

ISSN: 2326-4411 (Print)  
ISSN: 2326-442X (Online)

# International Journal of Big Data

**BIG  
DATA**



A Technical Publication of the Services Society

2015

VOL.

NO.

02 02

# ARCHITECTURE FOR INTELLIGENT BIG DATA ANALYSIS BASED ON AUTOMATIC SERVICE COMPOSITION

T. H. Akila S. Siriweera, Incheon Paik, Banage T. G. S. Kumara, C. K. Koswatta

School of Computing and Engineering, University of Aizu, Aizu-Wakamatsu, Fukushima, Japan

siriweera@gmail.com, paikic@u-aizu-ac.jp, btgsk2000@gmail.com, ckkoswatta@gmail.com

## Abstract

Big Data contains massive information, which are generating from heterogeneous, autonomous sources with distributed and anonymous platforms. Since, it raises extreme challenge to organizations to store and process these data. Conventional pathway of store and process is happening as collection of manual steps and it is consuming various resources. An automated real-time and online analytical process is the most cognitive solution. Therefore it needs state of the art approach to overcome barriers and concerns currently facing by the Big Data industry. In this paper we proposed a novel architecture to automate data analytics process using Nested Automatic Service Composition (NASC) and Cross Industry Standard Platform for Data Mining (CRISP-DM) as main based technologies of the solution. NASC is well defined scalable technology to automate multi-disciplined problems domains. Since CRISP-DM also a well-known data science process which can be used as innovative accumulator of multi-dimensional data sets. CRISP-DM will be mapped with Big Data analytical process and NASC will automate the CRISP-DM process in an intelligent and innovative way.

**Keywords:** Big data analytics; Architecture; Automatic Service composition; Data mining; CRISP-DM

---

## 1. INTRODUCTION

Big Data (BD) is flooding by mass amount of data, which are generated by two types of sources those are man and unmanned. Data is pouring from every conceivable direction. In 2011, it was reported that the amount of information created and replicated nearly as many as bits of information in the digital universe is same as stars in the physical universe (Hills, 2011). That means it shows exponential growth of digital data by various factors such as volume, velocity, variety, value, and veracity etc.

In the same time importance and challenge to manipulate BD is increasing exponentially too. It has to be considered various factors and very difficult to synchronize all factors to make a final solid solution. Nowadays BD Analytics (BDA) is doing by manually

accumulated tasks. Real-time analytical process is the most effective and efficient way of BDA.

BDA arises complex situation in its data mining process due to its diversified analytical requirements and multi-disciplinary data set's. We have to select a comprehensive data mining methodology to fulfil its diversified requirements in efficient way. We can choose a data-mining method for BDA and other necessary procedure using our experiences. Cross Industry Standard Platform for Data Mining (CRISP-DM) is a useful standard for BDA. However, the manual process of steps in CRISP-DM for BDA hinders faster decision making on real time application for efficient data analytics. Further, CRISP-DM has to pass thorough and rigorous steps to complete successfully the data-mining process.

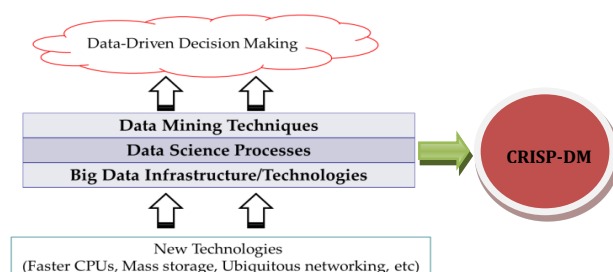
In this paper, we propose a novel architecture to automate BDA process with CRISP-DM. We have identified two critical factors of the BDA, which are suitable technical framework and architecture. We have selected Nested Automatic Service Computing (NASC) is key technology to automate CRISP-DM process. It has comprehensive capability to fulfil such multi-stepped process to automate while maintaining its scalability (Paik, Chen et al, 2014). We assume Intelligent and Innovative integration of above-mentioned technologies will result a scalable intelligent real-time BDA solution.

The rest of this paper is organized as follows. In Section 2 we discuss about preliminaries, in Section 3 we discussed about architectural designs, in Section 4 we discuss related works, in Section 5 we discuss about the experiment and evaluation. Finally, in Section 6 we conclude the paper.

## 2. PRELIMINARIES

### 2.1 BIG DATA ANALYTICS PROCESS

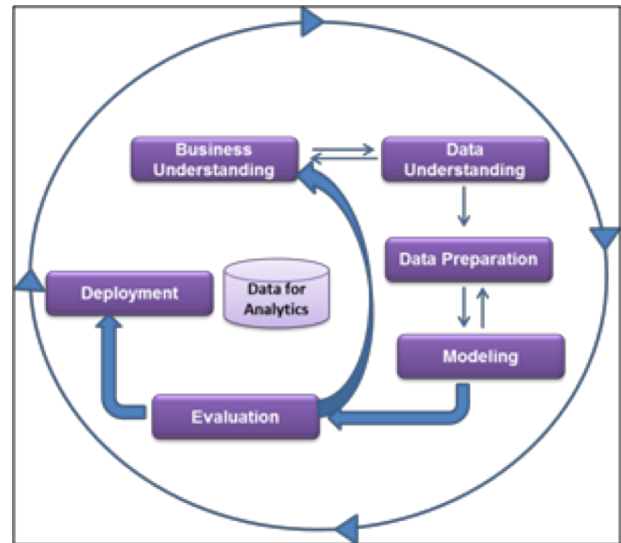
BDA is the process of collecting, organizing and analyzing large sets of data to discover patterns and other useful information. BDA helps organizations to better understanding the information contained within the data and help to identify the data that is most important to the business and future business decisions. Data warehouse will be processed by data science technology and mined by data mining techniques. Data manipulation will be done by a data science process and here we will use, CRISP-DM as shown in *Figure 1*.



**Figure 1.** Standard Process OF BDA

### 2.2 CRISP-DM

CRISP-DM has six stages and the stages fit to address effectively to complete data science requirements of the BD domain. *Figure 2* shows the graphical view of the model.



**Figure 2.** CRISP-DM Process

In **Business Understanding stage**, we understand the objectives and requirements from a domain perspective and a preliminary plan and designed to achieve the objectives.

**Data Understanding stage** initiates with given data set and continue with tasks up-to discover first insights into the data. **Preparation stage** prepares final purified and rectified data set needs to be moved to next stage. **Modeling stage** will apply various modeling techniques usually data mining techniques based on the requirements. **Evaluation stage** does thorough insight about the model with matured data. End of this stage it can take a decision whether it uses results of the mining process. In **Deployment stage**, the result will be organized to present in present in user friendly manner and finally deploy the project.

We note Business understanding and data understanding are manually confirmed and already done during this project. Next we have

to deal with rest of four stages to automate the process by NASC technology.

### 2.3 NESTED AUTOMATIC SERVICE COMPOSITION

NASC (Paik, Chen et al, 2014) is a technology derived from service oriented architectural design pattern. In this research, we use NASC to automate BDA. To automate BDA process in intelligent way, we need to define concepts of services for each step of the CRISP-DM process. After, each step will be matched to composition step logically. Development of intelligent BDA process includes these steps:

(1) Service types and instances development for the BDA; (2) Define workflow for the BDA; (3) Service discovery algorithm is developed for BDA; (4) Service selection algorithm is developed for BDA and (5) service algorithm is developed for BDA result.

## 3. ARCHITECTURE FOR INTELLIGENT BDA

Our problem domain is BDA and requirement is to make a comprehensive architectural solution to automate the BDA process. An Architecture is interpreting to real world problem into the technical language. Due to the BDA engineering perspective (size and concerns) and complexity of the solution, critical time-to-market needs demand new software engineering approaches to design software architectures (Nakagaw, Olivera et al, 2011). One of these approaches is software Reference Architecture (RA) that eases systematically reuse's knowledge and components when developing a concrete System Architecture (SA) (Cloutier, Mullar et al, 2011). After creating a concrete SA based on the RA finally we could be able to generate the implementation level UML class diagram.

### 3.1 REFERENCE ARCHITECTURE

RA is an architectural solution, which facilitates to make template solution for complex problem domain. RA for the BDA process is shown in *Figure 3*. It provides solid base to extract SA from that. SA is a conceptual

model that defines structure, behaviour and more views of a system. Simply RA is layered solution, which gives high-level view how each components and technologies of the product behave and how it maintains interactions between each of them. This layered pattern connected closely to an architectural principle "loose coupling" (Avgeriou, Zdun, 2005).

In addition, according to the BD architectural planning and designing perspective we have been studied five important observations from literature (Geerdink, Roep, 2013). (1) It is clearly define core of a BD architecture. (2) There is more than MapReduce: data sources, data mining processes, coordination and configuration engines, databases, monitoring etc. Further business intelligent systems and software components will have a place in a BD architecture. (3) Several architectural principles have been applied. Loop coupling and scalabilities are popular principles among them. (4) "Data pipeline approach" is the cognitive and truly stands out. This indicates that BD architecture is like a pipeline through that data flows. (5) There seems to be more consensus about the principles and best practices.

Angelov's framework is one of well-recognized frameworks to use in design and development of RA. We have been assisted that framework during the design and development of the RA for Intelligent BDA process (Angelov, Zdun et al, 2012). According to the Angelov's framework, it has to do following analysis before creation of the RA. Main three dimensions have to identify and clearly defined. Goal, Context, and Design are needed to clearly define. Next, it has to be determined and studied following sub dimensioned under above mentioned. Under the goal, it should be studied "Why is it defined?" dimension. Under the context, it has to be studied "Where will it be used?", "Who defines it?" and "When is it defined?" sub-dimensions. And under the design, it should be studied "What is described?", "Detailed: how is it described?", "Concreteness: How is it described?" and "Representation: How is it represented?". Here below given detail



information we have studied based on Angelov's framework of creating the RA.

### Discussion about Main three Dimensions:

#### Goal:

**Why is it defined?** Here it needs to clarify goal of the reference architecture. There are two possibilities of the Angelov et. al. defined, those standardization and facilitation. Our main ambition is to standardize concrete architecture. Therefore, goal of the RA for BDA solution is *standardization*.

#### Context:

**Where will it be used?** Here it needs to clarify the application context of the RA. This is for the organizations who are working for predictive analytics based analytical requirement data gathered from various data sources. Service computing will use as main technology. In addition, it should be able to generate concrete system architecture based on the RA. Therefore, context of this approach is to *multiple organizations*.

#### Design:

**What is it described?** Here it distinguishes, elements that can be defined in the RA. We need to distinguish components, connectors, interfaces, protocols, algorithms, policies and guidelines. We are using CRISP-DM as the data mining concepts behind the solution. Since we are based on the ASC as selected architectural principle of overall solution.

Above we have been discussed main three dimensions under the Goal, Context and Design.

Next we discuss the sub-dimension which are under main 2 dimensions as follows.

#### Context:

**Who defines it?** Here it distinguishes, who will define it?. It is about the stakeholders of the RA? Mainly 2 groups involved in. They are designers and providers. Mainly product will be use by who is doing predictive analytics in *standard organizations*.

**When is it defined?** Here it needs to clarify the timing aspects of the RA. That is our solution

needs to be time independent. That means RA will be outdated when components used in the RA are outdated. It has two possible artifacts, which are preliminary and classical. It is using Preliminary RA if there are no concrete components have used in the RA. We have to define and make clear idea about the components are used in, according to the system development and concrete system architecture development. It will result more practical and sustainable RA solution. Therefore, our approach is *classical*.

#### Design:

**Detailed, how is it described?** Here it distinguishes, number of levels can be defined in "What is it described?". It avoids more than two instances of usage of elements and components behind the solution. Further it will go through the *levels and stages* of above mentioned technologies in overall solution.

**Concreteness, how is it described?** Here it distinguishes, possible level of abstraction of RA. It is related to the level of choices made in an architecture in terms of technology, application and users etc. Here it has three values, which are abstract, semi-concrete and concrete. We plan to achieve a *concrete architecture*.

**Representation, How is it represented?** Here it distinguishes, possible levels of formalization of RA. We have planned to use ArchiMate 3.1.1 for development and design the RA. Three architectural layers planned to be used. ASC, that is one of main concepts behind the Service oriented architecture (SOA). That is we have *semi-formal and formal element specifications* about the using elements.

According to our initial analysis of RA on BDA domain is based on the Angelov's framework, *Table 1* is summarized information as follows. *Table 1* describes the dimension vs values we have identified during process. Each dimension represents identical value and this flow summarizes to way of conclude concreteness and representation of the RA. It is discussed the main technologies we used and organizational support etc.

Dimension	Values
<i>Why is it defined?</i>	<i>standardization</i>
↓	
<i>Where will it be used</i>	multiple organizations
<i>Who defines it?</i>	standard organizations
<i>When is it defined?</i>	classical
↓	
<i>What is it described?</i>	ASC, CRISP-DM
<i>Detailed</i>	Levels and stages of above technologies
<i>Concreteness</i>	concrete architecture
<i>Representation</i>	semi-formal and formal element specifications

**Table 1:** Dimensions and Values RA

Based on the Angelov's framework of RA, we have been concluded the Type 2 RA. That is the Standardization RA.

#### Development of the RA:

For the RA perspective, we have identified main 3 building block layers, top level layer called as Analytical Layer, middle layer called as Technology Layer and bottom layer called as Infrastructure Layer. Let start to identify each layers in summarily:

#### Infrastructure layer:

It mainly considers data warehouse and data mart layer. This contains Hadoop Eco system to manipulate BD infrastructure, web service pools and two relational data base managements (RDBMS) for data manipulation and to maintain analytical clusters. All of the above can exist in Intranet and Internet platforms. As an example, Hadoop cluster can be Geo distributed as data centres and then we have to deal with hadoop beyond the intranet level.

Since web services also can be distributed along the internet and local network. One of two RDBMS ready to accept export data from Hadoop and data processing facility to the technology layer. The other RDMS is used for handling analytical cluster and related activities of the analytical process.

#### Technology layer:

Mainly this layer is dominated by NASC and it supports technologies such as quality of services agent and intelligent planning agent to provide intelligent workflow automation facility.

Therefore it will identify the requirement to utilize respective resources distributed along the system to fulfil the functional and non-functional requirements of the project.

#### Analytical layer:

This layer is dominated by CRISP-DM to provide data mining process of the project. First two out of 6 stages of the CRISP-DM has been decided by manually and therefore ASC will be dealt with the rest of four stages to full-fill the data automation of mining requirement

### 3.2 SYSTEM ARCHITECTURE

#### Scenario 1:

*ABC Air Port Company requires analysis of the flight delay data to identify factors, which had been affecting to the flight delay. By the analysis, company hopes to take necessary decisions to reduce/ avoid flight delays.*

We derived SA based on the RA and applied our scenario to the SA. It is clearly shown how each layer will behave during the execution time and result (output) will be produced in it is (NASC) execution stage.

As shown in *Figure 4*, it is giving high-level view of derived SA based on the RA, which is applied our scenario. SA is displaying the existing technologies and their responsibilities. Moreover, it describes communication between layers across the total solution.

### 3.3 TOP LEVEL UML CLASS DIAGRAM

We are able to successfully integrated main technologies to achieve intelligent real-time analytics for BDA by using the RA. Next, we successfully derived SA for our scenario from

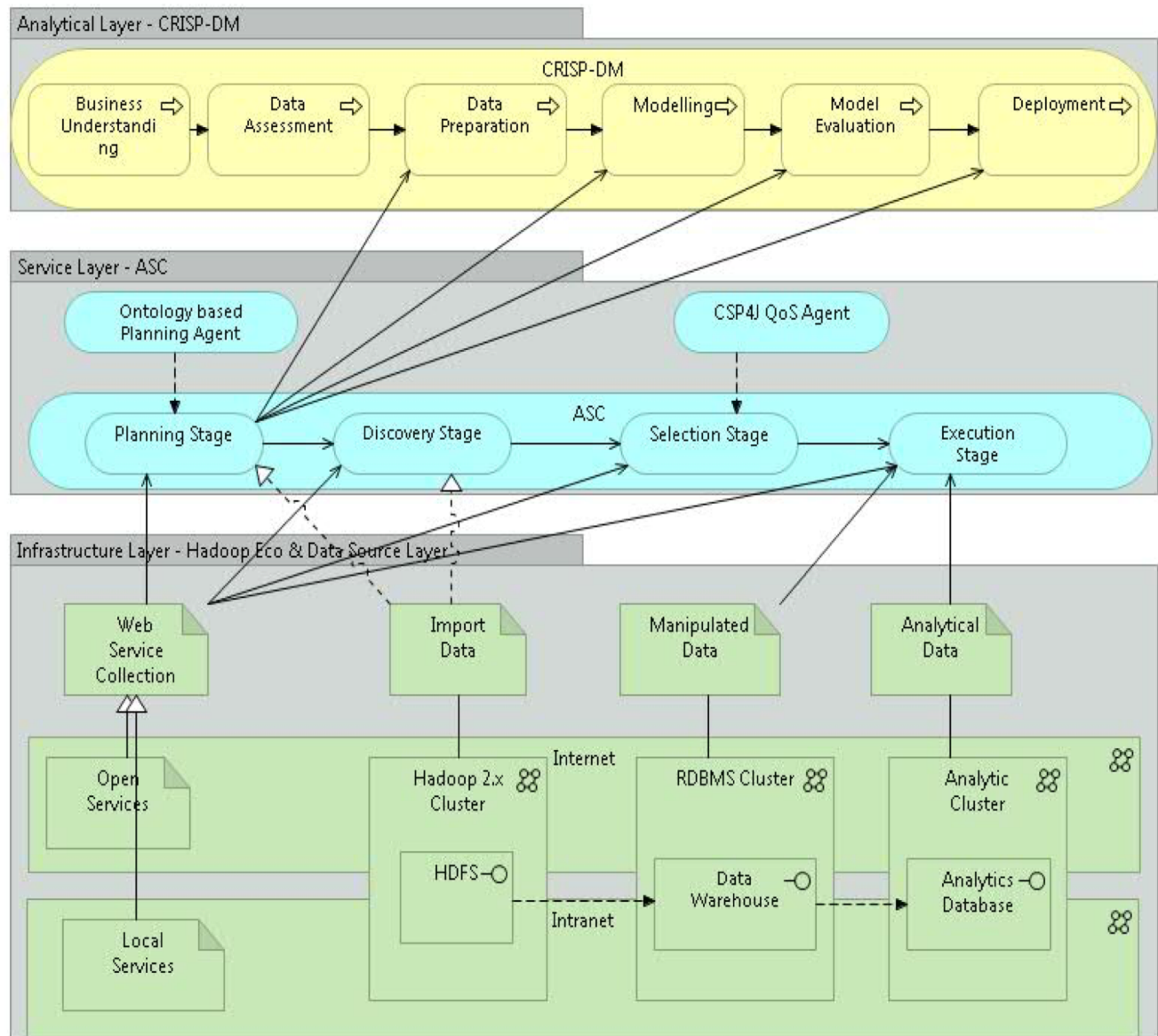
the RA. Finally we could be able to extract and designed top level detailed UML class diagram of scalable BDA for our scenario based on the RA and the SA.

Mainly we have identified 2 packages. One is for ASC and the other is for CRISP-DM related services.

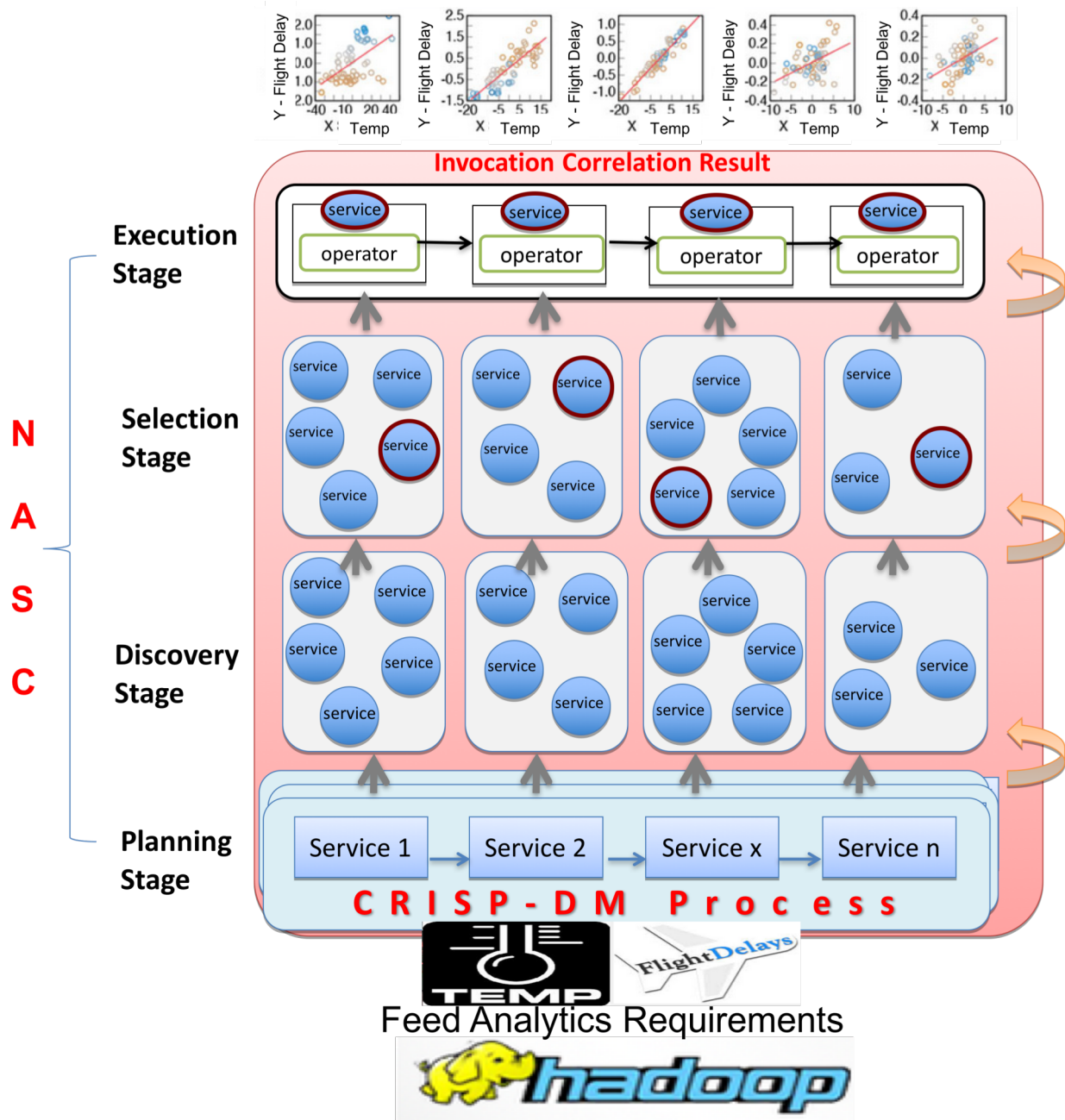
In addition to them, there are two utility packages to provide utility services to system. They are Planning Agent and Quality of Service (QoS) Agent.

Figure 5 displays high level view of the UML class diagram.

### NASC Process



**Figure 3.** Reference Architecture of the Intelligent BDA Solution



**Figure 4.** NASC vs CRISP-DM and Classification

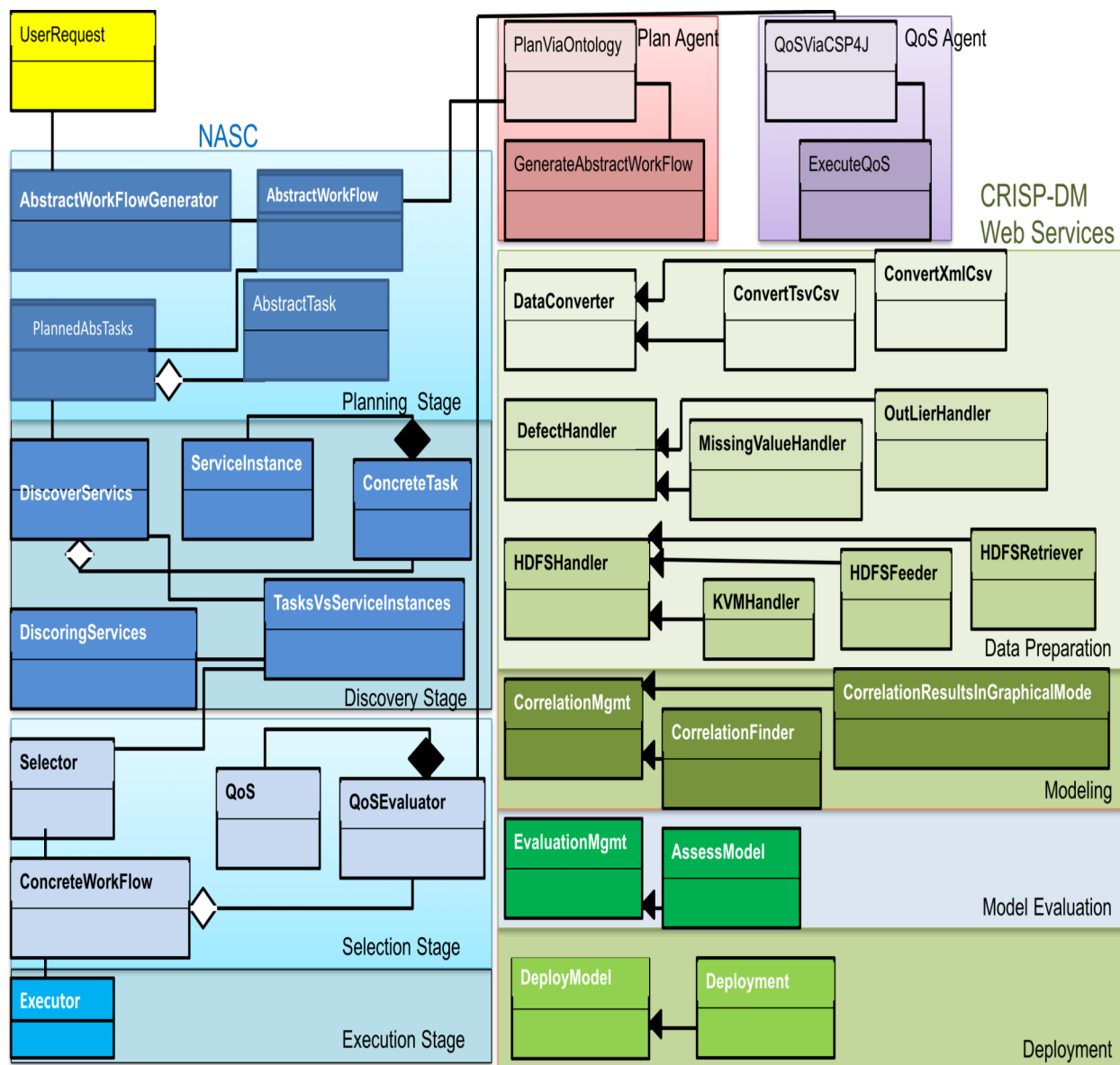
This is the based package of the solution. This allows identifying the functional and non-functional requirements for the analytics as follows.

R: Set of user's requests at the service level.

$W = \{t_1, t_2, t_3, \dots, t_l\}$ : Set of  $l$  abstract tasks in an abstract workflow  $W$ .

**Planning:**  $\Pi: R \rightarrow W$ .

$I_i = \{i_{i1}, i_{i2}, i_{i3}, \dots, i_{im}\}$ : Set of  $m$  service instances advertised in a service registry for an abstract task  $t_i$ .  $I$  is the set of  $I_i$ , for  $1 \leq i \leq l$ . If each task in a workflow has  $m$



**Figure 5.** Top-level UML class diagram of Intelligent BDA Process

instances, then the total number of service instances available for the workflow is  $l \times m$

Many researchers adopt the hierarchical task network (HTN) planner (Sirin, Parsia. 2004) technique to dynamically develop workflows. However, it is acknowledged that the formulation of HTN planning problem requires significant structural information.

In contrast to their work, we use ontology-based workflow generation method (Kumara, Paik et al, 2015). Our core idea is to utilize ontology to acquire hidden domain knowledge, in order to generate more application-specific abstract workflow. Based on ontology designed for the CRISP-DM, we have developed approach for the workflow

generation. A rule-based approach is developed for detailed inference. We have implemented SWRL rules to identify the abstract services according to the properties of dataset and user requirements.

**Discovery:**  $\Delta : W \rightarrow I$ .

$C_j = \{c_{j1}, c_{j2}, c_{j3}, \dots, c_{jp}\}$  : Set of  $P$  selected service instances to be executed from the service instance set  $I_j$ .  $C$  is the set of  $C_j$  where  $1 \leq j \leq l$

We applied ontology based discovery method to identify candidate services for abstract tasks in workflow. We compute the degree of semantic matching for a given pair of abstract task and web service instance by applying different filters.

**Selection:**  $\Sigma : I \rightarrow C$ .

$X = \{x_1, x_2, x_3, \dots, x_q\}$  : Set of  $Q$  executed service traces. Here it is identifying most cognitive services for each tasks from given set of services resulted by discovery stage.

**Execution:**  $E : C \rightarrow X$ .

All the selected services of given concrete tasks will be executed according to the concrete workflow.

### CRISP-DM Process

This package will be responsible to deal with web services (internet and locally distributed) related to services which are requesting by NASC to accomplish complex, dynamic and diversified tasks of the BDA process.

Note that according to the scenario, we have manually accomplished 1st 2 stages of CRISP-DM. Here NASC will deal with the rest of four stages.

### Utility Packages:

One of the two main packages is planning agent. It can be selected developer preferred planning agent to full-fill the planning

requirement such as HTN. However, we choose Planning Agent Ontology reasoning for the planning process.

Quality of Services (QoS) Agent: We have used Constraint Satisfaction Problem Solving Agent as our QoS agent to full-fill the planning requirement such as HTN. However, we choose Planning Agent by Ontology reasoning for the planning process.

### Quality of Services (QoS) Agent:

We have used Constraint Satisfaction Problem Solving Agent as our QoS agent.

## 4. RELATED WORK

The literature on scalable intelligent architecture for BDA is scarce. There is an architecture, which provided reference architecture for BDA, and result is indicative evidence (Geerdink, Roep, 2013). In addition, intelligent multi agent solution provided for particular domain (Ivan, Stula et al, 2014).

A memory centric real-time BDA also introduced and explained (Tao, Doshi et al, 2013). Health related real-time BDA solution for monitoring purposes discussed in detail (Moore, Baroli et al, 2013).

And it is providing predictive BDA for aviation industry with considering various factors of the aviation industry (Aylan, Pesce et al, 2013).

Wu et al. (Wu, Zhu et al, 2013) presented a HACE theorem that characterizes the features of the BD revolution, and proposes a BD processing model.

Most of solutions are industry specific and some of them are providing real time support to the analytical process. When it is comparing with our solution, we have introduced a industry independent scalable solution to the BDA process.

There are interesting approaches for the data science level in addition to the



architectural and design level approaches. These can be considered as the related works for the stages of the CRISP-DM process. Best practices in data preparation have been discussed and elaborated (Castanedo, 2015). It is discussing all five stages of CRISP-DM in different perspective. But it is not about the CRISP-DM. Open source BDA modelling level approaches are introduced. It is based on Apache Cassandra. It has proposed, defined, presented and demonstrated their methodology under four stages of the modelling process. (Chebotko, Kashlev et al, 2015)

## 5. EXPERIMENT AND EVALUATION

Architecture evaluation can do in more stages of the software development process (Mattson, Grahn et al, 2006). It can be used to evaluate various attributes, using various methods. There are four types of software architecture evaluation categories have identified. (1) **Experienced-based** evaluation is based on the previous experience and domain knowledge of developers or consultants, (2) **Simulation-based** evaluation relies on high level implementation some or all of the components in the architecture, (3) **Mathematical modelling**, which measures operational quality of requirements (Reusner, Schmidt et al, 2003) and (4) **Scenario-based** evaluation relies particular quality attributes by creating scenarios.

### 5.1. EXPERIMENTAL SETUP

Here, we use scenario-based evaluation category to evaluate our architecture for the BDA automation process. There are three types of key methods in scenario-based have identified. Software Architecture Analysis Method (SAAM), Architecture Trade-off Method (ATAM) and Architecture Level Modifiability Analysis Method (ALMAM). Here we use SAAM to evaluate our three staged architecture's designed and developed for the BDA automation. And it was used

experienced-based voting method to assess overall SAAM method.

SAAM consists of six main steps and we have followed our evaluation process according to them (MT Ionita, DK Hammer et al, 2002). Three types of roles should be involved during the SAAM process. Before starting the main six steps of the SAAM, we have to identify SAAM Roles.

### SAAM Roles

We have consulted fifteen industry experts, who have diverse expertization, which are needed in BDA process and used in this architectural process such as BDA, SOA, ASC, AI, Data Mining and Dynamic Workflow automation etc. These roles are involved in the evaluation process, **(a) External stakeholders:** three experts, **(b) Internal stakeholders:** six experts and **(c) SAAM team:** six experts.

### Step 1 – Develop Scenario

As a brainstorm exercise we have defined following scenario as for our first step.

*Scenario 1:* ABC airport Company plans to analyze the flight delay data to identify the factors that, led to the past delay. The company hopes to reduce airline delay through such data analytics. BDA engineers work to find the correlation between temperature and flight delay.

This scenario was used as base scenario and variation of this (by changing the parameters, such as collective ground operational time vs flight delay, Air traffic controllers mishaps vs flight delays etc.) are used as other respective variations of scenarios.

### Step 2 – Describe Architectures

We held a session to describe each candidate architectures invented to BDA process and also introduced our requirement by natural-language specification with a summary overview.

### Step 3 – Classify and Prioritize Scenarios

Here we identified direct scenario as Scenario 1 and indirect scenarios as variations, which are above mentioned. And levels of candidate architectures as stages of architectural design process. As the stage 1: RA was evaluated based on the given scenario, studied the way of achieve it and benefits and drawbacks of the candidate RA. Next, as of the stage 2: Studied the SA, the way of derived it from the RA. And let them to analyze, whether it satisfied given requirements of scenarios. Finally the high level UML class diagram was analyzed. We did the evaluation to studied ten key quality factors, which are Performance, Reliability, Availability, Security, Modifiability, Portability, Functionality, Variability, Subset-ability and Conceptual integrity.

### Step 4 – Individually Evaluate Indirect Scenarios

In the step3, we gave top priority to analyze the direct scenario. Next we gave other variations to further analysis and study the tallying with three stages of candidate architectures. Let them to study above mentioned quality factors.

### Step 5 – Assess Scenario Interactions

Next, we asked to evaluate architectures by considering with multiple scenarios vs changes over the components within architectures.

## 5.2. EVALUATION

The process of evaluation is laid on step 6 and we have followed following step 6 and received the below result.

### Step 6 – Create an Overall Evaluation

Finally evaluation team marked estimation of success weights of scenarios vs stages of candidate architectures and continued final report of estimation of quality factors based on their weights. We created a ranked

questionnaire of quality factors and produced it to the evaluation team.

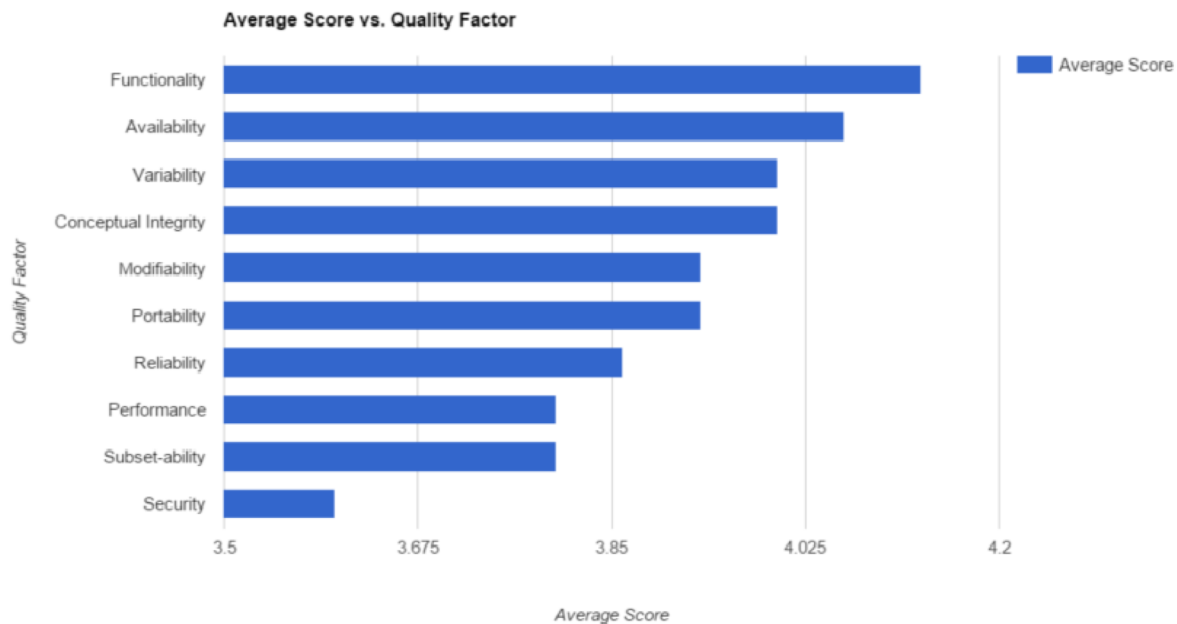
Ranks were given as follows their satisfaction of quality factors: 1 for poorly satisfied, 2 for minimally satisfied, 3 for averaged satisfied, 4 for good and 5 for perfectly satisfied. Figure 6 shows the sorted averaged scores vs quality factors.

All quality factors are scored above the 'Average' satisfaction rate. System Functionality and Availability perspective, architectures have been achieved above the state of the 'Good'. Variability and Conceptual integrity also achieved close to state of the 'Good'. Rest of all factors are scored beyond the 3.5 except security factor. It is the minimally scored quality factor. That implies architectural design needs to consider more on the security in addition to other quality attributes.

## 6. CONCLUSION

We have achieved successful design for Intelligent BDA using RA. Next, we derive the SA using that RA and simulate SA with our scenario. After that, we design UML class diagram for the software development process of the BDA. As the solution is scalable and we believe this architectural solution will work for our scenario effectively. Focusing on the planning stage of the ASC, we used a method of generating abstract workflow of the service composition for automating the BDA. We used an ontology-based workflow generation method. Further, in discovery stage we have achieved it ontology based web service discovery and global social service network (GSSN) based discovery methods.

SAAM is well known process of architecture evaluation method. Our main focus was to improve the quality and accuracy of the evaluation result. Therefore, it was involved only senior level experts in enterprise level such as senior consultants,



**Figure 6.** Sorted Average Scores vs Evaluated Quality Factors

architects, data mining experts and researchers in respective fields of the BDA domain. According to the overall results of the SAAM, it was scored above the state of the Good and this implies BDA community can be adopted this architectural design's smoothly for their requirements in the BDA domain. Results of the SAAM can consider as the good indicative proof of success our architectural design process for the BDA domain.

Now we are working on selection stage, which is the third stage of the ASC out of four stages. Once it is finished the development process of all four stages of the ASC in code base level, these architectures can be proved and justified scientifically.

In our future work, we need to do more experiments to evaluate about efficiency. Also plan to study and observe behaviors of the change of key technologies with relevant substitution in that time.

## 7. REFERENCES

- C. Hills, (2011 June 08) , IDC Digital Universe Study: Big Data is here. Now what? Retrieved February 06 2015. From [http://chucksblog.emc.com/chucks\\_blog/2011/06/2011-idc-digitaluniverse-study-big-data-is-here-now-what.html](http://chucksblog.emc.com/chucks_blog/2011/06/2011-idc-digitaluniverse-study-big-data-is-here-now-what.html)
- I. Paik, W. Chen, and M. N. Huhns. (2009) .“A scalable architecture for automatic service composition,” IEEE Transactions on Services Computing, vol. 7, no. 1, pp. 82–95.
- E. Nakagawa, P. Olivera, M. Becker. (2011). Reference architecture and product line architecture: A subtle but critical differences. ECSA, pp. 207-211
- R. Cloutier, G. Muller, D. Verma, R. Nichiani, E. Hole. (2010) . Bone The Concept of reference architectures, System Engineering 13(1), 14-27

P. Avgeriou en U. Zdun, (2005) ,Architectural Patterns Revisited – A Pattern Language, 10th European Conference on Pattern Languages of Programs, Irsee, Germany 2005

B. Geerdink, I.G Roep, (2013) , A Reference Architecture for Big Data Solutions – IEEE ICITST-2013

S. Angelov, P. Grefen, D. Greefhorst. (2012) , A framework for Analysis and Design of the Software Reference Architecture. Information and Software Technology 54 pp 417 - 433

E. Sirin, B. Parsia, D. Wu, J. Hendler, and D. (2004) .Nau, HTN Planning for Web Service Composition Using SHOP2. J. Web Semantics, vol. 1, pp. 377-396.

B. T. G. S. Kumara, Incheon Paik, Jia Zhang, T. H. A. S. Siriweera, Koswatte R. C. Koswatte, (2015). Ontology-Based Workflow Generation for Intelligent Big Data Analytics . IEEE International Conference on Web Services (ICWS2015), USA,

M. Ivan, M. Stula, J Maras. (2014) , Intelligent MultiAgent System for Insurance Industry – IEEE 2014 MIPRO, Croatia

Z. Tao, K. Doshi, X. Tang, T.Lou, Z. Lu, H. Li, (2013) ,"A Big Data Architecture for Real-time Analytics"– IEEE 2013 Big Data

P. Moore, L. Baroli, F. Xhafa, A.Thomas, (2013) , Monitoring and Detection of Agitation in Dementia. Towards Real- Time and Big-Data Solutions, IEEE 2013 Computer Society

S. Ayhan, J. Pesce, P. Comitz, D. Sweet, S. Bliesner, G. Gerberick, (2013) ,Predictive Analytics with Aviation Big Data" – IEEE 2013 ICNS Conference

X. Wu, X. Zhu, G. Q. Wu, and W. Ding, (2013) ,Data Mining with Big Data, IEEE

Transaction Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107

F. Castanedo, "Data Preparation in the Big Data Era. Best Practices for Data Integration", Published by O'Reilly, 2015

A. Chebotko, A. Kashlev, S. Lu, "A Big Data Modeling Methodology for Apache Cassandra", IEEE International Congress of Big Data, New York, USA, 2015

M. Mattson, H. Grahn, F. Martensson, "Software Architecture Evaluation Methods for Performance, Maintainability, Testability, and Portability," Conference on the Quality of Software Architecture (QoSA), 2006.

R. H. Reussner, H. W. Schmidt, I. H. Poernomo, "Reliability prediction for component-based software architectures,"Journal of Systems and Software, 66(3):241-252, 2003.

M. T. Ionita, D. K. Hammer, H. Obbink, "Scenario-Based Software Architecture Evaluation Methods: An Overview", International Conference on Software Engineering (Software Architecture Review and Assessment), 2002

## Authors



**T. H. Akila S. Siriweera** received the bachelor's degree from the University of Peradeniya, Sri Lanka, in 2005 and the reading for the master degree in University of Aizu, Japan. His research

Interests include Semantic Web, Big Data, Services and cloud computing. He has around 9 year industry experiences and finally served as software consultant.



**Incheon Paik** received the ME and PhD degrees in electronics engineering from Korea University in 1987 and 1992, respectively. During 1993-2000, he worked as an associate professor at

Soonchunhyang University, Korea. From 1996 to 1998, he was a visiting researcher at the State Key Laboratory, Beihang University, Beijing, China. He is currently an associate professor at the University of Aizu, Japan. His research interests include Semantic Web, web services and their composition, web data mining, awareness computing, security for e-business, and agents on Semantic Web. He served several conferences as a program chair and program committee member for numerous international conferences. He is a member of the ACM, the IEEE, and the IPSJ.



**Banage T. G. S. Kumara** received the bachelor's degree in 2006 from Sabaragamuwa University of Sri Lanka. He received the master's degree in 2010 from University

Peradeniya, Sri Lanka and he received Ph.D degree from School of Computer Science and Engineering, University of Aizu, Japan. His research interests include Semantic Web, Web Data Mining, Web Service Discovery and Composition.



**Koswatte R. C. Koswatte** received the bachelor's degree in 2006 from Sri Lanka Institute of Information Technology, Sri Lanka. She completed master degree in from School of Computer Science and

Engineering, University of Aizu, Japan.



[hipore.com](http://hipore.com)

# BIG DATA

A Technical Publication of the Services Society



ISSN: 2326-4411 (Print)  
ISSN: 2326-442X (Online)