



Data Replication and Synchronization Tool

**Ashish Sharma
Pradeeban Kathiravelu**

Introduction

- Data is huge
- Consumers often share a sub set of data with others.
 - Pointers to data, actually.
- Medical data is structured in hierarchies.

Motivation

- Creating and sharing pointers to interesting sub sets of data.
- Data Sharing Synchronization System
 - Fault-tolerant.
 - In-Memory.
- Generic, while targeting the medical images and meta data.
 - The Cancer Imaging Archive (TCIA)

Solution Architecture

- Users create, share, and update replica sets from a data source.
- Infinispan In-Memory Data Grid (version 6.0.2) to store the replica sets.

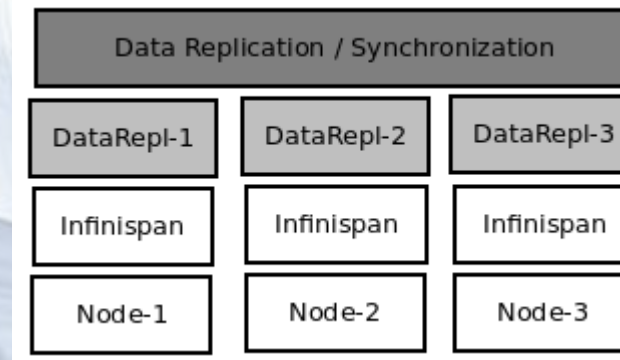


Fig 1. Deployment Architecture

Execution Flow

- Publisher-Consumer API to consume the replica sets and Data Provider API to communicate with the data source.

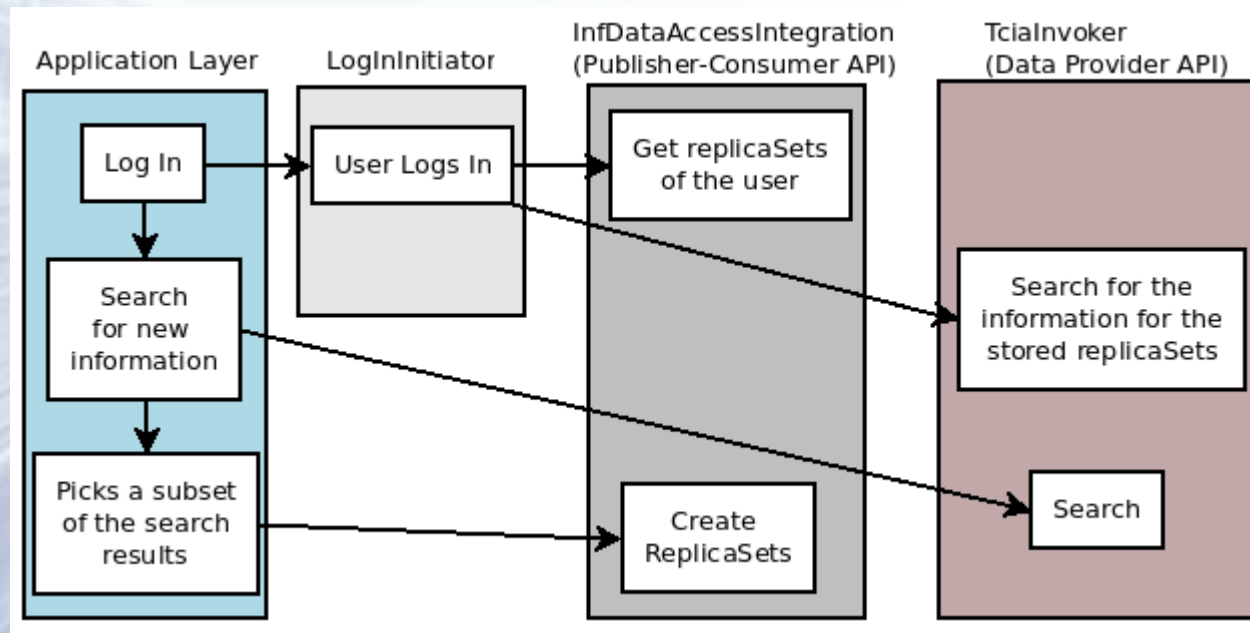


Fig 2. Execution Flow

Design

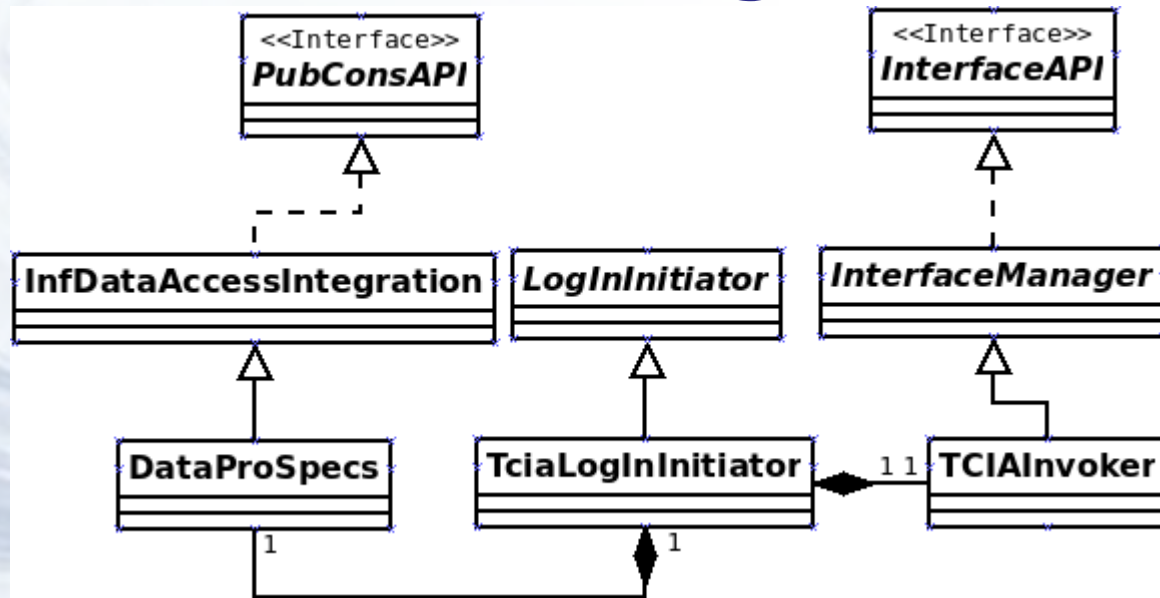


Fig 3. Back-end Class Hierarchy

- DataProSpecs API

- createReplicaSet
- getReplicaSet
- updateReplicaSet

- duplicateReplicaSet
- deleteReplicaSet
- getRawData

Extensibility

- Not tightly coupled to the technology.
 - Other data-grids
 - Hazelcast, Terracotta Big Memory, Oracle Coherence
 - Persistence
 - Integration to SQL or NoSQL solutions such as Mongo DB.
 - Data sources other than TCIA.

What Infinispan offers?

- High Performance and Scalability.
- Fault-tolerance
 - Multiple nodes with TCP-IP or Multicast based JGroups clustering configurations.
- Distributed Execution.
 - Optimized for single node as a local cache as well as a multiple-node execution.
- MapReduce Framework.

What Infinispan offers?

- High Performance and Scalable.
- Fault-tolerant
 - Multiple nodes with TCP-IP or Multicast based JGroups clustering configurations.
- Distributed Execution.
 - Optimized for single node as a local cache as well as a multiple-node execution.
- MapReduce Framework.

Thank you!

What Infinispan offers?

- High Performance and Scalable.
 - Fault-tolerant
 - Multiple nodes with TCP-IP or Multicast based JGroups clustering configurations.
- Thank you!**

- Distributed Execution.
 - Optimized for single node cache as well as a multi node execution.
- MapReduce Framework.

