DARPA XDATA

GORA

Introducing Object-to-Datastore Mapping

NASA Jet Propulsion Laboratory
xdata-dev@list.jpl.nasa.gov

Overview

Apache ***Gora*** speeds up data modeling, persistence and access using in-memory data structures and a friendly Java API. ***Gora*** allows your applications (such as MapReduce jobs) to access and interoperate with a host of underlying datastores efficiently.

Details

Object Relation Mapping (ORM) is a technology for converting software objects into database rows and vice versa. ***Gora*** extends this concept to introduce ***Object-to-Datastore Mapping***, which converts software objects into datastore objects (and vice versa) in non-relational databases. Non-relational databases are more flexible but less structured than conventional databases. By storing these software objects in memory, ***Gora*** eases high-speed distributed data analysis.

- ***Gora*** focuses specifically on non-relational datastores, but also works on relational databases.
- The main use case for ***Gora*** is to access and analyze big data using Apache Hadoop™.
- ***Gora's*** object-to-datastore mappings are specific to the type of datastore being used, so that users can leverage the datastore's full functionality.
- ***Gora*** greatly eases and simplifies the process of acquiring data from the datastore.

Benefits

To the developer . . .

- Provides Object Mapping for data that is large enough to require storage in NoSQL databases.
- Comes with MapReduce out of the box.
- Provides technology-agnostic storage methods for addressing data storage tasks. This allows users to access the data using different programming languages with no compatibility issues.

. . . and to the organization:

- Provides a flexible storage framework under the most liberal open source license available.
- Allows your organization to quickly set up and deploy applications on top of big data storage mediums.
- Helps your organization decide which data storage methods to use by testing how well your applications are suited to different data storage technologies.

Prerequisites

Platforms:

- Mac OSX
- Linux Mint
- Ubuntu

Languages/Technologies

- ***Gora*** is written in Java.
- Configuration requires a working knowledge of syntax for JSON and XML.
- You should be able to use the command line terminal.
- You should be able to use Apache Maven from the command line.

Future Enhancements

- Continue to extend support to other datastores, such as Oracle NoSQL datastores, Hive, Pig, and Hazelcast.
- Move towards providing ***Object-to-Datastore Mapping*** for streaming data.
- Develop the capability to parse streaming data into formats that can then be stored.
- Extend and improve data representation formats by integrating with other data formatting tools, such as Apache Parquet.

The XDATA Open Catalog is located at http://www.darpa.mil/OpenCatalog/XDATA.html
Get ***Gora*** from https://git-wip-us.apache.org/repos/asf/gora.git

Last updated: 8/21/2014