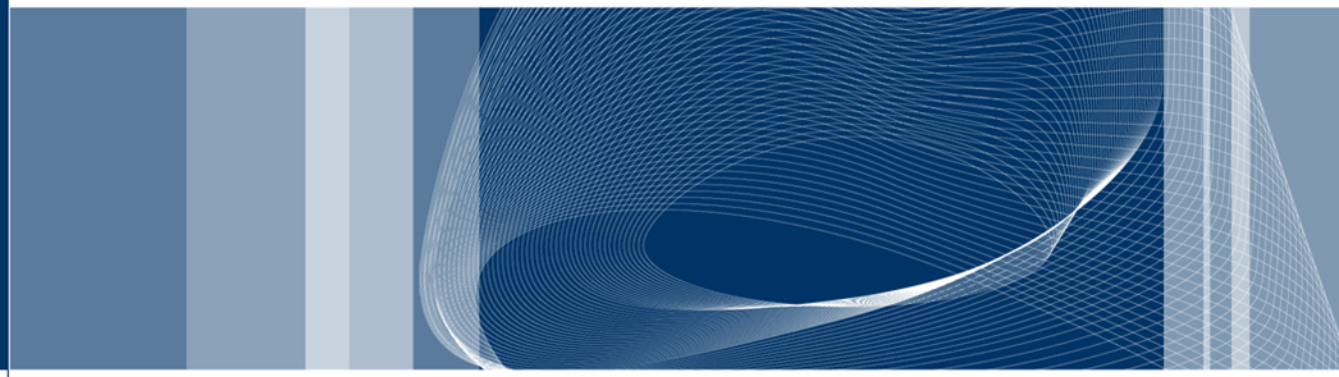
POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica

Anno Accademico 2013 - 2014



Avoiding CRUD operations lock-in in NoSQL databases: extension of the CPIM library

Candidato: Fabio Arcidiacono (799001)

Relatore: Prof.ssa Elisabetta Di Nitto

Correlatore: Ing. Marco Scavuzzo

Data management techniques

NoSQL

- Non-structured data
- Huge data volume
- Horizontal scaling
- No standard language
- BASE properties

RDBMS

- Well structured data
- Vertical scaling
- SQL
- ACID transactions

Common language approaches

Meta-model

- Apache MetaModel
- SOS platform

SQLification

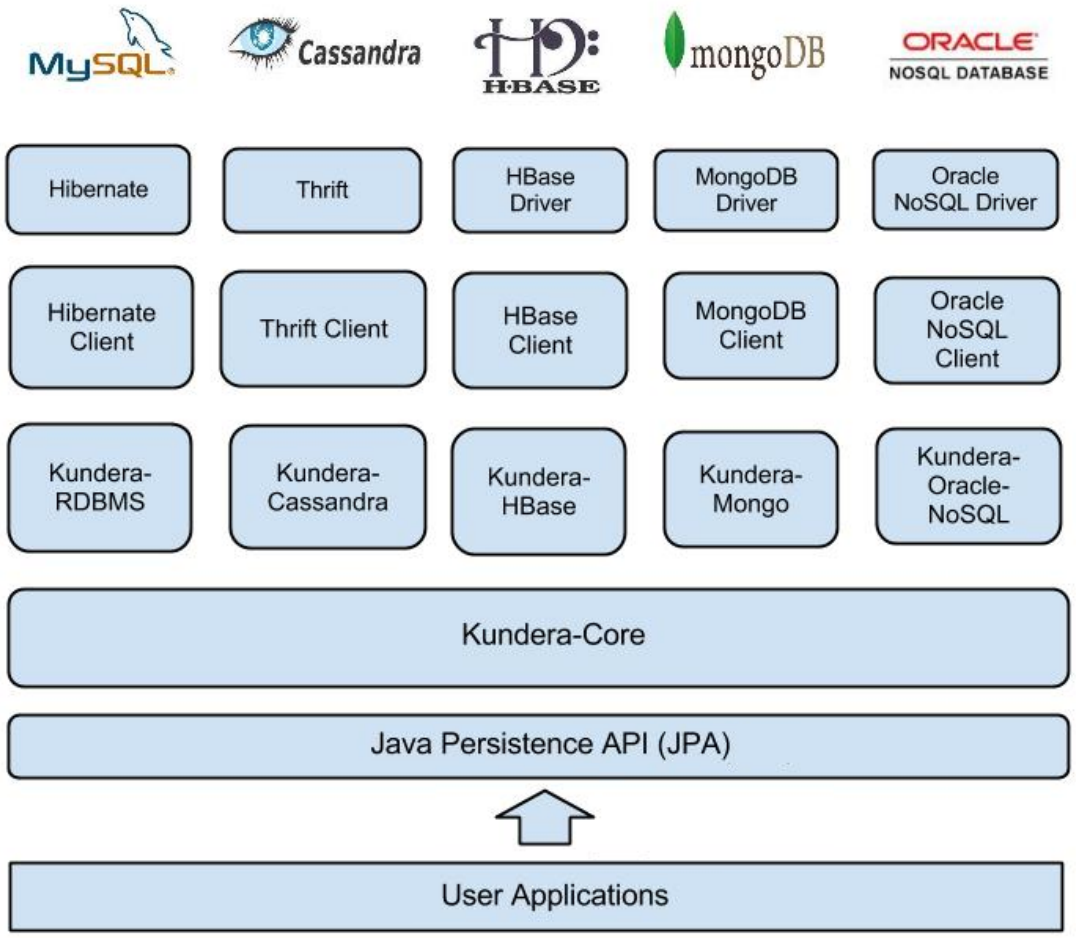
- Apache Phoenix
- UnQL

ORM

- Kundera
- PlayORM
- Spring-data
- Apache Gora

Kundera

A JPA 2.1 compliant ORM Library for NoSQL databases

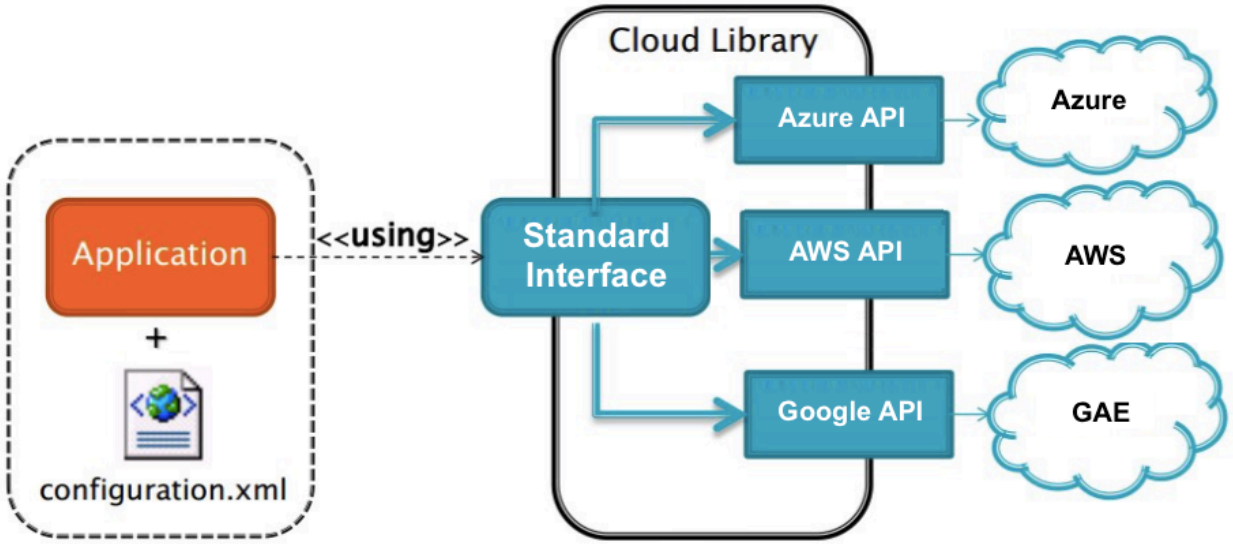


Cloud Platform Independent Model

Abstract application logic from the specific PaaS Provider to overcome the vendor lock-in

Many supported services:

- Blob
- NoSQL
- Memcache
- Queue
- Mail
- SQL

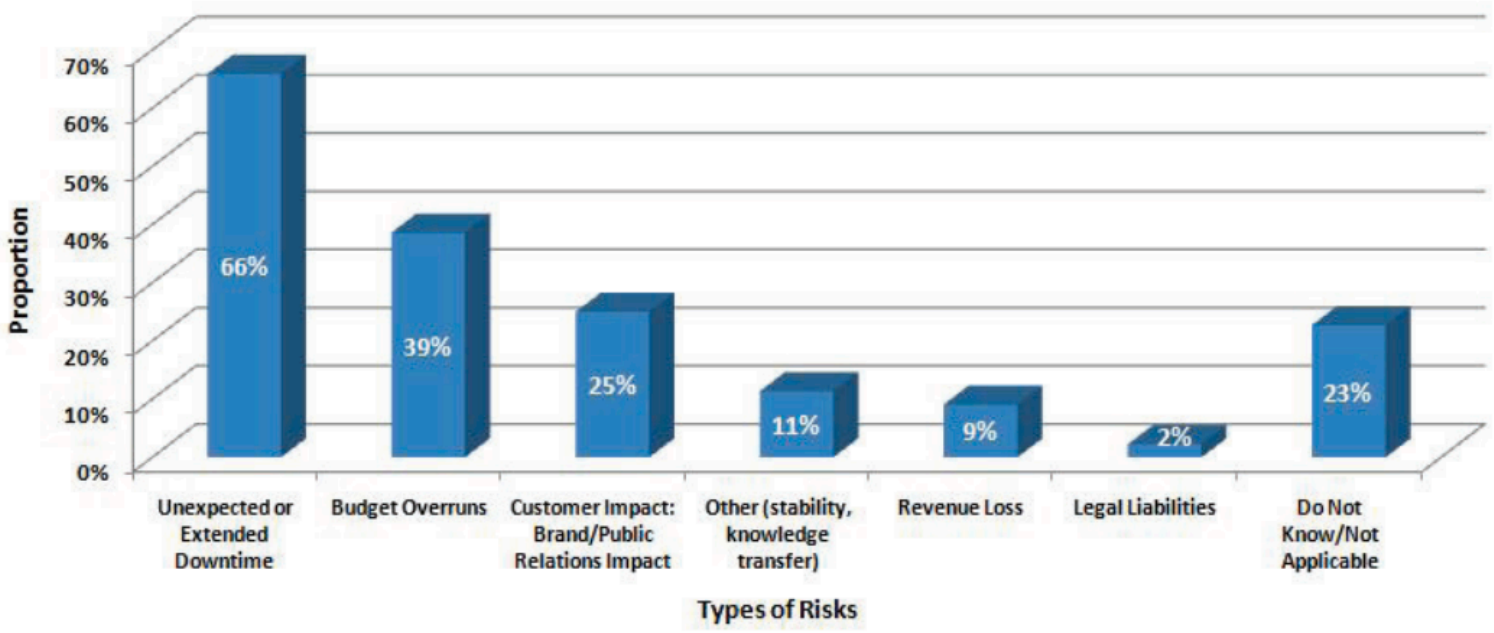


Work objectives

- Integrate Kundera in the CPIM library
  - extending the number of NoSQL databases supported
  - fix of the problems of the NoSQL service of CPIM
- Contribute to the open source project Kudera
  - developing a client for GAE Datastore
  - developing a client for Azure Tables
- Support data migration among NoSQL databases through the migration and synchronization system Hegira

Data migration

- Move application to another cloud provider
- load balancing, system expansion, failure recovery, etc.
- modern computer systems are expected to be up continuously
- data synchronization between the two involved systems

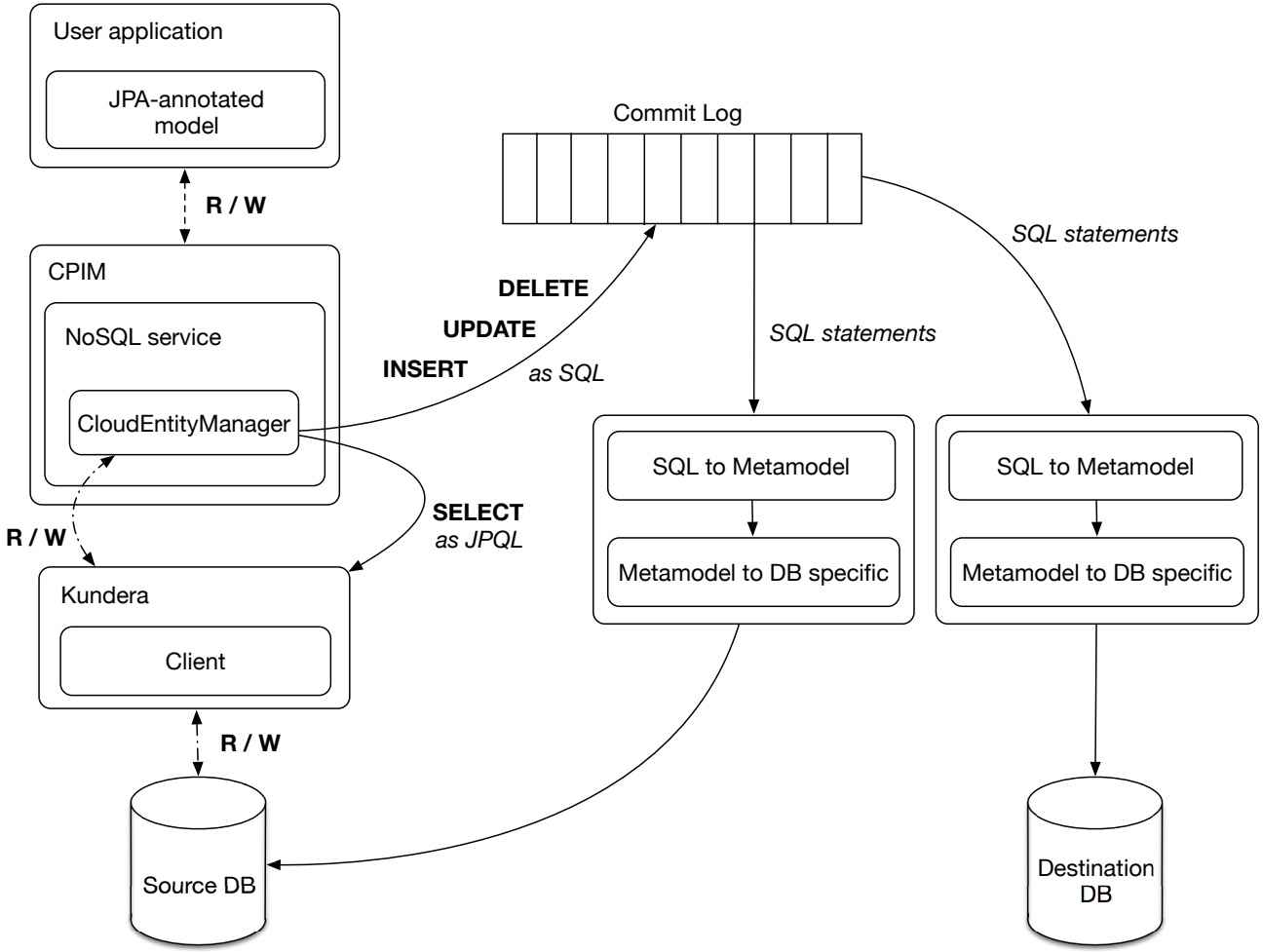


Why Kundera

- Open source
- Developed with extensibility in mind
- Ploygot persistency
- In the field since 2010 with an active community
- Already used in production
- Support to many different NoSQL databases

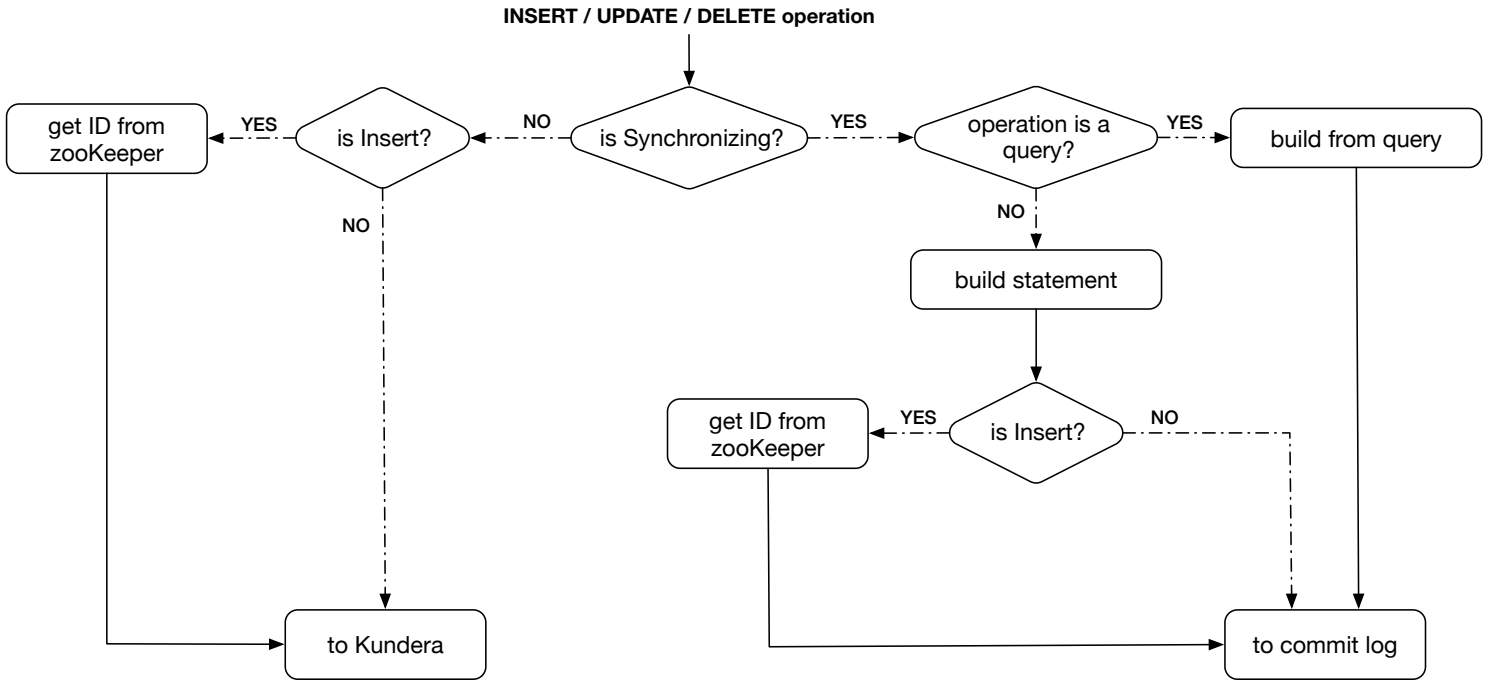
Supporto a Hegira

Migration service transparent to the user of the CPIM library



Hegira support

- Intercepting user operations (DMQ)
- Translate operations to SQL statements
- Send them to the Hegira commit-log



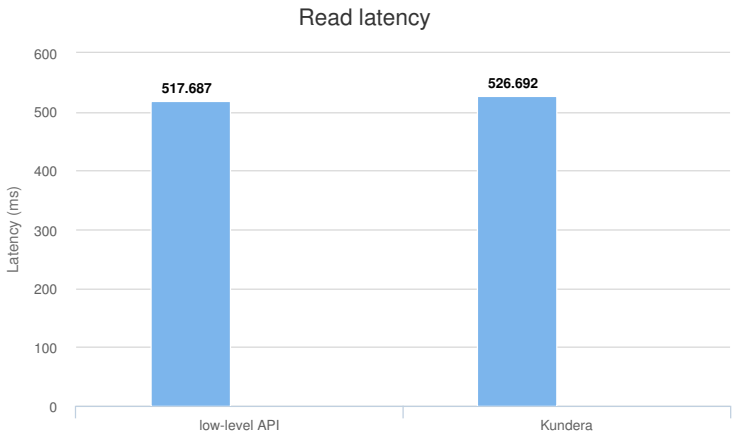
Performance

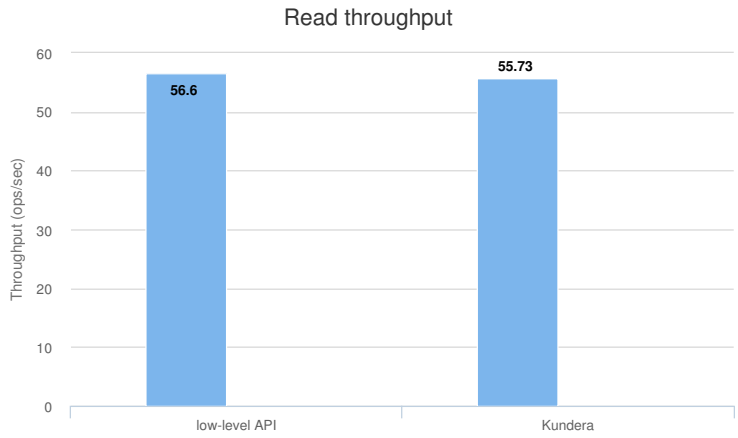
Performance test of the developed Kundera clients by using YCSB (Yahoo Cloud Serving Benchmark).

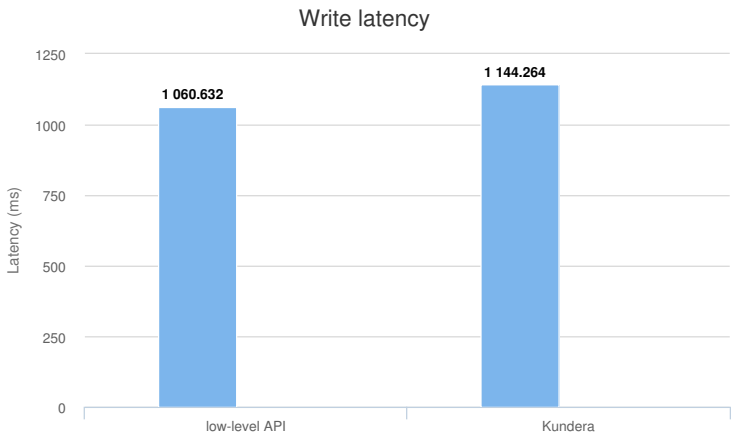
- Adapter YCSB for operations through Kundera
- Adapter YCSB for operations through low-level API

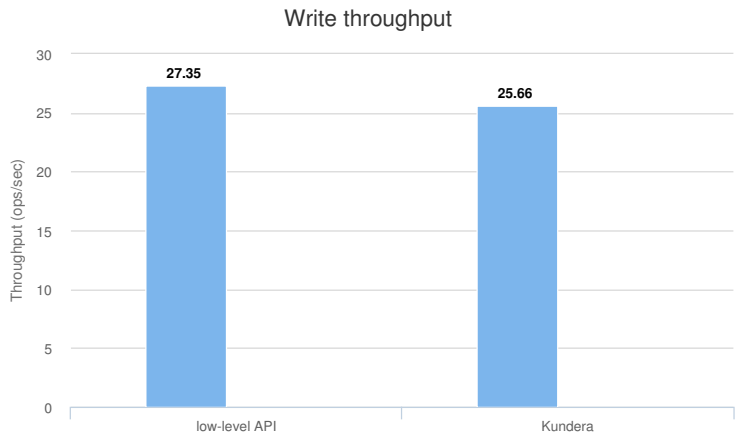
Workload of 100.000 operations, splitted in two phases (write/read) on remote instances of Google Datastore and Azure Tables.

Azure Tables results

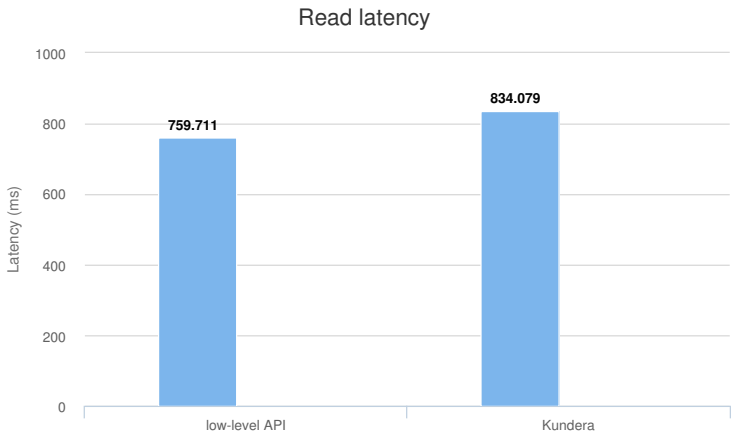


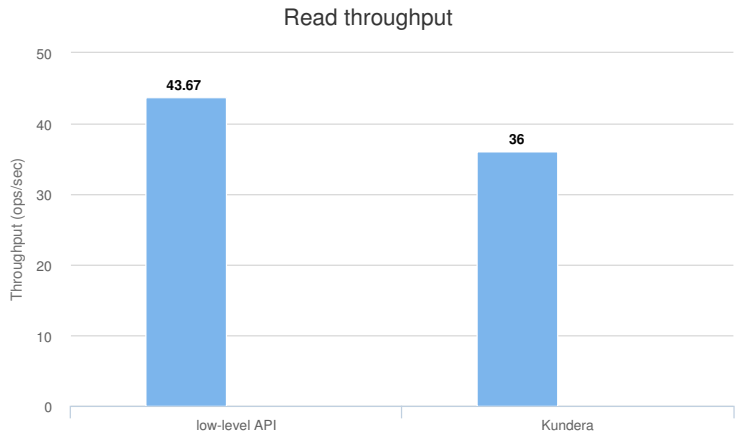


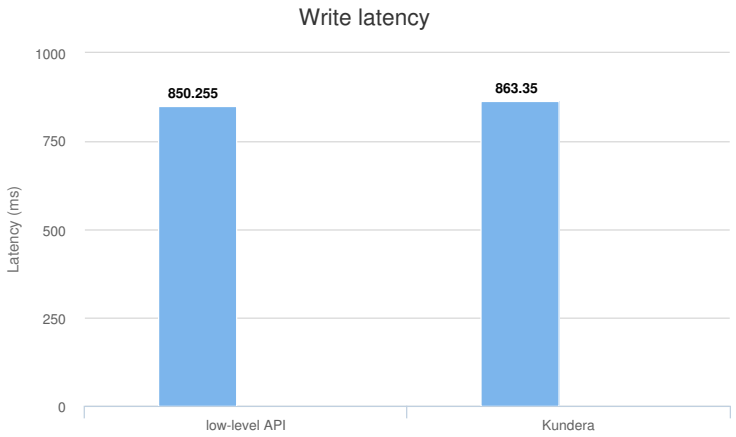


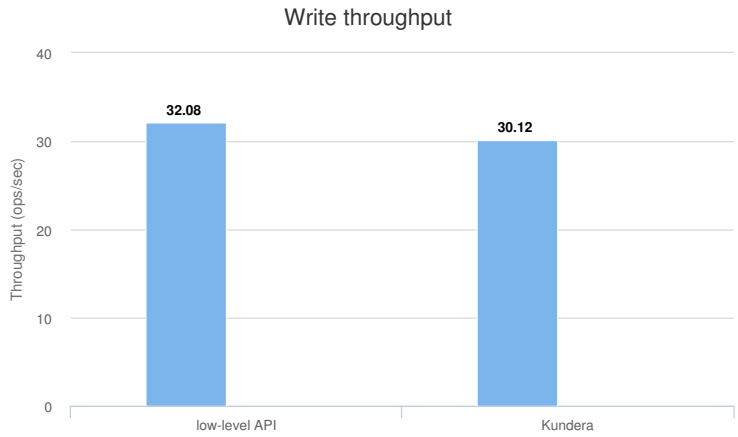


GAE Datastore results









Conclusions

Contributions:

- New clients forKundera to support Google Datastore and Azure Tables
- Hegira integration in the CPIM library

Sviluppi futuri:

- Extend the CPIM library to support more cloud providers and/or new cloud services
- Develop new extensions for Kundera to support more NoSQL technologies