

## Parallel Data Analysis for Fast Time-to-Insight

### Overview

ScaleOut StateServer Grid Computing Edition™ (GCE) combines scalable, in-memory distributed data caching with powerful computational features to deliver a distributed data grid for server farms and high performance computing environments. Advanced capabilities for rapidly searching cached data and quickly developing scalable grid applications make it perfect for data parallel computation. Also included is the ScaleOut Management Pack™ which provides comprehensive tools for observing, managing, and preserving grid-based data.

### Fast Data Access

Financial services and other data-intensive industries routinely demand real time processing and overnight batch analysis of large data sets. By storing fast-changing data in-memory in ScaleOut StateServer's distributed data grid using industry-leading distributed caching, these applications can dramatically reduce access latencies, avoid bottlenecks, and achieve peak performance.

### Powerful Parallel Cache Search

Applications can perform parallel queries to rapidly search grid-based data for selected objects based on properties of stored objects (or user-supplied metadata). Queries can be constructed using Microsoft's Language Integrated Query (LINQ) or Java query filters, providing powerful and familiar tools for developers. Employing patent-pending, parallel search and merging algorithms, GCE provides the fastest possible parallel query across all hosts within the distributed data grid.

### Data Parallel Computation

GCE takes data grid computation to the next level by enabling applications to execute user-defined methods in parallel on a selected set of objects and then combine the results using user-defined merge algorithms. Called "parallel method invocation," this feature simplifies application design and reduces data motion using familiar map/reduce semantics. It works in concert with the grid's job scheduler to further extend the distributed data grid's power to accelerate the performance of HPC applications.

Unlike traditional map/reduce implementations which process large, file-based data sets, GCE stores data sets in a scalable, in-memory, distributed cache within the distributed data grid. GCE automatically maps the user's "map" operation across multiple cores within each server and all servers within the grid. GCE's execution engine maximizes parallelism and minimizes data motion within the grid. In addition, GCE performs high-speed, parallel execution and merging of the user's "reduce" operation across all grid nodes. The result is very high performance for the complete map/reduce operation.

Importantly, the user does not need to write special code to harness the power of the grid's infrastructure. Grid Computing Edition's APIs enable map/reduce calculations to be written as in-memory methods that avoid explicit cache accesses, and it provides automatic parallel speedup. The need for traditional HPC message-passing within the user's application is eliminated.

### Powerful Management of Cached Data

Included with GCE, the ScaleOut Management Pack adds important functions that extend your ability to manage, analyze, and protect data stored in the distributed data grid. The Management Pack contains two components: an object browser for visually browsing and managing objects stored in the distributed data grid and a parallel backup and restore feature for quickly archiving its contents in the file system. The parallel backup and restore feature adds the capability to save the contents of a distributed data grid as a snapshot which later can be restored for analysis using parallel analysis algorithms.

