A SOA-Driven Business Intelligence Architecture

Lee-Kwun Chan[1], Yee-Wai Sim[1] and William Yeoh[2]

[1]Faculty of Information and Communication Technology, University Tunku Abdul Rahman, Perak, Malaysia

[2]School of Information Systems, Deakin University, Victoria, Australia

_____

Abstract

A business intelligence architecture comprises of different unique components to collect, transform, analyze and present the structured and unstructured raw data in simple formats to assist decision makers in making timely decisions. The introduction of service-oriented architecture (SOA) enables reusable services which are accessible over a network on demand. However, there is still a lack of academic literatures on the business intelligence architecture with service-oriented concept. Based upon various references on BI architectures from major vendors, a novel BI architecture that is built on service-oriented concept is presented and described in this paper. The proposed service-oriented architecture enables enterprises to deploy a more agile, flexible, cheaper, reusable, reliable and responsive BI applications in supporting decision making process.

Keywords: enterprise business intelligence, service-oriented architecture, business intelligence architecture
_____

Introduction

Recently, many enterprises are turning their effort towards implementing business intelligence (BI) systems to improve decision making process. A typical business intelligence system include functions such as reporting, multi-dimensional analysis, querying tools, online analytical processing tools, forecasting, data-mining, and advanced visualization capabilities (Microstrategy, 2007). Often, a typical enterprise-level BI architecture constitutes a data warehouse as a common back-end layer. In addition, distributed multi-tier architecture has been widely adopted for modern enterprise BI architecture (Davenport et al., 2007).

According to Erl (2005), service-oriented architecture (SOA) is a group of well-defined services which can be combined, reused and communicated with each other over networks. The introduction of service-oriented architecture changes many business operations into reusable services which are accessible over a network on demand (IBM, 2007). Within SOA environment, one of the outstanding advantages is individual services that can be accessed without knowing the underlying system platform (Kodali, 2005). Nickull (2005) further states that reusability and repurposing are the main reasons for adopting SOA in implementing enterprise-scale BI systems. SOA enables low-cost system development but with good system quality because it is able to provide a flexible and standardised architecture that supports data sharing and integration of various systems. Other advantages of SOA includes flexibility, responsiveness, reusability, ease of connection, development cost reduction and agibility (IBM Global Technology Services, 2007).

In the traditional approach, client-server architectures are tightly coupled. Any changes made to the client-side might

require relevant changes to be implemented on the server-side as well and vice versa (Cognos Corporation, 2008). However, in the ever evolving business environments, enterprises require an architecture that can easily include, remove or integrate additional services on the fly. In view of this, architectures built based on SOA principles, can separate implementation components from the underlying infrastructure. The separation enables the architecture to support the flexibility, interchangeability, resiliency, security, capacity and availability needed in a modern enterprise BI architecture (IBM Global Technology Services, 2007).

Therefore, this research aims to identify the shortcomings of existing enterprise BI architectures and puts forward a service-oriented approach. The authors develop a conceptual service-oriented BI architecture model based on the research findings reported in literatures from leading BI vendors such as IBM, Microsoft and Cognos.

In the remaining parts of this paper, the typical business intelligence architecture and its key components are introduced. The next section discusses service-oriented architecture and its benefits. Lastly, we study existing modern BI architectures and propose a new conceptual model on service-oriented business intelligence architecture.

Literature Review

According to Ponniah (2001), architecture is the combination of different unique components which provide a means to store data, and deliver information to users. Architecture is a template that contains rules and functions used for serving the business requirements. The various elements in architecture such as the standards, measurements, designs and other supporting techniques, aim to enable the smooth data flow from the source to the destination within a framework (Ponniah, 2001).

A BI architecture that comprises of different systems, applications, and processes in an enterprise enables business decision makers in accessing the valuable information for complicated analytical processes. The BI architecture should be able to provide the correct information promptly in supporting the decision making processes. Hence, the information must be distributed via multiple channels including emails and page alerts spreadsheets, analytic queries, scorecards and dashboards (Davenport et al., 2007). In other words, BI architecture is anything that enables the transformation of data into useful information in decision-making process and to acquire evolving business advantages.

According to Howson (2007), a fundamental architecture is the essential elemental for all BI system deployments. On one hand, the lower level BI architecture can be formed by the transactional systems and front-end tools. On the other hand, a higher level BI architecture deployment may consist of data marts, data warehouses, ETL (extract, transform and load) tools and BI front-end tools.

A study was separately performed on the BI architectures by two major BI vendors, namely, IBM and Microsoft. The authors found that both of the proposed architectures consist of five similar layers: data access/presentation layer, data analysis layer, data repository/data storage layer, data integration layer, and data source layer. Table 1 presents the result of comparison between IBM's BI Reference Architecture and Microsoft Conceptual System Architecture in accordance to layers.

Table 1: Comparison of IBM (2004) and Microsoft (2006) BI Architectures

Tier	BI Architecture Layer	Microsoft Conceptual System Architecture	IBM Reference Architecture
5	Data Presentation or Data Access	Navigation- drilling, pivoting, drag-drop capabilities Format- chart, graph, table, report, dashboard, or KPI Host- desktop PCs, Tablet PCs, and handheld mobile devices such as Pocket PCs and Smart-phones	Web browser Portals Devices Web services
4	Data Analysis	OLAP, Data Mining	Collaboration Query & reporting Data mining Modelling Scorecard Visualization Embedded Analytics
3	Data Storage or Data Repositories	Data Warehouse, Data Mart, Dimensional Modelling, Partitions, Indexes	Operational data stores Data warehouses Data marts Staging areas Metadata
2	Data Integration	Profiling, Extraction, Cleansing, Transformation, Loading, Staging	Extraction Transformation Load / apply Synchronization Transport /messaging Information integrity
1	Data Source	Different source environments with different systems, including different platforms and operating systems	Enterprise data, unstructured data, informational data, external data

It is found that both IBM's and Microsoft's architectures have similar functionality in each of the layers. Both architectures contain five layers with different names but similar characteristics in general. As shown in Table 1, the authors found that IBM and Microsoft define data presentation layer and data analysis with different components. IBM defines data presentation as placeholder for web browser, portals, devices, web services whereas Microsoft defines data presentation layer in wider scope which also includes navigational methods and report formats. In the next layer, Microsoft defines data analysis layer as Online Analytical Processing (OLAP) and data mining only while IBM defines this layer as a collective functions like collaboration, query and reporting, data mining, modelling, scorecard, visualization and embedded analytics.

Davenport et al. (2007) also introduced a BI architecture. However, the architecture has six layers instead of five layers. Operational tool is the layer which does not appear in Microsoft and IBM five-tier architecture. This layer mainly describes the administration issues. The six layers were summarized by the authors in Table 2.

Table 2: Six Layers BI Architecture Proposed by Davenport Et Al. (2007)

	Tools/Layers	Primary purpose
1.	Data management tools	To state the acquisition and management of data source
2.	Transformation tools	For extraction, cleansing, transmission and loading of data source
3.	Repository tools	To describe the storage of metadata and data
4.	Application tools	For data analysis
5.	Presentation tools	To describe about data accessing method, display format, visualization and manipulation
6.	Operational tools	To describe the significance of administration such as secrecy, data security, error handling and archiving

Cognos 8 business intelligence architecture is an example of modern architecture. It is a SOA-based BI platform. However, Cognos's business intelligence architecture deploys only three distinct tiers to deliver BI capabilities as follows (Cognos Corporation, 2008):

- A presentation tier that handles interaction with users over the network;

- An application tier to handle all BI processing with special-built services; and

- A data tier that provides access various data sources.

The Conceptual Model of SOA for Enterprise Business Intelligence

One of the main reasons for a BI project failure is largely due to the selection of inappropriate BI tools that fails to meet the specified business requirements. Therefore, Ponniah (2001) recommends that enterprises should design the architecture first, only then select the tools to match the functions and services stipulated for the architectural components. According to Friedman et al. (2004), there is a limited awareness of the value of architecture in the BI context. This creates challenges in data quality for many enterprises and the "hidden" aspects such as reliability, scalability and flexibility of BI are not being emphasized. According to Friedman et al. (2004), there is a limited awareness of the value of architecture in

the BI context. This creates challenges in data quality for many enterprises and the "hidden" aspects such as reliability, scalability and flexibility of BI are not being emphasized.
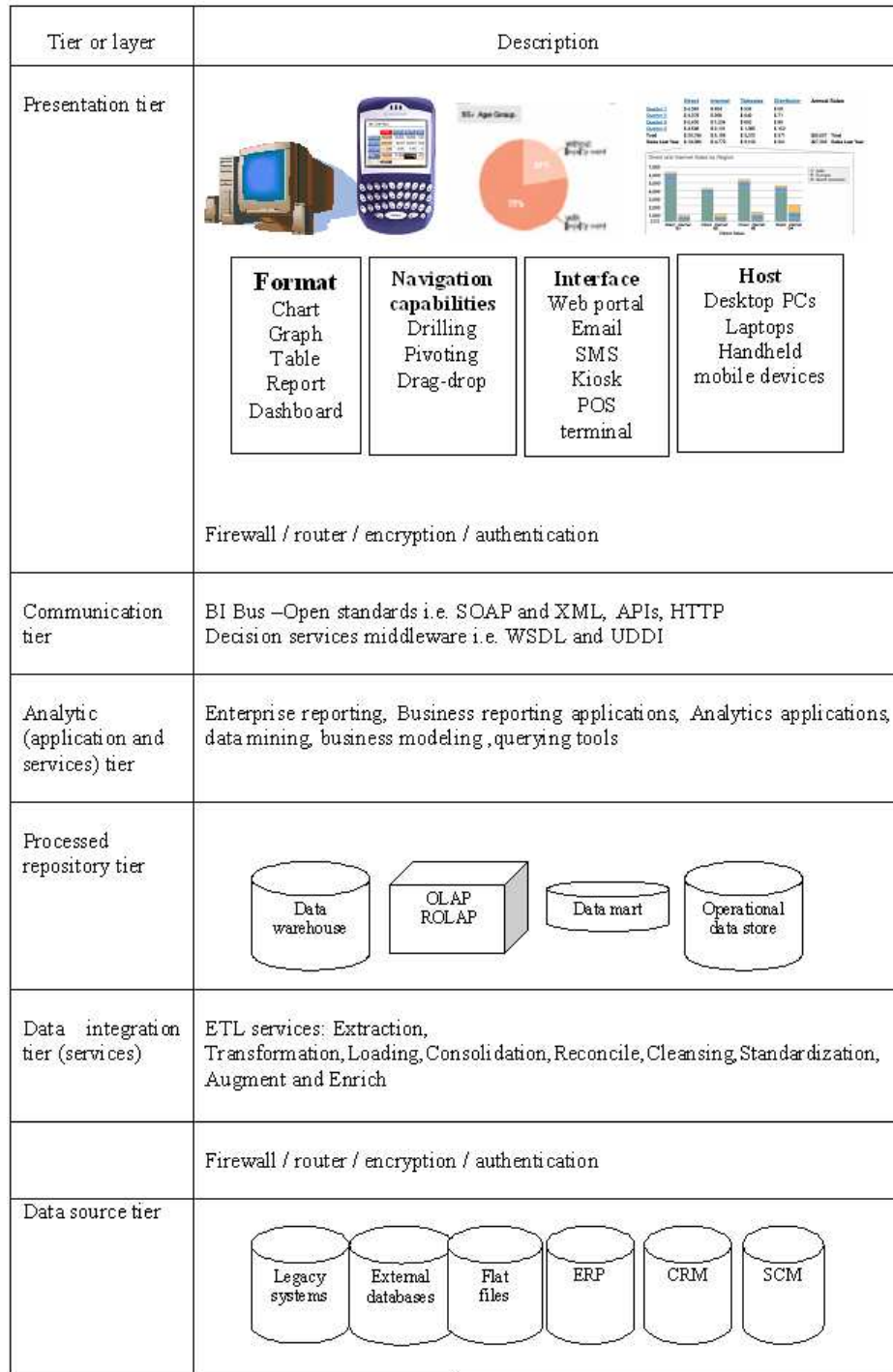
Davenport el al. (2007) assert that developing a robust BI system is more than just gathering and storing huge amount of data since it involves facets such as data quality, business processes, incentives, skills, organizational cultures, and sponsorships. Therefore, a BI architecture should be considered first when scoping a BI solution so that a more appropriate BI solution can be developed in meeting the actual needs of an enterprise. The finding was supported by Friedman et al. (2004), who state that architecture is the foundation of BI and paying attention to the architecture will ensure success in BI implementation.

Watson et al. (2005) also emphasize that architecture has significant impact when users can easily access the business data, which in turn leads to improved decision-making capabilities. In order to meet the requirements of an emerging business environment, a BI architecture should be designed in a flexible and adaptable manner (Davenport et al. 2007). Invariably, a rigid BI architecture design will be better off serving the needs of an ever-changing business environment.

In view of the challenges and limited capabilities of existing SOA-based architecture, a six-tier SOA-based BI

architecture conceptual model (Figure 1) is proposed to deliver BI capabilities in the constantly evolving business environments. The details of the model's tiers are described in the following sections.

Tier or layer	Description
Presentation tier	 **Format** Chart Graph Table Report Dashboard **Navigation capabilities** Drilling Pivoting Drag-drop **Interface** Web portal Email SMS Kiosk POS terminal **Host** Desktop PCs Laptops Handheld mobile devices Firewall / router / encryption / authentication
Communication tier	BI Bus –Open standards i.e. SOAP and XML, APIs, HTTP Decision services middleware i.e. WSDL and UDDI
Analytic (application and services) tier	Enterprise reporting, Business reporting applications, Analytics applications, data mining, business modeling , querying tools
Processed repository tier	 Data warehouse, OLAP ROLAP, Data mart, Operational data store
Data integration tier (services)	ETL services: Extraction, Transformation, Loading, Consolidation, Reconcile, Cleansing, Standardization, Augment and Enrich
	Firewall / router / encryption / authentication
Data source tier	 Legacy systems, External databases, Flat files, ERP, CRM, SCM

Fig1. Proposed Conceptual Model

Presentation Tier

The top layer is the data presentation or data access layer which provides access to the users via various means such as host computers, portal, web service or even handheld mobile devices. Furthermore, users are able to view reports in various formats such as chart, graph, table, report, dashboard, scorecard, metrics, key performance indicator (KPI). The reports are equipped with navigation capabilities like drilling, pivoting, drag-drop so that users can seamlessly analyze the data in any reports.

Communication Tier

Open interface is used by all services to communicate with each others and it should take place in the communication tier. This can achieve location transparency among services. Open web standards, such as Simple Object Access Protocol (SOAP), Extensible Markup Language (XML) and Web Services Definition Language (WSDL) should be adopted for communication purpose.

Analytic Tier

Since a BI system contains huge volume of data, tools and techniques from data integration layer, data analysis layer is critical for managing, summarizing, querying and analyzing the data. It provides analytics functions such as ad-hoc query, reporting, modelling, embedded analytics, Online Analytical Processing (OLAP), data mining and collaboration. Data is organized into multidimensional cubes based on a dimensional model which enables summarizing information for analytical queries in an efficient manner. Data analysis leverages the information stored in data storage.

Processed Repository Tier

Data Storage is the end result of Data Source and Data Integration layers. Data Storage can be termed as Data Warehouse (DW)/Data Mart (DM). According to Microsoft architecture, it involves elements such as dimensional modelling, partitions and indexes. Meanwhile IBM proposes operational data stores, staging areas and metadata in this layer.

Data Integration Tier

BI often involves analysis of aggregated and integrated data from various operational systems. Data is extracted, cleaned and filtered from the operational systems before it is loaded into a data warehouse. A data integration layer primarily consists of ETL (extracting, transforming, and loading) process, whilst it also involves processes such as profiling, cleansing, and staging (Howson, 2007).

Data Source Tier

The most fundamental data collection is done at the data source layer (bottom layer) where it includes different sources environment, and systems including different platforms and operating systems). Data are collected from various resources. It could be operational data which are generated by internal departments, unstructured data, informational data and external data. Operational systems, which serve as the first entry for quantitative data, can be sourced from any transaction processing systems, enterprise resource planning systems or any other systems. Operational data can be generated from internal enterprise operational systems such as production system, sales and marketing system, supply chain system, and accounting system. Data may also come from external sources such as from suppliers, customers or any information on the Internet.

The architectures proposed by Microsoft, IBM and Davenport et al. are not built on the SOA concept. Cognos 8 business intelligence architecture is built on SOA concept but it is insufficient to demonstrate architecture functional independence. Hence, a new SOA-based BI architecture (Figure 1) is proposed which enables flexible deployment and integration of independent services. These can be assessed without knowing the underlying computing platforms. The benefits of having SOA include flexibility,

responsiveness, reusability, ease of connection, cost reduction and agility (IBM Global Technology Services, 2007). In addition, we propose an extra separated layer which is called the communication tier to achieve functional independence. This layer solely describes the communication aspect in terms of open standards such as Simple Object Access Protocol (SOAP) and Extensible Markup Language (XML), Application Program Interfaces (APIs), Hypertext Transfer Protocol (HTTP) and decision services middleware like Web Services Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI).

The issues of confidentiality, integrity, and authenticity are the most important concerns in any computer system. Data communication over pervasive network increases the vulnerability to computer security risks like viruses and intrusion. Some of the basic computer security practices like firewalls, routers, encryption and authentication, thereby, help to protect and prevent BI systems from unauthorized access. Therefore, it is proposed to implement the mentioned security measures above in two tiers. First, a security tier between data source tier and data integration tier. Second, a security tier between communication tier and presentation tier. Data sources contain data from various resources. Therefore, it is extremely important to make sure that the input data is free of viruses and intruders. Before the information is presented to the presentation tier, it is proposed to allocate some security measures to clean the reports which will be delivered to any digital devices in various formats (Fenzel and Fenzel, 2004).

Conclusion

This paper presents a service-oriented architecture for enterprise-scale business intelligence environment. Based on reference BI architectures from major BI vendors, a modern BI architecture that is built on service-oriented concept is proposed and described in this paper. An effective service-oriented BI architecture enables enterprises to deploy more flexible and reliable BI applications to a greater number of business stakeholders in improving the decision support. With the proposed architecture, enterprises can incorporate BI elements without having to reengineer their existing IT infrastructure and associated complexities in systems administration, training, and server management.

References

Cognos Corporation (2008). "The Right Architecture for Business Intelligence: The Foundation of Effective Enterprise BI," White paper

Davenport, T. H. & Harris, J. G. (2007). Competing on Analytics: The New Science of Winning, *Harvard Business School Press*, pp.240.

Erl, T. (2005). "Service-Oriented Architecture: Concepts, Technology, and Design," Prentice Hall

Fenzel C. W. & Frenzel J. C. (2004). "Management of Information Technology," Course Technology

Friedman, T. & Strange, K. H. (2004). "Architecture: The Foundation of Business Intelligence," *Gartner Research*, AV-226453

Howson, C. (2007). Successful Business Intelligence: Secrets to Making BI a Killer App, McGraw-Hill.

IBM Corporation (2004). Introduction to the BI Architecture Framework and Methods, IBM Software Group, DB2 Data Management Software

IBM Global Technology Services (2007). "Infrastructure Considerations for Service-Oriented Architecture,".

Kodali R. R. (2005). "What is Service-oriented Architecture?," JavaWorld.com, [Online],[Retrieved October-7,2008], http://www.javaworld.com/javaworld/jw-06-2005/jw-0613-soa.html

Microsoft Corporation (2006). Business Intelligence Guidelines Conceptual Framework

Microstrategy (2007). "Enterprise Business Intelligence, Improving Corporate Performance through Integrated Reporting, Analysis, and Monitoring,"

Nickull, D. (2005). "Service Oriented Architecture Whitepaper," Adobe Systems, Inc, pp.2

Ponniah, P. (2001). "Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals," John Wiley & Sons, pp.127

Watson, H. J. & Ariyachandra T. (2006). "Benchmarks for BI and Data Warehousing Success,"