QoS and Customizable Transaction-aware Selection for Big Data Analytics on Automatic Service Composition

T. H. Akila S. Siriweera, Incheon Paik
Dept. of Computer Science & Information System,
University of Aizu, Aizu-Wakamatsu,
Fukushima, Japan
siriweera@gmail.com, paikic@u-aizu.ac.jp

Banage T. G. S. Kumara
Dept. of Computing & Information System,
Sabaragamuwa University of Sri Lanka,
Belihuloya, Sri Lanka
btgsk2000@gmail.com

Abstract: As services for Big Data Analysis (BDA) become prevalent, analysis services with intelligence and autonomy using automatic service composition (ASC) show bright prospects in the BDA market. Selection is one of the most important phases of successful ASC process. Moreover, it became competitive with the rise of demand for the services and criticalness of the BDA process. It is a challenge to accomplish a successful uninterruptable composition while serving diverse custom selection requirements. In the case of failure, it results in complete loss of time and resources. Traditional approaches are not applicable to handle failures during long running transactions. Instead, compensation suggests to being an error recovery. Therefore, analytics transactions scheduled as a composition of a set of compensable transactions. However, compensable services are a higher price and consume more time. Moreover, consumers equipped with diverse requirements. It is necessary to guarantee the critical stages of workflow using compensable services rather than whole workflow. Therefore, we proposed customizable Transaction and QoS-aware service selection approach under five user custom settings based on genetic algorithm (GA) to address above concerns. QoS-awareness facilitated by multi-objective QoS criteria and GA is used for multivariate optimization. We conducted a thorough evaluation, and it shows proposed method effectively and efficiently reach the global optimal of the overall selection criteria.

Keywords: Automatic Service Composition; Web service Selection; Big Data Analytics; Transactional Web Services; Genetic Algorithm

I. INTRODUCTION

Data analytics as a service is one of the leading service based industry, according to the industry survey, it increased the demand for the Amazon analytical services by 293% in 2016 compared to 2013. It has nearly 100% growth rate in each year for Amazon Web Services1. Services became the prevalent platform of the data analytics, especially Big Data Analytics (BDA). However, BDA process is heavily time and resource consuming job.

Therefore, we believe that the automation of the BDA process is the most desirable approach to the BDA domain. As the first step, we have proposed a comprehensive architectural design process for the BDA automation based on Automatic Service Composition (ASC) [1]. The ASC mainly contained four stages, planning, discovery, selection and execution [2].

Next, we have proposed novel approaches to achieve the planning and discovery stages [3], [4]. In this paper, we have addressed the selection stage of the ASC process while addressing concerns occurred in the BDA workflow.

BDA contained many tasks, and each task consumes extended periods and the considerably large amount of resources to accomplish given requirements. Nevertheless, these tasks are highly vulnerable due to the highly volatile environment such as bandwidth; infrastructure tends to result in unexpected terminations and errors.

Such interruptions of long running processes hamper the automation process and result loss of data, time and resources. In the BDA perspective, it will be a huge loss. Transactional properties of the services can be utilized to address above concerns in composition process [5]–[9].

However, conventional selection approaches do not consider the transactional awareness during the composition process [10]–[12]. They mainly focused on various aspects of multi-objective quantitative QoS-awareness (QoS\textsubscript{QN}) and qualitative QoS-awareness (QoS\textsubscript{Q}) for near optimal or global optimal during the selection. None of these approaches considering the risk of unexpected termination or errors during the composition process. However, the optimal composition does not guarantee the reliable composition during the execution.

Then transactional service (TS) coming into the topic. Here also, some approaches are only limited to transactional awareness other than QoS-awareness [13], [14]. However, Only TS aware approaches are not guaranteed the requirements of the multivariate QoS\textsubscript{QN} awareness. Especially it is a serious drawback of the BDA composition.

Few approaches are proposing composition models considering both TS and QoS\textsubscript{QN} awareness of the composition [5]-[9]. As per our literature review, only Haddad et. al. proposed ASC based TS and QoS\textsubscript{QN} aware composition model for the workflow automation under two key risk levels [5]. Moreover, they have proposed semantic TS properties (TSP) identification methodology. It allows defining transactional requirements for the workflow in general not domain specific. It proposed to find the local optimal not global optimal and not flexible for custom user requirements. Z. Ding et. al. proposed genetic algorithm (GA) based approach and it focused global optimal solution [6]. They employed GA with

1www.statista.com/statistics/233725/development-of-amazon-web-services-revenue/
a penalty function for given workflow. Nevertheless, as per our study, it has not considered the workflow automation and he proposed algorithm does not provide the flexibility to customization. J. Li et al. proposed composition model for the Directed Acyclic Graph models but not guaranteed either automation and custom settings [7]. J. Cao et al. proposed Ant Colony-based TS and QoS selection for near optimal solution [8]. Y. Cadinal et al. proposed TS and QoS aware selection based on Petri net unfolding algorithm [9]. However, none of these approaches considered the user custom setting of TS for the given workflow or domain specific solution during the composition.

TS services are pricier and consume more time than regular services. However, it provides recovery mechanism to avoid unexpected failures. In the real world, especially BDA service users equipped with diverse requirements, due to the budget and time constraints. Existing transaction aware service selection approaches provides TS and QoS aware for complete workflow. Risk levels of TS’s and priorities of TP’s are essential components in customization of TS-awareness. None of the existing methods are flexible to custom user selection settings for TS awareness to desired locations of the workflow such as critical stages. To overcome such concerns, we proposed novel,
- **TS Risk and TP prioritization** for the BDA process and
- **Customizable TS and QoS aware service selection** algorithm for the BDA planned to do in ASC. We used GA as based approach of multivariate optimization.

The remainder of this paper structured as follows. In Section II, we discuss the preliminaries. In Section III, we present the proposed method for the selection stage of the ASC process for BDA automation. In Section IV, we describe our implementation for evaluation of the proposed method. Section V concludes the paper.

II. PRELIMINARIES

In this section, we discuss the key techniques and technologies behind the overall solution. In this paper, we discuss only QoS preferences of the non-functional properties (NFP’s) and as QoS preference of the NFP as the TS awareness according to our study and experience in BDA domain. First, we explain the TSP for web services and composite web services. Next, we explain the multivariate QoS properties of the BDA selection process. Here onwards, we use QoS to imply the QoS.

A. Transaction aware Compensable Service (TS)

TSP is the behavioral NFP of the web services (WS). It measures the ability to level of successful accomplishment of given task and secure interactivity locally or globally. S. Mehrrota and H. Korth proposed a solid fundamental study on the transactional models [15]. Moreover, J. Hadad et al. discussed the semantic interpretation and how this transactional model utilizes to service selection domain [5]. Therefore, we inspired by studies of both [5], [15]. Based on their proposals, we lay the foundation of TS awareness of the services as follows under the section A.

**TSP’s of Web Services:** We have identified following properties as TS properties of WS. **Property 1:** Pivot WS (p): A WS is p if it can be accomplished its task successfully; however, it cannot be rollback and fixed after effects. If it failed, no effects at all. That means result cannot undo semantically. **Property 2:** Compensatable WS (c): A WS is c if it is available another service, which can semantically undo the execution of the given service. **Property 3:** Retriable WS (r): A WS is r, it guarantees the successful execution of given number of invocations. Then we can combine these primary properties and can make the secondary properties as follows;

**Property 4:** Pivot Retriable (pr): A WS, which is pivot and can be retrievable as well. **Property 5:** Compensatable Retriable (cr): A WS, which is compensatable and retrievable as well.

**TSP’s of Composite Web Services (CWS):** It is making a CWS to attain a common goal by coordinating set of WS’s. We have identified following TSP’s of the CWS as follows. **Property 6:** Pivot CWS (p): A CWS is p, if it is completed successful completion, it effects cannot be semantically undone, fixed and if it failed no effects at all. **Property 7:** Compensatable CWS (c): A CWS is c, if its component WS’s are compensatable. **Property 8:** Retriable CWS (r): A CWS is r, if its component WS’s are retrievable. Then we combine these primary TSP’s of the CWS and define secondary TSP’s of the CWS as follows; **Property 9:** Pivot Retriable CWS (pr): A CWS is pr, if it is pivot and retrievable. **Property 10:** Compensatable Retriable CWS (cr): A CWS is cr, if it is compensatable and retrievable.

Building Blocks of the Workflow

We have identified three fundamental components of the workflows occurred in BDA process. The **Sequential pattern, AND pattern** and **XOR pattern** as basic patterns used to make complex patterns in BDA as of our experience’s and literature review [16], [17]. Fig. 1 shows their graphical representations. We define these three types as follows.

**Sequential pattern (CWS):** Tasks T1 and T2 Sequentially executed. That means, T1 executes after successful completion of T1. Then WS1 and WS2 assigned to T1 and T2, these results in the CWS pattern, and it represents as WS1: WS2.

**AND pattern (CWS):** Tasks T1 and T2 execute in parallel. That means, it executes both T1 and T2 at the same time. WS1 and WS2 assigned to T1 and T2. It results in the CWS pattern, and it represents as WS1 || WS2.

![CWS Patterns](image-url)

Figure 1: Fundamental Workflow Patterns
**XOR pattern (CWS^j):** Tasks T_i and T_j, it executes either T_i or T_j. Then WS_i and WS_j assigned to T_i and T_j. This results in the CWS^j pattern, and it represents as WS_i | WS_j.

B. Quantitative QoS Properties of Service (QoS_QN)

Here below Table 1 summarized the QoS_QN properties of WS proposed to consider during the BDA service composition. These QoS criteria’s consider for the multivariate QoS optimizations using the GA.

- Assume, a workflow of a BDA is WF_i; it contains a set of CWS’s, WF_i = {CWS_1i, CWS_2i, \ldots, CWS_{ni}}, here n is the total number of CWS’s in the workflow.
- Available WS’s for each of the CWS, CWS_{12} = {WS_{11}, WS_{12}, \ldots, WS_{1y1}, WS_{21}, WS_{22}, \ldots, WS_{2y2}, \ldots, (WS_{ni}, WS_{n2}, \ldots, WS_{nyn})}, where WS_{ij} represents the jth candidate web service of CWS, and yn is the total number of candidate web services of CWS_i
- QoS values of WS_{ij} are q_{av}(WS_{ij}), q_{ub}(WS_{ij}), q_{re}(WS_{ij}), q_{pr}(WS_{ij}) and q_{pr}(WS_{ij}) for each of the candidate web services of S_{ij}, V_{ij1}, V_{ij2}, V_{ij3}, V_{ij4} and V_{ij5}, where 1 \leq i \leq n and 1 \leq j \leq yn.
- Weights for QoS criteria, w_1, w_2, w_3, w_4, and w_5 for q_{av}(WS_{ij}), q_{ub}(WS_{ij}), q_{re}(WS_{ij}), q_{pr}(WS_{ij}) and q_{pr}(WS_{ij}), respectively, where w_1 + w_2 + w_3 + w_4 + w_5 = 1.

(1), (2) and (3) use in GA algorithm in fitness and fittest value calculation process of the WS’s.

\[
F(ws_i) = 1 - \frac{3}{2} \left( \frac{v_{ij}^{max} - v_{ij}^{(ws_i)}}{v_{ij}^{max} - v_{ij}^{min}} \right) + w_i \left( 1 - \frac{2}{3} \left( \frac{v_{ij}^{(ws_i)} - v_{ij}^{min}}{v_{ij}^{max} - v_{ij}^{min}} \right) \right) (1)
\]

\[
F(cws_i) = \sum_{j=1}^{n} ws_i (2)
\]

\[
F(WF) = \sum_{j=1}^{m} cws_i (3)
\]

<table>
<thead>
<tr>
<th>QoS_QN Criteria of the WS</th>
<th>Functional Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability q_{av}(WS)</td>
<td>T_{t} / T_{a}</td>
</tr>
<tr>
<td>Throughput q_{ub}(WS)</td>
<td>R_{m} / T_{u}</td>
</tr>
<tr>
<td>Reliability q_{re}(WS)</td>
<td>N_{s} / N_{sf}</td>
</tr>
<tr>
<td>Price q_{pr}(WS)</td>
<td>Execution cost per unit</td>
</tr>
<tr>
<td>Time q_{ti}(WS)</td>
<td>T_{t} - T_{r}</td>
</tr>
</tbody>
</table>

Acronyms used in the table as follows.
- T_t represents the total amount of time available, and T_r represents the possible time during last T_t time
- R_m represents completed maximum requests and T_u represents unit time
- N_s represents the total number of successful invokes and N_{sf} represents the total number of successful and fail invokes during given time.
- T_t represents the request sent time and T_r represents results received time.

III. CTQS: CUSTOMIZABLE TRANSACTION AND QoS AWARE SERVICE SELECTION

Here we discuss the two proposed approaches under three sections. First, we explain the proposed architecture of the selection method. Next, we explain the custom transaction awareness for the service selection. Finally, we explain the multivariate optimization algorithm of the TS and QoS based on GA.

A. Architecture of the proposed CTQS

Fig. 2 displays the proposed architecture for the BDA automation [1]. We already achieved the planning and discovery stages of the ASC as of our previous works [3], [4]. In the beginning, user feeds the BDA requirement to the planning stage, and it creates a workflow for the data analysis. Next, this workflow forwards to the discovery stage of the ASC to discover the candidate services to the each task of the workflow. After that, the output of the discovery stage goes to the selection stage. In selection stage, our aim is to find the global optimal plan, which satisfies multivariate QoS of NFP’s and assured uninterrupted, error free, successful composition for available budget and time. Finally, the output of the selection stage, which is the ensured global optimal plan put forward to the execution stage to accomplish given BDA requirement. In this paper, we discuss only selection stage of the ASC process considering the concerns we have identified in BDA workflows and ASC process, neither execution nor other previous stages.

Next, we elaborate a sample workflow derived for the BDA. Fig. 3 displays the sample workflow with minimum complexity, which contained only basic patterns. Our data mining process is based on cross industry standard process for data mining. ASC proposed to automate last four stages of the data mining process. As shown in Fig. 3, data preparation stage, contained CWS_1^S and CWS_2^A patterns. In modelling stage, it contains CWS_1^X pattern, finally evaluation and deployment linked by CWS_2^S. We continue the explanation...
Data Mining Last 4 Stages of CRISP-DM

![Diagram of Big Data Analytics Workflow](image)

Figure 3: Big Data Analytics Workflow with Three basic Composition Patterns

and evaluation of our proposal based on this workflow. Let name this as the $WF_{BDA}$. Our aim is to derive the global optimal plan for the $WF_{BDA}$. Then, $WF_{BDA} = \{CWS_1^1, CWS_1^2, CWS_1^3, CWS_1^4\}$. CTQS method handles each composition of CWS’s as identical critical stages and allows BDA user to select which parts he needs to be assured by TS-awareness and QoS-awareness. Next section we discuss the proposed way to approach customization of TS-awareness.

B. Customizable Transactional aware Service Selection

In section II part A, we explained fundamental behind the TSP’s. Here, we explain the proposed method for the customizable TS (CTS) aware selection. Let define the risk levels and TSP priority identification of the for the BDA composition.

Risk and TSP Prioritization

According to the TSP’s explained in section II, A; define the Lemma to identify the levels of the risks of TP’s. Let define the risk levels and TSP priority identification of the for the BDA composition.

**Lemma 1 (Risk levels of WS):** TSP set of the WS, $S_{WS} = \{p, c, r, pr, cr\}$ for all candidate services which belong under each of $T_i$ tasks of the given workflow. Here $i \leq n$. $n$ is a number of tasks in the workflow. The descending order list of the risk level of the TS priorities of WS is $p, c, r, pr, cr$.

**Proof:** We prove this by contradictory reasoning.

  - According to the Property 1 of the WS, $p$ cannot rollback. Moreover, it tends to have an unsuccessful termination. It is only this property contained these two drawbacks. This is the worst-case can happen to a service. Then TSP’s $p$ is the worse TSP of the WS.
  - Property 2, c does not guarantee the successful execution rest of three other minimally guaranteed the successful execution due to retrievable property. However, it can undo. Therefore, it is better than $p$ but worse than other three of set $S_{WS}$
  - Property 3, $r$ does not guarantee the pivot retrievable or compensatable retriebility. Therefore, it is worse than $pr$ and $cr$. However, it ensures the successful execution after given number execution. Therefore, it is better than $p$ and $c$ properties.
  - Property 4, $pr$ does not guarantee the compensatable retriebility. It guaranteed pivot retriebility. Therefore, it is worse than $cr$. However, it consists of retriebility and pivot property. Therefore, it is better than other three properties of SWS except for $c$.
  - The Property 5, $cr$ it guaranteed retriebility and compensability. It is the best assurance can expect from given service. Therefore, $cr$ is the best TCP property of the set of properties contained in the SWS.

Then based on Lemma 1, we define the priorities of TP’s of the WS for BDA workflow as follows.

**Definition 1 (Priorities of the TSP’s of the WS):** Priority levels of TS awareness of the TSP’s of the WS’s according to ascending order as follows: $p < c < r < pr < cr$. That means A WS $cr$ is the most TS aware property of the WS and $p$ is the least TS aware property of the WS. Here onwards, we ignore the pivot ($p$) and retrievable ($r$) properties for BDA automation. Because $p$ does not provide assurance for the tasks of the workflow and $r$ guarantees the successful execution only after given number of executions. Then both of properties reduce the effective of the TS awareness of the BDA.

**Lemma 2 (Risk level of CWS):** TSP’s set of the CWS, $S_{CWS} = \{\bar{p}, \bar{c}, \bar{r}, \bar{pr}, \bar{cr}\}$ for all composite services which are belongs under each of $WP$ workflow. Here $WP$ contained $T_n$ number of tasks in the workflow. The descending order list of risk level of the TS priorities of CWS is $\bar{p}, \bar{c}, \bar{r}, \bar{pr}$ and $\bar{cr}$.

**Proof:** We prove this by deductive reasoning.

  - According to the Property 6 of the CWS, $\bar{p}$ means the collective results of individual $p$’s of WS’s of the CWS. Then collective combination of $p$’s result the $\bar{p}$. Then it inherited all $p$ qualities. Which are inability to undone and consistent execution.
  - Likewise, according to the rest of Properties (7, 8, 9, 10) of the CWS, $\bar{c}, \bar{r}, \bar{pr}$ and $\bar{cr}$ means the respective collective results of individual $c, r, pr$ and $cr$ of WS’s of the CWS. Then collective combination of individual $c, r, pr$ and $cr$’s result respective, $\bar{c}, \bar{r}, \bar{pr}$ and $\bar{cr}$. And, they inherited those TSP qualities as well. Therefore, according to the Lemma 1, CWS’s risk quality should be behaved as descending order of $\bar{p}, \bar{c}, \bar{r}, \bar{pr}$ and $\bar{cr}$.

Then based on Lemma 2, we define the priorities of the
Priorities

Definition 2 (Priorities of the TSP's of the CWS): Priority levels of TS awareness of the TSP's of the CWS's according to ascending order as follows: \( p < c < \overline{c} < \overline{p} \). That means, A WS \( \overline{c}r \) is the most TS aware property of the WS and \( \overline{p} \) is the least TS aware property of the CWS.

As we done like to TS of the WS, here also, for the same reasons we ignore the pivot (\( \overline{p} \)) and retrievable (\( \overline{r} \)) properties for BDA automation.

Based on definition 1 and definition 2, we elaborated the composition rules of the TSP's of WS and CWS to on behalf of guaranteed TS awareness of the BDA process.

- TS awareness of WS: Table 2 shows the summarized information on a combination of TS awareness of the TSP’s based on definition 1. We ignore the composition patterns, which are resulted by XOR pattern. Because, during the XOR, it results in one out of two based on the criteria. This applies to the CWS too. Then it results in another Sequential combination. Therefore, we consider only Sequential and AND patterns.

Rule 1 (Composition Rules of TSP’s of WS): According to the Table 2, we have shown the acceptable composition patterns for the successful WS composition of the BDA process based on ASC. Therefore, here we conclude the acceptable composition patterns as follows; Acceptable Sequential patterns are \( c : c, c : cr, cr : cr \) and \( cr : c \) only. Moreover, acceptable AND patterns are \( c // c, c // cr, cr // cr \) and \( cr // c \). These are only patterns that guarantee the end to end TS-awareness for the BDA process.

- TS awareness of CWS: Table 3 shows the summarized information of a combination of TS awareness of the TSP’s based on definition 2. Here we ignored all the combinations resulted by \( \overline{p}r \), which are not guaranteed the TS awareness for the BDA process.

Rule 2 (Composition Rules of TSP’s of CWS): According to the Table 3, we have shown the acceptable composition patterns for the successful CWS composition of the BDA process based on ASC. Therefore, here we conclude the acceptable composition patterns of the CWS as follows; Acceptable Sequential patterns are \( c : c, c : cr, \overline{c} : cr, \overline{cr} : cr \) and \( \overline{cr} : c \) only. And, acceptable AND patterns \( c // c, \overline{c} // \overline{c}, \overline{cr} // \overline{cr} \) and \( \overline{cr} // \overline{cr} // \overline{c} \). These are only patterns that guarantee the end-to-end TS awareness for the BDA process.

<table>
<thead>
<tr>
<th>TSP1</th>
<th>TSP2</th>
<th>Result of CWS</th>
<th>Result of CWS</th>
<th>Consideration to BDA on ASC</th>
</tr>
</thead>
<tbody>
<tr>
<td>pr</td>
<td>pr</td>
<td>( \overline{pr} )</td>
<td>( \overline{pr} )</td>
<td>No</td>
</tr>
<tr>
<td>pr</td>
<td>c</td>
<td>( \overline{pr} )</td>
<td>( \overline{pr} )</td>
<td>No</td>
</tr>
<tr>
<td>pr</td>
<td>cr</td>
<td>( \overline{pr} )</td>
<td>( \overline{pr} )</td>
<td>No</td>
</tr>
<tr>
<td>c</td>
<td>pr</td>
<td>( \overline{cr} )</td>
<td>( \overline{cr} )</td>
<td>Yes</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
<tr>
<td>c</td>
<td>cr</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
<tr>
<td>cr</td>
<td>pr</td>
<td>( \overline{cr} )</td>
<td>( \overline{cr} )</td>
<td>No</td>
</tr>
<tr>
<td>cr</td>
<td>c</td>
<td>( \overline{cr} )</td>
<td>( \overline{cr} )</td>
<td>Yes</td>
</tr>
<tr>
<td>cr</td>
<td>cr</td>
<td>( \overline{cr} )</td>
<td>( \overline{cr} )</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2: Composite TSP consideration of WS to the BDA

<table>
<thead>
<tr>
<th>TSP1</th>
<th>TSP2</th>
<th>Result of CWS</th>
<th>Result of CWS</th>
<th>Consideration to BDA on ASC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c )</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
<tr>
<td>( \overline{c} )</td>
<td>( c )</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
<tr>
<td>( \overline{cr} )</td>
<td>( \overline{c} )</td>
<td>( \overline{cr} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
<tr>
<td>( c )</td>
<td>( \overline{cr} )</td>
<td>( \overline{c} )</td>
<td>( \overline{c} )</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 3: Composite TSP consideration of CWS to the BDA

Level 1 (L1) Custom Sequential, AND Patterns and QoS\(_{QW} \) awareness: It guarantees the TS awareness for the selected Sequential and AND patterns in the BDA workflow. It will contain at least one of two patterns or multiples of given patterns.

Scenario 1: BDA user has WS\(_{BDA} \) and he needs to find a global optimal composition plan, which assured the TS-awareness in only during CWS\(_2 \) and CWS\(_3 \) places in the workflow while satisfying his multivariate QoS requirement. Then we proposed, L1 to satisfy scenario-1 selection requirement.

Level 2 (L2) Custom Sequential, XOR Patterns, and QoS\(_{QW} \) awareness: It guarantees the TS awareness of the selected Sequential and XOR patterns in the BDA workflow. It will contain at least one of two patterns or multiples of given patterns.

Scenario 2: BDA user has WS\(_{BDA} \) and he needs to find a global optimal composition plan, which assured the TS-awareness in only during CWS\(_2 \) and CWS\(_3 \) places in the workflow while satisfying his multivariate QoS requirement. Then we proposed, L2 to satisfy scenario-2 selection requirement.

Level 3 (L3) Custom AND, XOR Patterns and QoS\(_{QW} \) awareness: It guarantees the TS awareness of the selected Sequential and AND patterns in the BDA workflow. It will contain at least one of two patterns or multiples of given patterns.

Scenario 3: BDA user has WS\(_{BDA} \) and he needs to find a global optimal composition plan, which assured the TS-awareness in only during CWS\(_2 \) and CWS\(_3 \) places in the workflow while satisfying his multivariate QoS requirement. Then we proposed, L3 to satisfy scenario-3 selection requirement.
Level 4 (L4) Complete workflow assured by TS awareness and QoS\textsubscript{QN} awareness: It guarantees the TS awareness to the complete BDA workflow. Scenario 4: BDA user has WF\textsubscript{BDA}, and he needs to find a global optimal composition plan, which assured the TS-awareness in in the complete workflow while satisfying his multivariate QoS requirement. Then we proposed, L4 to satisfy scenario-4 selection requirement.

Level 5 (L5) Except TS awareness, only QoS\textsubscript{QN} awareness: It guarantees QoS\textsubscript{QN} awareness throughout the workflow automatically, not considered TS awareness. Scenario 5: BDA user has WF\textsubscript{BDA} and he needs to find a global optimal composition plan. He does not expect to the TS-awareness to the workflow and only looks forward to satisfying his multivariate QoS requirement. Then we proposed, L5 to satisfy scenario-5 selection requirement.

C. CTS and QoS\textsubscript{QN} based Genetic Algorithm

We elaborate the algorithm to find global optimal of the proposed method using Algorithm 1. In the selection stage of the ASC, it uses the result of discovery stage use as one input [4] and the custom level (L1, L2, L3, L4 or L5) of the selection criteria as the next. Feed the input to the proposed GA based Algorithm 1.

- Initialize Population: Line number 1 to 6, at the beginning it populates generation based on custom level requirements. Gene’s encoding scheme during initialization is shown in Fig. 4. Here, 1 ≤ i ≤ n; n is the number tasks in the workflow. j and k are j\textsuperscript{th} and k\textsuperscript{th} candidate services of respective tasks. It follows the composition rules definitions 1 and 2.

![Gene Encoding Scheme](image)

<table>
<thead>
<tr>
<th>Algorithm 1 QoS\textsubscript{QN} and Custom TS based Genetic Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> - Result Service and Tasks set of Discovery stage of ASC</td>
</tr>
<tr>
<td>- Selection Level 1/2/3/4/5</td>
</tr>
<tr>
<td><strong>Output:</strong> - Optimal Plan, - fittest_Qos</td>
</tr>
<tr>
<td><strong>BEGIN</strong></td>
</tr>
<tr>
<td>1. Initialize the Population</td>
</tr>
<tr>
<td>2. for each Individual in Population do</td>
</tr>
<tr>
<td>3. if check selection Level (1/2/3/4/5) then</td>
</tr>
<tr>
<td>4. populate individuals based on the Level requirement</td>
</tr>
<tr>
<td>5. end if</td>
</tr>
<tr>
<td>6. end for</td>
</tr>
<tr>
<td>7. Find the Best Individuals from the Initial population</td>
</tr>
<tr>
<td>8. Evolve Population</td>
</tr>
<tr>
<td>9. for n Individuals ∈ Population do</td>
</tr>
<tr>
<td>10. if elitism true then</td>
</tr>
<tr>
<td>11. elitismOffset set to 0 and save individual</td>
</tr>
<tr>
<td>12. end if</td>
</tr>
<tr>
<td>13. for each Individuals start from elitism offset to crossover do</td>
</tr>
<tr>
<td>14. create parent1 and parent2 from tournament selection</td>
</tr>
<tr>
<td>15. Probabilistically crossover parent1, parent2 &amp; create new child and add in to population</td>
</tr>
<tr>
<td>16. end for</td>
</tr>
<tr>
<td>17. for each Individuals start from elitism offset to mutate do</td>
</tr>
<tr>
<td>18. Probabilistically cross over the each pairs</td>
</tr>
<tr>
<td>19. end for</td>
</tr>
<tr>
<td>20. return new population</td>
</tr>
<tr>
<td>21. end for</td>
</tr>
<tr>
<td><strong>END</strong></td>
</tr>
</tbody>
</table>

IV. EXPERIMENT AND EVALUATION

To evaluate the behavior of our selection approaches, we conducted the extensive experiment on our approach. We focused on the accuracy and the efficiency of our approach. As of our studies in the domain, this is the first approach to QoS criteria’s and custom TS-aware.

For the accuracy evaluation, we have two main perspectives. First, it evaluates the accuracy of the QoS optimization of our proposed method. Next, we have to evaluate the accuracy of custom level methods. To simulate the evaluation of the first user case, we conducted an evaluation of our approach with a leading swarm intelligent multivariate optimization algorithm, which is the D. Karbagoda and B. Basturk proposed artificial bee colony algorithm (ABC) [18]. It is one of the most cited AI based multivariate optimizers. We prepare multivariate QoS based selection approach based on D. Karbagoda proposed method and conduct the overall accuracy of QoS optimization (scenario 5) of the proposed approach. For the second user case, we conduct each custom levels L1, L2 and L3 then find their average and compare their average with L4 to evaluate their Transaction and QoS aware ability.

For the efficiency evaluation, we consider evaluating the processing time. For that, at first, we conduct L5 method with ABC to show the performance of the comparative methods. Moreover, to evaluate the internal efficiency, we conduct above-mentioned second test case with changing a number of tasks and candidate services in the workflow.

We used our experiment platform as Intel Core i7, Windows 8.1, 16GB RAM computer and Java 1.8 Enterprise edition. We used the 1000 services information that we already used in the previous method [4]. We used Fig. 3 shown scenario as of our workflow pattern for the evaluation. In the algorithms, for ABC, we set run time as 100, colony size to the two times of a number of plans, number food sources is equal to the number of plans, a number of employee bee and onlookers are equal to the colony size, one scout bee and 100 as the max cycle. We found these are optimal for the multivariate optimization for our selection. For our GA method, we use 100 as of initial population; tournament size 50 to 100, and mutation rate is 0.01. We conducted five times for each test cases and user the average as that the result of the
particular test case.

A. Evaluate the Accuracy of the proposed method

Here we focused on evaluate the overall accuracy of the QoS criteria’s except compensability of the proposed method. According to the level information, it’s the L5. We compare the L5 with the ABC method as shown in Fig. 5. To find the accuracy of the result, we measure deviation from the pre-identified global solution of multivariate QoS criteria’s. For that, first, we set the candidate services according to the ascending order of (1). Next, we calculate the error of deviation from global solution by measuring the deviation from each resulted candidate service from given approach and aggregate each error deviations. It results in a percentage of total error deviation. This reduced by 100 and find accuracy. Fig. 5 indicates proposed method outperform the ABC method by accuracy. It clearly shows that our method is performing very well in the multivariate optimization of the QoS criteria’s.

Next, we compared the average of L1, L2, and L3 vs. L4. Here L1, L2, and L3 represent critical stages, and we simulated scenario 1, 2 and 3. We maintained two critical stages for each scenario for L1, L2, and L3 user cases. Moreover, take an average of them to make a single representation of them. As we measure the accuracy of the above method here, also we find the accuracy with the pre-identified global solution of TS and QoS awareness. Next, it calculates the error deviation and finally gets the accuracy. Fig. 6 shows accuracy comparison of this user case. It shows the accuracy of compensability and QoS between L1, L2, and L3 vs. L4. It shows the highest accuracy scored by L4 mode. That means when it is increasing number of critical stages of the workflow, our proposed method doing better. It ensures the proposed method perform better to find global optimal when it perform with more critical stages.

B. Evaluate the Efficiency of the proposed method

To assess the processing efficiency of the proposed method, we conduct an above-mentioned 1st experiment and find the execution time. For the next experiment, we consider L1, L2, and L3 one instance. Then we compare the execution time of that with L4 and L5 by changing the number tasks and candidate services. Respective Fig. 7, 8 and 9 shown results in those user cases.

Fig. 7 shows the ABC with combined L1, L2, L3 vs. L4, vs. L5 while increasing number candidate services with fixed number tasks. All L test results laid on each other. It shows significantly lowest maintained in all L test cases while increasing candidate services. However, ABC shows the exponential growth and significant gap between all L cases. That means, our proposed method remain almost unchanged. This caused; because it remains unchanged, the number of tasks in the workflow, population size and number of the evolution of the proposed GA based method. On the other side, a number of candidate services for each task is the main reason to affect the execution time of the ABC.
Fig. 8 shows the combined $L_1$, $L_2$, and $L_3$, vs. $L_4$ vs. $L_5$ with fixed number of tasks and increasing number of services. All remained lower execution time. However, $L_4$ shows the highest execution time. Moreover, $L_5$ shows lowest possible. This caused; $L_4$ scenario, it needs to find the services, which are satisfying composition rules 1 and 2 to all of the tasks to make a complete individual from existing services. However, the $L_5$ case does not have such a constraints to make the individual genomes. Therefore, it shows lowest possible execution. $L_1$, $L_2$, and $L_3$ have to limit 2 critical stages to satisfy their TS-awareness during preparing individuals. Therefore, it maintains lower execution time compared to the $L_4$ and higher than $L_5$.

Fig. 9 shows above experiments with fixed number of services and increasing number of tasks. It gives same results as shown in Fig. 8. However, Fig. 8 maintains convex shape while Fig. 9 has concave shapes in combined $L_1$, $L_2$, $L_3$, and $L_4$. That means Fig. 9 has higher derivatives of the derivatives than Fig. 8. This implies number tasks are more effective to increase execution time than candidate services.

According to the above experiments results, CTQS is more effective and efficient in the perspective of accuracy and processing time. And, it works better to find the global optimal in the BDA automation based on ASC process.

V. CONCLUSION

We proposed novel approach to customizable TS awareness for the BDA automation on ASC by considering diverse requirements of the BDA process. Experiment results show we proposed method is outperform existing multivariate optimization method and reaches the goals while consuming minimal resources. That means, TCQS perform better and reach the global optimal, very efficient manner. Our aim was to proposed flexible and assured composition method for BDA automation based on ASC. Experiment results show that our method is highly effective and efficient to achieve the desired goal. As the future works, we are working to address the uncertainty and concerns occurred in trustworthy of the service domain and workflow adaptability for more effective BDA automation based on ASC.

REFERENCES