DataSpaces: An Extreme Scale Data Management Framework

Technology #2017-059

Summary:

High performance computing infrastructures enable large-scale scientific (and other) applications and workflows to run with increased complexity and improved accuracy. As the size of these workflows grow, they are generating massive amounts of data that must be processed and analyzed. Traditional methods of running these types of workflows do not work at this scale.

Researchers at Rutgers Discover Informatics Institute have developed a software solution, DataSpaces which is a programming system and data management framework targeting coupled application workflows running on very large-scale systems. The framework supports dynamic interaction and coordination between component applications and services that are part of an application’s workflow.

DataSpaces enables live data to be extracted from running simulation components, it indexes the data online, and allows it to be monitored, queried, and accessed by other component and services using semantically meaningful operators. The underlying data transport is asynchronous, low-overhead, and largely memory-to-memory.

DataSpaces also provides a distributed in-memory associative object store, scalable messaging, as well as runtime mapping and scheduling of online data analysis operations.

Benefits

- Increases overall data access performance utilizing memory layers across distributed computing nodes to support in-memory persistent data storage.
- Adaptively places data across server nodes and different storage levels (e.g. DRAM, NVRAM, SSD).
- Dynamic code placement and mapping, which can intelligently execute code in-situ and/or in-transit based requirements and constraints.

Improves overall performance leveraging Remote Direct Memory Access (RDMA) technology.

Market Applications:

Big Data and Simulation processes for Manufacturing, Financial Services and Large Pharma. Scientific computing (such as simulations) in the areas of physics, chemistry, material sciences, etc. Data scientists/engineers working on data analysis, visualization, and monitoring.
**Intellectual Property & Development Status:**

DataSpaces is available under open source license.

**Inventors**

**Manish Parashar, Ph.D.**

Director, Rutgers Discovery Informatics Institute Distinguished Professor, Department of Computer Science