A Study on SE Methodology for Design of Big Data Pilot Platform to Improve Nuclear Power Plant Safety

Junguk Shin*, Jae-Min Cha, Jun-Young Kim, Sung-Ho Park, Choong-Sub Yeom
Plant SE Team, Institute for Advanced Engineering (IAE), Yong-in, Gyeong-gi, South Korea
**Corresponding author: paulshin@iae.re.kr*

1. Introduction

A big data concept is expected to have a large impact on the safety of the nuclear power plant (NPP) from the beginning of the big data era. Though there are high interests on the NPP safety with the big data, almost no studies on the logical and physical structures and the systematic design methods of the big data platform for the NPP safety have been conducted [1].

For the current study, a new big data pilot platform for the NPP safety is designed with the main focus on the health monitoring and early warning systems, and a tailored design process based on the systems engineering approaches is proposed to manage inherent high complexity of the platform design.

The proposed design process is composed of several steps from the stakeholders identification to the integration test as shown in Fig. 1, and a part of the design processes are studied for this research as a feasibility study.

2. Design of pilot platform based on SE approach

2.1. SE approach applied for design of pilot platform

Systems engineering (SE) is a structured approach and process to develop system development systematically, and the SE approach has been embodied with the international standards such as ANSI/EIA-632 [2], IEEE 1220 [3], and ISO/IEC 15288 [4]. However, the process and method of the international standards could be unsuitable for the pilot platform development from the viewpoint of low cost and time effectiveness, resulting in the need of a simplified method and process with the basic SE approach characteristics [5]. Thus a development process based on the tailored SE approach method is defined to develop a pilot platform, with the co-work of the specialists in the fields of NPP, big data, and SE. Fig. 1 shows a proposed SE based design process for the pilot platform, and the sequential steps in the dotted lines are studied for this research.

2.2. Stakeholders identification and operational concept definition

The system lifecycle along with the steps of design, development, operation and maintenance is classified, and the main stakeholders are derived from each step. '*As-is operational concept*' and '*To-be operational concept*' are derived for the main stakeholders.

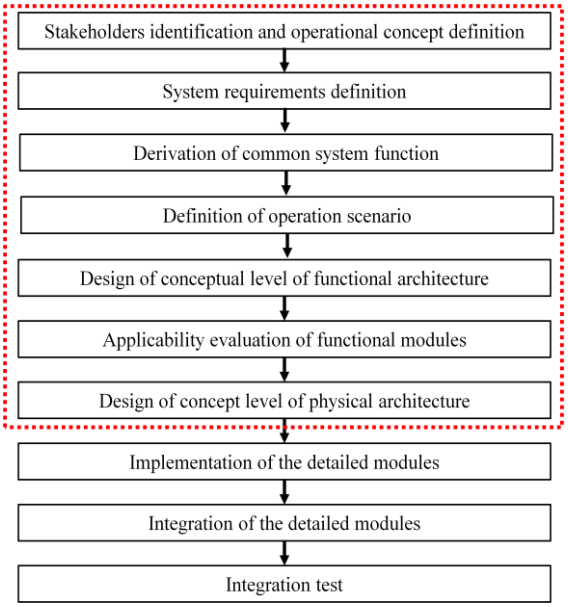

Fig. 1. SE based design process for pilot platform

2.3 System requirements definition

It is necessary to define the system requirements for the successful development of the pilot platform. Detailed functions for system requirements of the pilot platform are defined, and the operation scenario for the operation concept is described with the specialists. Also the additional system requirements are defined from the requirements classified at the scenario description process.

2.4 Derivation of common system function

A pilot platform which is used for various applications as well as a specific big data platform solution is designed, and common system functions are derived by analyzing related materials including standards, reports, papers for big data platforms from National Institute of Standards and Technology (NIST), National Inform ation Society Agency (NIA), Telecommunications Technology Association (TTA), and etc. Fig. 2 shows derived common functions.

2.5 Definition of operation scenario

After definition of system functions, operation scenario for the platform is defined from the six main common functions of pilot platform. The defined scenario consists of detailed steps as follows.

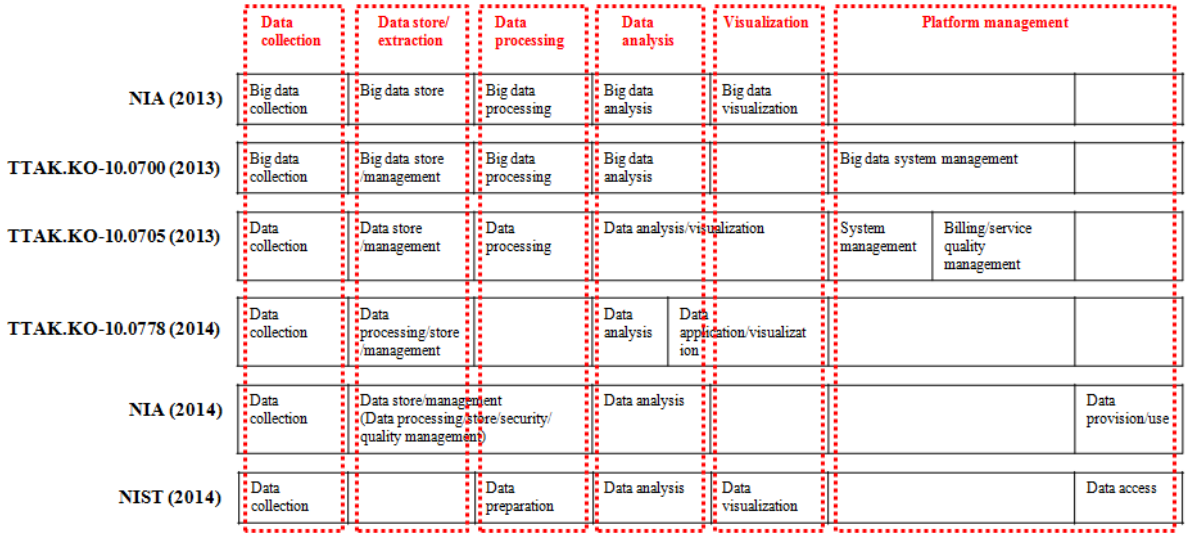

Fig. 2. Six common functions for big data platform

PI system (which is NPP simulator instead in this study) generates data in real-time and generated data is collected, stored, processed, and analyzed sequentially. During the operation, if abnormal conditions are earlier detected, then alarm is provided to NPP operators.

2.6 Design of conceptual level of functional architecture

Structured functional architecture based on the defined system requirements is derived from the consensus of internal development specialists. The main functions and description for the system are derived as below.

[Big data pilot platform server for NPP]

- Data collection: Generated big data from sensors in NPP are collected in real time.
- Data store and extraction: Distributed storage on the data collected in real time and extracts the required data.
- Data processing: Removed not require the collected data prior to saving and the analysis items. Before and after treatment to change to the required data format and compression.
- Data analysis: Analysis of the abnormal diagnosis from massive data sets stored in the distributed data storage.
- Platform management: Management such as monitoring for implementation of the full function in the pilot platform.

[Big data pilot platform client for NPP]

- Data query/extraction: Data query and extract to CSV, XLS files from pilot platform server storage.
- Basic statistical analysis of data: Analysis for the quantity of data, the mean, variance, standard deviation, the distribution, minimum, maximum, etc.
- Real-time data monitoring: Stored data monitoring data in accordance with the time axis.
- Visualization of data analysis results: visualization of the results derived from the data analysis.
- Abnormal alarm: Early notification to the user when abnormality found through data analysis.
- Abnormal diagnosis: Analysis and diagnosis for the cause of the abnormality data.
- Platform server access: The authentication process and access to the users who are allowed access to the pilot platform server.
- Function/performance monitoring and management of platform modules: Real-time status monitoring for modules in the pilot platform.
- HW resource management: Management for the server CPU utilization, storage utilization and network utilization rate.
- Platform module install/uninstall/update managem ent: install/remove/update functions for supported modules that make up a pilot platform.

3. Applicability evaluation of functional modules

There are lots of the existing modules that support each function out of the six main functions of the physical architecture. The key point of the physical architecture design for the big data platform is to select the appropriate module from the existing modules through the applicability evaluation. When the appropriate module does not exist, then it is necessary to develop a module which can support a function. The applicability of functional modules is evaluated for the data collection, processing, store and extraction, analysis and visualization in this part.

3.1 Applicability evaluation of function module for data collection

		Apache Flume	Facebook Scribe	Apache Chukwa	Apache Sqoop	FluentD
Vendor		Apache	Facebook	Apache	Apache	Treasure Data Inc.
Web site		http://flume.apache.org/	https://github.com/facebookarchive/scribe	http://chukwa.apache.org/	http://sqoop.apache.org/	http://www.fluentd.org/
Current version (Updata)		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)
Detailed driteria for alternatives evaluation						
Real-time		20,000 data/sec Various data types supported	20,000 data/sec Various data types supported	20,000 data/sec Various data types supported	20,000 data/sec Various data types supported	20,000 data/sec Various data types supported
Flexibility		High	High	High	High	High
Reliability		Support	Support	Support	Support	Support
Compatibility	Hadoop v2.4	Platform System integration evaluation in the implementation phase				
	Storage module	Platform System integration evaluation in the implementation phase				
	Processing module	Platform System integration evaluation in the implementation phase				
	Analysis module	Platform System integration evaluation in the implementation phase				
	Visualization Module	No relationship in compatibility				
Usability		normal	normal	normal	normal	normal
Supportability	Related data	Rich	Poor	Rich	Rich	Rich
	Programming language	Java	C++, Python	Java	Java	C++, Ruby
	Programming support language	Java	Python	Java	Java	Ruby
Potentials		High	Low	Medium	High	High
License		Apache License 2.0	Apache License 2.0	Apache License 2.0	Apache License 2.0	Apache License 2.0

Fig. 3. Applicability review of data collection module candidates

Five functional modules of Apache Flume, Facebook Scribe, Apache Chuckwa, Apache Sqoop, and FluentD are selected as candidates for the data collection module, and Apache Flume, Apache Sqoop, and FluentD are chosen from the applicability evaluation. Fig. 3 shows an example for applicability review of data collection module candidates [5].

3.2 Applicability evaluation of functional module for data storage/ extraction function

The applicability evaluation of functional module for the data storage/extraction is processed through two steps. Lots of modules are screened to save time and effort for the applicability evaluation in the first step, and then the chosen modules in the first step are reviewed from the practical viewpoints of real-time performance, flexibility, reliability, compatibility, usability, and supportability in the second step.

MongoDB, Cassandra, Redis, and Apache HBase are selected as an alternative in the first step, and then Cassandra and HBase are finally chosen for the functional modules in the second step.

3.3 Applicability evaluation of functional module for data processing function

While an alternative method is chosen by the review of requirements with the inner specialists, the data processing module is not processed for the applicability evaluation unlike the other functional modules. MapReduce and YARN which are core components modules in Hadoop version 2.4 for the safety and compatibility are selected for the pilot platform. Additionally Apache Spark and Apache Storm which are compatible with Hadoop and suitable for real-time data processing are also chosen for the alternative candidates.

3.4 Applicability evaluation of the functional module for data analysis

Five functional modules of Mahout, Spark MLib, RHadoop, RHive, and RHIPE which are known as data analysis modules are selected as an alternative for the applicability evaluation, and then RHive which is used for statistics and data mining is chosen for the data analysis modules. Also it is determined to develop a module for data analysis function with RHive.

3.5 Applicability evaluation of functional module for data visualization

The big data visualization function is recently operated by several modules. Thus it is necessary to evaluate the applicability of functional modules that satisfy the requirement for data visualization.

The applicability evaluation of functional module for the data visualization is processed through two steps. A lot of modules are screened to save time and effort for

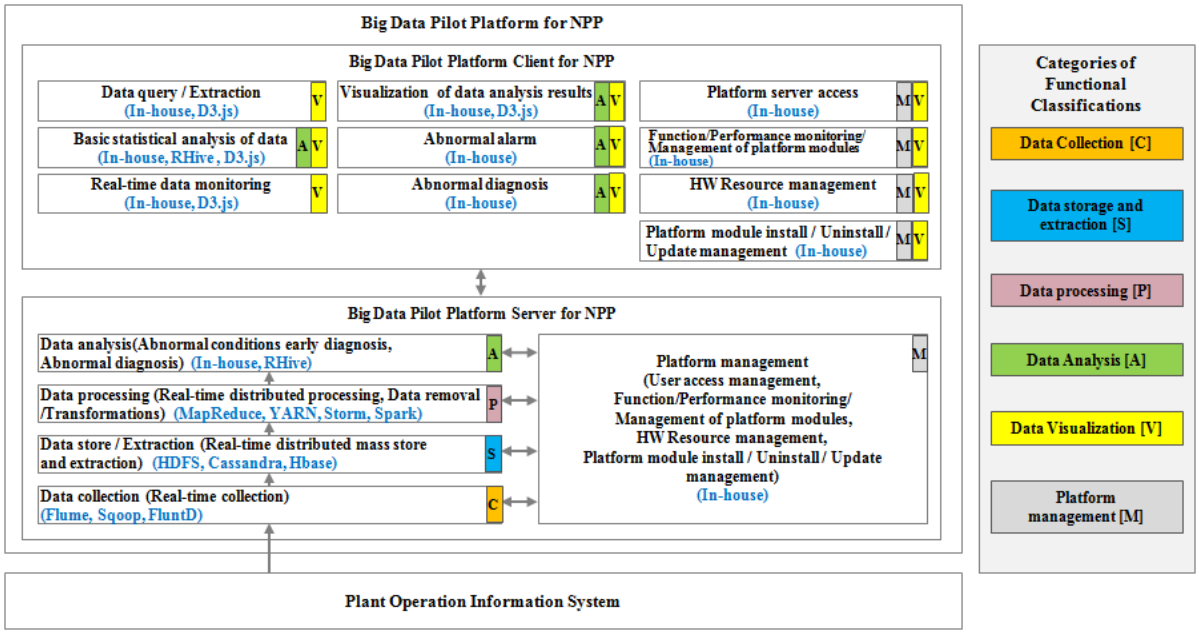

Fig. 4. Conceptual level of physical architecture of pilot platform

the applicability evaluation in the first step, and then the chosen modules in the first step are reviewed for the applicability in the second step.

D3.js, Flot, Google Charts, and gRaphaël are selected as an alternative in the first step, and then D3.js is chosen for the functional modules in the second step.

3.6 Applicability evaluation of functional module for platform management

For the convenience of platform operation and maintenance, the platform management modules perform various functional activities such as user access management for the pilot platform, monitoring and management of the platform module, HW resource management, and installation/uninstallation/update managements. It was hard to find appropriate modules that support these various functions among the Hadoop eco systems, thus it is determined to develop a module that support the requirements for the current study.

4. Design of conceptual level of physical architecture

The conceptual level of physical architecture is derived by assigning the selected functional modules into the functional architecture. The derived physical architecture can be a blueprint for detailed design and implementation of the pilot platform, and also can reduce design changes that occur in the detailed design and implementation phases. Fig. 4 shows the conceptual level of physical architecture of the pilot platform.

5. Conclusion

The big data concept is expected to have a large impact on the safety of the NPP. So, in this study, the big data pilot platform for the health monitoring and early warning of the NPP is designed. For this, the development process based on the SE approach for the pilot platform is proposed and the design results along with the proposed process are also presented. Implementation of the individual modules and integrations of those are in currently progress. Also, demonstration test will be conducted in the near future.

Acknowledgements

This research is 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] S. K. Kang, Development of Big Data Demonstration Platform to Improve the Safety of NPP (2014520400090) 1st yearly report, Korea Institute of Energy Technology Evaluation and Planning (KETEP), 2015.
[2] ANSI/EIA-632, Processes for Engineering a System, American National Standard Institute(ANSI)/Electronic Industries Alliance(EIA), 2003.
[3] IEEE 1220, IEEE Standard for Application and Management of the Systems Engineering Process, IEEE Computer Society, 2005.
[4] ISO/IEC 15288, Systems and Software Engineering - System Life Cycle Processes, International Organization for Standardization, 2015.
[5] J. M. Cha, J. Shin, C. Y. Son, D. S. Hwang, and C. S. Yeom, 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), Journal of the Korea Society of Systems Engineering, 11(2), 2015.