Wikibon

Wikibon is a *professional community* solving technology and business problems through an *open source sharing* of free advisory knowledge.

tour

NOTICE TO ALL USERS

You are now viewing a snapshot of Wikibon.org. As of November 14th, 2015, all recent content has been migrated to our new site: Wikibon.com. The new site has been optimized for ease of use and for better access on mobile devices. Please click on this banner to go to Wikibon.com.

Become a Member!

Sign Up!

Why Register?

Login

User Name

Password

Remember Me

Login

Forgot Password?

Featured Research

- Big Data
- SLI
- Cloud
- infographics
- theCUBE
- Peer Incite Archive

Announcements

- All New Content on Wikibon dot com
- Wikibon Big Data Capital Markets Day
- TheCUBE 2013 Schedule
- TheCUBE 2014 Schedule

- July 30 Peer Incite: Enhancing Cloud Services with Hybrid Storage

Technology Events

Big Data - Is your Data Warehouse a Dinosaur?

Last Update: Feb 12, 2013 | 03:00
Viewed 33123 times | Community Rating: 5
Originating Author: **Wikibon Daemon**

#memeconnect #emc #Big Data

- Currently 5/5 Stars.
- 1
- 2
- 3
- 4
- 5

rate this

Storage Peer Incite: Notes from Wikibon's March 1, 2011 Research Meeting

Recorded audio from the Peer Incite:

Listen

Video replay

On March 1, 2011 the Wikibon Community convened to consider whether the advent of big data with its focus on multiple kinds of data including unstructured data-types has turned the data warehouse into a dinosaur. The center piece of the discussion was a recent Wikibon study of the advantages and disadvantages of the Oracle Exadata IT appliance and similar offerings expected from IBM and HP, and a survey of early big data users. David Floyer presents the results of this study, which he led, in his piece below titled "Financial Comparison of a Big Data MPP Versus a Data Warehouse Appliance Solution".

The conclusions of the meeting were first that the data warehouse is not becoming an obsolescent dinosaur, and second that Exadata, while it has some tradeoffs, offers some impressive advantages when properly applied.

The Wikibon survey shows that big data and the data warehouse will coexist in the enterprise. They will be used for different but equally important purposes, with the data warehouse often supplying data to big data projects. Early big data users predict that eventually some of those projects will in turn feed data back to the data warehouse, which will remain the central repository for the enterprise's important structured business data.

The study also demonstrates that when applied properly, a data appliance such as Oracle Exadata can result in an impressive 15%-30% savings over roll-your-own data warehouses while greatly simplifying maintenance and updating. This if anything means that data warehouses are evolving to become more efficient, an analog of the dinosaur evolution into birds. It also implies that data appliances will become the infrastructure-of-choice for many data warehouse projects. The main tradeoffs are first that an appliance locks users into the vendor and second that adding storage capacity to the box creates complexities that reduces the operational benefits of the IT-in-a-box

architecture. Therefore, data appliances provide the greatest advantage in situations where the data warehouse contains a rolling copy of the last 18-24 months of transactions and less advantage where the data warehouse houses much larger amounts of historical data that requires more storage than the appliance provides.

Overall, the meeting concluded that both big data and the IT appliance are important new developments that CIOs need to consider carefully. Neither, however, invalidates or replaces the other. If anything, big data makes the data warehouse more valuable. G. Berton Latamore, Editor

Contents

- 1 Big Data Update: Your Data Warehouse is Not a Dinosaur
  - 1.1 Key Findings on Data Warehouse Trends
  - 1.2 Impact of Oracle Exadata/Exalogic
  - 1.3 The Big Data Tsunami
  - 1.4 Case Study in Big Data Using MPP Architectures
  - 1.5 Is Your Data Warehouse a Dinosaur?
- 2 CIOs Need to Organize Big Data Teams
- 3 Financial Comparison of Big Data MPP Solution and Data Warehouse Appliance
  - 3.1 Executive Summary
  - 3.2 Introduction
  - 3.3 Traditional Data Warehouse Approach
  - 3.4 Big Data Approach
  - 3.5 IT Cost Comparisons
  - 3.6 Business Benefit Assumptions
  - 3.7 Conclusions
- 4 IT Appliances, Big Data, Both Challenge the Traditional IT Organization
- 5 The Emerging Big Data Vendor Ecosystem
- 6 Leverage the Cloud for Big Data to Avoid Traditional Traps

Big Data Update: Your Data Warehouse is Not a Dinosaur

David Vellante

Traditional data warehouse infrastructure and emerging big data applications, while distinctly different in terms of business drivers, architectures, operational procedures, skills sets, and use cases, in many ways complement each other and over time will both become sources of enormous leverage for many enterprises.

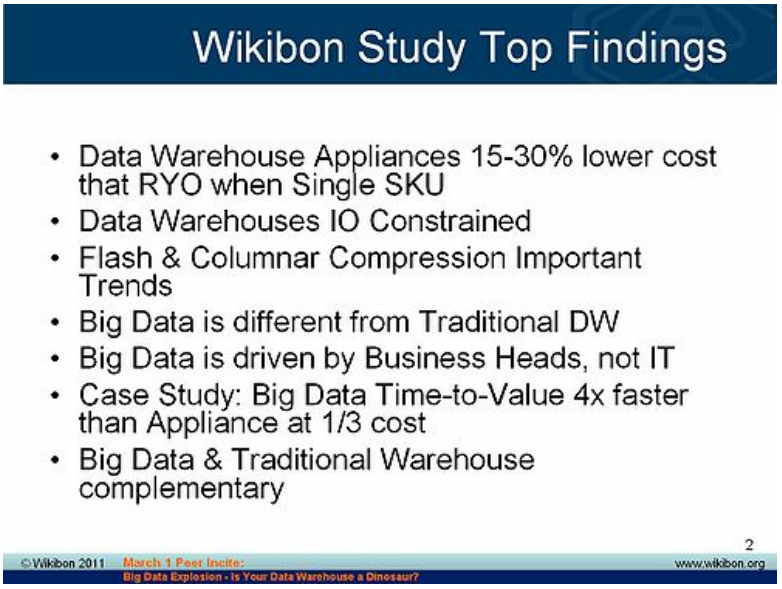
Today's data warehousing space is characterized by legacy system complexity, performance woes, and generally an operational support and reporting-driven model. This mindset is shifting however toward increased focus on finding ways to improve efficiencies, cut costs and generate revenues. From an architectural perspective, this market is being disrupted by the introduction of appliance-based data warehouse systems generally and Oracle Exadata specifically. The aggressive moves by Oracle, which is using its Sun acquisition and a vertical integration strategy to compete, has caused a chain reaction in the industry and led to strong user awareness of appliance solutions; and competitive acquisition responses from EMC (Greenplum), IBM (Netezza) and HP (Vertica) as well as emergent partnerships such as HP and Microsoft.

At the same time the so-called "Big Data" explosion" is being driven by Internet giants such as Google, Yahoo, Facebook, Twitter, Bit.ly, LinkedIn, etc., and startups such as Cloudera, ClickFox, Membase, Karmasphere, and many others. The big data movement is spilling into mainstream enterprises (e.g. BofA, GE, ComScore and the U.S. Army) and is bringing critical organizational questions to CIOs and CTOs related to how best exploit this emerging trend.

These were some of the key themes discussed at the March 1, 2011 Wikibon Peer Incite Research Meeting on big data. The agenda of the research meeting was to present key findings and conclusions from a recent study of 40 Wikibon community members that included dozens of IT practitioners, data scientists, academics, business leaders and technology "alpha geeks." The findings were presented by Wikibon CTO David Floyer and assessed by members of the Wikibon community.

Key Findings on Data Warehouse Trends

A main catalyst for the study was the disruptive force of appliances and specifically Oracle Exadata. Based on economic models and discussions with data warehouse practitioners, Wikibon estimates that appliance-based solutions are between 15%-30% less expensive on a total-cost-of-ownership (TCO) basis than "roll-your-own" solutions. Generally the cost savings trend toward the high end of this scale, but the high expense of Exadata licenses tempered the savings for many users.



Wikibon Top Findings

In addition, the study revealed that:

- Virtually all practitioners indicated that they were I/O constrained. Users are looking to flash storage and columnar compression to ease performance bottlenecks and lower costs by allowing more data to be stored longer.
- While big data and traditional data warehousing are quite different, practitioners indicate that data warehouses are already feeding big data apps, and over time they believe that big data apps will in turn feed traditional data warehouses.

Practitioners almost universally cited performance, capacity and complexity challenges. For example, here are some direct quotes from data warehouse professionals:

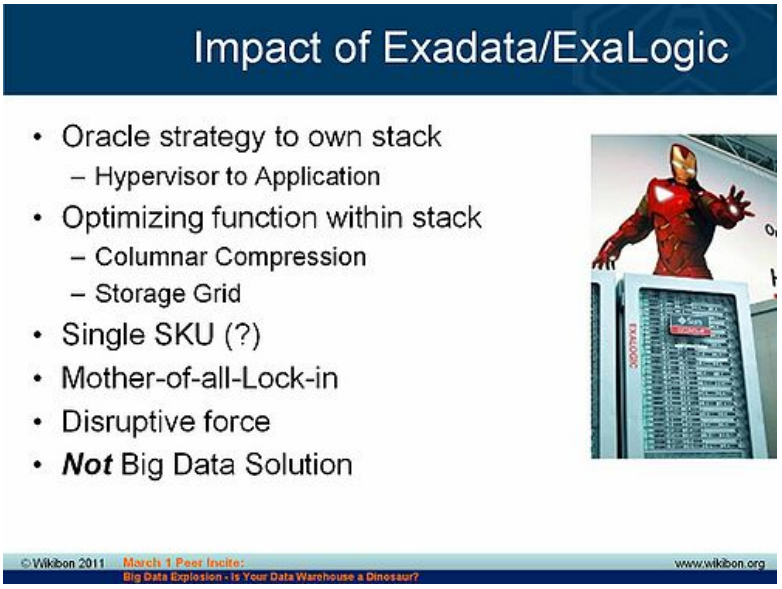
"We are constantly chasing the chips. If Intel puts out a new microprocessor, we can't get it in here fast enough." Financial Services

"Our data warehouse is like a snake swallowing a basketball." -Financial Services

"We're running hard to stay still. We have a patchwork infrastructure with one of everything. We still have not gotten traditional DW/BI working and we're not even thinking about big data." –Mortgage Co

Impact of Oracle Exadata/Exalogic

Exadata is a disruptive industry force. Oracle's acquisition of Sun has hastened the vertical integration movement, and Exadata has been brilliantly positioned by Oracle to gain share on Oracle's major systems and storage competitors. Oracle is optimizing function and is attempting to own the stack from systems to applications. Additionally, Oracle is streamlining functions within the stack, leveraging its middleware extensively and at the same time pushing certain functions to its storage grid. Oracle's extensive use of Infiniband as a systems-to-systems and systems to storage interconnect is a strong networking play designed for low latency and high performance switching.



Impact of Exadata/Exalogic

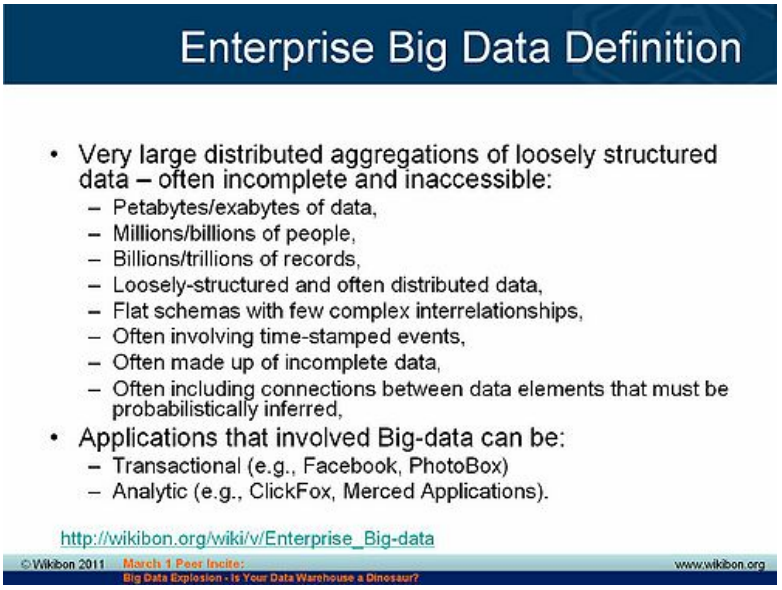
- The single SKU approach of an appliance is alluring because it minimizes complexity, speeds deployment, and simplifies patches. However some practitioners questioned the merits the Exadata single SKU approach because of concerns over accommodating diverse operating systems beyond Linux. In addition, to scale resources in a single SKU model they need to buy a full appliance (for example to scale storage independently of compute). While this can be done through an Ethernet connection off of an appliance, it breaks the single SKU model.
- Exadata is a masterful lock-in strategy by Oracle. It's appeal is very high because it's essentially the iPhone of data warehouse appliances. At the same time, Oracle's high degrees of vertical integration fully expose customers to lock-in.
- As a result of these factors, RYO data warehouse solutions will still command a good portion of the marketplace. These solutions integrate best-of-breed servers (dominated by IBM, HP and Dell) and Storage (dominated by EMC and followed by IBM and HP).
- Exadata and products like it are not well-suited for big data. High degrees of hierarchy and locking make it ideal for traditional DW use cases while big data apps, in contrast, will often thrive with a shared-nothing approach with lots of inherent parallelism (e.g. Greenplum).

"If you look at the Exadata in its current iteration, architecturally it's very promising because they're using solid state not as solid state storage but as flashed memory…" -Insurance

The Big Data Tsunami

The term "big data" refers to data that is too large, too dispersed, and too unstructured to be handled using conventional methods. The big data movement was hatched out of efforts by search companies (e.g. Google and Yahoo) to handle massive volumes of unstructured information on the Web. This vast amount of information was too unwieldy to reside in traditional relational database management systems, and search companies developed techniques to process, manage, and analyze this largely unstructured and disbursed data. Unlike traditional data

warehousing systems that rely on a centralized set of compute, storage, database, and analytic software resources, so-called big data systems move smaller amounts of function (megabytes) to massive volumes of dispersed data (petabytes) and process information in parallel. While trends in big data are alluring, traditional data warehousing remains a $10B+ market worldwide consuming the vast majority of spending (90%) compared to emerging big data markets.

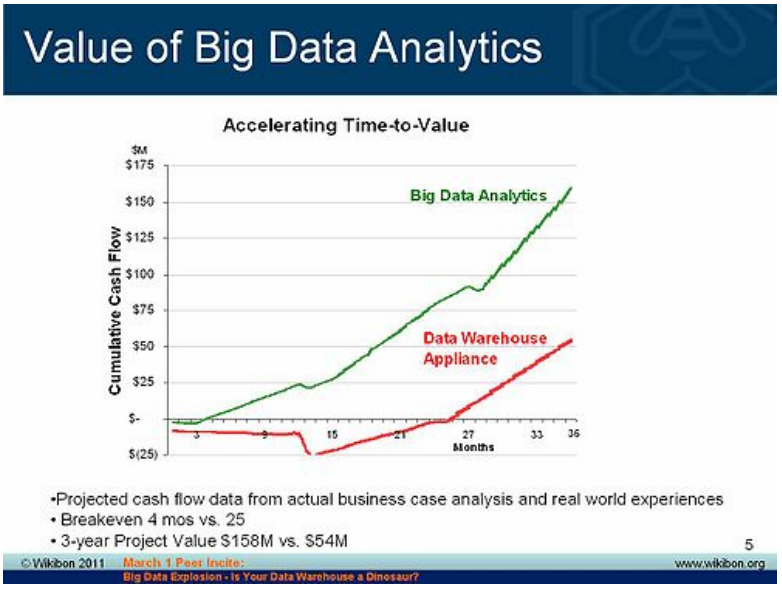


Enterprise Big Data Definition

- Big data applications can be both transactional (e.g. uploading Facebook pictures) or analytical (e.g. ClickFox, LinkedIn, Bit.ly, etc).
- Big data apps have traditionally been built by techies, but as the trend goes mainstream, business heads are driving ways to monetize information and build data products.
- Big data apps are often also very industry specific – for example focused on geological exploration in energy, genome research, medical research applications to predict disease, predicting terrorist threats, etc.

Case Study in Big Data Using MPP Architectures

Wikibon extensively modeled the economic impact of appliances as compared to RYO systems. It also studied the impact of shared nothing / MPP appliances in big data apps as compared to traditional data warehouse appliances. The impact is dramatic as shown in the chart below. The data depicts the influence of data warehouse architectures in big data applications. The vertical axis shows cumulative cash flow and the horizontal axis displays time. In this case, a mobile phone operator used an MPP architecture from Greenplum to dramatically accelerate time-to-value and break-even time tables.



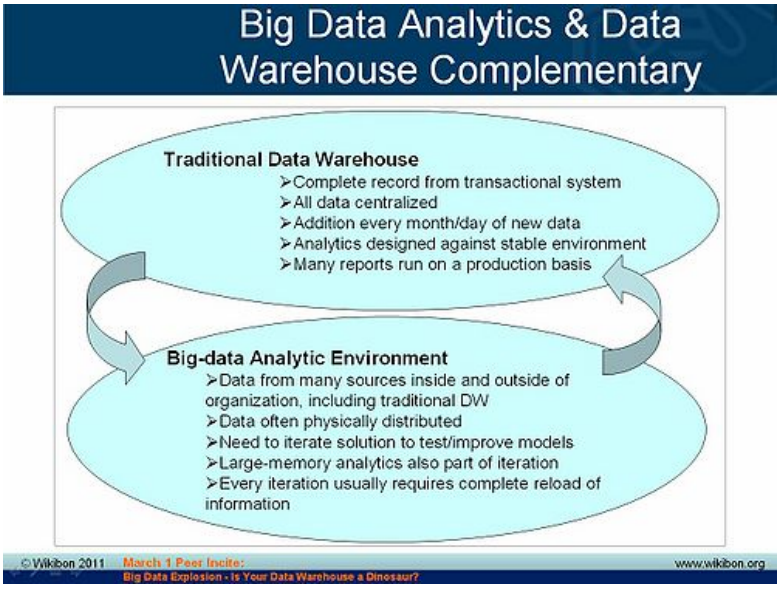
Value of Big Data Analytics

- Big data is more real-time in nature than traditional DW applications.
- Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps.
- Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps (but not so much for traditional DW use cases).

"With Greenplum we could analyze and implement change within a quarter compared to taking a year with traditional DW. This was worth $10s of millions in revenue" -Mobile Phone Operator

Is Your Data Warehouse a Dinosaur?

Big data and traditional DW apps are distinctly different, and key issues remain for CIOs regarding how to organize teams to exploit the big data opportunity. Initially, successes are being seen as skunkworks blossom, but eventually big data initiatives will require strong governance and oversight, especially as they are monetized.



Big Data Complementary to Data Warehousing

The following additional key findings are relevant:

- Traditional DW apps predominantly see incremental growth (with intermittent step functions). In other words, at the end of each month, traditional DW apps might ingest 10% more data – unless there's been some regulatory or other event requiring a step-function increase in database size.
- Big data apps typically require a complete re-load of the data and operate on the entire data sets (no sampling).
- Traditional DW data sources tend to be from internal sources, while big data apps take feeds from internal and external sources and very large Internet databases

The bottom line is traditional data warehouses are not dead, they are being complemented by new and emerging big data apps. These newer applications will take feeds from corporate data warehouses and feed back analytics to the main enterprise warehouse over time.

Action item: While traditional data warehouses remain the lion's share of investment for IT organizations, increasingly, emerging big data applications will deliver value streams of an accelerating nature. Business leaders, CIOs, and IT organizations must collaborate to tap this new source of value and organize to exploit information as a new competitive differentiator. Initially, these initiatives should be separate from traditional DW/BI efforts so as not to dilute innovation. Over time, a data value group will emerge that will play a critical role in delivering future revenue to enterprises. CIOs should plan accordingly and construct a five year plan to evolve this role and the skill sets needed to thrive in this new world.

CIOs Need to Organize Big Data Teams

David Vellante

Even though purists don't like the term, "Big Data" is a reality that's coming to an enterprise near you. Big data is different from traditional data warehousing and business analytics initiatives.

Traditional DW/BI analytics systems tend to be highly centralized with infrastructure built around a 'data temple' consisting of an RDBMS (typically from Oracle, IBM/DB2 or Microsoft SQL Server), high performance storage from EMC (the DW/BI roll-your-own market leader), IBM or HP and analytics software from the likes of Cognos, SAS, Microstrategy, etc. Lately, the market has been disrupted by Oracle with the Exadata appliance that has brought single SKU simplicity to the marketplace

Big data apps on the other hand leverage a diverse set of semi-unstructured or unstructured data types that are disbursed on the Internet. Data scientists perform mathematical, statistical, and data hacking operations to mash up data and create enormous databases typically accessible over the Internet. Increasingly big data applications are becoming a source of competitive value for enterprises as firms monetize information by building data products and services. Going forward, the exploitation of data will become of increasing importance to enterprises.

CIOs need a big data strategy and an answer to the question: "What is our organization doing around so-called big data." The questions CIOs should be asking are:

- What is the potential value of data to my organization?
- Beyond traditional data warehousing, which opportunities exist for monetizing new data models?
- How can data be best exploited, and how should we go about leveraging the new data realities?
- What skill sets do we need to compete effectively?
- Which partnerships should we develop with the line-of-business?
- Which ISVs are best positioned in our market space?

Despite uncertainty within many organizations, several things are clear:

1. Big data requires new thinking – it's not an simply an extension of traditional DW/BI.

2. Big data requires new skill sets revolving around math geeks and data scientists as well as business skills that can construct revenue models around data.
3. Big data is largely unstructured or semi-structured.
4. Big data requires new architectures (e.g. MPP, Hadoop, new database approaches new tools, etc.)
5. Big data presents monetization opportunities that are specific to many industries and not necessarily cross industry.

In addition, major opportunities exist to partner with customer care professionals and ask new questions that previously couldn't be answered (e.g. what are my customers doing, where are they going, what's the experience like, and how can it be improved).

While traditional DW efforts will not die, they have often failed to deliver the predictive monetization value that many hoped for. Big data applications promise to deliver this value as vast Internet data reservoirs are tapped and made available through open APIs, new architectures, and emergent business models.

Action item: While many unknowns exist around so-called big data, CIOs need to position their organizations to take advantage of the new data reality by developing a strategy around big data. CIOs in data-rich industries should organize an elite team consisting of data scientists, programmers, and business professionals that can monetize data. This team should be tasked with developing a comprehensive data strategy and identifying an industry-specific ecosystem that can evolve with both internal and external partners.

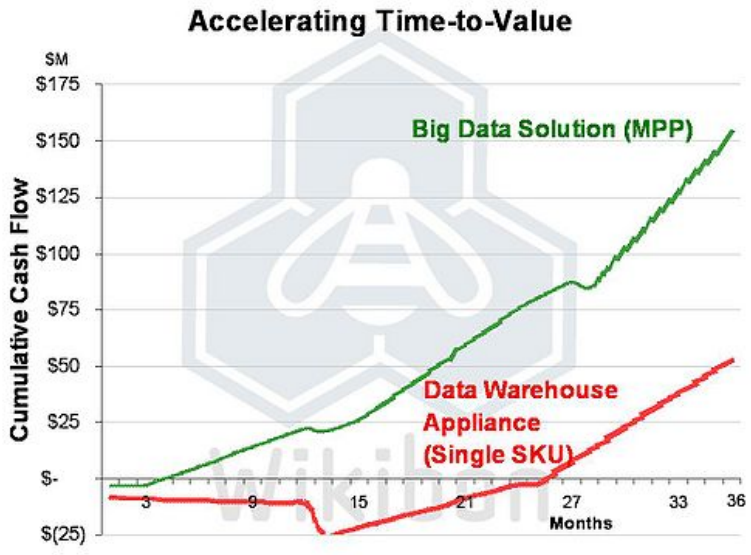
Financial Comparison of Big Data MPP Solution and Data Warehouse Appliance

David Floyer

Executive Summary

Big data is a topic of significant interest to users and vendors at the moment. Wikibon has completed significant research in this area to define big data, to differentiate big data projects from traditional data warehousing projects and to look at the technical requirements. In this paper Wikibon looks at the business case for big data projects and compares them with traditional data warehouse approaches.

The bottom line is that for big data projects, the traditional data warehouse approach is more expensive in IT resources, takes much longer to do, and provides a less attractive return-on-investment. However, big data projects are using new and less mature technologies and carry more risk. As well, big data technologies are unlikely to be suitable for traditional data projects and vice versa – as is so often the case, it is a question of horses for courses.



Comparison of Cumulative Cash Flows of Customer Experience Project using a Big Data Solution (MPP) vs. a Traditional Data Warehouse Appliance
Source: Wikibon 2011

The results of a composite case study are shown in Figure 1, which compares the cumulative cash flows for a project for evaluating customer experience for two different strategies:

1. A traditional warehouse approach using a best-of-breed data warehouse appliance (Oracle Exadata) for the data warehouse and data analytics (this composite analysis was done after the project was completed).
2. A big data approach that used CR-X to define the model and data requirements iteratively, an MPP database (Greenplum) to load the data quickly after each iteration, and big data analytic tools (ClickFox and Merced).

The project favored the big data approach because:

- The data was distributed through many systems both inside and outside the organization.
- The data scheme a simple and “flat”, using event times to inference to establish the customer experience.
- The quality and availability of data was unknown at the start and needed many iterations before the right data could be selected and transformed.
- The MPP database engine was very fast to load and run as the processing was done where the data was stored.
- Very, very large amounts of data needed to be extracted. It was not possible to centralize the data before analysis except by taking a very restricted sampling approach, unsuitable for this particular project.

The financial metrics of the two approaches were overwhelmingly in favor of the Big Data approach:

- Big Data Approach:
  - Cumulative 3-year Cash Flow - $152M,
  - Net Present Value - $138M,
  - Internal Rate of Return (IRR) - 524%,
  - Breakeven – 4 months.
- Traditional DW Appliance Approach:
  - Cumulative 3-year Cash Flow - $53M,
  - Net Present Value - $46M,
  - Internal Rate of Return (IRR) - 74%,
  - Breakeven – 26 months.

The conclusion is that for big data projects different IT tools and approaches are needed. When used, these tools can dramatically reduce the time-to-value – in this case from more than two years to less than four months. The result is that many more speculative projects can be run and abandoned if necessary.

Introduction

Wikibon talked to a number of Wikibon members who had traditional data warehouses and some that had initiated big data solutions using MPP architectures. This composite case study compares different analytical solutions to a big data problem.

The core of the problem is to understand the true customer experience. Most organizations have multiple customer touch points, including call operational systems, call centers, Web sites, chat services, retail stores, and partner services. Customer are free to and do use all these touch points. In the case of a mobile phone operator, each can be measured individually, but the measurement systems do not necessarily reflect the overall customer experience, or show the combined effects of all the touch systems.

Traditional Data Warehouse Approach

Many hundreds of systems are distributed throughout the organization and partners. Each system is largely independent, and any customer experience data is concentrated within that system. The traditional data warehouse system approach would have required extensive data definition work with each of the systems and extensive transfer of data from each of the systems. Many of the data sources are incomplete, do not use the same definitions, and not always available. Copying all the data from each system to a centralized location and keeping it updated is unfeasible. Sampling the data would have been very problematic, as the objective was to construct a customer experience view over time from all the events that took place. Sampling by specific customers would have been very difficult. From a traditional data warehouse point-of-view, this would have been a project from hell. The timescale for implementing this project, revising it, and implementing any results was estimated to be at least one year.

Big Data Approach

The alternative big data approach is essentially to iterate to a result. In this case a modeling tool called CR-X was used to define potential relationships to customer experience from the data; data was extracted from the disparate sources using traditional extract tools (newer techniques such as Hadoop may be considered in the future), and loaded into an MPP database (Greenplum). The data schema was fairly simple and “flat”, which was suited to a database architecture where the processing is done where the data resides. This allows much faster data loading and analysis that traditional data warehouse appliances. Specific customer experience analytical packages (ClickFox and Merced) were used to analyze the data as part of the iterative process.

IT Cost Comparisons

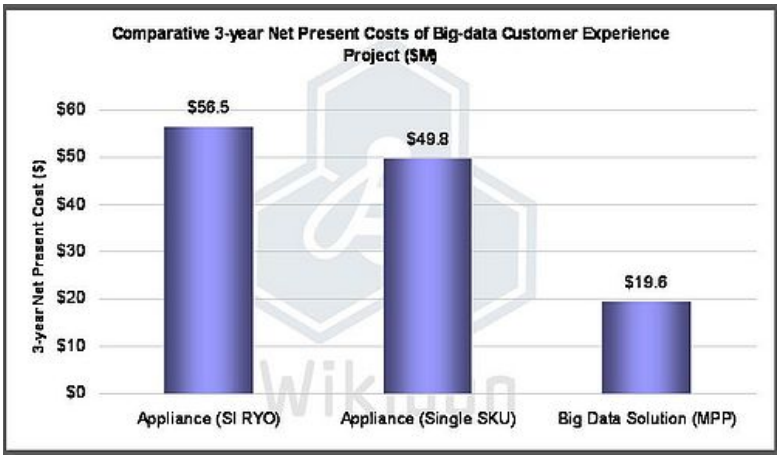
The core assumptions for IT costs are shown in Table 1:

Cost Assumptions	
Cost Reduction Hardware per Year	35%
Cost Reduction Internal Support/PB/Year	30%
*Cost of Maintenance Big Data Solution (MPP)	15%
*Cost of Maintenance Appliance (SI RYO)	22%
*Cost of Maintenance Appliance (Single SKU)	15%
*Cost of Internal Support Big Data Solution (MPP)	70%
*Cost of Internal Support Appliance (SI RYO)	70%
*Cost of Internal Support Appliance (Single SKU)	35%
Growth PB Year-on-Year	78%
*Costs calculated as a function of Hardware Cost	

IT Cost assumptions for alternative approaches to a Customer Experience Project
Source: Wikibon 2011

Three alternative approaches were analyzed:

- A traditional data warehousing approach using a roll-your-own (RYO) approach supplied by a systems integrator (SI). This required 20% less initial IT capital cost that a single SKU solution but was more expensive in support costs as the maintenance of each component had to be done by the customer. The reference model was normalized to an Oracle database. (There were multiple installed alternatives that could have been used.)
- The second case used data warehousing appliance provided by the supplier as a single SKU, including all the software. The software was based on Oracle Exadata, and components included a hypervisor, Linux operating system, and database operational middleware. Support from Oracle would have been from a single update to all components simultaneously. This system was not directly assessed by the customer because it was unavailable at the time. However, as the results show in Figure 2 below, it would have been significantly more cost-effective that the RYO alternative.
- The third approach considered was a big data solution using an MPP database (Greenplum). The cost of the hardware and software was about 40% of the cost of a traditional SI RYO data warehousing system.



Comparison of 3-year IT Costs for Customer Experience Project using a Big Data Solution (MPP), a Traditional Data Warehouse Appliance RYO solution from a Systems Integrator and an Single SKU Appliance solution
Source: Wikibon 2011, based on data in Table 3 in the footnotes

Figure 2 shows the IT cost results of the three approaches over five years.

The source of this data was the detailed five-year table shown in Table 3 in the footnotes. The big data solution was the least-cost solution for this project and about 40% of the next best single SKU appliance solution.

Business Benefit Assumptions

The core assumptions for the business benefits are shown in Table 2:

Benefit Assumptions		
Customer Experience Phase 1 Benefits/year ($M)	$	36.0
Customer Experience Phase 2 Benefits/year ($M)	$	72.0
Time to Benefit Big Data Solution MPP (Months)		3
Time to Benefit DW Appliance Single SKU (Months)		12

Business Benefit Assumptions
Source: Wikibon 2011

Only the best two from the IT cost comparisons were analyzed for business benefits. The project had two phases. The business benefits were considered confidential by the customer and were not discussed in detail. From the information given, the benefits for phase one are conservatively assumed to be $3M /month, rising to $6M/month after the implementation of phase two. The same customer experience benefits were applied to both IT approaches. The key difference was that the big data solution (MPP) could start achieving benefits in three months, whereas the time taken to start accruing benefits with the data appliance was assessed to be 12 months.

Conclusions

The main financial conclusions are shown in Figure 1 in the executive summary. The comparison between the big data approach and the traditional DW appliance approach can be seen by comparing the key financial metrics:

- Cumulative 3-year Cash Flow - $152M vs. $53M,
- Net Present Value - $138M vs. $46M,
- Internal Rate-of-Return (IRR) - 524% vs. 74%,
- Break-even – 4 months vs. 26 months.

The project would probably not have been started using the traditional data warehousing techniques, as the IRR of 74% would have been below the hurdle rate for high-risk projects, and the break-even of 26 months too long for the current economic environment.

The main conclusions drawn from this study are:

- Appliances are best when they have a single SKU, and are supported by single, tested updates to all the components of the appliance;
- Appliances will increasingly become the way that traditional data warehouses are provisioned;
- Big data projects require different IT tools and approaches. When used, these tools can dramatically reduce the time-to-value – in this case from more than two years to less than four months;
- Big data projects will tend to be more speculative and will need tight management review and a willingness to abandon them when necessary;
- Data warehouses will be a significant source of data for big data projects;
- Successful big data projects are likely to be folded back into the data warehouse as data extraction capabilities are built into operational systems;
- In the era of big data, businesses and suppliers will need to adapt to shorter and more intense projects where the outcome is less certain and the IT resources are much more likely to be provided by service providers.

Action item: Big Data projects are real and can lead to enormous business benefits in a short period of time. These projects are likely to be led by the business, and IT should separate these projects from the traditional data warehousing groups to ensure that new big data thinking and approaches can be adopted.

Footnotes: Table 4 below shows the five year IT cost analysis of the three approaches, and is the source of IT costs Figues 1 and 2.

Cost Case Scenario: Total 5-year Costs for Big-data Customer Experience Project (Billions of records)	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Capacity (PB at end of period)	1.0	1.4	6.6	9.2	12.9	18.0	
Big Data Solution (MPP)							
Scaling Factor (Concurrency, Compression)	1.0	1.0	1.0	1.0	1.0	1.0	
Initial Appliance Cost	$ 2,000,000	$ 520,000	$ 4,380,200	$ 1,441,232	$ 1,311,521	$ 1,836,130	$ 11,489,083
Set up & Implementation	$ 300,000						$ 300,000
Maintenance		$ 378,000	$ 1,032,030	$ 1,248,215	$ 1,444,843	$ 1,720,362	$ 5,823,550
Internal Support		$ 1,372,000	$ 4,500,160	$ 4,410,157	$ 4,321,954	$ 4,235,515	$ 18,839,785
Total Cost	$ 2,300,000	$ 2,270,000	$ 9,892,390	$ 7,099,604	$ 7,078,418	$ 7,792,007	$ 36,432,418
Cumulative Cost	$ 2,300,000	$ 4,570,000	$ 14,462,390	$ 21,561,994	$ 28,640,411	$ 36,432,418	
Net Present Value (5%)							$ 31,496,171
Appliance (SI RYO)							
Scaling Factor (Concurrency, Compression)	1.0	1.0	1.3	1.2	1.2	1.2	
Initial Appliance Cost	$ 5,000,000	$ 1,300,000	$ 14,170,650	$ 4,323,696	$ 3,934,563	$ 3,580,453	$ 32,309,362
Set up & Implementation	$ 600,000						$ 600,000
Maintenance		$ 1,386,000	$ 4,503,543	$ 5,454,758	$ 6,320,380	$ 7,108,080	$ 24,772,710
Internal Support		$ 3,430,000	$ 11,250,400	$ 11,025,392	$ 10,804,884	$ 10,588,786	$ 47,099,463
Total Cost	$ 5,600,000	$ 6,116,000	$ 29,924,593	$ 20,803,844	$ 21,059,828	$ 21,277,299	$ 104,781,543
Cumulative Cost	$ 5,600,000	$ 11,716,000	$ 41,640,593	$ 62,444,437	$ 83,504,245	$ 104,781,543	
Net Present Value (5%)							$ 90,535,669
Appliance (Single SKU)							
Scaling Factor (Concurrency, Compression)	1.0	1.0	1.3	1.2	1.2	1.2	
Initial Appliance Cost	$ 6,000,000	$ 1,560,000	$ 17,004,780	$ 5,188,435	$ 4,721,476	$ 4,296,543	$ 38,771,234
Set up & Implementation	$ 300,000						$ 300,000
Maintenance		$ 1,134,000	$ 3,684,717	$ 4,462,982	$ 5,171,204	$ 5,815,685	$ 20,268,588
Internal Support		$ 2,058,000	$ 6,750,240	$ 6,615,235	$ 6,482,930	$ 6,353,272	$ 28,259,678
Total Cost	$ 6,300,000	$ 4,752,000	$ 27,439,737	$ 16,266,653	$ 16,375,610	$ 16,465,500	$ 87,599,500
Cumulative Cost	$ 6,300,000	$ 11,052,000	$ 38,491,737	$ 54,758,390	$ 71,134,000	$ 87,599,500	
Net Present Value (5%)							$ 76,139,510

5-year Comparison of IT costs for a Customer Experience Project using a Big Data Solution (MPP), a Traditional RYO Data Warehouse Appliance from an SI and a Traditional Single SKU Data Warehouse
Source: Wikibon 2011

IT Appliances, Big Data, Both Challenge the Traditional IT Organization

G. Berton Latamore

The new IT appliances such as Oracle's Exadata and the big data phenomenon are coincidental revolutionary trends in IT that only intersect incidentally. However, they do have one important thing in common. They both challenge the traditional siloed IT organizational structure, and they reinforce the pressure on the organization generated by the need to treat applications holistically to provide better service to business end-users.

IT appliances such as Exadata have several important advantages, including their much lower cost. But they create an organizational headache. Because they are essentially “IT-in-a-box” containing processing, storage, network connectivity, operating system, and applications in a single package, they do not fit into the traditional IT organization. Early adapters Wikibon has talked to have tended to assign their Exadata boxes to the storage group, but that is really a forced fit because the box is not only or even basically a storage device. It is an everything device, closer akin to a PC in that sense than to traditional IT components. Assigning it to storage forces the data gurus to deal with application, network, and other issues outside their expertise, but the same would be true of any other IT group.

What makes more sense is to build a multifunctional team to support whatever application is running on the appliance. This team should include skill sets covering all the basic IT infrastructure components plus the application itself. This cross-functional team would be dedicated to providing specific functionality to the end-user at a specific service level.

Some visionaries have advocated this kind of organization for several years to solve the problems of providing better service. The problem is a basic disconnect between IT professionals, who measure performance in terms of server uptime, network transmission speeds, disk seek times, etc., and business end-users who see IT in terms of specific functionality that they need to do their jobs. The argument is that IT organizations need to reorganize around the applications and functionality sets that they provide and see those holistically rather than focusing on the different speeds and feeds of the components. When all those components are in one physical box, the idea of reorganizing into cross-functional teams around specific applications or end-user functionality becomes even more logical.

Big data may have an even greater impact on the IT organization in the end for two reasons. First, big data projects are almost always stand-alone efforts, largely isolated from normal IT. They often rely heavily on third-party, sometimes publicly available data sets and other resources. And because they are often experimental, short-term, and high-risk, and the next big data project may require very different resources than the present one, it often may make sense to rent cloud resources rather than building a system in-house. And they are usually completely driven by

business need and a business executive champion rather than by the CIO. They also are heavily analytic in nature – the idea is not just to capture and store data but to create business value through advanced data analysis. This means that they often will require teams that include data "quants", analytics geeks, cloud service providers, and business people, none of whom are part of the IT organization. Ironically the IT team supporting these efforts is once again likely to be multi-skilled. A siloed organization will be unable to provide what these projects require.

Action item: CIOs should start experimenting with cross-functional teams to support specific applications or end-user groups, particularly if they are also bringing in IT appliances, becoming involved in big data projects, or simply feel pressure to improve the quality of service to end-user groups. By looking at applications holistically IT can better understand the end-user view of computing.

The Emerging Big Data Vendor Ecosystem

Stu Miniman

The big data trend has intersections with traditional data warehousing, analytics, and large file systems. While the definition of big data is still being debated, enterprise IT vendors are already jockeying (including plenty of M&A) for appliance and big data positioning.

- IBM has a strong background in analytics and data warehousing and has considerable assets that relate to this space including: Cognos, SPSS, DB2, scale out NAS (SONAS) and Netezza (through acquisition).
- HP is playing catch-up but is motivated with the acquisition of Vertica and other assets like IBRIX.
- Oracle is disrupting with Exadata/Exalogic and heavy vertical integration.
- EMC's acquisition of Greenplum and Isilon put it squarely in the big data space.
- Microsoft has a solid story (e.g. parallel DW and shared nothing) but doesn't seem to have clout beyond SMB.
- Cisco is also one to watch with location-aware services, routing services, and cloud network infrastructure.
- *Update 3/3/11* Teradata moves to acquire Aster Data (rumors had Dell paired with Aster Data, expect more activity here)

Generally, startups such as Cloudera, ClickFox, Membase, Karmashpere, DataStax, and Internet companies like Google, Yahoo, LinkedIn, and Bit.ly are driving innovation.

Vendors must go beyond simply supplementing product lines with new offerings. Traditional enterprise software pricing models, which are based on compute and/or storage resource installed, are not appropriate for these new markets.

Similar to what is being done for cloud environments, enterprise vendors need to fine-tune solutions and balance traditional pricing models with the new data reality (elastic, on-demand, distributed) while competing with startups that have no legacy base to protect. Vendors need to look at open source stack models (e.g. Cloudera) to determine where/how they fit, should they compete or complement, and how can they monetize their IP. Open source or other methods that allow customers to try the secret sauce before buying is required in this space. Greenplum's Community Edition is a good model to observe. The bottom line is traditional pricing and go-to-market models won't work.

Action item: Vendors need to differentiate solutions for the new data realities from traditional product offerings. Beware simplification of the trend (see *cloud-washing*); it's about creating value from information and today most solutions are specific by industry and use-case. CIOs should look at success stories for guidance in this nascent space.

Leverage the Cloud for Big Data to Avoid Traditional Traps

Nick Allen

One of the most cited benefits of big data applications is flexibility. It can often take months or years to change a data warehouse application, while changes to big data applications may only of few hours or days. Moreover, big data

apps typically require a complete re-load of the data, operate on entire data sets, pull distributed information not only from inside the firewall but increasingly from the cloud, while data warehouse apps normally see incremental growth of data from internal sources.

Fortunately this highly dynamic nature of big data apps presents big savings opportunities. Like the data, the IT infrastructure can also be dynamic by leveraging cloud computing and cloud storage. In essence the infrastructure can be rented from a bevy of smaller players or big ones such Amazon, Google, and Yahoo. Indeed, many offer HADOOP-based cloud services along with their own added value. In addition, these providers often offer data buyers in search of some particular data set a way to visualize the data before purchasing it, thereby ensuring buyers don't waste money on data they don't want. An early example is Google Books, which provides sample pages, but cloud service providers data visualization services are far more sophisticated.

Increasingly, 'rented' cloud infrastructure will be the norm for big data analysis, reducing the reliance on in-house IT, providing faster time to solution and significant cost savings.

Action item: If in doubt, outsource infrastructure to the cloud. For big data apps, IT organizations should take an experimental mindset, avoid investments in in-house infrastructure where possible, and rent from the cloud.

categories

Big Data, CIO, ClickFox, Data, Data warehousing, EMC, Greenplum, HP, IBM, Information Management

Contributors

Wikibon

Comments (0)

Comments on 'Big Data - Is your Data Warehouse a Dinosaur?'

There are currently no comments. Be the first!

Post A Comment

You must be logged in to post a comment, please Sign in

Revision ID	Author	Timestamp	Comment
41555	Wikibon Daemon	12 Jun 21 16:48:51	
32878	Wikibon Daemon	11 Mar 18 13:28:28	
32872	Wikibon	11 Mar 18 12:33:00	
32842	Wikibon Daemon	11 Mar 09 16:54:36	
32841	Wikibon Daemon	11 Mar 09 16:53:10	
32840	Wikibon Daemon	11 Mar 09 16:52:51	
32839	Wikibon Daemon	11 Mar 09 16:52:22	
32838	Wikibon Daemon	11 Mar 09 16:51:52	

32834	Wikibon Daemon	11 Mar 09 16:45:37	
32833	Wikibon Daemon	11 Mar 09 16:45:03	
32832	Wikibon Daemon	11 Mar 09 16:44:27	
32831	Wikibon Daemon	11 Mar 09 16:43:20	
32830	Wikibon Daemon	11 Mar 09 16:36:13	
32829	Wikibon Daemon	11 Mar 09 16:34:48	
32828	Wikibon Daemon	11 Mar 09 16:24:15	
32826	Wikibon Daemon	11 Mar 09 16:12:34	Protected "[[Big Data - Is your Data Warehouse a Dinosaur?]]" ([edit=sysop] (indefinite) [move=sysop] (indefinite))
32825	Wikibon Daemon	11 Mar 09 16:11:57	Created page with ""Storage Peer Incite: Notes from Wikibon's March 1, 2011 Research Meeting"' Recorded audio from the Peer Incite: {{Click||image=Icon_listen.gif|link= Media:Peer Incite 03...'

Cu

most recent wikibon articles

- Citrix Licensing
- Aws Partner Group
- Managing the Risk of Hidden Data
- All New Content on Wikibon dot com
- Big data analytics

latest wikibon blog posts

- Breaking Down the Hortonworks S-1
- Industry Vet Rich Napolitano Joins SDN Startup Plexxi as CEO
- Dell World 2014: No to Wall St, Yes to Cash
- #BigDataNYC and Wikibon Capital Markets Day is Almost Here
- HP Invests in Hortonworks to Jumpstart its Transformation

more »

company profiles

- VMware
- Oxygen Cloud
- Objectivity
- Avere Systems
- Cleversafe

- Q1 Labs

all »

wikibon community information

- About Wikibon
- Tour Wikibon
- Wikibon Tutorial
- Wikibon Publishing FAQ
- Wikibon Contributor Center
- Wikibon Help Section

Browse best practices . publish tips . access project tools . collaborate with peers . get help on RFP's . use privacy settings to control who sees your info . join a group and share experiences with colleagues . review case studies . read professional alerts

© **Wikibon 2008-2019**