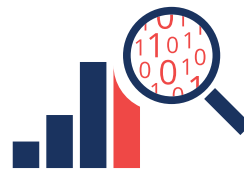


Data Warehouse and Hadoop Management



Analytics for Intelligent Data Management

Information Technology (IT) managers are under siege to meet multiplying business demands while controlling data delivery costs. Data warehouses in large enterprises now routinely reach hundreds of terabytes or petabytes. Hadoop extends the move to multi-platform environments, making it more difficult for enterprises to manage their data ecosystems holistically.

Organizations cannot continue to increase data consumption without a prohibitive impact on infrastructure cost and complexity. They need to measure what they manage. Just as CFOs need metrics to run their business, IT needs metrics to run its infrastructure. To make effective decisions, IT needs to understand how business users are consuming application and data assets.

With Attunity Visibility, enterprises can gain insight into business activity, data usage and performance across platforms including Hadoop to improve operational efficiency and cost.

Data Usage Analytics

Challenges:

- Inability to assess how much history is actually used in specific tables
- Lack of insight into usage of fast-growing structured and unstructured data in EDW and Hadoop
- Difficulty in tracking user activity with sensitive data to minimize risk and compliance exposure

Capabilities:

- Identify unused data that can be archived rather than consuming valuable EDW capacity and loading cycles
- Track key Hadoop file system metrics, including files, owners, file owners group, directory sizes, and correlate with users and growth trends to assist capacity planning
- Assess how much historical data is being used in large tables to guide data retention policies
- Ensure trusted data usage for easier compliance – track users, apps, tables and columns

What does Attunity Visibility provide?

- A single view across multiple platforms, including Teradata, Oracle Exadata, IBM DB2, IBM PureData (formerly Netezza) and Hadoop
- Equal insights into EDWs and Hadoop across storage levels.
- Mapping of users and application activity to data usage and data warehouse performance
- Continuous collection, storage, and analysis of all queries and applications against data warehouses
- Data consumption and workload performance metrics for the entire data warehousing environment

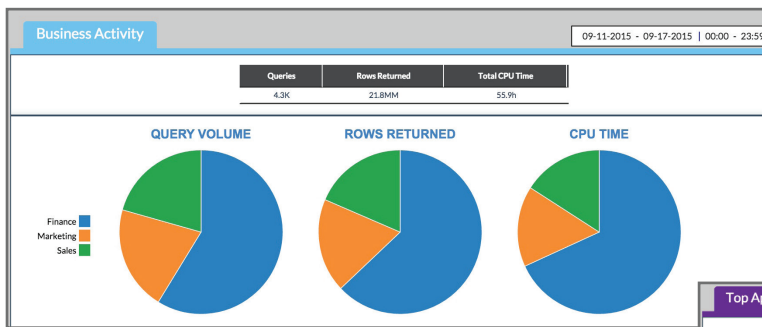
Business Activity

Challenges:

- Rising business demands, but inability to assess return on current IT investments
- Lack of visibility into business usage of applications and data for chargeback
- Difficulty in identifying underutilized applications and unused data by business line

Capabilities:

- Measure activity and usage trends by relevant business group to enable chargeback and show-back
- Justify and assess return on investments; prioritize resources
- Identify the most frequently used applications and data in the business context and assess the impact on your entire data warehouse
- Discover infrequently used applications and unused data to eliminate wasted resources and reduce costs via rebalancing



This table highlights activity and usage trends by line of business.

Workload Performance

Challenges:

- Diagnosing performance issues quickly and without major resource impact
- Difficulty and delay in identifying sources responsible for problematic workloads
- Lack of insight into frequently used EDW and Hadoop datasets to guide performance tuning

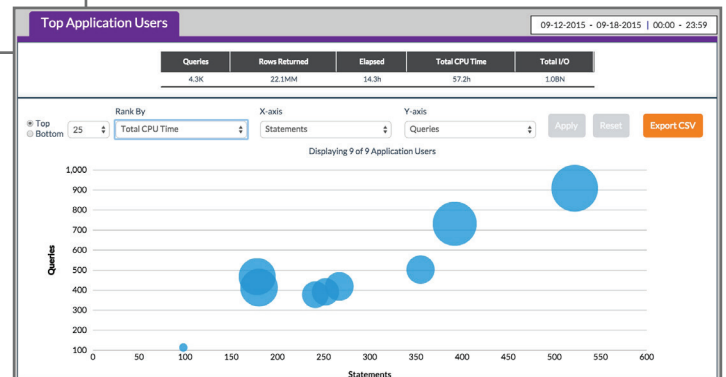
Key Features:

- Easily identify individual application users and BI reports that impact EDW performance
- Monitor workloads and performance for Hadoop and Cloudera Impala
- Identify frequently used data associated with problem SQL; quickly identify associated users and applications
- Discover and isolate repetitive and frequent operations; top joins, macros, procedures; and complex SQL structures

Key Information:

Examples of insights given by Attunity Visibility:

- Finance generates less than 15% of query volume but consumes 40% of CPU – identify the top users and what they are doing
- Over 60% of server time is consumed for loading and ingesting data – discover the most intense repetitive workloads
- Over 30 TB of data is not being used by the business – identify the specific tables that can be offloaded or archived
- The batch load for the key scoring table did not complete – find the users that need to be notified immediately
- The database contains sensitive customer data – pinpoint the BI users and reports used to retrieve unusual amounts of customer data.
- The Data Lake is growing 40% per year – measure file, file group and directory size and usage to assist capacity planning and chargeback



This chart ranks BI users by CPU time, number of queries and statements.