SonarG: Big Data Warehousing for Guardium Data

SonarG is the industry's first Big Data warehousing solution specifically built for optimizing data aggregation, storage and reporting for Guardium Database Activity Monitoring environments. By combining the best of Hadoop, web-scale technologies, MPP performance and NoSQL, SonarG dramatically increases the value of your Guardium deployment by improving visibility, security and compliance while also reducing operating costs.

Long-term, highly granular activity information can now be consolidated into an inexpensive, centralized data warehouse, providing improved access to rich activity information, simplifying data collection processes and significantly reducing operational overhead.

SonarG is a purpose built Big Data solution for Guardium. It makes use of SonarW, an ultra-efficient and ultra-usable NoSQL warehousing solution that has been adapted specifically for Guardium DB activity information in terms of its data schemas, data ingestion and reporting facilities. The result is an inexpensive yet highly effective Big Data solution for Guardium that can cost effectively manage multi-year data for hundreds of Guardium collectors regardless of the audit policies and processes used by different organizations.

KEY BENEFITS

Ultra-efficient and inexpensive storage
Activity data is efficiently captured, compressed and aggregated from tens or even hundreds of Guardium collectors into a data warehouse that easily handles terabytes of data for long-term retention and for reporting in the most demanding regulated environments.

Ultra-fast querying and reporting
Using advanced Big Data techniques, SonarG provides immediate and easy access to large Guardium data sets spanning years of fine-grained activity, without waiting hours for batch reports to be available or daily aggregation processes to complete.

Easy, Fast Deployment
SonarG provides an out of the box Big Data warehouse solution fully optimized to easily and rapidly integrate with existing Guardium deployments. Instead of fragile and lengthy aggregation processes, you simply point collectors at SonarG and focus on extracting the high value data needed for improved security and compliance reporting.

Improved HW utilization
Since SonarG eliminates the need for Guardium aggregators, you can shut off expensive aggregation processes on collectors, and simplify the Guardium HW footprint and operational overhead via the centralized data warehouse. Collector processing capacity and storage efficiency are increased, enabling collectors to support more databases and/or a richer set of data collection.

Expanded functionality while reducing costs
Exploiting next generation data warehousing technology enables SonarG to dramatically reduce the level of ongoing effort and costs associated with overseeing the collection and management of massive datasets, while also paving the way to efficiently expand activity collection policies in order to capture a more comprehensive view of database activity.
Key to the power of the SonarG solution is the use of SonarW, the world's first JSON-native Big Data warehousing solution. This architecture merges the best features of Hadoop, web-scale technologies, MPP architectures and NoSQL to create the ideal data warehouse platform for massive scale data collection and analytics at a fraction of the cost of traditional warehouse solutions. SonarW matches or exceeds the performance of various Big Data alternatives, but does so using a substantially simpler and lower cost HW implementation. And unlike competing Big Data architectures that demand highly specialized coding skills to operate and data scientists to perform analytics, SonarG delivers an out of the box solution that empowers all users with easy access to desired data.

Data from Guardium collectors are exported using CSV extracts – a fast and preferred data extraction pattern. This is the only methodology that can be performed effectively without impacting collector operation. In fact, when CSV extracts replace aggregator exports, collectors run more efficiently and can be used to monitor more database servers and/or expand activity collection. In addition, collector data can be purged more frequently, further reducing storage requirements and enabling the use of smaller virtual and physical collectors.

CSV files are copied to the SonarG host where they are ingested. Ingestion is fast even when CSVs from tens of collectors are shipped to a single SonarG host concurrently due to parallel Big Data ingestion techniques that are built into the Sonar Collector. Data is columnized, indexed and partitioned to ensure fast retrieval time and then loaded into SonarW – a columnar NoSQL data warehouse.

A typical SonarG implementation includes a primary node and possible a secondary node when HADR is required. SonarG nodes can scale to any size – even to petabytes of data. Important – by using SonarW, scale does not mean high costs. SonarW scales in a way that requires less hardware and costs far less than any other Big Data technology solution, as shown via results of Big Data Benchmark in Figure 3.

SonarG provides custom-built analytics and reports for Guardium data. The solution comes pre-packaged with a schema that supports efficient access to data required for regulatory processes as well as analytics required for anomaly detection and outlier identification.
Native columnar data
Data is ingested from Guardium collectors and stored as compressed columnar data, enabling large scale storage at a fraction of the cost of other solutions. This format also makes data available for massively parallel processing, which dramatically accelerates analytics and reporting performance.

Indexed and Partitioned for Speed
Data is indexed and partitioned so that queries that require looking at specific date/time ranges and/or specific dimensions (such as user, source program, SQL verb or even SQL statements) run quickly even when storing tens of terabytes of raw data.

Parallel and Distributed Processing
Key processing is always done in parallel, such as ingesting data from Guardium collectors, executing complex analytics queries or generating reports.

Simple and Effective Reporting and Advanced Analytics Tools
SonarG leverages the jSonar Analytics Platform, an advanced NoSQL Big Data platform that includes the industry’s first NoSQL Data Warehouse (SonarW) and best-of-breed NoSQL analytics tools JSON Studio. These tools make reporting simple, fast and ultra easy to use, distribute and share within any organization.

Figure 3: Big Data Benchmark
ABOUT SONARW AND THE JSONAR ANALYTICS PLATFORM

SonarW is a key component of the jSonar Analytics Platform, which is focused on providing rich analytics, gateway and productivity tools to let developers and analysts access the warehouse easily. As shown in Figure 4 below, the platform includes a number of key building blocks instrumental to the construction and operation of a high performance, yet low cost Big Data warehouse solution.

The platform comes with the following tools and gateways:

- **JSON Studio** – A high productivity analytics tool built to enable users to quickly and easily write queries, visualize data, analyze results and more. JSON Studio was purpose-built for JSON analytics and also works with other Big Data technologies such as MongoDB.

- **Sonar Gateway** – a middleware layer that allows users and applications to call RESTful APIs in order to run queries and access data stored in SonarW.

- **SonarSQL** – A SQL-enabling layer that allows users to write queries in SQL to access data stored in SonarW.

- **SonarR** – A gateway that allows analysts and data scientists to access data from SonarW from within the “R” environment.

SUMMARY

Guardium is the most widely used Database Activity Monitoring solution in the world and proven to assist industry leading organizations in meeting their database security and audit objectives. As reporting requirements become more stringent in support of improved security and expanded compliance controls, the challenge of efficiently storing, managing and mining this vast ocean of audit information can be overwhelming. SonarG is the industry’s first Big Data solution purpose-built to warehouse and analyze Guardium audit data while providing unparalleled efficiency, cost reduction and data accessibility.