항	플랫폼은 원전 빅데이터 분석자에게 센서 로그 데이터를 딜레이 없이 다양한 출력 포맷 (CSV, txt) 등으로 출력할 수 있어야 한다.
	플랫폼은 관리자에게 운영되고 있는 플랫폼의 리소스 (CPU, 저장공간, 네트워크 등) 현황을 모니터링 할 수 있도록 지원해야 한다.
	플랫폼은 플랫폼에 이상이 발견되었을 때 무슨 문제가 생겼는지 관리자에게 제공할 수 있어야 한다.
	플랫폼은 플랫폼의 사용이 가능한 사용자를 제한할 수 있어야 한다.
	플랫폼은 데이터의 전송 시 데이터가 변경되거나 사라지지 않도록 보장해야 한다.
기타 요구사항	플랫폼은 버전을 업그레이드 할 때, 플랫폼 관리자에게 손쉬운 절차를 제공해야 한다.
	플랫폼은 버전을 업그레이드 할 때, 현재 운영되고 있는 기능에 영향을 미치지 않아야 한다.
	플랫폼은 범용적으로 주로 사용되는 오픈 소스 분산 처리 플랫폼인 Hadoop v.2.4 이상에서 문제없이 운영될 필요가 있다.
	플랫폼은 기반이 되는 Hadoop과 높은 호환성을 가질 수 있도록 Linux 환경에서 운영 가능해야 한다.
기타 요구사항	플랫폼 개발자는 플랫폼 개발 시 기반이 되는 Hadoop과 호환성을 위해서 개발 언어로 Java를 사용할 필요가 있다.
	플랫폼은 데이터 전송 부하의 최소화를 위해서 RESTful 방식을 고려할 필요가 있다.

사항, 데이터 분석 요구사항, 데이터 시각화 요구사항, 플랫폼 관리 요구사항으로 구성되며, 더하여 개발 및 운영상에서의 제약 등을 포함하는 기타 요구

사항으로 구성된다. 수집 된 시스템 요구사항은 Table 3과 같다.



[Figure 7] Conceptual functional architecture of pilot platform

### 3.3.4 개념 수준의 기능 아키텍처 설계

앞서 도출된 시스템 요구사항을 바탕으로 내부 개발 전문가들과의 미팅을 통해 시스템 수준에서 플랫폼이 수행해야 하는 기능을 정형화 한 기능 아키텍처를 도출하였다. 도출 된 시스템의 주요 기능 과 각 기능에 대한 설명은 다음과 같으며, Figure 7 은 시범 플랫폼의 개념 수준의 기능 아키텍처(Conceptual functional architecture)를 보인다.

#### [원전 빅데이터 시범 플랫폼 서버]

- **데이터 수집:** 원전 내 수천 개 이상의 센서로부터 실시간으로 발생하는 빅데이터를 실시간으로 수집
- **데이터 저장/추출:** 원전 내 수천 개 이상의 센서로부터 실시간으로 수집된 데이터를 처리 및 분석하기 위하여 대용량 분산 저장 공간에 분산 저장하고, 데이터 처리 및 분석에 필요 시 분산 저장된 데이터로부터 필요한 데이터를 추출
- **데이터 처리:** 실시간으로 수집되는 데이터를 분산 데이터 저장소에 저장하기 전이나 혹은 데이터 분석 전에 저장/분석에 필요 없는 항목을 삭제, 저장/분석에 필요한 데이터 형태로 변경, 데이터의 압축 등의 데이터 전/후처리
- **데이터 분석:** 분산 데이터 저장소에 저장된 대량의 원전 데이터 집합으로부터 원전 조기 이상 상태 진단을 위한 데이터 분석

- **플랫폼 관리:** 시범 플랫폼의 데이터 수집에서부터 저장/추출, 처리, 분석, 시각화에 이르는 전체 기능의 수행 작업을 모니터링하고, 수행 기능에 문제가 있는지 분석 및 보고하며, 인가된 사용자만 접근 가능하도록 제어하는 등 플랫폼이 지속적으로 문제없이 운영될 수 있도록 관리

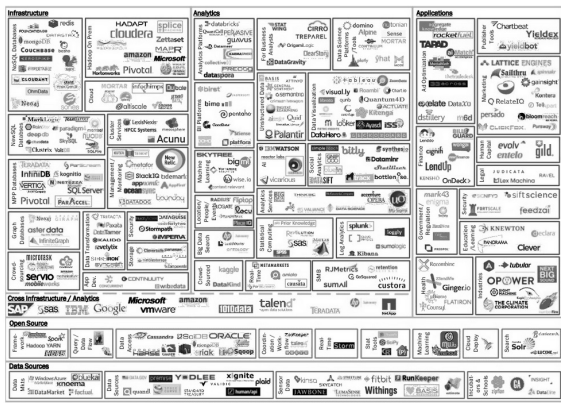
#### [원전 빅데이터 시범 플랫폼 클라이언트]

- **데이터 조회/추출:** 시범 플랫폼 서버 저장소에 저장되어 있는 데이터 집합으로부터 원하는 센서와 기간에 대하여 조회하고 CSV, XLS 파일로 추출
- **데이터 기본 통계 분석:** 시범 플랫폼 서버 저장소에 저장되어 있는 데이터 집합에 대하여 수량, 평균, 분산, 표준편차, 분포(점도, 왜도), 최소값, 최대값 등에 대하여 분석
- **실시간 데이터 모니터링:** 시범 플랫폼 서버 저장소에 실시간으로 수집 및 저장되고 있는 데이터 집합에 대하여 실시간으로 시간 축에 따라 모니터링
- **데이터 분석 결과 시각화:** 시범 플랫폼 서버의 데이터 분석 기능을 통해 도출된 분석 결과에 대하여 시각화
- **이상 상태 알림:** 시범 플랫폼 서버의 데이터 분석 기능을 통해 원전 데이터에서 이상 상태가 발생 시 조기에 사용자에게 알림

- **이상 상태 원인 진단:** 시범 플랫폼 서버의 데이터 분석 기능을 통해 원전 데이터에서 이상 상태가 발생 시 그 원인을 분석 및 진단
- **플랫폼 서버 접속:** 시범 플랫폼 서버로의 접근이 허가된 사용자는 권한 인증 과정을 통해 시범 플랫폼 서버에 접속
- **플랫폼 모듈 기능/성능 모니터링 관리:** 시범 플랫폼을 구성하고 있는 수집, 저장, 처리, 분석, 시각화, 플랫폼 관리 기능 지원 모듈들의 기능 및 성능의 실시간 상태 모니터링
- **HW 리소스 관리:** 시범 플랫폼 서버의 HW의 자원(서버 CPU 사용률, 저장 공간 사용률, 네트워크 속도 및 사용률 등)의 관리
- **플랫폼 모듈 설치/삭제/갱신 관리:** 시범 플랫폼을 구성하고 있는 수집, 저장, 처리, 분석, 시각화, 플랫폼 관리 기능 지원 모듈들의 설치/삭제/갱신

### 3.3.5 세부 기능 모듈 대안 선정 및 모듈별 적용성 평가

최근 빅데이터 기술이 빠르게 발전하면서, 빅데이터 플랫폼의 핵심 엔진인 하둡(Hadoop)과 호환성을 가지면서 빅데이터의 수집, 저장/추출, 처리, 분석, 시각화, 플랫폼 관리 기능 등을 지원하는 기술과 패키지화된 모듈이 빠르게 개발되고 있다. 이러한 모듈들은 흔히 하둡 에코시스템(Ecosystem) 혹은 빅데이터 에코 시스템으로 지칭된다. Figure 8은 Truck(2014)이 빅데이터 에코 시스템을 표시한 Big Data Landscape



[Figure 8] Big Data Landscape v3.0

v.3.0이며, 그림에서 보는 바와 같이, 수많은 선택 가능한 모듈들로 구성됨을 알 수 있다.

기능 아키텍처를 통해 개념적으로 설계된 시범 플랫폼을 물리적으로 구현하기 위해서는 기능 아키텍처의 주요 기능을 지원하는 물리적 모듈을 직접 구현하거나, 기존의 모듈 중에서 선택하는 것이다. 즉, 빅데이터 플랫폼의 물리 아키텍처 설계의 주안점은 기능 아키텍처의 6개 주요 기능별로 각 기능을 지원하는 수개~수십 개 이상의 기존의 모듈 중에서 기능 목적에 적합한지 적용성 평가를 통해 적합한 모듈을 선택하는 것이며, 이를 충분히 지원하는 모듈이 없는 경우에는 필요한 기능을 자체적으로 개발하는 과정이 필요하다. 본 연구에서는 데이터 수집, 처리, 저장/추출, 분석, 시각화 기능의 구현을 위한 기능 모듈의 적용성 평가 과정과 선정 결과를 보인다.

#### 3.3.5.1 데이터 수집 기능 구현을 위한 기능 모듈의 적용성 평가

데이터 수집 기능을 지원하는 모듈은 다양하지만, 본 연구에서는 주요 수집 모듈로 널리 알려져 있는 Apache Flume(2015), Facebook Scribe(2015), Apache Chuckwa(2015), Apache Sqoop(2015), FluentD(2015)의 5개 기능 모듈을 대안으로 선정하였으며, 이들을 대상으로 적용성을 검토하였다.

데이터 수집 기능을 위한 기능 모듈의 적용성 평가는 전문가들을 대상으로 실시간성, 유연성, 신뢰성, 호환성, 사용성, 지원성, 발전 가능성, 라이선스의 기준을 바탕으로 적용성을 검토하였으며, 적용성 평가를 통해 종합적인 측면에서 5개의 수집 모듈 대안 중 Apache Flume, Apache Sqoop, FluentD가 선정되었다. Figure 9는 데이터 수집 기능 모듈의 적용성 평가 결과의 예시를 보인다.

#### 3.3.5.2 데이터 저장/추출 기능 구현을 위한 기능 모듈의 적용성 평가

최근 빅데이터 저장/추출을 지원하는 모듈이 빠르게 개발 진행 중이며, 현재 MongoDB(2015),

기준	Apache Flume	Facebook Scribe	Apache Chukwa	Apache Sqoop	FluentD	
개발사	Apache	Facebook	Apache	Apache	Treasure Data, Inc.	
웹사이트	http://flume.apache.org/	https://github.com/facebook/archive/scribe	http://chukwa.apache.org/	http://sqoop.apache.org/	http://www.fluentd.org/	
최신 버전 (업데이트 날짜)	1.5.2 (2014.11.18)	2.2 (2010)	0.6 (2014.11.22)	Sqoop1: 1.4.5 (2014.8.12) Sqoop2: 1.99.5 (2015.2.24)	0.12.5 (2015.2.9)	
<b>세부 대안 평가 기준</b>						
실시간성	20,000개/Sec 데이터 전송 지원	20,000개/Sec 데이터 전송 지원	20,000개/Sec 데이터 전송 지원	20,000개/Sec 데이터 전송 지원	20,000개/Sec 데이터 전송 지원	
유연성	다양한 데이터 유형 지원	다양한 데이터 유형 지원	다양한 데이터 유형 지원	다양한 데이터 유형 지원	다양한 데이터 유형 지원	
신뢰성	높음	높음	높음	높음	높음	
호환성 (운영환경, 플랫폼, 보안, 호환성)	Hadoop v2.4	지원	지원	지원	지원	
	저장 모듈	플랫폼 시스템 통합 구현 단계에서 증가				
	처리 모듈	플랫폼 시스템 통합 구현 단계에서 증가				
	분석 모듈	플랫폼 시스템 통합 구현 단계에서 증가				
	시각화 모듈	독립적/이므로 호환성 관계 없음				
사용성	보통	보통	보통	보통	보통	
지원성	관련 자료 (영문, 번역, 보통, 전문, 풍부)	풍부	빈약	풍부	풍부	풍부
	개발 언어	Java	C++, Python	Java	Java	C++, Ruby
	개발 지원 언어	Java	Python	Java	Java	Ruby
발전 가능성 (낮음, 보통, 높음)	높음	낮음 (프로젝트 종료)	보통	높음	높음 (최근 가장 빠르게 업데이트)	
라이선스 (상용/오픈소스, 비유, 수정가능성, 수정 후 소스 공개/비공개)	Apache License 2.0 (오픈소스, 무료, 수정가능, 수정 후 소스 비공개)	Apache License 2.0 (오픈소스, 무료, 수정가능, 수정 후 소스 비공개)	Apache License 2.0 (오픈소스, 무료, 수정가능, 수정 후 소스 비공개)	Apache License 2.0 (오픈소스, 무료, 수정가능, 수정 후 소스 비공개)	Apache License 2.0 (오픈소스, 무료, 수정가능, 수정 후 소스 비공개)	

[Figure 9] Applicability review of data collection module candidates

Cassandra(2015) 등 수십 개 이상의 모듈이 가용하다. 이러한 다양한 데이터 저장/추출 모듈은 각기 다른 특징과 기능을 제공하고 있기 때문에, 시범 플랫폼이 요구하는 기능을 만족시킬 수 있는지 적용성을 검토할 필요가 있다.

적용성 평가는 2단계를 통해 진행되었다. 1단계에서는 수십 개 이상의 개별 모듈의 적용성 평가를 위한 시간과 노력을 줄이기 위하여 스크리닝을 수행하였으며, 2단계에서는 1단계에서 선정된 모듈에 대하여 적용성을 검토하였다. 구체적으로, 1단계에서는 전 세계 데이터베이스의 정보의 요약과 순위를 제공하고 있는 DB Ranking(2015) 자료를 참고하여, 최근 1년간 상위 순위를 유지하고 있는 모듈을 선정하였으며, 이를 통해, MongoDB, Cassandra, Redis(2015), Apache HBase(2015)의 4개 모듈이 선정되었다. Table 4는 1차 스크리닝 과정을 보인다.

다음으로, 2단계에서는 전문가들을 대상으로 실시간성, 유연성, 신뢰성, 호환성, 사용성, 지원성, 발

전 가능성, 라이선스의 기준을 바탕으로 데이터 저장/추출 기능을 위한 기능 모듈의 적용성 평가를 수행하였다. 적용성 평가를 통해 종합적인 측면에서 4개의 1차 저장/추출 모듈 대안 중 Cassandra와 HBase가 선정되었다. 데이터 저장/추출 기능 모듈의 스크리닝과 적용성 평가 결과 예시는 지면의 제약으로 생략한다. 추가적으로, 하둡 기반의 빅데이터 플랫폼에서는 대용량 파일의 분산 저장 및 처리를 지원하는 파일시스템인 Hadoop Distributed File System(HDFS)을 기본으로 이용하기 때문에, 데이터 저장/추출 기능 모듈에 포함되었다.

### 3.3.5.3 데이터 처리 기능 구현을 위한 기능 모듈의 적용성 평가

데이터 처리 기능 모듈의 경우, 다른 기능 모듈과는 달리 별도의 적용성 평가를 수행하지 않고, 내부 전문가들을 대상으로 요구사항을 만족할 수 있는 대안을 선정하였다. 구체적으로, 앞서 정의된 요구사항에서 시범 플랫폼은 안전성과 호환성을 위해 하

<Table 4> Screening process of major data store/retrieval modules

Rank		DBMS	Database Model	Score (Apr 2015)
Apr 2015	Apr 2014			
1	1	Oracle	Relational DBMS	1446.13
2	2	MySQL	Relational DBMS	1284.58
3	3	Microsoft SQL Server	Relational DBMS	1149.11
<b>4</b>	<b>5</b>	<b>MongoDB</b>	<b>Document store</b>	<b>278.59</b>
5	4	PostgreSQL	Relational DBMS	268.31
6	6	DB2	Relational DBMS	197.65
7	7	Microsoft Access	Relational DBMS	142.19
<b>8</b>	<b>9</b>	<b>Cassandra</b>	<b>Wide column store</b>	<b>104.89</b>
9	8	SQLite	Relational DBMS	102.30
<b>10</b>	<b>13</b>	<b>Redis</b>	<b>Key-value store</b>	<b>94.55</b>
11	10	SAP Adaptive Server	Relational DBMS	86.69
12	11	Solr	Search engine	82.00
13	12	Teradata	Relational DBMS	70.26
14	19	Elasticsearch	Search engine	64.66
<b>15</b>	<b>16</b>	<b>HBase</b>	<b>Wide column store</b>	<b>61.14</b>
...	...	...	...	...

둑 2.4 버전을 사용하는 것으로 결정되었으며, 이에 따라 하둑 2.4의 핵심 구성 모듈인 MapReduce와 YARN을 기본적으로 사용하기로 결정하였다. 하지만, MapReduce의 경우 실시간 처리보다는 대용량 데이터의 배치 처리에 적합하다고 알려져 있기 때문에, 시범 플랫폼의 실시간 처리 기능을 만족하기 어려울 수 있다. 따라서, 이를 만족시키기 위하여 하둑과도 호환되며 실시간 데이터 처리에 적합하다고 알려진 Apache Spark와 Apache Storm 역시 대상 후보 모듈로 선정하였다.

### 3.3.5.4 데이터 분석 기능 구현을 위한 기능 모듈의 적용성 평가

데이터 분석 기능을 지원하는 모듈은 다양하지만, 본 연구에서는 주요 분석 모듈로 널리 알려져 있는 Mahout(2015), Spark MLlib(2015), RHadoop(2015), RHive(2015), RHIVE(2015)의 5개 기능 모듈을 대안으로 선정하였으며, 이들을 대상으로 적용성을 검토하였다.

데이터 분석 기능을 위한 기능 모듈의 적용성 평

가는 전문가들을 대상으로 기능성, 신뢰성, 호환성, 사용성, 지원성, 발전 가능성, 라이선스의 기준을 바탕으로 적용성을 검토하였으며, 적용성 평가를 통해 종합적인 측면에서 5개의 분석 모듈 대안 중 RHive가 선정되었다.

RHive는 학계와 산업계에서 통계, 데이터 마이닝 분야에서 널리 사용되고 있는 언어이자 프로그램인 R을 하둑과 호환되어 대용량 분산 처리를 가능하게 하는 기능 모듈로써 통계적 혹은 데이터마이닝 기반의 범용적인 데이터 처리를 수행하는데 적합한 것으로 판단되었다. 하지만, R에서 제공하는 다양한 알고리즘 패키지들이 시범 플랫폼의 분석 요구사항의 기능을 구현할 수 있는지 보다 상세한 검토가 필요하며, 상황에 따라 개발자 입장에서 자체적으로 개발하는 것이 효과적일 수 있다. 이러한 점을 고려하여, 데이터 분석 기능 모듈은 RHive와 함께 자체적으로 개발하는 것을 고려하였다. 데이터 분석 기능 모듈의 적용성 평가 결과 예시는 지면의 제약으로 생략한다.

### 3.3.5.5 데이터 시각화 기능 구현을 위한 기능 모듈의 적용성 평가

최근 빅데이터 시각화를 지원하는 모듈이 빠르게 개발 진행 중이며, 현재 Polymans(2015), NodeBox(2015), Flot(2015), Processing(2015), D3.js(2015), gRaphaël(2015) 등 수십 개 이상의 모듈이 가용하다. 이러한 다양한 데이터 시각화 모듈은 각기 다른 특징과 기능을 제공하고 있으며, 지원하는 시각화 방법, 시각화 모델링 지원 프로그래밍 언어 등에도 많은 차이가 존재하기 때문에, 시범 플랫폼이 요구하는 기능을 만족시킬 수 있는지 적용성을 검토할 필요가 있다.

적용성 평가는 2단계를 통해 진행되었으며, 1단계에서는 수십 개 이상의 개별 모듈의 적용성 평가를 위한 시간과 노력을 줄이기 위하여 스크리닝을 수행하였으며, 2단계에서는 1단계에서 선정된 모듈에 대하여 적용성을 검토하였다. 구체적으로, 1단계에서는 Luire(2014), Suda(2014), SocialCompare(2015)가 각각 추천하고 있는 빅데이터 시각화 모듈 후보 중에서 공통으로 추천하고 있는 모듈을 선정하였으며, 이를 통해, D3.js, Flot, Google Charts, gRaphaël의 4개 모듈이 선정되었다.

다음으로, 2단계에서는 전문가들을 대상으로 실시간성, 유연성, 신뢰성, 호환성, 사용성, 지원성, 발전 가능성, 라이선스의 기준을 바탕으로 데이터 시

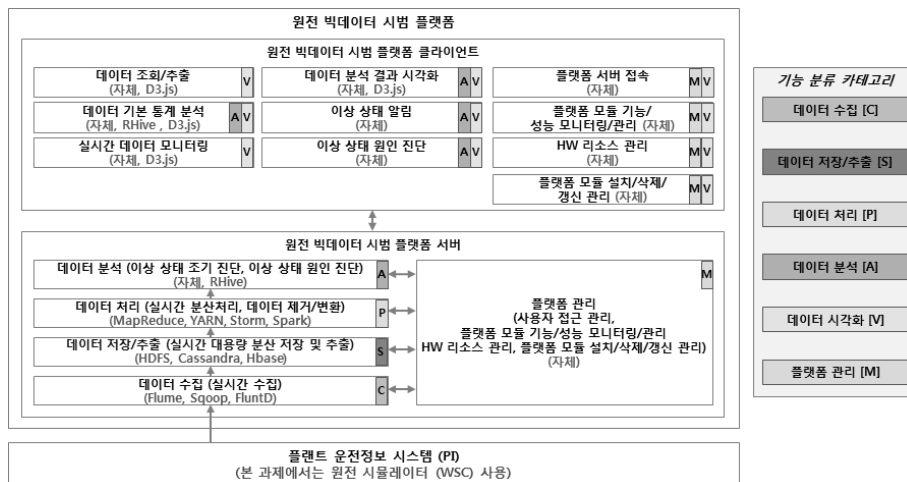
각화 기능을 위한 기능 모듈의 적용성 평가를 수행하였다. 적용성 평가를 통해 종합적인 측면에서 4개의 1차 저장/추출 모듈 대안 중 D3.js가 선정되었다. 데이터 시각화 기능 모듈의 스크리닝과 적용성 평가 결과 예시는 지면의 제약으로 생략한다.

### 3.3.5.6 플랫폼 관리 기능 구현을 위한 기능 모듈의 적용성 평가

빅데이터의 수집에서부터 분석 및 시각화를 수행하는 다른 기능들과는 달리 플랫폼 관리는 시범 플랫폼의 사용자 접근 관리, 플랫폼 모듈의 기능/성능 모니터링 및 관리, HW 리소스 관리, 플랫폼 모듈의 설치/삭제/갱신 관리 등 시범 플랫폼을 운영 및 유지보수하는 관점에서 편의성을 제공하기 위한 다양한 기능을 수행한다. 조사 결과 기존 존재하는 하둡 에코 시스템에는 이러한 다양한 기능을 지원하는 기존 모듈을 찾기 어려웠으며, 요구사항에 적합한 기능을 구현하기 위해서는 자체적으로 개발하는 것으로 결정하였다.

### 3.3.6 개념 수준의 물리 아키텍처 설계

3.3.5.1~6절의 주요 기능 모듈별 적용성 평가를 통해 선정된 기능 모듈들을 앞서 기능 아키텍처에 할당함으로써 개념 수준의 물리 아키텍처를 도출하였다. 도출된 물리 아키텍처는 Figure 10과 같다.



[Figure 10] Conceptual physical architecture for pilot platform

도출된 물리 아키텍처는 시범 플랫폼의 상세 설계 및 구현을 하는데 있어서 청사진이 될 수 있으며, 상세 설계 및 구현 단계에서 일어나는 설계 변경을 줄일 수 있다.

#### 4. 결론 및 향후 계획

빅데이터 시대의 도래와 함께 원전 안전성 분야에도 빅데이터 분석의 효과가 높을 것으로 기대를 받고 있다. 본 연구에서는 높은 관심에도 불구하고 아직까지 국내에서 개발된 적이 없는 원전 상태 감시 및 조기 경보를 위한 원전 빅데이터 시범 플랫폼을 설계하였다. 더하여, 개발의 복잡성이 높은 시범 플랫폼의 설계를 위해 복잡성을 다룰 수 있는 SE 접근 방법에 기반한 개발 절차를 제안하였으며, 제안된 개발 절차에 따라 설계 결과를 보였다.

현재 본 연구는 시범 플랫폼의 아키텍처 설계를 마치고, 플랫폼의 개발 및 구현을 위해 상세 기능을 구현 및 테스트하고 있는 단계이며, 나머지 연구에서 아키텍처의 설계 최적화, 개별 기능 모듈 구현, 기능 모듈 간 통합 및 테스트와 현장에서의 실증 테스트를 수행할 예정이다. 이에 본 논문에서는 시범 플랫폼의 개발 과정 중 아키텍처 설계 단계까지만 다루고 있으며, 다음 논문을 통해 본 연구의 실증 결과를 보일 예정이다.

#### Acknowledgements

This research was supported by the Energy Technology Development Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, and Republic Korea (No. 2014520400090).

#### References

1. Alam, J. R., Sajid, A., Talib, R., and Niaz,

M., A Review on the Role of Big Data in Business, 2014.

2. Apache Chuckwa, Retrieved Sep. 1, 2015, from <http://chukwa.apache.org/>

3. Apache Flume, Retrieved Sep. 1, 2015, from <http://flume.apache.org/>

4. Apache HBase, Retrieved Sep. 1, 2015, from <http://hbase.apache.org/>

5. Apache Mahout, Retrieved Sep. 1, 2015, from <http://mahout.apache.org/>

6. Apache Sqoop, Retrieved Sep. 1, 2015, from <http://sqoop.apache.org/>

7. Baaziz, A. and Quoniam, L., How to use Big Data technologies to optimize operations in Upstream Petroleum Industry, 21<sup>st</sup> World Petroleum Congress, 2014.

8. Beyer, M. A. and Laney, D., The importance of 'big data' : a definition. Stamford, CT: Gartner, 2012.

9. Brulé, M., Tapping the power of Big Data for the oil and gas industry, IBM Software White Paper for Petroleum Industry, 2013.

10. BSR-AIAA G-043A-201, Guide to the Preparation of Operational Concept Documents, 1992.

11. Cameron, D., Big Data in Exploration and Production: Silicon Snake-Oil, Magic Bullet, or Useful Tool?, In SPE Intelligent Energy Conference & Exhibition, Society of Petroleum Engineers, 2014.

12. Cassandra, Retrieved Sep. 1, 2015, from <http://cassandra.apache.org/>

13. Cloudera, Retrieved Sep. 1, 2015, from <http://www.cloudera.com/>

14. D3.js, Retrieved Sep. 1, 2015, from <http://d3js.org/>

15. Dijcks, J. P., Oracle: Big data for the enterprise, Oracle White Paper, 2012.

16. ANSI/EIA-632, Processes for Engineering a System, American National Standard Institute(ANSI)/Electronic Industries Alliance (EIA), 2003.
17. Facebook Scribe, Retrieved Sep. 1, 2015, from <http://github.com/facebookarchive/scribe/>
18. Feblowitz, J., The Big Deal About Big Data in Upstream Oil and Gas, IDC Energy Insights, 2012.
19. Flot, Retrieved Sep. 1, 2015, from <http://www.flotcharts.org/>
20. FluentD, Retrieved Sep. 1, 2015, from <http://www.fluentd.org/>
21. Gantz, J. and Reinsel, D., Extracting value from chaos, IDC iView, 1142: 9-10, 2011.
22. Google Chart, Retrieved Sep. 1, 2015, from <http://developers.google.com/chart/>
23. gRaphaël, Retrieved Sep. 1, 2015, from <http://raphaeljs.com/>
24. Hems, A., Soofi, A., and E. Perez, How innovative oil and gas companies are using big data to outmaneuver the competition, 2013.
25. Hortonworks, Retrieved Sep. 1, 2015, from <http://www.hortonworks.com/>
26. IEEE 1220, IEEE Standard for Application and Management of the Systems Engineering Process, IEEE Computer Society, 2005.
27. INCOSE, What is systems engineering, Retrieved Sep. 1, 2015, from <http://incose.org/practice/whatisystemseng.aspx/>
28. ISO/IEC 15288, Systems and Software Engineering – System Life Cycle Processes, International Organization for Standardization, 2015.
29. Kang, S. K., Development of Big Data Demonstration Platform to Improve the Safety of NPP (2014520400090) 1<sup>st</sup> yearly report, Korea Institute of Energy Technology Evaluation and Planning (KETEP), 2015.
30. Kei, F., Masaya, K., Isao, H., Satoshi, T., Takayuki, Y., and Toshiyuki, O., Failure sign monitoring system for large-scale plants applying system invariant analysis technology (SIAT), NEC Technical Journal, 9(1), 2014.
31. Kim, M., Risk Communication about Nuclear Power in Korea: One-Year Descriptive Analysis on Twitter, Science Education International, 24(3), 324-343, 2013.
32. Kim, D. S., & Kim, J. W., Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter, Int. J. of Multimedia and Ubiquitous Engineering, 9(11), 373-384, 2014.
33. Korea Atomic Industrial Forum, The 18<sup>th</sup> survey on the status of nuclear industry in 2012, Research Program for Nuclear Policy 2013-032251, 2014.
34. Lurie, A., 39 Data Visualization Tools for Big Data. In Cloud Computing, Retrieved from Sep. 1, 2015, <http://blog.profitbricks.com/39-data-visualization-tools-for-big-data/>
35. Manyika, J., Chui, M. Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H., Big Data: the next Frontier for Innovation, Competition and Productivity, McKinsey Global Institute, 2011.
36. MapR, Retrieved Sep. 1, 2015, from <http://www.mapr.com/>
37. MongoDB, Retrieved Sep. 1, 2015, from <http://www.mongodb.org/>
38. NIA, Classification and current status of big data technology, National Information Society Agency (NIA) Report, 2013.
39. NIA, Work process and technology application manual on each big data usage



- phases ver. 1.0: Providing considerations for work process and technology application on each big data collection · store · application, National Information Society Agency (NIA) Report, 2014.
40. NIST, Draft NIST Big Data Interoperability Framework: Vol. 2, Big Data Taxonomies, 2014.
  41. NodeBox, Retrieved Sep. 1, 2015, from <http://www.nodebox.net/>
  42. Polymaps, Retrieved Sep. 1, 2015, from <http://polymaps.org/>
  43. Processing, Retrieved Sep. 1, 2015, from <http://processing.org/>
  44. Redis, Retrieved Sep. 1, 2015, from <http://redis.io/>
  45. RHadoop, Retrieved Sep. 1, 2015, from <http://github.com/RevolutionAnalytics/RHadoop>
  46. RHIPE, Retrieved Sep. 1, 2015, from <http://github.com/tesseractdata/RHIPE>
  47. RHive, Retrieved Sep. 1, 2015, from <http://github.com/nexr/RHive>
  48. Ryu, J.-H., Kang, S.-H., Cha, J.-M., ICT application of plant industry technology, KIC News, Vol. 18, No. 5, 2015.
  49. Sagioglu, S., and Sinanc, D., Big data: A Review, Proc. of Collaboration Tech. and Systems (CTS), pp. 42–47, 2013.
  50. Seshadri, M., Big Data Science Challenging, The Oil Industry, CTO Global Services, EMC Corporation, 2013.
  51. SocialCompare, Javascript Graphs and Charts libraries, Retrieved April 15, 2015, from <http://socialcompare.com/en/comparison/javascript-graphs-and-charts-libraries/>
  52. Spark MLlib, Retrieved Sep. 1, 2015, from <http://spark.apache.org/mllib/>
  53. Suda, B., The 37 best tools for data visualization, Retrieved Nov. 11, 2014, from <http://www.creativebloq.com/design-tools/data-visualization-712402/>
  54. Truck, M. and Zilis, S., The Big Data Landscape 3.0, 2014.
  55. TTA.KO-10.0700, Functional Requirement Analysis for Big Data Service in Cloud Computing, TTA, 2013.
  56. TTA.KO-10.0705, Framework for Big Data Service in Cloud Computing, TTA, 2013.
  57. TTA.KO-10.0778, Big Data Real-time Processing Technology – Part 1: Overview, TTA, 2014.
  58. Ward, J. S., and Barker, A., Undefined by data: a survey of big data definitions, arXiv preprint arXiv:1309.5821, 2013.