

Now Use MapReduce for Operational Intelligence on Your Live Data

Overview

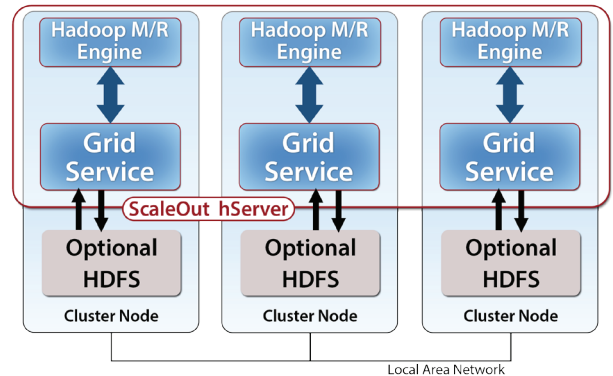
ScaleOut hServer® unlocks the power of Hadoop to analyze live, fast-changing data in production environments. Now applications can extract important patterns and trends and generate immediate feedback within seconds to provide operational intelligence for live systems. This unique combination of an in-memory data grid (IMDG) and an integrated, data-parallel compute engine delivers real-time capabilities previously unavailable with Hadoop or Spark. Standard MapReduce applications now can track and analyze live data with object-oriented access, in-memory execution speed, and high availability – all without the need to install and manage the Hadoop stack of software. Alternatively, ScaleOut hServer easily can be integrated into existing Hadoop clusters to add in-memory data storage and computing.

In-Memory Hadoop for Live Data

Unlike other real-time analytics architectures, such as Spark, ScaleOut hServer incorporates a fast, scalable in-memory data grid that enables data to be continuously saved, updated, and analyzed using ScaleOut hServer’s built-in MapReduce engine. ScaleOut hServer’s IMDG stores key/value pairs across an elastic set of networked servers, ensuring fast data access, linear scalability, and high availability. ScaleOut hServer’s in-memory compute engine executes standard Hadoop MapReduce programs and Hive queries directly in the IMDG, delivering results in seconds or less so that applications can spot important trends in live data as they occur.

Here are some key advantages of using ScaleOut hServer for real-time analysis:

- It is designed to hold fast-changing, operational data in memory as key/value pairs, enabling fast updates to in-memory models of live systems.
- Stored data is automatically replicated to keep it continuously available and protect from server or network failures.
- The built-in MapReduce engine has demonstrated 40x faster execution of in-memory Hadoop in benchmark tests, and no code changes are required to run MapReduce under YARN.
- Performance linearly scales just by adding servers to increase memory capacity and throughput; ScaleOut hServer automatically rebalances the workload.
- ScaleOut hServer seamlessly integrates into existing Hadoop environments and serves as a low-latency MapReduce execution framework. It eliminates Hadoop’s batch scheduling overhead and delivers sub-second start-up times. Data can be streamed to/from either ScaleOut’s in-memory data grid or HDFS.
- The MapReduce engine provides numerous automatic optimizations for memory-based data sets, such as automatic setting of key parameters (splits, partitions, and slots) and distributed combining. These features simplify development while ensuring fast performance.
- Standard Hive queries can be executed at in-memory speeds.
- ScaleOut hServer is fully compatible with Apache Hadoop, Cloudera, Hortonworks, and IBM InfoSphere BigInsights.

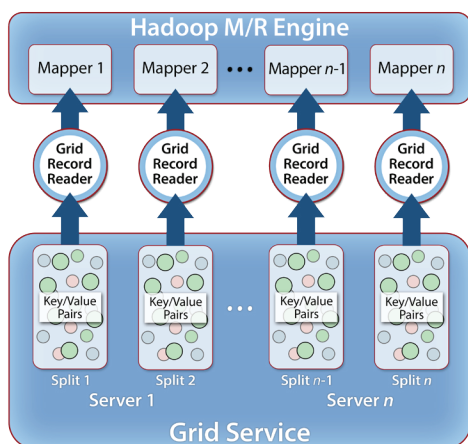
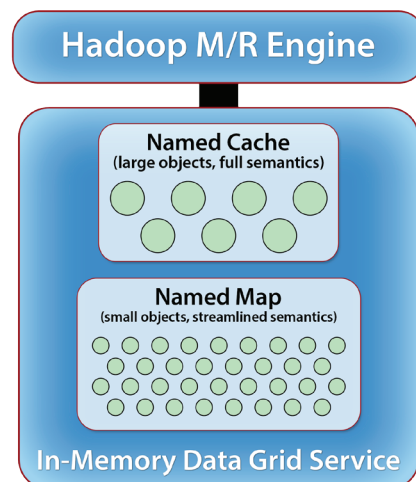


Optimized, In-Memory Data Grid

ScaleOut hServer installs in minutes, and it automatically discovers other servers, self-aggregating into an in-memory data grid spanning a cluster of servers. Using ScaleOut hServer's Java APIs, applications can create, read, update, and delete key/value pairs in the IMDG to manage fast-changing data and track the behavior of a live system. MapReduce jobs execute in seconds or less, providing immediate results that can be used to steer the system's behavior.

IMDGs traditionally host complex, in-memory objects with rich semantics. However, MapReduce jobs typically require storing and analyzing huge numbers of very small key/value pairs, such as sensor data or tweet streams. To handle these divergent requirements, ScaleOut hServer supports two object storage mechanisms in its IMDG:

- Designed for large, complex objects, the **Named Cache** supports rich functionality such as property-oriented query, dependency relationships, timeouts, optimistic and pessimistic locking, transparent access from remote data grids and backing stores, and more.
- With the new **Named Map**, ScaleOut hServer adds Java ConcurrentMap semantics to efficiently manage large populations of small key/value pairs and enable fast analysis with MapReduce. Objects stored in a named map can be queried in parallel and can be cached in the client using user-adjustable coherency policies. For fast loading and updating of key/value data, the Named Map provides bulk insert and bulk update functions.



The Hadoop developer now has the choice to store and analyze either heavyweight objects with rich semantics or lightweight objects depending on the type of data being analyzed.

ScaleOut hServer includes a Grid Record Reader to input key/value pairs to Hadoop's mappers with minimum latency. The Grid Record Reader works with both named caches and named maps. Likewise, a Grid Record Writer enables pipelined output of results from Hadoop's reducers back to a named cache or named map in the grid.

Finally, ScaleOut hServer provides a simple, easy-to-use debugging environment for developing MapReduce applications. This means that developers can rapidly iterate on MapReduce code as they refine their applications.

Community Edition

ScaleOut hServer is available in both a free Community Edition and several commercial editions. The Community Edition is licensed for either evaluation or production use and supports up to a four-server IMDG with a maximum data set size of 256GB. Support for the Community Edition is provided via the ScaleOut Community Forum, where developers can ask questions and exchange ideas with other users and ScaleOut Software's experts.

Please download the community edition at www.scaleoutsoftware.com/products/hserver.

Open source client libraries are available on GitHub at <https://github.com/scaleoutsoftware/hServer>.



ScaleOut hServer®



ScaleOut Software

www.scaleoutsoftware.com

For more information, call us at
(503) 643-3422 or download a
free product trial from our website.