
A Krish Krishnan White Paper

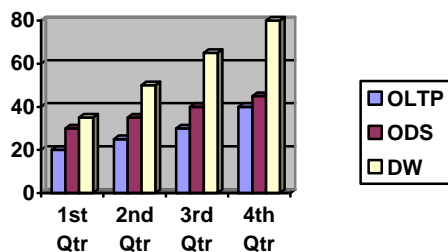
**Using the Data
Warehouse Appliance for
Operational BI**

Using the Data Warehouse Appliance for Operational BI

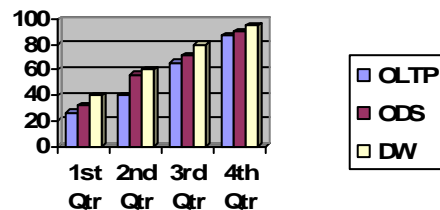
One of the hottest trends in managing your business, whether it is a retail chain or a health insurance company, is to understand how your business runs on a daily basis through operational analytics and reporting. At the push of a few buttons the power to access critical data to drive decisions can be very useful in managing your business. The proliferation of a slew of pre-configured analytics that can be run from your PDA or your Cell phone has got the attention of the business world. However, the amount of data and the frequency of its availability pose IT challenges, as business decision makers' expectations have changed toward on-demand access to the most granular level of information.

The data management problem is not confined to the data warehouse world alone, it transcends to the operational systems and to the non-bi applications like backup and recovery too. We are talking about drinking-from-a-garden-hose-versus-a-fire-hose problem. What we mean by this analogy, is the data growth in the industry is doubling every 6 months and sometimes even every quarter. With that unprecedented growth, handling concurrent data loading and querying in the operational BI environment becomes just as large a challenge as it does for the data warehouse itself from an infrastructure and data processing perspective.

Data Growth By Level - Pre 2003



Data Growth By Level - Current Trends





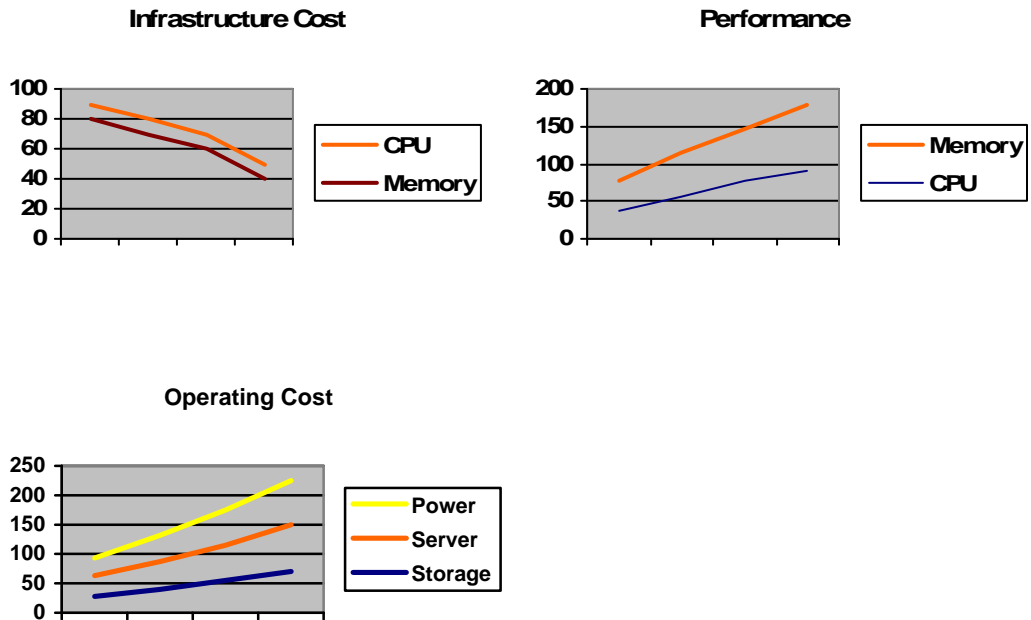
What is the driver behind this data explosion? To understand the key drivers let us take a look at the current trends.

1. Global economy – The world is flat when we talk of businesses. Whether you are the world's largest retailer or the best coffee chain in the USA or the best resort in the Caribbean, there is a connection between your worldwide suppliers and customers. Today business value is driven by opportunity cost. If you lose time in accessing key information, you are left playing catch up all the time.
2. Competition – The drive to capture market and wallet share has propelled businesses to demand access to detailed data in a short turnaround from the transactional system.
3. Business Intelligence – Increased adoption of business intelligence solutions has contributed to more demand for near real time data
4. Event Driven Processing – Increased web traffic is another big contributor to the combined increase of data volumes and the need to analyze the same as close to the occurrence of the event as possible.
5. Data Mining – In its second coming data mining is getting more attention. The ability of companies to analyze large data volumes is providing information that was not feasible to access and analyze earlier. But this also means more data requirements from a historical and current data perspective.
6. Analytical Applications – Analytical applications have become trendsetters in the data warehouse industry. The analytical applications too demand data in detail *and* aggregated structures.
7. Legal Mandate – SOX, Basell II and HIPPA compliance demands more details than previously gathered. Legal mandates are a big contributor to the data explosion.

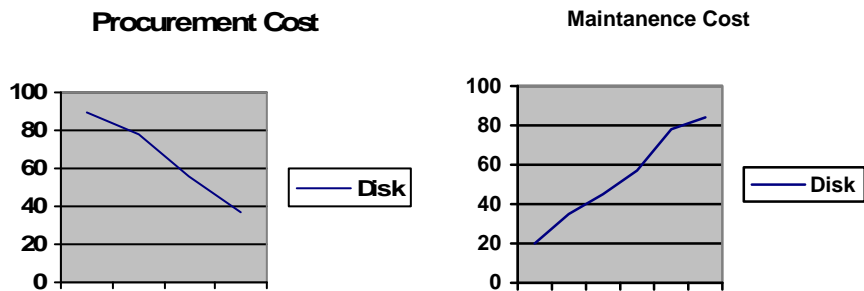
This increased demand for data is not confined to just the data warehouse, it extends beyond the data warehouse into the operational systems on one end, and datamart and analytical databases on the other side. This paper will focus on operational BI.

Let us look at the different impacts arising out of the data explosion and the demands that are imposed by Operational BI.

1. Hardware Costs - To support the exploding data volumes, companies need to add servers. New servers mean additional software licenses, additional data center floor space, additional performance service levels, and additional overhead all round. This overhead is not free; it comes with a big bill. CPU and Memory are cheap, but not the cost of operating a server, power consumption and data center heating/cooling costs.



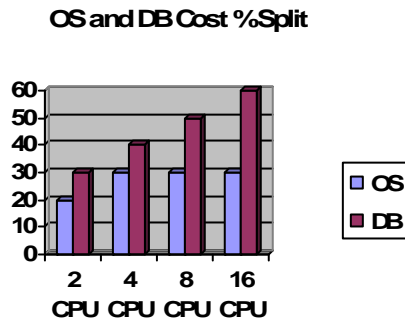
2. Storage Costs – Storage costs are not increasing due to the underlying disk hardware; that technology is getting cheaper. What is driving the cost is the configuration of the standard storage infrastructure to handle the data volume and operational BI demands.



- a. Online – More data volumes mean more storage. IT increases the amount of disks and eventually the infrastructure needs expansion. Now disk is cheap, right! Wrong? Think again, disk is cheap but it's not just the cost of disk that needs to be calculated. The cost ripples through your SAN/NAS configuration which increases due to the SAN costs in its infrastructure and setup.
- b. Offline – Increase in data volume means increase in offline storage costs too. You will need more tape or disk to store data offline.

•
•
•
•
•
•
•

3. Software Costs – In order to support the data volumes and the operation BI needs you need more servers and storage. This drives additional layers of infrastructure, which in turn drive the need for licenses of software. The graph below shows the cost implications on software as infrastructure increases.



4. Maintenance Costs – Maintenance costs for hardware, software and storage, add to the pieces of the puzzle that make up IT spend in all organizations. The additional demands placed by users for operational BI, keep impacting the bottom line across the budget. Which maintenance cost takes precedence? Often due to this issue, companies choose to keep certain applications on older versions of software and thus end up with higher support costs overall.

Now that we have discussed the different aspects of what portions of IT are impacted by Operational BI, let us take a step back and look at how operational BI works.

Operational BI

Operational BI is all about reporting on current business trends, which will provide closed loop feedback to the business about critical behaviors or trends that can lead to increase in productivity and efficiency. An increasing number of user segments across the enterprise are using operational BI to manage and influence changes to the business environment. The table below lists some of the areas in operational reporting and operational BI and the data consumers.



Business Requirement	Data Consumers
Parameterized and Ad hoc Reporting	<ul style="list-style-type: none">• ERP / CRM• Sales• Financial Systems
Key Performance Indicators	<ul style="list-style-type: none">• Dashboards• Scorecard Analysis
Dynamic Aggregations	<ul style="list-style-type: none">• Data Patterns and Exploration

Why operational BI is gaining momentum BI has been the domain of executives and analysts who use it to make sense of a company's ever-increasing volume of data and to gain a comprehensive view of the business. The driver behind operational BI is to get the *right information into the right hands at the right time*, no matter where in the corporation a worker resides. Why does this matter? For example, if you want to provide a better service to customers, the call center representative having access to the most recent data about the customer will benefit both the customer and the company. This is a direct impact to the bottom line of organization, since a happy customer is a brand ambassador too.

Integrating the Appliance to the infrastructure

One way to mitigate and control costs across the data warehouse infrastructure to support the growing data volumes and operational BI, is to look at alternative solutions like the data warehouse appliance. But what is the differentiator between the appliances; there are many vendors and offerings. Before we go there, let us step back for a moment, the bottom line is the current solution of hardware and software bundles are effective at what they do and have been built for which is transactional data processing and historical reporting. It is a stretch to deploy the same stack in areas where the strength of the bundled solution is not robust enough to handle the data volumes and user demands. This is what we referred to as "drinking from garden versus fire hoses."

The goal of the data warehouse appliance is a purpose built platform to manage and address the data management and data architecture needs for organizations.



In order to achieve the best performance there are two kinds of appliance architecture that can be leveraged.

1. The replacement architecture – In this setup you can deploy the appliance as the new platform on which you deploy the applications. The drawback of this approach is how often can you have the bandwidth to implement this kind of a solution? How quickly can you realize the ROI of this deployment? The challenge with this architecture is that you need the resources (time, staff etc) to port from current to new architecture. Often the limitation of these resources stretches the implementation cycles and cause problems upstream and downstream both with management and users. This approach often becomes expensive and sometimes the program is deemed a failure since the original ROI that was calculated as the savings in deploying the appliance, diminishes due the delays in the deployment processes.
2. The augmentation architecture – In this setup you can deploy the appliance as an augmentation to offload data intensive operations from a host computer. The offloaded workload may involve operational, specialized analytics, or archival processing. The benefit of this deployment is all the original applications and the underlying infrastructure can be left intact. The appliance is actually deployed as the true augmented extension to the existing infrastructure. By adapting to this approach, you can actually predict and measure the ROI since the risk associated with the deployment is minimal compared to the replacement approach.

The augmentation architecture is the category where the offering from Dataupia, called the Dataupia Satori Server fits.

Dataupia Satori Server Overview

The Dataupia Satori Server is an integrated bundle of commodity hardware, operating system, storage, database interfaces and a dynamic aggregation engine all bundled into a single blade. A system can be built from 2 to n blades, depending on scalability requirements.

Satori Server is built from the ground up to handle mixed workloads of queries. The underlying SQL is all ANSI 92 standards and will support all BI applications from vendors including open source.



A fundamental differentiator of the appliance offering from Dataupia is the driving principle behind the design, which is to manage the data workload and architecture without being complex. The key benefits from this architecture are

- Transparency – The appliance integrates with an existing database and operates within environment in such a way that applications do not interact with appliance at all. From the database's perspective, the appliance is seen as remote tablespace.
- Self managing – The appliance is built to be self managing, which means that it can recognize the addition of a new “blade”; manage disk outage by switching between disks etc. This is a fundamental architecture differentiator in the appliance world, what makes Satori a desirable platform is the ability to scale without administration limitations or resource dependencies.
- Non-disruptive – This is a key driver to consider when you bring in augmentation architecture. The business needs to keep driving its daily deliverables while IT can build a robust backplane. When you deploy an appliance to your IT infrastructure you want to have minimal disruption in your business and IT teams. This goal is very critical when selecting an appliance.
- Continuously scalable – The new additions in the backplane need to be continuously scalable, which means you can add more disk and horsepower without a system downtime, planned or unplanned.
- Highly cost effective – The augmentation architecture is cost effective, because you can buy as needed and the integration of the additional “node” is self managed.
- Dynamic aggregation – An additional component that the Dataupia platform provides is a powerful pre-calculation engine. This engine can build multi dimensional aggregates in matter of minutes to hours, not days for customer use, with complete benefits including drill down across dimensions, drill across time while supporting thousands of users concurrently.

With the architecture benefits clearly seen and the overall picture of augmenting with the current investments, it is clear that to manage the data problem outside the data warehouse; the data warehouse appliance can be deployed in the augmentation architecture as a data management appliance.

The Dataupia Satori Server has been built for mixed workload query management. This plays well with the ability to integrate the architecture into the operational environment. Mixed workload is fundamental to Operational BI where many users are using systems in many ways, often 24 by 7. There are a few specific industry segments that can integrate this architecture today to drive daily business decisions at multiple levels in a company from senior management to customer-facing staff. In the next sections of this document we will be discussing the implementation scenarios for the Satori server in operational reporting and BI types of applications.



Implementation Scenarios

The following scenarios illustrate the uses of the Dataupia Satori Server in operational business intelligence applications across the enterprise.

1. Telco – Dynamic CRM

A common issue with wireless phone and data users is dropped calls or data packets. This is a problem that happens due to wireless infrastructure issues. Using predictive analytics and the ability to analyze data in large volumes using a data warehouse appliance, wireless providers can implement a program to dynamically predict which infrastructure component will fail and ensure that a proactive management of such failures will improve the level of service that can be provided to customers.

The Dataupia Satori Server has been implemented by infrastructure providers for wireless carriers to solve such a problem. Data packets from the towers are collected and sent to the appliance. The data is loaded using the built in multi loader and available for analysis almost instantly. By being able to look at near real time data and report on it with minimal latency, it is possible to build proactive reports and monitor the operational aspects of the infrastructure. This operational intelligence is driving the ability to maintain the infrastructure in a 99.999% uptime. This results in retaining customer loyalty.

The overall ability of the platform to load and build massive volumes of data for use in a near real time fashion at a fraction of the cost compared to traditional data warehouse platforms, makes this an attractive solution offering. A big plus point in this approach is the ability to use existing master data and metadata management solutions and extend the same to the new platform.

2. Healthcare – Smarter and faster clinical analysis

Clinical trials of drugs by pharmaceuticals and physicians generate a lot of data from patient studies. The current data set from a clinical trial is compared with historical data from past trials to check and measure the effectiveness of changes added to the drug and the population on which it was tested. The current challenge faced in this area is the need to process data at a faster pace.

This is where Dataupia provides a very effective platform the data from historical loads can be loaded into a designated set of blades and the data from the current loads can be loaded to another area of designated blades. The data can be loaded very fast using the data loader.



On the analytical end, a multidimensional data cube can be built and refreshed everyday or as needed. This cube can be used to perform all the desired analysis, detect patterns and behavior trends, calculate results etc.

Since Satori is compliant with both Oracle and SQL Server or as a stand alone platform, it supports ANSI 92 SQL which means that you can run any standard reporting tool on top of this database to report on the results.

3. Historical Data and Customer Loyalty –

It is a well known fact that casinos know their customers the best. A lot of casinos actually have a special customer card which you are awarded as a customer. The customer activity when the customer is inside the casino floor and playing at any tables is monitored. Based on the customers winning situation, customers are often offered special gifts like dinner and show tickets etc, to keep them from losing more today and thereby ensuring they come back from next day to recoup today's losses.

Well this demands a lot from the infrastructure perspective. For a casino the cost may not be as important as the ROI, but if you want to implement a similar program in the retail sector, what are your options?

Imagine that as a retailer, you want to identify when the top 10% of your customers shop in the retail chain, you want to offer them special discounts and products or services. To do this, you need to be able to access the customers, their buying trends and be able to prompt the clerk who is completing the customer sale to award the customer.

In order to implement such a program, the volumes of data that we need at the lowest level and an aggregate level are large and deep. The Satori server can be used to implement the data layer of this infrastructure. The dual requirement of lower grain of recent data and the aggregated grain of historical data can be built using the dynamic aggregation and the data can be integrated with pattern and modeling algorithms to produce the desired result.



Conclusion

As you can see there are numerous other operational business intelligence applications that can gain by adapting to the data warehouse appliance platform. As we conclude this whitepaper, the overarching goal here is to highlight the growing volume of data and the demand for its availability, integration and analysis in the operational business intelligence and how an appliance like the Dataupia Satori Server can be deployed to solve such requirements. This approach though radical and bleeding edge, will be a very effective approach to solving such types of problems. With its ease of integration, reuse of existing master and metadata solutions and ability to deploy and scale, companies have started using Dataupia in these areas and have become successful in implementing the solution and realized benefits sooner than later and in some cases benefits that were never possible to achieve due to prohibitive costs of traditional problem solving approaches.

About the Author

Mr. Krishnan is an expert in the strategy, architecture and implementation of high performance data warehousing solutions. He is a recognized data warehouse thought leader, writing and speaking at industry leading conferences, user groups and trade publications. He is a certified Bill Inmon professional and holds all major DBA certifications. In his 18 years of professional experience he has been solving complex solution architecture problems spanning all aspects of data warehousing and business intelligence for fortune 1000 clients. With his 'get it done' approach he has implemented data warehouse solutions ranging up to hundreds of terabyte data volumes and drives performance tuning into existing BI/DW investments to realize greater than 90% performance gains.

Krish leads the [Data Warehouse Appliance Expert Channel](#) at BEyeNetwork.com and is helping drive and mature the data warehouse appliance market. Krish also serves as Associate Vice President of Programs, DAMA Chicago.