ATTUNITY

EGUIDE / **ATTUNITY VISIBILITY**

A practical guide to data warehouse offload and optimization with Hadoop

Executive Summary

Data warehouses have been the foundation for business analytics for many years, and have grown to support increasing data volumes as well as analytics and ETL workloads. Yet over time, some data becomes older and is used infrequently or not at all. And ETL workloads implemented as transformations inside the data warehouse can grow to occupy significant CPU cycles, impacting resources that support critical analytics processes.

Hadoop provides a modern complement to the enterprise data warehouse, presenting significantly lower cost for storing and processing large volumes of data. With Hadoop, enterprises have the opportunity to offload less valuable data from their data warehouse as well as some workloads like ETL, freeing up valuable resources in the data warehouse while reducing total cost of ownership.

With Hadoop, enterprises have the opportunity to offload less valuable data from their data warehouse as well as some workloads like ETL, freeing up valuable resources in the data warehouse while reducing total cost of ownership.

To capitalize on this opportunity, enterprises need to accurately identify the right data and workloads that can be offloaded, and understand the potential impact of moving them on their existing users, reports, and IT resources. This can be challenging given the complex and multi-layered enterprise data warehouse environment.

Attunity Visibility identifies these key areas and offers a unique migration solution to offload unused or infrequently used ("cold") data from enterprise data warehouses to Hadoop.

Attunity Visibility solution for EDW data offload to Hadoop is a proven, cost-effective, and low-risk solution to rebalance data and workloads from traditional data warehouses to Hadoop. Attunity Visibility provides the level of robust intelligence required to run enterprise data environments effectively and efficiently.

Table of Contents

Key Challenges

The enterprise data warehouse today is rapidly filling with rising volumes of data from increasingly varied sources. While this data flood creates new and rich analytics opportunities, it also can overwhelm data warehouses with data that sits idle, sapping premium space and resources. Many analysts estimate that companies on average only use a third of their structured data for analytics. Most IT departments suspect this is the case, but don't know how to identify such data and thereby measure and address the problem.

Most BI and data warehousing inefficiencies arise from incomplete and complex procedures for tracking the use of applications and data across organizations. Legacy application monitoring tools are designed for tracking transactional applications, rather than revealing how the business uses analytical data. Monitoring analytics data and workloads with custom scripts can be cumbersome.

Enterprises can greatly improve their monitoring efficiency, and thereby the efficiency of their data warehouse, by instead relying on automated collection and presentation of data usage metrics through intuitive graphical dashboards and charts. This can enable a proactive, flexible, and forward-thinking approach to system performance thanks to incisive views into how the business uses data. In short, you can manage more effectively because you can measure.

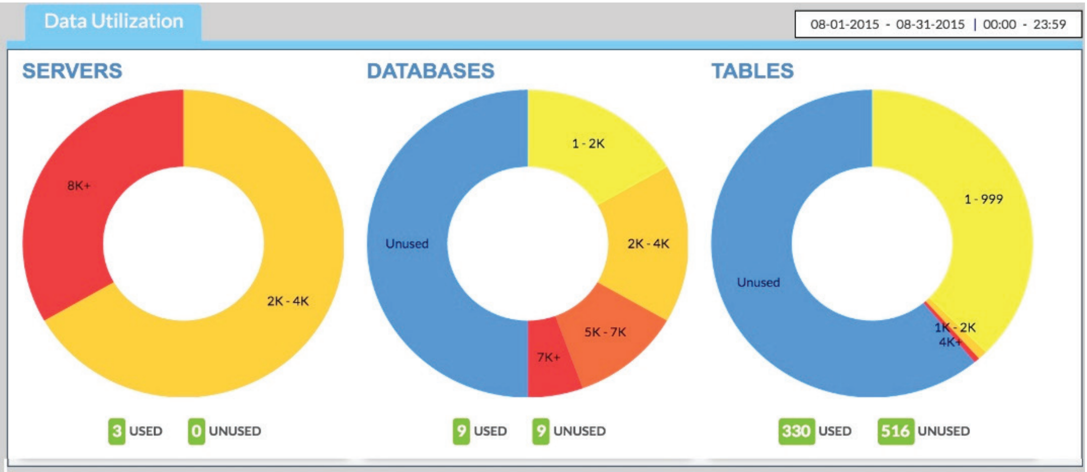
Attunity Visibility's rich charts and configurable metrics provide a uniquely effective tool for accomplishing this. Attunity Replicate provides an efficient and automated method of executing the migration itself.

Checklist for Offloading Data and Workloads to Hadoop

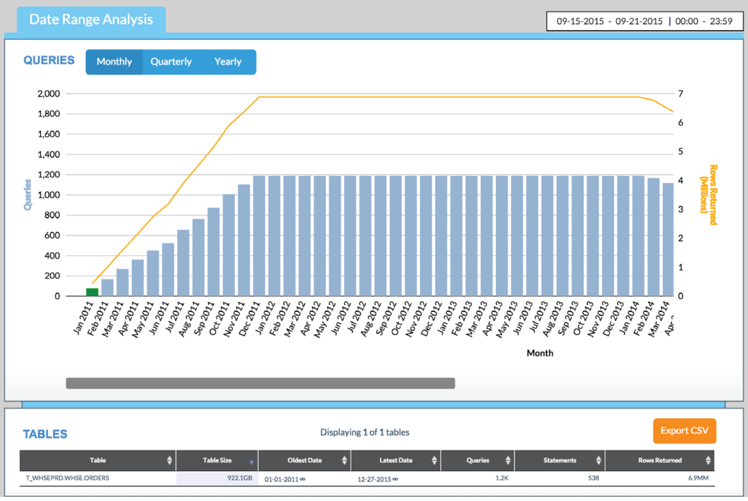
Below is a three-step process which will help in overcoming the challenges highlighted so that you can start getting the benefits of offloading data and workloads from your data warehouse to Hadoop:

Step 1. Identify data to offload

Cold data: Data Warehouse managers can sort data sets by usage frequency over configurable time periods, measuring the utilization of servers, databases and tables. For example, they can measure server utilization to identify hot spots, then drill into logical data constructs such as databases and tables to locate the less used and untouched data that are ideal candidates for offload to Hadoop. The chart below indicates that as much as half of databases and 2/3 of tables have not been used in the defined time period, flagging a significant opportunity for cost savings via Hadoop offloads.



Historical Data: Data warehouse managers can measure even more precisely what data within a table is really being used. In the example histogram below, one can count how many queries are run on various data sets, broken out by creation date on the horizontal axis. This example shows clearly that any columns created before January 2011 are no longer being queried. Columns for 2010 and earlier could be offloaded to Hadoop, archived or retired. This frees up additional capacity and resources in the data warehouse.



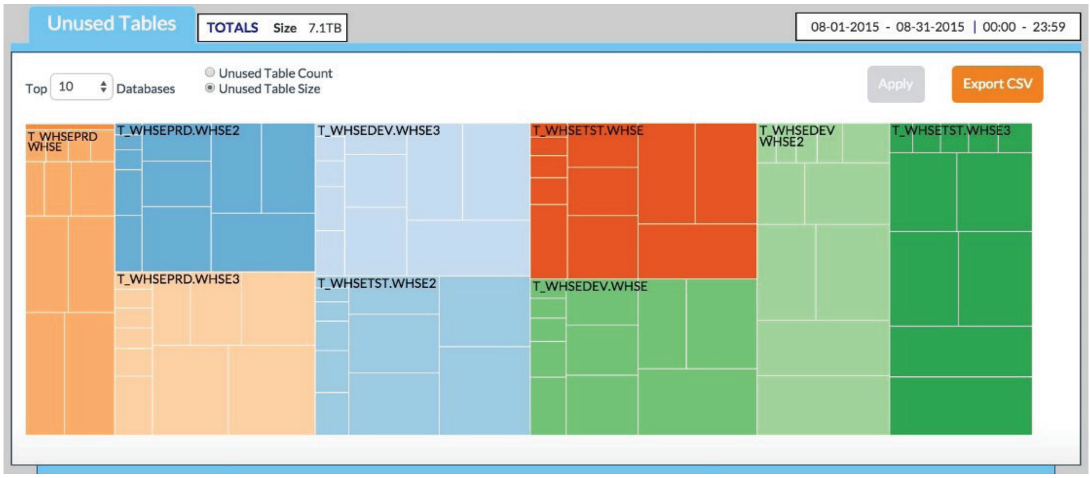
TABLES

Displaying 1 of 1 tables

Table	Table Size	Oldest Date	Latest Date	Queries	Statements	Rows Returned
T_WHSEPRD.WHSE.ORDERS	922.1GB	01-01-2011 ∞	12-27-2015 ∞	1.2K	538	6.9MM

Step 2. Prioritize unused data and related workloads for offload

Attunity Visibility will display the relative sizes of databases and tables as shown below and thereby help data warehouse managers prioritize which should be offloaded for the greatest capacity savings. Attunity Visibility also tracks a variety of resource consumption metrics, including CPU time and CPU % load, to help identify and prioritize the most expensive processing, transformations and other workloads that can be offloaded to Hadoop for performance benefits.



TOTALS Size 7.1TB

08-01-2015 - 08-31-2015 | 00:00 - 23:59

Top 10 Databases

○ Unused Table Count
● Unused Table Size

For example, daily ETL updates to unused tables can consume significant CPU cycles, slowing overall data warehouse performance with no benefit. Moving that data and the associated workloads to the Data Lake can ease the CPU strain.

Step 3. Offload the identified data and workloads to Hadoop

Data ingest and replication software facilitates the process of moving data efficiently between data sources, either in bulk or real-time with enterprise-class change data capture (CDC) technology. With Attunity Replicate, enterprises can accelerate data offload by replicating data from the data warehouse to Hadoop with an intuitive, automated interface that requires no manual coding or setup of agents on either source or target. Attunity Replicate also can ingest data to Hadoop from many data sources, enabling IT teams to feed data to Hadoop to support offloaded ETL processes.

What is the Attunity Visibility solution and how does it work?

The Attunity Visibility EDW Offload solution is a comprehensive end-to-end solution of software and services.

Attunity Visibility ties users and application activity to data usage and workload metrics for data warehouses. The software continuously collects, stores, and analyzes all queries generated by end users and applications against data warehouses without disruption. These queries are correlated with data usage and workload metrics in the Visibility Data Usage Analytical Store. Then Visibility's intuitive graphical interface provides data warehouse managers with a detailed and comprehensive view into data and resource usage.

Key activities include the following:

- Discover unused and infrequently used data
- Monitor all activity on data warehouses to assess resource-consuming ETL and analytical workloads
- Find history of data used from large tables by analyzing calendar dates used in queries
- Identify repetitive and intensive workloads
- Fine tune and optimize data across platforms
- Run impact analysis to predict performance result of potential offload decisions

Attunity product and services experts utilize a repeatable methodology to deliver a focused engagement to capture and report on the key analytics to be used for decision making. This will then be followed by analyzing the data usage across the environment, identifying potential opportunities for rebalancing or retiring data. Our service experts also can reduce Hadoop migration project risk by helping employ proven best practices and automated data checks for data.

Key Benefits of your engagement with Attunity:

- Easily and rapidly identify cold data to be offloaded, without manual coding
- Increase the value of your existing data warehouse and save on processing and storage costs
- Reduce Hadoop migration project risk through the use of proven best practices and automated checks for identifying data usage patterns.

Organizations that use Attunity Visibility appreciate the fact that they can rebalance data and workloads to the right platform by identifying unused data and resource intensive workloads. In addition, IT teams can measure and audit utilization of information assets. Auditing can be automated for data governance and user activity related to sensitive tables and columns can be tracked. Storage and performance can also be optimized, since data can be tiered based on its frequency of use.

What your peers have achieved by working with us

Organizations in diverse industries have implemented Attunity Visibility to address their most pressing enterprise data offload use cases. Here are a few examples:

- **Large online travel company saved $6M by identifying and offloading unused data:** This Company was experiencing exponential growth and an explosion in the volume of data. To better manage their valuable data assets, our customer was looking to scale and optimize their existing analytics data system with Hadoop. They were unable to measure the different data usage patterns within their multi-platform shared services environment in order to rebalance the infrequently used data and workloads to Hadoop.Attunity Visibility EDW Offload solution enabled them to defer $6M in EDW capacity upgrade costs by freeing up existing capacity. They were able to identify the most demanding processing, transformation and unused data, all of which were ideal candidates for offload to Hadoop. About 75 users access Visibility reports and analytics, gaining insights on multiple EDW platforms through a single unified console.
- **Large financial institution saves close to $5M per year by rebalancing and optimizing their Data Warehouse:** A large financial institution's card services data warehouse was nearly at capacity. They were expecting their data to grow by over 100TB in one year. They estimated that in just 18 months they would have had to spend over $15 million, adding new capacity. Instead they began using Attunity Visibility to identify unused data to show the statistics to the business on a monthly basis. By having the ability to collaborate with the business with factual information on data usage, they started to eliminate loading data that was not needed, archive data that was not required, and offloaded data into Hadoop.In addition, every month they identify frequently and infrequently used data that is moved to Tier 1 and Tier 2 storage for optimal data placement. Since customer data is exposed on the data warehouse, they use Visibility to generate weekly reports on which users are accessing sensitive customer data. These reports are sent to the data owners in the business as part of their process to meet risks and compliance requirements.

Conclusion

More and more IT organizations recognize the opportunity to lighten the data and processing load on their data warehouses by balancing EDW data and workloads with Hadoop. The typical methods to identify unused data and resource-intensive workloads, however, involve cumbersome manual coding that undermines the economic benefits of such offloads. The Attunity Visibility solution automatically and plainly presents to EDW managers the data sets and workloads that would better reside in Hadoop. Attunity can guide you through this offload process end to end, starting with the initial EDW assessment and where appropriate, finishing with completing the actual data migration via Attunity Replicate. Attunity can help you build the right offload plan to contain costs and minimize execution risk, thereby incorporating new levels of fact-based intelligence into your data management practices.

About Attunity

Attunity is a leading provider of data integration and Big Data management software solutions that enable access, management, sharing and distribution of data across heterogeneous enterprise platforms, organizations, and the cloud. Our software solutions include data replication, test data management, change data capture (CDC), data connectivity, enterprise file replication (EFR), managed file transfer (MFT), data warehouse automation, data usage analytics, and cloud data delivery.

Attunity has supplied innovative software solutions to its enterprise-class customers for over 20 years and has successful deployments at thousands of organizations worldwide. Attunity provides software directly and indirectly through a number of partners such as Microsoft, Oracle, IBM and Hewlett Packard Enterprise. Headquartered in Boston, Attunity serves its customers via offices in North America, Europe, and Asia Pacific and through a network of local partners. For more information, visit www.attunity.com.

Over 2,000 companies, including half of the Fortune 500, trust Attunity solutions

Attunity focuses on getting the right data to the right place at the right time. The company engineers high-performance data integration and Big Data management solutions that are fast to deploy and easy to operate, empowering enterprises to simply and cost-effectively ensure business-critical information is accessible and manageable when, where and how it's needed to become a more agile, intelligent enterprise.

Contact Us

Americas
866-288-8648
sales@attunity.com

Europe / Middle East / Africa
44 (0) 1932-895024
info-uk@attunity.com

Asia Pacific
(852) 2756-9233
info-hk@attunity.com

Learn More

www.attunity.com