

Business Intelligence on the Cloud Overview, Use Cases and Rol

Lekha Menon
HiTech Solutions Central
Tata Consultancy Services
Mumbai, India

Bhawna Rehani
HiTech Solutions Central
Tata Consultancy Services
Pune, India

Sudharshan Gund
Tata Consultancy Services
Pune, India

ABSTRACT

The use of Business Intelligence (BI) in the cloud is a game-changer, as it makes BI affordable and easily available as compared to traditional BI. It is expected that customers will slowly but surely migrate from in-house BI to BI in the cloud. Cloud Computing was a big topic of discussion in the early part of 2010 and many vendors developed strategies and solutions around Cloud BI in 2011. Customers are expected to continue their shift to the cloud in 2012, encouraged by big vendors enabling their traditional BI solutions as a cloud offering and having seen successful implementations where the movement to Cloud has shown substantial benefits. Vendors are also committed to further strengthen their cloud BI offerings with a view to addressing current inhibitors that customers have and refine the migration approach to promote better adoption.

This paper touches upon the drivers and challenges related to Cloud BI, and can act as a source of information related to leading Cloud BI vendors. It also covers a few sample use cases where Cloud BI could be evaluated and recommended solutions for each scenario. A real life implementation case study detailed towards the end of the paper would provide more information on the benefits accrued by migrating to a Cloud Data warehouse.

General Terms

Cloud Computing, Business Intelligence.

Keywords

Cloud, Business Intelligence, BI, Software as a Service, SaaS, PaaS, Platform as a Service, Datawarehouse, DW.

1. INTRODUCTION

Today companies are increasingly looking to cloud computing as an integral component of their computing strategy. Companies now understand the various benefits that cloud computing offers in terms of reduced total cost of ownership, better scalability and elasticity and the ease of setting up, configuring, and deploying new systems. Cloud computing is characterized by ability to consume resources as required in an elastic manner and scaling the consumption arbitrarily as required. Cloud Computing provides several benefits in terms of Faster time to value, lower capital expenditure, better scalability, better hardware utilization etc.

2. CLOUD BI

2.1 Cloud BI – Drivers

There are several operational and financial factors that work in favor of Cloud Business Intelligence (BI), prime of them being:

- **Speed of Implementation and Deployment** – Immediate availability of environment without any dependence on the long periods associated with infrastructure procurement, application deployment etc. drastically reduces the BI implementation time window.

- **Elasticity** – Leverage the massive computing power available on the web, Scale up and scale down based on changing requirements
- **Focus on Core Strength** – Outsource running of BI apps to professionals and focus on their core capabilities
- **Lower Total Cost of Ownership** – Convert some part of Capital expenditure (capex) to Operational expenditure (opex), Cost effective pricing models, pay per use model etc.
- **On-demand Availability** – Support mobile and remote users, Browser based access to control everything from the cloud platform to database management to the data warehouse layer to the analytics platform

2.2 Cloud BI – Challenges

There are also many inhibitors which have resulted in a very slow adoption rate to Cloud BI so far. A few common and leading concerns are mentioned below, along with recommendations on addressing the concerns[4][5]:

- **Data Security** - Security concerns including confidentiality, integrity and availability of the data continues as the top concern for utilizing the Cloud. For some organizations, the concerns over security may be a barrier that is impossible to overcome today. However, as more organizations move to the Cloud it is expected that the concerns will lessen. In many cases, the Cloud vendors provide a more secure environment than what exists at customer sites.
- **On premise Integration** – Sudden movement to cloud is not feasible and a phased approach is usually recommended. There will be a co-existence model until the cloud BI market is more mature.
- **Lack of control** – Tough to get SLAs from cloud providers. Data control and data ownership, reliability of service challenges are some of the main reasons for client concern. To mitigate this, organizations should already have in place thorough IT governance and service delivery standards and models
- **Vendor Maturity** – Many cloud BI vendors, hosting providers with varying offerings etc. makes choosing the right vendor based on desired needs and vendor capabilities a bit confusing
- **Performance** - Limits to the size and performance of data warehouses in the Cloud, Significant latency if BI applications exist in the Cloud but the data exists at a client site, especially when processing and returning large amounts of data.
- **Pricing models** – Lack of standardized pricing models makes it difficult for customers to select the right one

3. CLOUD BI PRODUCT LANDSCAPE

There are many options, both from traditional BI vendors and from newcomers, providing different BI functionality based on different architectures and platforms. The table below lists the major vendors offering Cloud BI solutions[2][3].

Table 1: Cloud BI Product Landscape

Architecture and platform	Representative vendors
Software AS A Service (SaaS) BI Solutions	Actuate, IBM Cognos, Information Builders, Jaspersoft, Kognitio, MicroStrategy, Oracle, Panorama Software, Pentaho, QlikView, SAS Institute, TIBCO Spotfire
BI/ Datawarehouse (DW) Platform as a Service (PaaS)	AsterData MPP on Amazon EC2 IBM Cognos Express on Amazon EC2 Teradata Express on Amazon EC2 RightScale/ Talend/Vertica/Jaspersoft on Amazon EC2
Data warehouse as a service (DaaS)	Vertica, Kognitio
Data Integration as a Service (DIaaS)	IBM Infosphere DataStage, IBM CastIron, Informatica Powercenter Cloud Edition, Snaplogic Dataflow, Talend Integration Suite on RightScale Managed Platform
SaaS BI Packaged Analytic Application Solutions	Cloud9 Analytics, IBM Cognos Analytic Applications, PivotLink, Rosslyn Analytics, SAS, SAP BusinessObjects BI On-demand for Salesforce

There are Software-as-a-Service (SaaS) providers where the BI application is hosted by the vendor either in their own premises or through hosting providers, and made available to customers. Some of them also host their solutions on the public cloud. Many of these companies focus on specific verticals and provide pre-built data models that reduce the amount of time required from data loads to reporting.

There are Platform as a Service (PaaS) offerings where vendors provide a platform on the cloud for developers to design and develop BI applications quickly. Here are single vendor offerings as well as cases where multiple vendors work together to provide a single platform offering end to end services. An example of a multi-vendor option is the RightScale/Talend/Vertica/Jaspersoft PaaS offering on Amazon EC2. In this offering, data integration is very much file based such that one could upload files of data and then there is some processing of that data to load it into the PaaS DW/BI database. Several single-vendor PaaS offerings give only fairly lightweight data integration once data is uploaded and not full blown ETL with built-in data quality that might be used to in a data centre.

Database vendors are also starting to leverage the Cloud. Vertica and Kognitio offer options for customers to deploy their database in the Cloud. Vertica offer fixed monthly pricing based on the amount of data and number of nodes required. Both utilize dedicated, hosted servers rather than on-demand,

virtualized servers, and allows running large-scale data analytics projects (marketing campaigns, business reporting etc) on a pay-per-use basis. There is no overhead of implementing a data warehouse onsite or for the in-house IT department to service, maintain and support the database and its users.

Data Integration as a Service (DIaaS) providers offer solutions to integrate between cloud and on-premise applications, databases, flat files, file feeds, and even social networking sites. DIaaS prescribes the use of Web services, well-defined interfaces, and calls between and among service-enabled applications and data sources to deliver a loosely-coupled integration experience.

Finally there are vendors who focus on packaged analytic applications. BI vendors (e.g., Cognos, SAS, Business Objects, Hyperion and SPSS), ETL vendors (Informatica) and ERP vendors (SAP, PeopleSoft and Oracle) are a few of those who increasingly provide prepackaged analytic applications.

4. RETURN ON INVESTMENT (ROI) FOR A CLOUD BI IMPLEMENTATION

For a given use of money in an enterprise, the ROI is how much profit or cost saving is realized. We need to analyze and predict the ROI for any implementation to understand if the implementation will be beneficial for the organization. When an organization plans to move to a Cloud BI solution or a Hybrid solution, they need to compare the expected ROI of this solution with the ROI of a pure on-premise implementation. (Hybrid solution is where some of the components of BI stack are moved to cloud, keeping the rest on premise) This will help in making the decision on what components of BI should be moved to cloud for the most beneficial scenario [6][7].

4.1 ROI Metrics

Business Benefits can be categorized as

- Financial Benefits: include impacts on the organization's budget and finances, e.g., cost reductions or revenue increases
- Non-financial Benefits: include impacts on operations or mission performance and results, e.g., improved customer satisfaction, better information, shorter cycle-time

4.1.1. Financial ROI

The below 3 metrics are the key measure of the Financial ROI of an organization

- Payback Period: The amount of time required for the benefits to pay back the cost of the project. This does not take into account discounted cashflows.
- Net Present Value (NPV): The value of future benefits restated in terms of today's money.
- Internal Rate of Return (IRR): The benefits restated as an interest rate.

4.1.2. Non Financial ROI

These are the intangible benefits of a BI implementation

Some of the 'non-financial' metrics can be Customer Satisfaction Index (CSI), Ratings by Independent agencies, Better time utilization by business & IT users, Quality Data for Downstream/Upstream applications, Public relations, Risk avoidance, Greater confidence in the decision making process

In the below Section we will focus on the calculation of 3 Financial ROI metrics:

To calculate the metrics, we first need to list down the cashflows of the organization. This includes Cost and Returns.

Cost and Returns are estimated for x years where x is the life of BI implementation in no of years. Generally the life span where a BI implementation brings benefit is: 3 to 5 years

• **Cost**

The Cost for a BI implementation can have two components

- Capital Expense (Capex): e.g. one time hardware cost, software license procurement, software services cost for Implementation etc
- Operational Expense(Opex) : e.g. Support services cost, Annual Maintenance Contract for Software, hardware

• **Returns**

The benefits achieved from the BI implementation are of two types

- Savings: For e.g Time saving for data consolidation, Time saving for data fixing etc
- Revenue: For e.g. Earnings due to marketing campaigns derived exclusively from the DW

Once we know the cashflows, the ROI metrics can be calculated. Below example shows the calculation these metrics

4.2 Comparison of on-premise and Cloud based ROI

Though the process for calculation ROI remains the same in on-premise ad cloud/hybrid implementations, but the cost and returns will vary. Here are some of the factors which will be different.

Hardware cost: For Cloud components, there will be no hardware cost required for production servers. Organization may still have to buy development servers; this cost will have to be considered. Also, in a hybrid system, the cost of production servers will be zero for components on cloud but for on-premise components it will remain the same.

Software License Cost: The License cost will be different for both the solutions. The cloud cost will be based on factors like no of users, specific features required. This needs to be considered separately. In a hybrid system, the cost of licenses will be considered separately for on-premise and cloud components

IT Services Cost: IT implementation skills required for cloud based applications may be different, hence effort can be different

Support cost: In cloud based solution, primary support of application will be provided by cloud vendor. This cost will not be required in cloud application

Upgrade: Generally within a span of BI implementation (typically 3 to 5 years), there will be atleast one software upgrade required. This cost will come as Opex cost in on premise solution, but will not be required in cloud solution as this is handled by cloud vendor

Based on the above differences, the ROI metrics can be calculated for cloud and on-premise solutions. The value of the three metrics will help in deciding which is a better cost effective solution for a given problem statement

Sample out of the ROI calculator is as given below:

Cashflow Calculation Worksheet		On Premise						vs. On cloud/Hybrid					
Cost Component	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	
Outflow													
Capex-Hardware	140270	122	122	122	0	0	30270	57	57	57	0	0	
Capex-Software	251	56	56	56	0	0	251	56	56	56	0	0	
Capex- Cloud Licenses	0	0	0	0	0	0	66000	66000	66000	66000	0	0	
Capex-IT Services	100	0	0	0	0	0	30	4	4	4	0	0	
Opex-AMC (Annual Maintenance Contract) for TCO period	225	5	5	5	0	0	225	5	5	5	0	0	
AMC - Hardware for TCO period	63	3	3	3	0	0	40	0	0	0	0	0	
IT Support for TCO Period	128	0	0	0	0	0	128	0	0	0	0	0	
Total outflow	141037	186	186	186	0	0	96944	66122	66122	66122	0	0	
Inflow													
Savings	190	102000	102000	102000	0	0	6140	157000	157000	157000	0	0	
Revenue	40	0	0	0	0	0	40	0	0	0	0	0	
Total Inflow	230	102000	102000	102000	0	0	6180	157000	157000	157000	0	0	
Cashflow	-140807	101814	101814	101814	0	0	-90764	90878	90878	90878	0	0	

ROI Metrics	Formula	Value (in Years)	
		Cloud	On Premise
Payback Period	No of years when Total Inflow = Total Outflow	0.99	1.4
Net Present Value (NPV)	The value of future benefits restated in terms of today's money.	\$122,941.94	\$102,172.14
Internal Rate of Return (IRR)	The benefits restated as an interest rate.	84%	13%

Fig. 1: Sample ROI Calculation

5. CLOUD BI USE CASES

Here are few of the most frequently seen use cases for BI on Cloud and recommended solution ^[1].

Table 2: Cloud BI Use Cases

Use Case	Analysis	Recommended Solution
Coexistence case: SaaS BI application for transactions along with on-premises BI solution. For example, Salesforce used for sales and marketing.	SaaS BI application data needs to be brought into the enterprise data warehouse for integrated information analysis. Using on premise integration solutions to extract cloud-based application data has its challenges.	ETL on Cloud
SaaS centric Enterprises: Organizations are moving many of their key systems to cloud. For example, CRM, HR, ERP	CRM, HR, ERP are some of the key sources for any BI application. If the sources are moving to cloud, it makes sense to store the integrated information to the cloud as well.	SaaS BI DW on Cloud
Small mid-sized business enterprises: Organizations which are growing from very small business to middle size, or expanding location from one to two/three.	Such organizations, need a BI system but will not like to spend on their own fully functional BI systems	SaaS BI model
Elasticity Requirements: Companies with strong variation in transactional data periodically. For example, high data activity when analysis required for yearend data, or high transaction volumes in Point-of-Sales in the peak festival season.	The Customer will not like to maintain a high capacity warehouse for the whole year. There is a need to provide flexibility in paying for only the used capacity.	DW on Cloud Analytics on Cloud

Use Case	Analysis	Recommended Solution
Consolidation Scenario: Organizations with multiple departments working in silos, each having its own BI system	Multiple BI tools, servers may coexist, with each not being used to the fullest of capacity. Need for an integrated/consolidated BI system which is available enterprise wide. The server capacity and user licenses can be shared among the departments for better utilization of costs incurred in BI systems.	Private Cloud

6. CASE STUDY - ENTERPRISE DATA WARE HOUSE IMPLEMENTATION ON CLOUD

In mid-2010 one of the TCS clients took a big decision to move their enterprise data ware house to cloud platform. And because of very long and credible relationship going back to over a decade, TCS was given responsibility of end to end implementation of data ware house on the external cloud platform. And in the month of Mar-2012, after 1.5 years of migration and development efforts TCS has enabled client to move from in house data ware house to cloud hosted data warehouse. Before client has one of biggest enterprise data ware house in USA, now probably they have biggest and very efficient data ware house hosted on the cloud.

Below were the drivers for moving to cloud hosted data ware house

- Reduce the Hardware licensing cost.
- Reduce the Software licensing cost
- Reduce the maintenance cost.
- Decommission most part of in-house developed ETL code to avoid person dependencies.
- Increase availability business intelligence platform for various business analytical needs.
- Increase the accessibility of Data Warehouse data. Business should able to access the data even from their internet browser enabled mobile handsets.
- Improve the performance of adhoc and as well as batch reports.
- Improvement in terms of data quality and integrity.
- Overcome the data storage and maintenance related issues

6.1 Implementation Approach

It was decided to just host the data on the cloud platform and then point all current reporting applications to data source on the cloud. By this approach end users were least impacted.

Exception to this was sensitive data like labor, HR and some customer related data was decided to be kept in house on sql server 2008.

Data ware House Size

- Database Size – 15+ TB of data was migrated to cloud vendor and daily there is growth of 20+ GB of data.
- Application Integration –
 - 70+ mainframe job routines are sending data daily on the cloud. Data includes sales data, inventory, advertising, marketing etc.
 - 400+ SSIS routines are pulling data that is hosted on the cloud
 - 20+ SSAS cubes are built from data pulled from cloud
 - 20+ front end reporting applications are based on this data
 - 70+ feeds are sent outside clients systems to external vendors and third parties

6.2 Challenges during project - related to cloud hosting

- **Data Transmission:** As huge data needed to be transmitted daily, utility called ZIP390 was used to send data on the cloud. This utility use to encrypt, compress and ftp the data securely to cloud vendors ftp site.
- **SSIS connectivity:** This was big challenge at the start of the project, because SSIS package was not able to pull large data by directly connecting to data on the cloud. So in the first step data was unloaded in the flat file using utility provided cloud vendor. Once data was on flat file it was taken for processing by SSIS package. Lot of initial time was spent in trying different approaches of pulling data from source before coming up with this solution. (SSIS - SQL Server Integration Services)
- **Cognos Connectivity :** Just so that end user were not affected, cognos reports that were pulling data from in-house data ware house needed to be pointed to data source on the cloud. But here also there were issues in ability to pull data from cloud data source. Here also collaborative efforts from cloud vendor, cognos vendor and our developers were able to find solution to connectivity issues. In some case cognos provided customized ini file, in some cases cloud vendor changed their design and in some cases TCS developers gave workaround solutions.
- **Data Controls:** As the data started increasing on the host platform and lot many feeds were sending data, there were concerns about the reliability of the data. So we have come up with controls for every feed. Once the data was loaded different counts, amounts etc. was verified against the data that was sent. This made sure complete and correct is loaded on host system.

6.3 General cloud hosting concerns and resolutions

- **Data protection during transmission** – Data was transmitted to vendor serves using ZIP390 utility, which took care of encryption of data transmitting it using secure ftp.
- **Privacy and Confidentiality of customer's data** – Cloud vendor's facility and staff was made HIPPA act compliant and it was audited by external HIPPA auditor. (HIPPA-Health Insurance Portability and Accountability Act). This made sure that enough safeguards are in place at cloud vendor's facility to protect sensitive and confidential customer information.
- **Disaster Recovery** - Cloud vendor use to conduct regular disaster recovery drills and give update to client. This made sure data is accessible from some other place in case facility is not available for some reason.

6.4 End Result

At the end of Mar-2012, all historic data was successfully migrated to cloud platform and also daily feeds were in place which were loading that days data on cloud every night. From client's point they have achieved whatever they expected from making this move of hosting their enterprise data warehouse on the cloud platform.

On high level below are the benefits that are being enjoyed by the client

No Hardware and Software Maintenance Cost – Client is no longer paying huge licensing and annual maintenance cost to supplier. Hardware has been returned to the supplier.

Resources which were supporting Data ware house are being used to other critical requirement.

Business Analyst are very satisfied that batch reports that were running for hours are executing in 5-10 mins, making data readily available. Also by running simple queries directly on the cloud vendor's portal they can quickly get any information from the 4-5 years of data, which helps them in analysis and decision making.

7. CONCLUSION

Cloud is a big part of future Business Intelligence and offers several advantages in terms of cost benefits, flexibility of implementation, availability and speed of implementation. It is very relevant to BI/DW implementations as typical BI/DW applications demand high infrastructure requirements, handle unpredictable load volumes, involve high upfront investment, high development & maintenance costs, takes a longer duration for provisioning and has so far displayed a high reliance on IT.

Although initially cloud-based solutions were designed for small- to mid-size companies that didn't have available IT resources or capital to spend on creating and managing a software and hardware infrastructure, today, many large companies are investigating the cloud as a way to add new business solutions quickly and augment existing data center capacity.

While considering Cloud BI, organizations are recommended to follow a few risk mitigation steps and strategies[1]:

- Perform due diligence for security, backup, and disaster recovery - Check whether the BI SaaS provider complies with emerging SaaS standards such as the SSAE 16 Type II audit.

- Thoroughly understand the BI SaaS pricing and contract matters - Understand the various pricing models offering by the vendor and choose the one that is definitely needed, study the service-level agreements (SLAs) agreed upon by the vendor and keep track of actual application usage
- Evaluate true long-term total costs of ownership - Perform a detailed RoI calculation to calculate long-term total ownership costs based on the specific environment and requirements
- Double-check whether any additional source data licenses are needed - Understand if any additional licenses need to be procured for other enterprise applications that the SaaS BI application would need to interface with.
- Plan for the worst - Have a detailed cloud to on-premise migration strategy in place in case the Cloud Vendor fails to perform as per desired expectations

8. REFERENCES

- [1] Forrester Report - BI In The Cloud? Yes, And On The Ground, Too
- [2] http://www.servicetechsymposium.com/soa_archive/pdf_berlin/Mike_Ferguson_BI_and.pdf
- [3] <http://intelligentbusiness.biz/wordpress/?p=281>
- [4] <http://tdwi.org/blogs/wayne-eckerson/2009/06/implementing-bi-in-the-cloud.aspx>
- [5] <http://tdwi.org/Articles/2011/04/27/BI-and-Cloud-Computing.aspx?Page=1>
- [6] <http://searchbusinessintelligence.techtarget.in/news/2240022273/BI-ROI-calculation-Issues-and-methodology>
- [7] <http://www.information-management.com/news/2487-1.html>.