

# A Service Oriented Analytics Framework for Multi-Level Marketing Business

Rich C. Lee<sup>1,2</sup>

<sup>1</sup>System Technology Group, IBM, Taipei, Chinese Taipei; <sup>2</sup>Department of Information Management, National Sun Yat-Sen University, Taipei, Chinese Taipei.  
Email: rich.chih.lee@gmail.com

Received May 2<sup>nd</sup>, 2012; revised May 31<sup>st</sup>, 2012; accepted June 10<sup>th</sup>, 2012

## ABSTRACT

Today's enterprises have accumulated vast amount of data and keep exploding by business activities. These datasets may contain potential undiscovered business strategies as a key basis of competition; underpin new waves of productivity growth, innovation, and consumer surplus. Data analysis is crucial in making managerial decisions. Although there are many Business Intelligence (BI) software of commercial and open source, but serving statistical purpose as exuberant as GNU-R (R) is rare. R is a highly extensible language and environment for providing a variety of statistical and graphical features. In enterprise environment, the source data are stored in various forms such as files, database, and streaming data. Currently analysts conduct data analysis in offline mode using statistical software. The offline mode means analysts 1) extract the desired data; 2) store extracted data into files; 3) manipulate software; 4) draw analytical results; 5) generate the inferences. Automating the statistical procedures by directly pulling source data will make critical decision sooner and less costly. It is a common practice that enterprise adopts Service-Oriented Architecture (SOA) to achieve its operation excellence. Since the business applications populate the source data during the operation processes, pulling the source data directly under SOA is the most effective way of data analysis. This paper demonstrates how service-oriented statistics engine was developed and how such a system benefits the business decision-making.

**Keywords:** Software Engineering; SOA; Business Analytics; Big Data; GNU-R

## 1. Introduction

The *Big Data* is a short hand term for advancing technology trends that open the door to a new approach to disclose the meaning behind the business activities and then based on the findings to make decisions with quality. Enterprise performance relies on the responsiveness and the quality of decision-making. Decision-making is based on rationale of the knowledge which enterprise has possessed. Enterprise knowledge is accumulated and compiled by the information, and the information is generated from the raw data by business activities. The decision makers have to develop a feasible approach to acquire sufficient knowledge and on the situation where makes a confident enough decision to support business needs [1]. Today's good decisions are driven by reliable data. Business managers and professionals are increasingly required to justify decisions based on data. They need statistical model-based decision support systems [2]. Enterprise strives on process improvement and variation reduction; it requires a disciplined, data-driven approach to reach business goals. Statistics helps enterprise describing in quantitative way about how well the processes are performing, transform raw data to valuable in-

formation, and later become useful knowledge.

Analyzing data, deriving information, and accumulating knowledge is not handy in most enterprise's information environment currently. First analysts must know where the desired data is, and ask IT professionals to retrieve that data, save it into files, store them under a shared folder over network, and then develop statistical models. If data is considerably large and fast-updated, it will make the statistical process more inconveniently. If other analysts concern just a portion of the data or the data is not fully covered, IT professionals must repeat these resource-exhausting tasks again for each of them respectively. Another drawback is the poor reusability if the statistical procedures are not shared. In fact, enterprise knowledge does not just cover the information but also the processes of how information was derived. These processes ensure the reproducibility of knowledge that is a part of the enterprise intelligent properties. To reach this objective, a software application—the Business intelligence (BI)-is used to analyze enterprise data; it requires expertise to design and manage those analytical models, it is usually rather expensive and complicated requiring IT professionals and software vendor's inten-

sive assistance. Not like commercial BI software, GNU-R is a flexible and extendable language and environment for statistical computing and graphics; it has been widely used in many applications for years. The analysts who are planning to use GNU-R can find abundant resources over the Internet to ease their learning curves. Thus, GNU-R is more cost-effective as a statistical engine automating analytical processes than commercial BI software. It is worthy to develop a statistical service engine solution using R to improve enterprise knowledge generating and reusability.

However, developing such statistical service engine solution to meet time-to-market objective within budget, it is important to know why most software projects failed: 1) unrealistic or unarticulated project goals; 2) inaccurate estimates of needed resources; 3) badly defined system requirements; 4) poor reporting of the project's status; 5) unmanaged risks; 6) poor communication among customers, developers; and users 7) use of immature technology; 8) inability to handle the project's complexity; 9) sloppy development practices; 10) poor project management; 11) stakeholder politics; 12) commercial pressures [3]. The third factor attracts the most concerns; mainly because during the software development it has not considered the non-functional requirements during the development. These non-functional requirements are to ensure the software capability does meet business goal with less Total Cost of Ownership (TCO).

The Statistical Service Engine solution helps analysts offer their business's performance perspectives to enterprise, and enhance their daily decision making by reusing these statistical inferences, in a more effective way. Although GNU-R has the ability to access the databases directly from analysts' desktops, but such an approach might cause security breach if database schemas are disclosed, and potentially jeopardize the performance of databases to serve the designated business activities if databases were inappropriately accessed. On the other hand, the analysts expect the solution will eventually execute the submitted GNU-R scripts on behalf of them at most appropriate timing. There for the solution must be under a unified flexible reliable robust mechanism integrating the backend processes—running GNU-R scripts on distributed servers.

To adopt such a unified flexible reliable robust mechanism, the Enterprise Service Bus (ESB) is a proven approach to meet the requirements; it is an infrastructure which underpins a fully integrated and flexible end-to-end Service-Oriented Architecture (SOA). The ESB enables SOA by providing the connectivity layer between services [4]. The ESB combines event-driven and service oriented approaches to simplify the integration of business units, bridging heterogeneous platforms and environments. The ESB acts as an intermediary layer to en-

able communication between different application processes. A service deployed onto an ESB can be triggered by a consumer or an event. It supports synchronous and a synchronous, facilitating interactions between one or many applications (One-to-One or Many-to-Many communications) [5], this is a vital scheme to the statistical service engine solution.

To maximize the reusability and the influence of statistical inferences to enterprise, the user-friendliness of the solution is very crucial to the acquisition and dissemination of business knowledge. The user-friendliness increases user's perceived ease-of-use against a system. Accumulating knowledge is a complex and dynamic process; it needs continuously reshaping the appearance of the knowledge by enhancing the solution's usability. Usability, a synonym of user-friendliness, is a core term in human-computer interaction. The platform portability is another important consideration of technology selection for the solution. Java is well-known for better portability, and its *portlet* technology—a proven approach of better usability—can enable users to dynamically construct and reconstruct Web applications of information convergence in run time to resolving urgent and unplanned business requirements [6].

## 2. Quality Attributes of Statistical Service Engine Solution

It is obvious and important that the Statistical Job Engine software requires high quality because business decisions depend on it. Software quality needs attributes to measure and evaluate to meet business requirements. In Software Engineering perspective, the requirements cover 1) functional; 2) non-functional. These Quality Attributes are for those non-functional requirements covering: 1) functionality; 2) usability; 3) reliability; 4) performance; 5) supportability [7]. The solution especially focuses on the reliability and supportability attributes. Reliability is “a set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time”, it consists of three sub-factors: 1) Maturity is the frequency of software faults; 2) Fault Tolerance is the ability of software to deal with software faults or infringement of its specified interface; 3) Recoverability is the capability to recover data affected in case of a failure and measured by the time and effort needed for it [8]. Supportability includes 1) Testability; 2) Extensibility; 3) Adaptability; 4) Maintainability; 5) Compatibility; 6) Configurability; 7) Serviceability; 8) Installability; 9) Localizability; 10) Portability [7]. **Table 1** exhibit the most concerned attributes as follows.

## 3. Statistical Service Engine

GNU-R provides a powerful programming language and

**Table 1. Quality attribute requirements.**

<i>Quality Attribute</i>	<i>Requirement</i>	
<b>Reliability</b>	Maturity	Using Proven Technologies.
	Fault Tolerance	Job Execution at All Times.
	Recoverability	Job Automatically Resumed When Failed.
<b>Supportability</b>	Testability	Feature Self-Tests at Any Time.
	Extensibility	Adding More Computing Resources at Any Time.
	Adaptability	Standard Protocols of Service Integration.
	Maintainability	Providing a Management Console.
	Compatibility	None
	Configurability	Providing a Management Console.
	Serviceability	7 × 24 Operations.

a statistical environment; it is very extensible, with over hundreds of add-on packages obtainable from CRAN (the “Comprehensive R Archive Network”), providing high quality of graphical output [9]. The *r-base-core* is a universe/math package described as a GNU R core of statistical computation and graphics system. GNU R script can access databases by *r-cran-rodnc* package through *Open Data Base Connectivity* (ODBC). The package should be platform independent and provide access to any database for which a driver exists [10]. **Table 2** shows how R script access database through ODBC. In many occasions, the source data for statistical analysis is sophisticated and requires complex process to generate it. Using Java *Servlet* technology is a common practice. Using *Servlet* to populate complex data from various sources and transform them into desired format. The HTTP output format should be text/plain with tab delimiter for data columns. **Table 3** shows how R script gets data from web. Graphics help user to comprehend the trend and the meaning behind statistical figures intuitively. Normally statist uses tool to visualize the statistical graphics; the Statistical Job Engine runs GNU-R scripts on behalf of the statist, it must save the graphics to image file so that the statist can visualize the image file later. **Table 4** shows how R script saves graphics into image file. Java program has two approaches of executing GNU-R script 1) using *Runtime.getRuntime().exec()*; 2) using *JRI* as the interface to GNU-R. It allows Java program to take control over R. **Table 5** shows how Java program interacts with R.

#### 4. The Architecture of Statistical Service Engine Solution

Based on the above-mentioned required Quality Attributes, Statistical analysis takes time and consumes computing resources; it means the response time varies from

the complexity of the task. Asynchronous messaging tactics is the solution for time-consuming requests. In asynchronous messaging, the process response does not return immediately but some time later in a different communication session. The user does not wait for the process result at the request time but retrieve the result later after the called process finished or failed from another request [11]. Asynchronous messaging relies on message-oriented middleware (MOM). The MOM creates a “software bus” for integrating heterogeneous applications [12]. Messages are distributed across applications running on different servers; its reliability and robustness is crucial to MOM. An Enterprise Service Bus is an open standards, message based, distributed integration infrastructure that provides routing, invocation and mediation services to facilitate the interactions of disparate distributed applications and services in a secure and reliable manner [13]. **Figure 1** illustrates the system architecture of the Statistical Service Engine.

The Statistical Job Portal server provides a number of predefined statistical procedures and ad-hoc analysis to users. The Enterprise Service Bus server receives messages, the statistical job requests, from the Statistical Job Portal server and dispatches them accordingly. GNU-R Engine is a set of Blade servers executing GNU-R scripts designated by the messages. The script may retrieve data from the Database servers or the File Repository server. The application servers are user process engines populating source data on the Database servers or the File Repository server. The File Repository server also stores the statistical results expecting users to retrieve later. The statistical job request is in XML format covering the following fields shown in the **Table 6**.

**Figure 2** illustrates the Activity Diagram of the solution. There are two routes: 1) Predefined; 2) Ad-hoc Statistical Job. When user select a predefined job from the

**Table 2. GNU-R uses ODBC example.**


---

1	<code>myChannel &lt;- odbcConnect("ResellerTransaction")</code>
2	<code>sqlSave(myChannel, PerfInfo, rownames="EventDate", verbose=TRUE)</code>
3	<code>sqlQuery(myChannel, paste("SELECT EventDate, CPU from PerfInfo", "WHERE EventDate &gt; '2012/04/27' ORDER BY EventDate"))</code>
4	<code>close(myChannel)</code>

---

**Table 3. GNU-R reads data from HTTP.**


---

1	<code>mySourceData &lt;- readtable("http://172.16.5.23/getResellerData")</code>
---	---

---

**Table 4. GNU-R saves graphics to image file.**


---

1	<code>setwd('/data/images')</code>
2	<code>reset &lt;- par(no.readonly=TRUE)</code>
3	<code>par(mfrow=c(1,3))</code>
4	<code>plot(x[,1],x[,2],xlim=c(min(x[,1],x[,2]),max(x[,1],x[,2])), ylim=c(min(x[,1],x[,2]),max(x[,1],x[,2])))</code>
5	<code>plot(x[,1],x[,3],xlim=c(min(x[,1],x[,3]),max(x[,1],x[,3])), ylim=c(min(x[,1],x[,3]),max(x[,1],x[,3])))</code>
6	<code>plot(x[,2],x[,3],xlim=c(min(x[,2],x[,3]),max(x[,2],x[,3])), ylim=c(min(x[,2],x[,3]),max(x[,2],x[,3])))</code>
7	<code>savePlot('myPlot',type="jpg")</code>
8	<code>par(reset)</code>

---

**Table 5. JRI interacts with GNU-R example.**


---

1	<code>Engine re=new Engine (args, false, new TextConsole());</code>
2	<code>REXP x;</code>
3	<code>re.eval("print(1:10/3)");</code>
4	<code>System.out.println(x=re.eval("iris"));</code>
5	<code>RVector v = x.asVector();</code>
6	<code>if (v.getNames()!=null) { System.out.println("has names:"); for (Enumeration e=v.getNames().elements(); e.hasMoreElements(); ) { System.out.println(e.nextElement()); } }</code>
7	<code>if (true) { System.out.println("Console:"); re.startMainLoop(); } else { re.end(); System.out.println("end"); }</code>

---

Statistical Job Portal, the responded portal event sends the Statistical Job Request message to Message Queue. One of the GNU-R Engine server gets the routed message from Enterprise Message Bus and does the follow-

ing tasks:

- 1) Predefined Statistical Job:
  - a) Loads predefined GNU-R script.
  - b) Replaces the paramters (*i.e.* ResultDestinamtion,

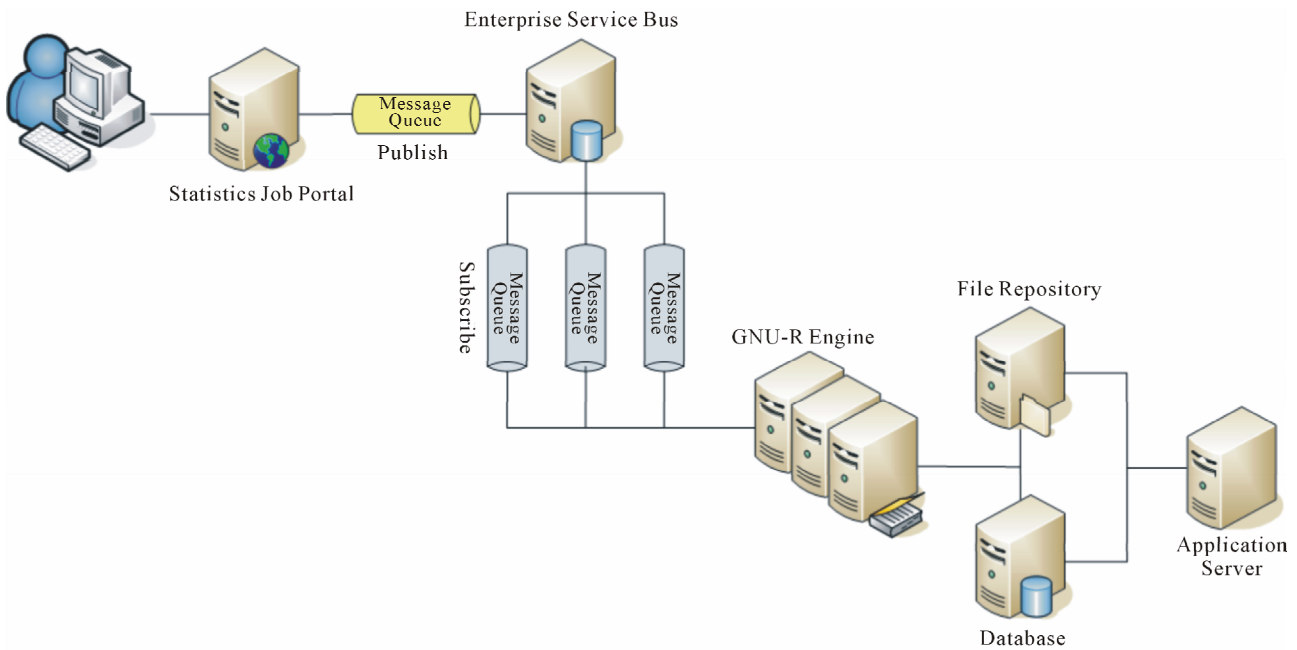


Figure 1. Statistical service engine solution.

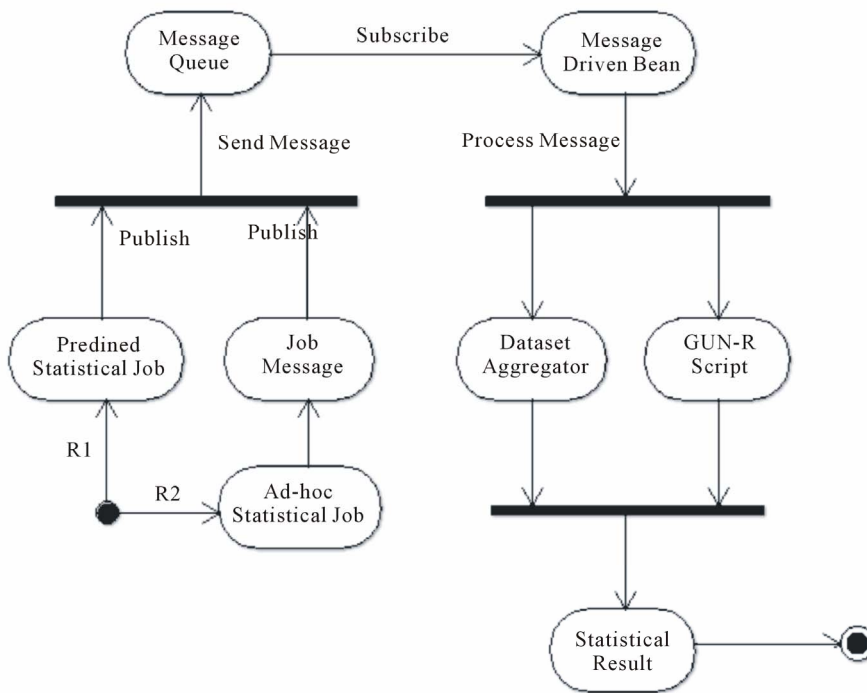


Figure 2. Activity diagram of statistical service engine solution.

ServerID, and DateRange, etc.).

c) Spawns GNU-R interpreter to run the prepared script.

d) Retrieves data from File Repository and Database.

e) Populates the statistical result on File Repository.

f) If *Result Destination* > 0 emails the statistical result to the requested user.

2) Ad-Hoc Statistical Job:

a) Loads the GNU-R script from the message received.

b) Spawns GNU-R interpreter to run the prepared script.

c) Retrieves the designated data from File Repository and Database.

d) Populates the statistical result on File Repository.

**Table 6. Statistical job request.**

<i>Field</i>	<i>Description</i>	<i>Example</i>
<b>User ID</b>	User Identification	richlee@tw.ibm.com
<b>Submit Date Time</b>	Job Submitted Date/Time	20120427-164530
<b>Job ID</b>	Statistical Job Identification	Statistics Summary: 00: Ad-hoc 01: Reseller 02: Level 03: Sales Regression 11: Sales~Level Moving Average: 21: Total Sales 22: Sales by Level
<b>Ad-Hoc Script</b>	GNU-R Script	If Job ID==00
<b>Source Data</b>	Statistics Source Data	If Job ID==00, /DB1205/20120531.RData
<b>Result Destination</b>	Statistics Result Destination	0: Web 1: Email 2: Both
<b>Server ID</b>	Server Identification	DB1205
<b>Date Range</b>	Date Range of Performance Counters	20050901 ~ 20120531

e) If *Result Destination* > 0 emails the statistical result to the requested user.

The user can view the statistical result from the Statistical Job Portal under [My Job Results] web page if *Result Destination!* = 1. The web page [My Job Results] lists down the statistical results of previous submitted jobs with hyperlinks.

## 5. Empirical Case

More people are accepting the self-care concept in natural health regimens as the complementary and alternative medicine has become a key driver for wellness foods industry business. Many wellness food companies rely on multi-level marketing—directly selling products through individuals' social networks—for their business development. However, promoting wellness foods via multi-level selling model has been a challenge for years to business owners including both companies and the associated resellers. The research subject is a very successful biological company selling a series of products containing Reishi and uses the multi-level marketing as their primary tool of business development. This company faced high competition from the wellness foods market, and the products sales did not grow as expected. So, the company wished to disclose which link had gone wrong; based on the research findings, to develop associated business policies to secure its market position. This company encourages the resellers to pursue higher level position for better product discount rates. In the beginning, the resellers who were also consumers too, they move up to higher levels—bronze, silver, gold, platinum, and titan—in turn by the increasing the sales amount and developing deeper levels in distributing channels. Regardless in which level, some resellers became hibernate

for not having any transaction for at any giving period. Once a hibernated reseller resumes the sales, the reseller is back to the original level and gets the discounts accordingly. Few resellers were able to move up to next level position successfully, while most resellers remain staying in the current level. The research objective was to identify whether the current distributed channels have any pattern; if there was any, what would be the pattern.

Since this company was looking for a possible pattern among these discrete states—the resellers role change, the paper mapped the business questions onto the Discrete-Time Markov Chain (DTMC) problem domain. The DTMC is a random process that moves among various states in a “memory-less” fashion. The analysis of DTMC problem depends on a Transition Diagram and maps the diagram onto a corresponding *Transition Matrix*, so that the researcher can solve the problem by manipulating the matrix calculation. The DTMC concerns about whether the state will become stable and how many terms that each state needs to shift to others respectively in the run-run. Let  $P_{ij}$  denote the probability of moving from state  $i$  to state  $j$  during a time step. The *Transition Matrix*  $P$  is in  $m \times n$  dimension,  $1 \leq i \leq m$ ,  $1 \leq j \leq n$ . If there exists an  $n$  elements steady-state row vector  $\mu = [\mu_1 \ \mu_2 \ \dots \ \mu_n]$ , such that:

$$\mu \cdot P = \mu \quad (1)$$

$$\sum_1^n \mu_j = 1 \quad (2)$$

If there is a solution for the linear Equations (1) and (2), then the steady-state vector exists. On the other hand, three column vectors of the same  $m$  elements  $M$ ,  $M_0$ , and  $I$ ; while  $M$  contains the terms required for state  $S_i$  changes to other states  $S_j$  respectively;  $M_0$  is derived from

$M$  that excludes the term for recurrence of state  $S_i$ ; and all the elements of  $I$  are ones, such that:

$$M = \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{ii} \\ \vdots \\ m_{ij} \end{bmatrix} M_0 = \begin{bmatrix} m_{i1} \\ m_{i2} \\ 0 \\ \vdots \\ m_{ij} \end{bmatrix} I = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \quad (3)$$

$$M = I + P \cdot M_0$$

By solving the linear equations derived from (3), the terms required for state  $S_i$  shifting to other states  $S_j$  will be disclosed respectively. If  $S_i$  is an absorbing state, it means that there is a probability of  $P_{ii}$  for continuously staying in the state  $S_i$ .

This company has several datasets with vast amount of records about the resellers and the associated transactions. This paper exported these records for past decade into separate annual aggregated dataset files; the files were named after the year of the exported records respectively. There were two additional marks as the file name suffix; using  $P$  was for new product introduced,  $B$  was for bonus plan changed, and  $PB$  if both cases happened. Each annual aggregated dataset consists of the following information: 1) *Reseller ID*; 2) *Reseller Level*; 3) *Date of Level Promoted*; and 4) *Sales Amount*. By issuing the following SQL statement against each year’s dataset file to get the number of resellers and the sales amount for each level:

```
“SELECT RESELLER_LEVEL, COUNT (RESELLER_ID) AS NUM_ALL_RESELLER, SUM (SALES_AMOUNT) AS LEVEL_AMOUNT FROM RESELLER_TRANS GROUP BY RESELLER_ID”
```

This paper conducted a series of statistical processes of ANOVA (Analysis of Variance) to see if there was any difference between ordinary datasets and those files with additional marks. The result showed that it did make difference when new products was introduced to the market or the bonus plan was changed by company’s policy. By issuing a series of SQL statements as follows against each year’s dataset file to get the proportions of each level:

```
“SELECT RESELLER_LEVEL, COUNT (RESELLER_ID) AS NUM_HIBERNATE_RESELLER FROM RESELLER_TRANS WHERE SALES_AMOUNT = 0 GROUP BY RESELLER_ID”
```

```
“SELECT RESELLER_LEVEL, COUNT (RESELLER_ID) AS NUM_PROMOTED_RESELLER FROM RESELLER_TRANS WHERE YEAR (PROMOTED_DATE) = @DATA_YEAR GROUP BY RESELLER_ID”
```

Therefore, the number of resellers remained in their current levels— $NUM\_RESELLER$ —was calculated by the following formula:

$$NUM\_RESELLER = NUM\_ALL\_RESELLER - (NUM\_HIBERNATE\_RESELLER + NUM\_PROMOTED\_RESELLER)$$

By using these three numbers— $NUM\_RESELLER$ ,  $NUM\_PROMOTED\_RESELLER$ , and  $NUM\_HIBERNATE\_RESELLER$ —this paper got the probabilities of each state respectively. **Figure 3** illustrated the Markov Chain Transition Diagram and **Figure 4** illustrated the associated Transition Matrix for the business problem. According to Formula (2), the probability of Hibernate for a state can be derived by one minus the sum of other states’s probability. To solve this business problem, this paper applied the GNU-R’s capabilities of Matrix Calculation and Linear Equations Solution. **Table 7** illustrated how GNU-R manipulate the matrix and find the solution of linear equations. Another example showed a GNU-R script to calculate the visiting times of each state and the average travel times before shifting into an absorbing state, and the probabilities of non-absorbing states shifting to those absorbing states. The absorbing state in DTMC is a state that never shifts out once entering it. To do the calculation, the analyst needs to transform the transition matrix— $m$  denotes the number of transition states and  $n$  is number of absorbing states—into an absorbing form consisting of four quadrants: 1)  $Q$  is a square matrix of  $m$ ; 2)  $R$  is non-zero matrix of  $m \times n$ ; 3)  $I$  is an identity matrix of  $n$ ; and 4)  $0$  is a zero-matrix of  $n \times m$ . The example does the rest of the calculation automatically. **Table 8** illustrated the GNU-R script for the calculation.

This paper disclosed the business problem which had been a long-time mystery to their operation of the company. By taking advantage of the automation of the whole analytical process, the company is able to adjust

**Table 7. Linear equation solution sample.**

Solv	$a_{11}x + a_{12}y = a_{13}$	$a_{21}x + a_{22}y = a_{23}$
1	myEq <- array(0, c(2, 2))	
	myEq[1, 1] = 1	# $a_{11}$
	myEq[1, 2] = 1	# $a_{12}$
2	myEq[2, 1] = 1	# $a_{21}$
	myEq[2, 2] = -3	# $a_{22}$
3	myEqVal <- array(0, c(2, 1))	
	myEqVal[1, 1] = 1	# $a_{13}$
4	myEqVal[2, 1] = 0	
5	myAnswer <- solve(myEq, myEqVal) print(myAnswer)	

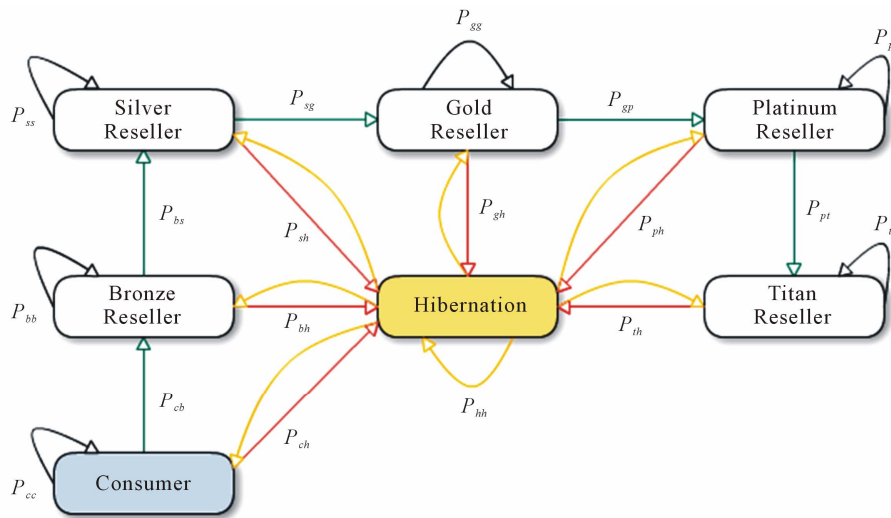


Figure 3. State transition diagram.

	Consumer	Bronze	Silver	Gold	Platinum	Titan	Hibernate
Consumer	$P_{cc}$	$P_{cb}$	0	0	0	0	$P_{ch} = 1 - (P_{bb} + P_{cb})$
Bronze	0	$P_{bb}$	$P_{sg}$	0	0	0	$P_{bh} = 1 - (P_{bb} + P_{bs})$
Silver	0	0	$P_{ss}$	$P_{gp}$	0	0	$P_{sh} = 1 - (P_{ss} + P_{sg})$
Gold	0	0	0	$P_{pp}$	$P_{pt}$	0	$P_{gh} = 1 - (P_{gg} + P_{gp})$
Platinum	0	0	0	0	$P_{tt}$	$P_{pt}$	$P_{ph} = 1 - (P_{pp} + P_{pt})$
Titan	0	0	0	0	0	$P_{tt}$	$P_{th} = 1 - (P_{tt})$
Hibernate	$P_{hc}$	$P_{hb}$	$P_{hg}$	$P_{hp}$	$P_{ht}$	$P_{ht}$	$P_{hh} = 1 - (\sum P_{hi})$

Figure 4. Transition matrix.

Table 8. Example of calculating an absorbing matrix.

1	<pre>I &lt;- array(0, c(n, n)) # Assign Values to I Q &lt;- array(0, c(m, m)) # Assign Values to Q</pre>	
2	$N = \text{solve}(I - Q)$	$N = (I - Q)^{-1}$
3	$T = N \%*\% I$	$T = N \cdot I$
4	<pre>R &lt;- array(0, c(m, n)) # Assign Values to R B = N \%*\% R</pre>	$B = N \cdot R$
5	<pre>Print(N) Print(T) Print(B)</pre>	

its policy to better business result. These business findings inspired the company to dedicate more resources to develop and explore the undisclosed business intelligence. The proposed Statistical Service Engine Solution makes this automation more seamless and handy for the analysts and the executives to make business decision swiftly with quality.

### 6. Conclusions and Future Work

To increase business competitiveness requires continuous

innovation and operation excellence. Re-examining the data business activities and finding patterns and trends in various statistical perspectives help enterprise making rational decisions swiftly. Knowledge Management offers a collaborative platform to acquire, compile, disseminating, and reuse the knowledge to elicit creative and improve operation efficiency. A more intense use of knowledge management platform has both a direct and indirect (innovation-mediated) positive effects on enterprise performance [14]. To make knowledge management platform success, the users' perceived usefulness and user satisfaction is the key [15]. Based on useful statistical results, it disclosed the implications of business competitiveness improvement will stimulate and inspire employees' further finding by reusing these statistical procedures and the data. This reinforced process makes employees use knowledge management platform more intensively and help them making business decision more rational and swiftly. To realize this goal, a more convenient solution is called to help analysts retrieve data, reuse statistical procedures, and disseminate the findings easier. This paper answered such need and presented a reliable robust architecture on top of Enterprise Service Bus; it also showed how GNU-R scripts access vivid data from



databases, generate statistical graphics and save them to image files. In software engineering perspective, this paper demonstrated how to define and measure the quality attributes of the solution to ensure the solution will meet business requirements during the development cycle.

The proposed solution is still very primitive but workable. More useful statistical models should be developed and categorized. The solution has more room in the reuse and dissemination of statistical work products. To integrate with existing proven knowledge management platform product seems to be a practical approach via Web Services. First is to configure appropriate taxonomy and document hierarchical folders for statistical results. Second is to apply Ajax technology to embed the statistical job submission form from the proposed solution. Third is to design an interface to check in the statistical results from the file repository to the knowledge management platform by Web Service calls. Fourth since the computing resources (*i.e.* CPU, MEM, and I/O) consumption is determined by the complexity of statistical model and the volume of data, the GNU-R Engine must re-dispatch the job to another available message queue if resource-shortage failure occurs. To resolve that, designating a First-in-First-out (FIFO) sequential processing queue on a dedicated blade server to handle those earlier failed requests should be a feasible approach. Lastly, this paper urges employees to take advantage of existing data and transform them into useful knowledge to enhance the quality of decision making in their enterprises and let statistics become the core of strategy reasoning.

## REFERENCES

- [1] B. P. Kumar, J. Selvam, V. Meenakshi, K. Kanthi, A. Suseela and V. L. Kumar, "Business Decision Making, Management and Information Technology," *Ubiquity*, Vol. 2007, 2007, p. 4. [doi:10.1145/1226690.1232401](https://doi.org/10.1145/1226690.1232401)
- [2] H. Arsham, "Statistical Thinking for Managerial Decisions," 9th Edition, 2011. <http://home.ubalt.edu/ntsbarsh/Business-stat/opre504.htm>
- [3] R. N. Charette, "Why Software Fails," *IEEE Spectrum*, Vol. 42, No. 9, 2005, p. 36. [doi:10.1109/MSPEC.2005.1502528](https://doi.org/10.1109/MSPEC.2005.1502528)
- [4] M. T. Schmidt, B. Hutchison, P. Lambros and R. Phippen, "The Enterprise Service Bus: Making Service-Oriented Architecture Real," *IBM Systems Journal*, Vol. 44, No. 4, 2005, pp. 781-797. [doi:10.1147/sj.444.0781](https://doi.org/10.1147/sj.444.0781)
- [5] J. L. Maréchaux, "Combining Service-Oriented Architecture and Event-Driven Architecture Using an Enterprise Service Bus," *IBM Developer Works*, 2006.
- [6] Y. Wei, Z. Sun, X. Chen and F. Zhang, "A Service-Portlet Based Visual Paradigm for Personalized Convergence of Information Resources," *2nd IEEE International Conference on Computer Science and Information Technology*, Beijing, 8-11 August 2009.
- [7] L. Chung and J. do Prado Leite, "On Non-Functional Requirements in Software Engineering," *Conceptual Modeling: Foundations and Applications*, Vol. 5600, 2009, pp. 363-379.
- [8] B. Behkamal, M. Kahani and M. K. Akbari, "Customizing ISO 9126 Quality Model for Evaluation of B2B Applications," *Information and Software Technology*, Vol. 51, No. 3, 2009, pp. 599-609. [doi:10.1016/j.infsof.2008.08.001](https://doi.org/10.1016/j.infsof.2008.08.001)
- [9] A. T. Yalta and R. Lucchetti, "The GNU/Linux Platform and Freedom Respecting Software for Economists," *Journal of Applied Econometrics*, Vol. 23, No. 2, 2008, pp. 279-286. [doi:10.1002/jae.990](https://doi.org/10.1002/jae.990)
- [10] M. Lapsley and B. Ripley, "ODBC Database Access," 2011. <http://cran.r-project.org/web/packages/RODBC/RODBC.pdf>
- [11] A. Friis-Christensen, N. Ostländer, M. Lutz and L. Bernard, "Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions," *Transactions in GIS*, Vol. 11, No. 6, 2007, pp. 799-818. [doi:10.1111/j.1467-9671.2007.01075.x](https://doi.org/10.1111/j.1467-9671.2007.01075.x)
- [12] C. F. Liao, Y. W. Jong and L. C. Fu, "Toward a Message-Oriented Application Model and Its Middleware Support in Ubiquitous Environments," *Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering*, IEEE Computer Society, Washington, 2008, pp. 298-303. [doi:10.1109/MUE.2008.24](https://doi.org/10.1109/MUE.2008.24)
- [13] F. Menge, "Enterprise Service Bus," 2007.
- [14] A. Vaccaro, R. Parente and F. M. Veloso, "Knowledge Management Tools, Inter-Organizational Relationships, Innovation and Firm Performance," *Technological Forecasting and Social Change*, Vol. 77, No. 7, 2010, pp. 1076-1089. [doi:10.1016/j.techfore.2010.02.006](https://doi.org/10.1016/j.techfore.2010.02.006)
- [15] U. R. Kulkarni, S. Ravindran and R. Freeze, "A Knowledge Management Success Model: Theoretical Development and Empirical Validation," *Journal of Management Information Systems*, Vol. 23, No. 3, 2007, pp. 309-347. [doi:10.2753/MIS0742-1222230311](https://doi.org/10.2753/MIS0742-1222230311)