Verteilung - Data Grid

Hazelcast:

- Applikationen skalieren
- Daten über Cluster verteilen
- Daten partitionieren
- · Nachrichten senden und empfangen
- Lasten verteilen
- Parallele Tasks verarbeiten
- ...

CAP-Theorem: Es ist in einem verteilten System unmöglich, gleichzeitig die drei Eigenschaften zu garantieren:

- Consistency
- Availability
- Partition Tolerance

Hazelcast stellt die Verfügung über die Konsistenz

Wo sind die Daten?

Topologie:

Nodes sind VMs.

- Embedded: Hochleistungs-Computing/asynchrone Ausführung
- Client/Server: Für Cluster von Server-Knoten, die erstellt und skaliert werden können.

In-Memory Data Grid (IMDG):

- Daten im RAM von Cluser Nodes (schneller, bessere Reaktionszeit)
- Punkt-zu-Punkt (Socket)
- · Redundanz: Datenkopien in versch. Nodes
- Skalierbarkeit: Nodes hinzufüg-/entfernbar
- Persistenz: Können in relationalen- oder NoSQL-Datenbanken gespeichert werden

Cluster: Verbund von Rechnern

Datenpartitionierung in Cluster:

- Fixe Anzahl Partitionen (Default 271)
- Für jede Partition gibt's einen Schlüssel

partitionId = hash(keyData) % PARTITION COUNT

Alle Partitionen möglichst gleichverteilt (Backups/Redundanz)

Bei Hazelcast kommt eine leere Partition hinzu. Durch Migration übernimmt diese Partition von Knoten A, dann Backup Partition C und B und dann Backup Partition von C.

Migration komplett (12 Partitionen)

Knoten B stürtz ab. Wiederherstellung mit Hilfe der Backups // TODO!!

Zusätzliches

- · Hazelcast ist threadsafe
- · Viele Instanzen auf gleicher JVM
- Alle Objekte müssen serialisierbar sein

Cluster Interface: zeigt Informationen über Mitglieder des Clusters

Distributed Map: Collection Map mit der Möglichkeit der verteilten Speicherung von Daten

Distributed Queue: Auch Queues können verteilt gespeichert werden und z.B. für verteiltes Ausführen von Tasks genutzt werden

Distributed Lock: Verteilungsmechanismus für das Veröffentlichen von Nachrichten, an mehrere, registrierte Abonnenten

// TODO folie 43

Executor Service