jSONAR
The NoSQL Warehouse

SonarW: The NoSQL Warehouse

SonarW is the industry's first JSON-native Big Data warehousing solution. It is fast and simple to use, merging the best from Hadoop and Web-scale technologies, from MPP architectures and from NoSQL. It reduces costs through ultra-efficient analytics execution as well as through simplicity, ease of use and developer productivity. Cost savings over alternative warehousing platforms come from:

ULTRA-EFFICIENT parallel processing so that today's analytic workloads can run using less hardware. jSONAR uses commodity hardware – but even commodity nodes cost money and management of large clusters is expensive. jSONAR's platform efficiently uses resources so that fewer nodes are needed as compared with other alternatives.

ULTRA-USABLE software and high developer productivity. jSONAR brings NoSQL simplicity to the Big Data world and is the industry's first NoSQL data warehouse. It is 100% MongoDB compatible and in the same way that MongoDB is easier to use than relational databases, SonarW is easier to use than relational MPP platforms and much easier to use than other Hadoop services.

SonarW can run standalone or as part of Hadoop and merges Hadoop's scale with NoSQL's ease-of use and with MPP's database performance.

KEY ARCHITECTURAL CHARACTERISTICS

SonarW shares many architecture properties with MPP warehouses, but is purpose-built for today's modern Big Data applications that require semi-structured data, sparse data and use cases requiring schema-on-read.

Native columnar JSON

JSON is always stored as columnar data, is always compressed and kept in a way that can be processed in parallel for every operation. It supports flexible schema and is therefore useful for modern applications and especially for Big Data projects. Because it is native JSON it inherently supports schema-on-read, but it does so efficiently because of the columnar nature of the database.

Everything is indexed and partitioned for speed

All data is indexed without the need for the user to request indexes, but indexing is per chunk and is not memory-hungry so that it can scale to any size.

Parallel and Distributed Processing

All processing is always done in parallel - both across nodes and within a node. Parallel processing is not limited to query only, but also to enrichment and ELT/ETL workloads.

Efficient code, algorithms and data structures defer cluster growth

Code is highly optimized C/C++ designed from the ground-up for efficient columnar JSON processing. Data Structures are lock-less and/or built for multi-thread, multi-core, SIMD and GPU acceleration. It is an ultra-efficient platform that defers the need for large clusters. It scales to any size but does not require an unreasonable number of nodes as do some Big Data solutions.

NoSQL Warehousing on Hadoop

SonarW inherits ease-of-use and simplicity from the NoSQL world, and specifically, is 100% MongoDB compatible. It allows Big Data teams to be more productive and focus on business applications rather than upkeep of the Big Data platform and code. When running on Hadoop it runs on data nodes and uses HDFS.

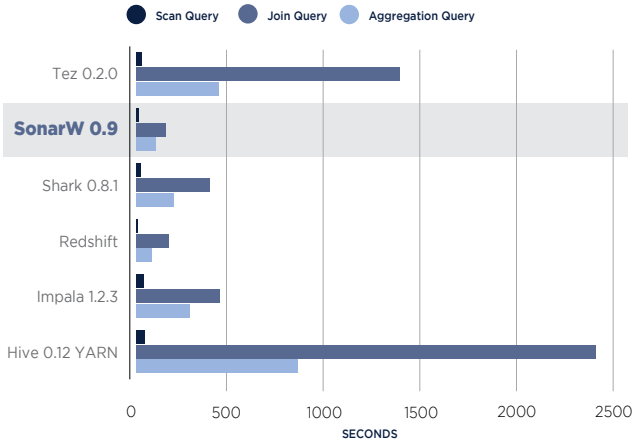
THE JSONAR ANALYTIC PLATFORM

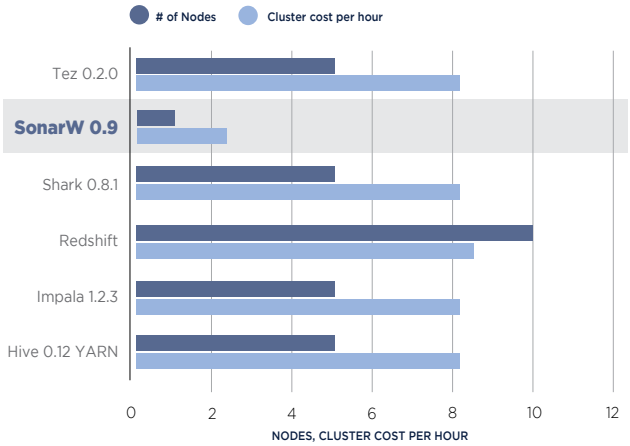
SonarW is the core of the jSONAR NoSQL Analytics Platform. The platform also provides analytics, gateway and productivity tools to let developers and analysts access the warehouse easily. In addition to being 100% MongoDB compatible thus enabling any driver and tool built for MongoDB, the platform comes with the following tools and gateways:

- **JSON Studio** *A high productivity analytics tool built for JSON data allowing users to write queries, visualize data, analyze results and more. JSON Studio was purpose-built for JSON analytics and also works with MongoDB (it received the MongoDB Innovation Award from MongoDB Inc.)*
- **Sonar Gateway** *A middleware layer that allows users and applications to call RESTful APIs to run queries and get 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.*
- **Sonar Collector** *An ingestion framework that supports fast parallel data ingest from multiple sources including JSON files, CSV files, MongoDB (through files, dumps or replication), RDBMSs and more.*

SPEED AND EFFICIENCY

SonarW performs at least as well and often much better than other Big Data technologies, but does this using less hardware (and thus less cost). For example, the following figure shows the results when running the Big Data Benchmark as maintained by the AMP lab at Berkeley (https://amplab.cs.berkeley.edu/benchmark/).





Performance and Cost when running the Big Data Benchmark