



Data Grid Case Study

OGF19

29 January 2007

Cameron Purdy
CEO
Tangosol



Company

Tangosol, Inc.

- Leading Provider of Data Grid Solutions
- Key Player in Virtualization Space

Coherence™ Data Grid

- **Reliable**, Distributed, In-Memory Data Management
- Data Services for
 - SOA, Virtualization, Compute Grids

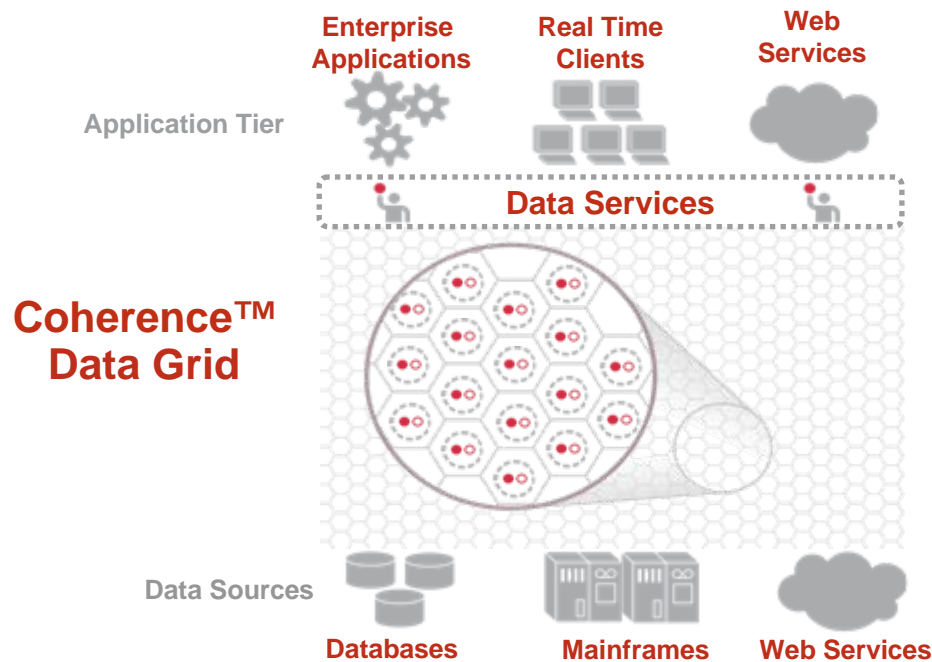
Quick Facts

- Founded in 2000
- 100% Employee Owned
- 20 Quarters of Profitability
- Over 1000 production applications using Coherence



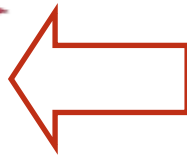
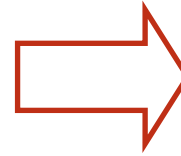
Data Grid: Introduction

- Provides a reliable data tier with a single, consistent view of data
- Enables dynamic data capacity including fault tolerance and load balancing
- Ensures that data capacity scales with processing capacity



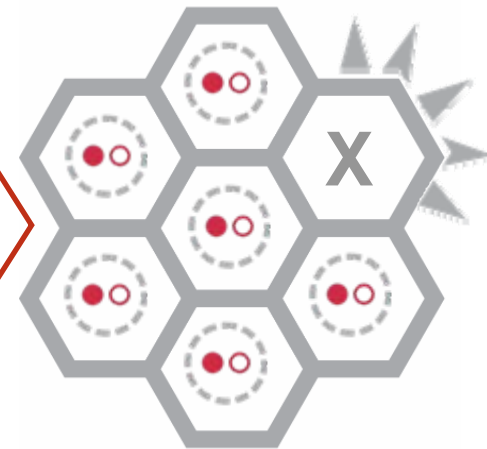
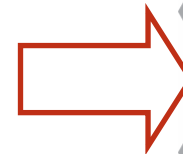
How does Coherence™ Data Grid work?

- Cluster of nodes holding % of primary data locally
- Back-up of primary data is distributed across all other nodes
- Logical view of all data from any node



- All nodes verify health of each other
- In the event a node is unhealthy, other nodes diagnose state

- Unhealthy node isolated from cluster
- Remaining nodes redistribute primary and back-up responsibilities to healthy nodes



Data Grid: Indicators

- **Data (Not Just Algorithms)**
 - Applications tend to always assume data is always available, immediately
- **Real-Time (Not Yesterday's Data)**
 - Applications make decisions on live data
 - Lots of examples in financial services, telcos
- **Transactional (Micro/Nano/Pico Jobs)**
 - Application decisions reflect a one-time reality
 - Durability required
- **Virtualization (Share your toys)**
 - Grid infrastructure viewed as a shared IS resource
 - Information viewed as a shared grid resource

Data Grid: Uses



Caching

Applications request data from the Data Grid rather than backend data sources



Analytics

Applications ask the Data Grid questions from simple queries to advanced scenario modeling



Transactions

Data Grid acts as a transactional System of Record, hosting data and business logic



Events

Automated processing based on events

Data Grid: Case Study

• Customer Background

- Large regional bank
- Experience with compute grid infrastructure (e.g. risk)
- Low average server utilization made virtualization into a priority

• Goals

- Shared, virtualized infrastructure to increase server utilization and reduce long-term infrastructure costs
- Eliminate data bottlenecks & cost of data infrastructure
- Make big compute tasks into near “real time” operations
- Package the result as an easy to use/deploy service

Data Grid: Case Study

• General Benefits

- Elimination of all but incremental ETL
- Information always available, real-time
- Data capacity is scalable (e.g. in-memory capacity)
- Automated “Locality of Reference” guaranteed processing scalability (goal: eliminate “IO wait”)

• Business Benefits

- “Data as a Service” – the information is available to additional tasks, jobs and applications, such as providing internal market data feeds
- Previously “impossible” processes become realistic, such as a 50-day risk calculation ... *in under one hour!*

Data Grid: Challenges

- **Metering, Charge Back**

- Shared infrastructure is by its nature ... *shared*
- Lots of parallel, thread-level concurrent execution
- Based on Clock cycles? RAM? Network bandwidth?

- **Holistic Function Mindset**

- Dominance of imperative, iteration- and recursion-based programming models
- Lack of declarative models

- **Granularity, Rate and Reliability**

- Application scenarios require different QoS
- Qualities can “cost” an order of magnitude *each*



Data Grid Case Study

OGF19

29 January 2007

Cameron Purdy
CEO
Tangosol

