

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica

Anno Accademico 2013 - 2014

 POLITECNICO DI MILANO



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
- Horizontal scaling
- BASE properties
- No standard language

RDBMS

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

Common language approaches

Meta-model

- Apache MetaModel
- SOS platform

SQLification

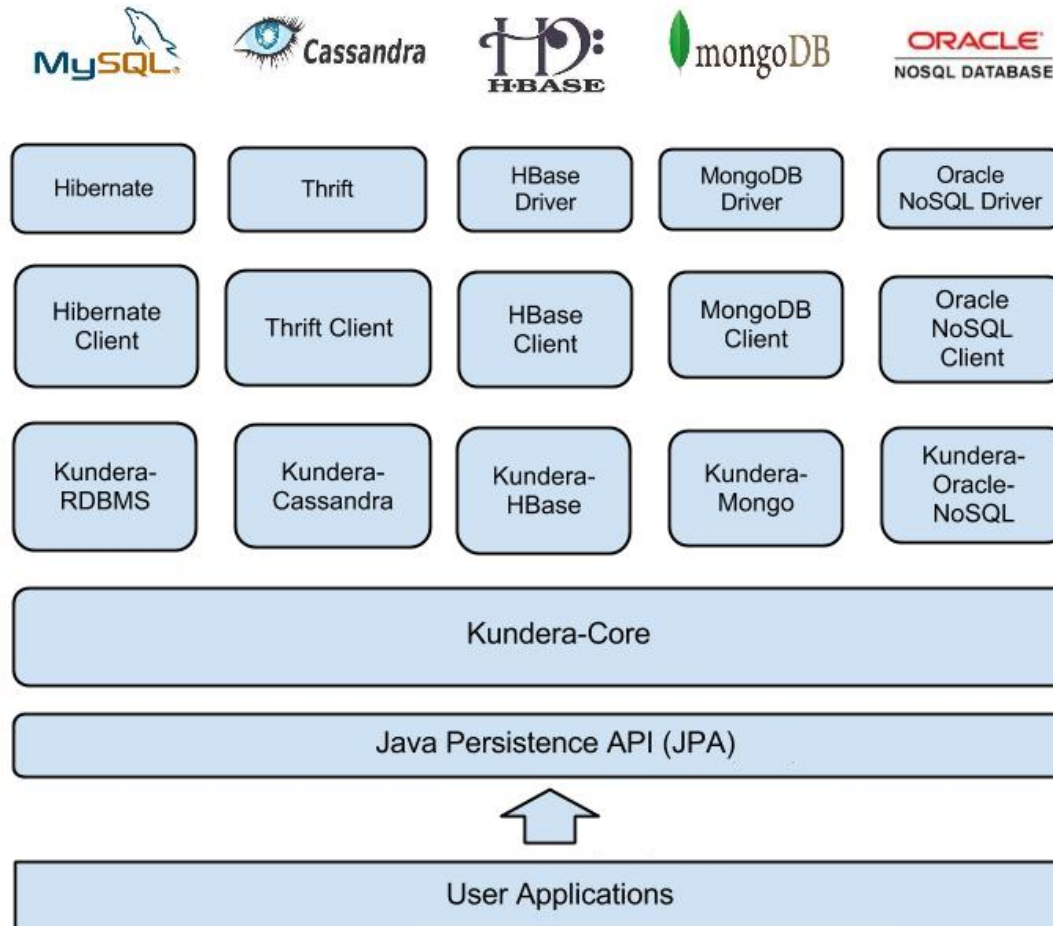
- Apache Phoenix
- UnQL

ORM

- Kundera
- PlayORM
- Spring-data
- Apache Gora

Kundera

A JPA 2.1 compliant ORM Library for NoSQL databases

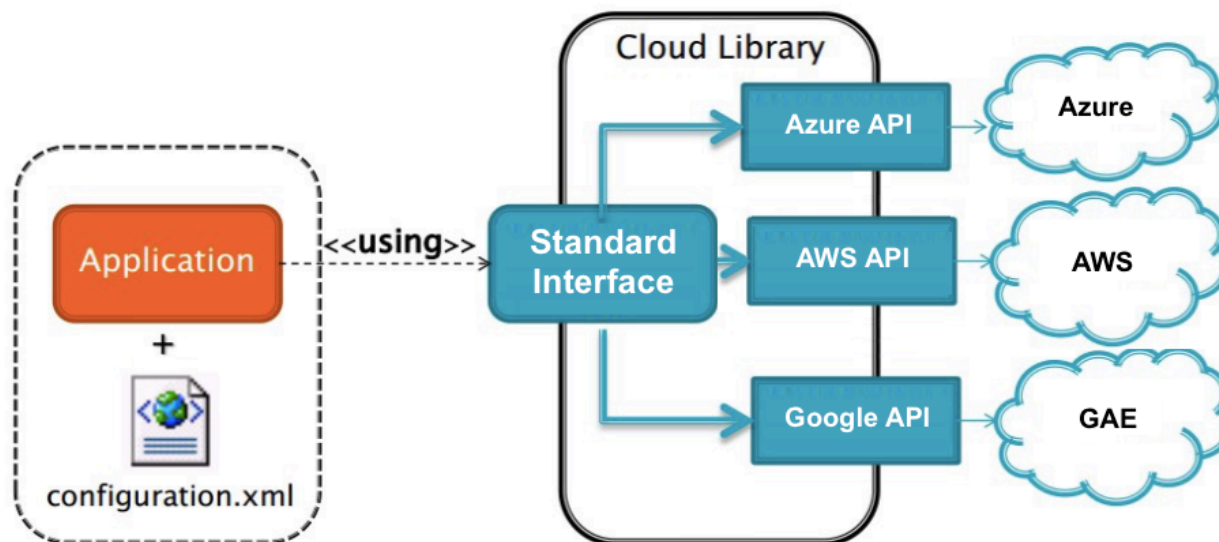


Cloud Platform Independent Model

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

Many supported services:

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

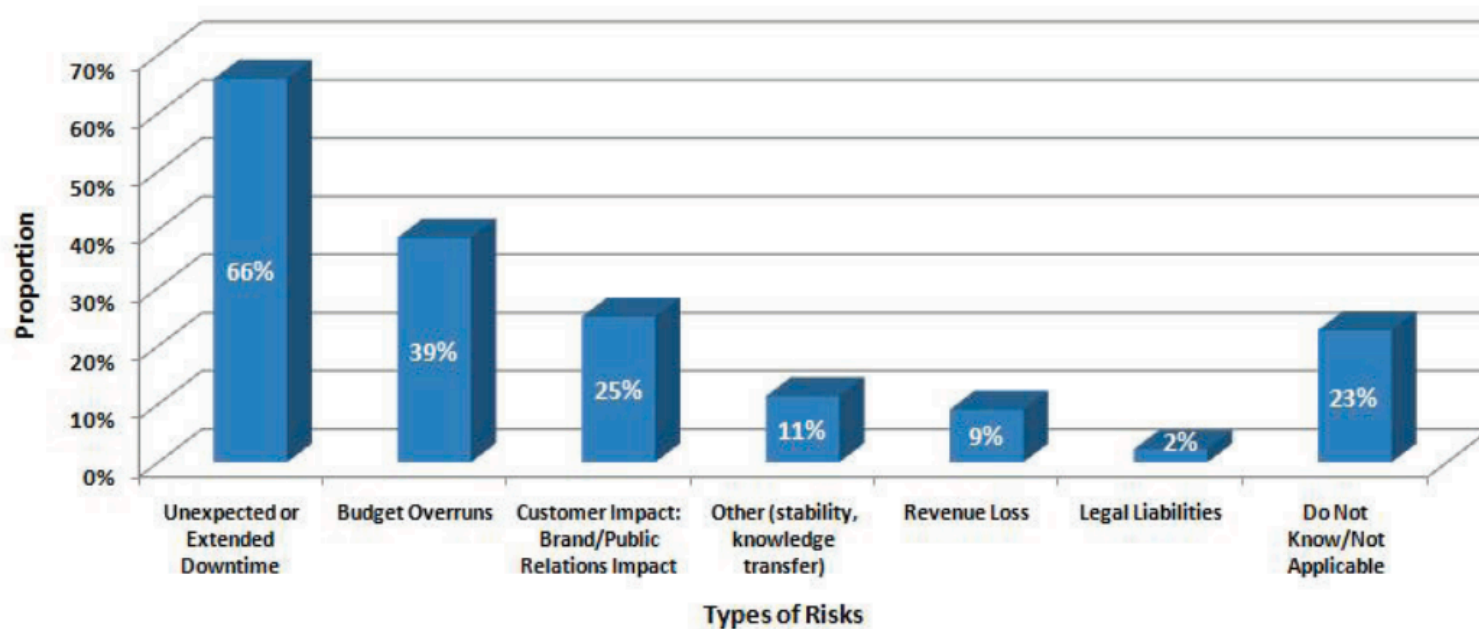


Work objectives

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

Data migration

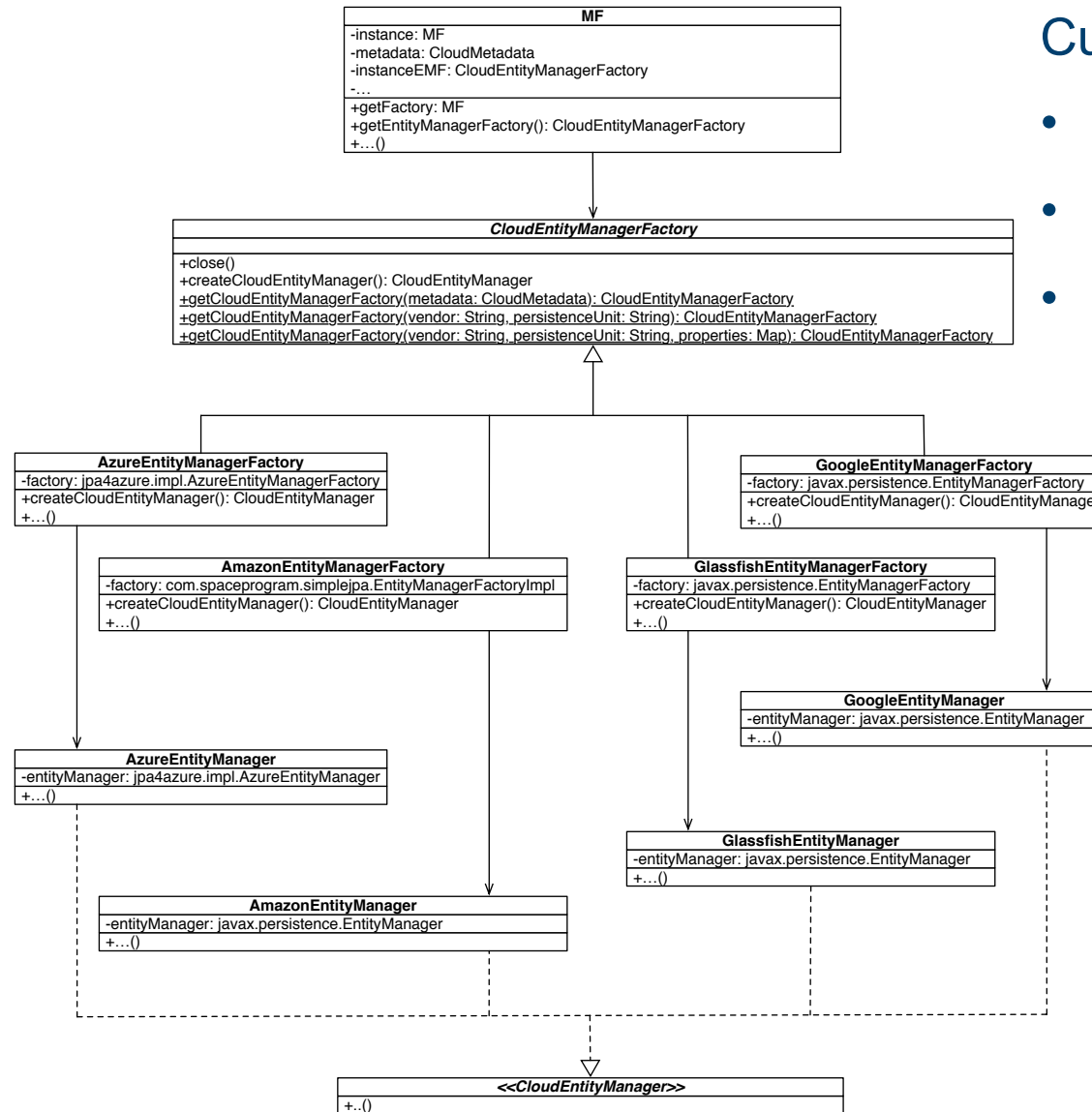
- Move application to another cloud provider
- load balancing, system expansion, failure recovery, etc.
- modern computer systems are expected to be up continuously
- data synchronization between the two involved systems



Why Kundera

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

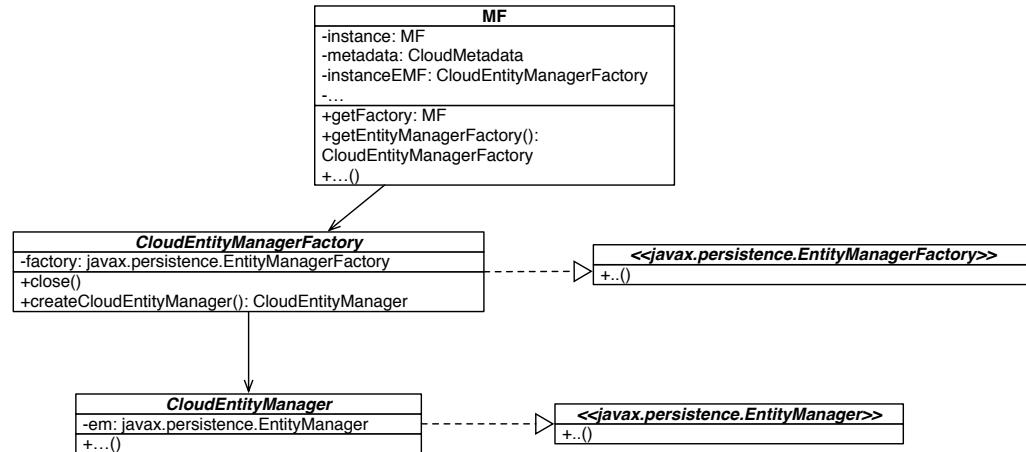
Kundera integration (1)



Current implementation

- Many JPA provider
- Duplicated code
- No complete code portability

Kundera integration (2)

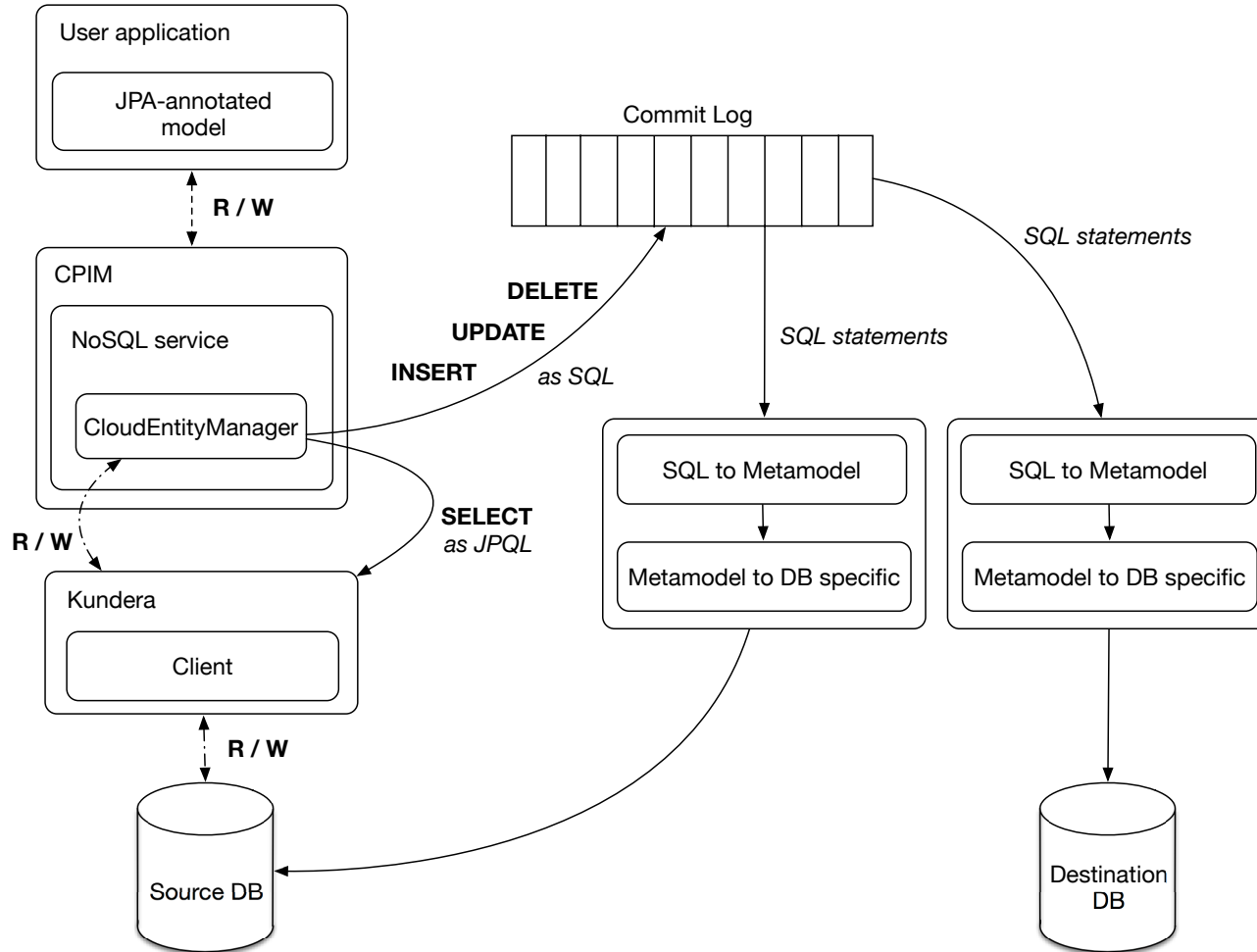


```

<persistence-unit name="pu">
  <provider>com.impetus.kundera.KunderaPersistence</provider>
  <class>it.polimi.kundera.client.datastore.entities.Department</class>
  <class>it.polimi.kundera.client.datastore.entities.Employee</class>
  <class>it.polimi.kundera.client.datastore.entities.Project</class>
  <exclude-unlisted-classes>true</exclude-unlisted-classes>
  <properties>
    <property name="kundera.keyspace" value="gae-test"/>
    <property name="kundera.client.lookup.class"
      value="it.polimi.kundera.client.datastore.DatastoreClientFactory"/>
  </properties>
</persistence-unit>
  
```

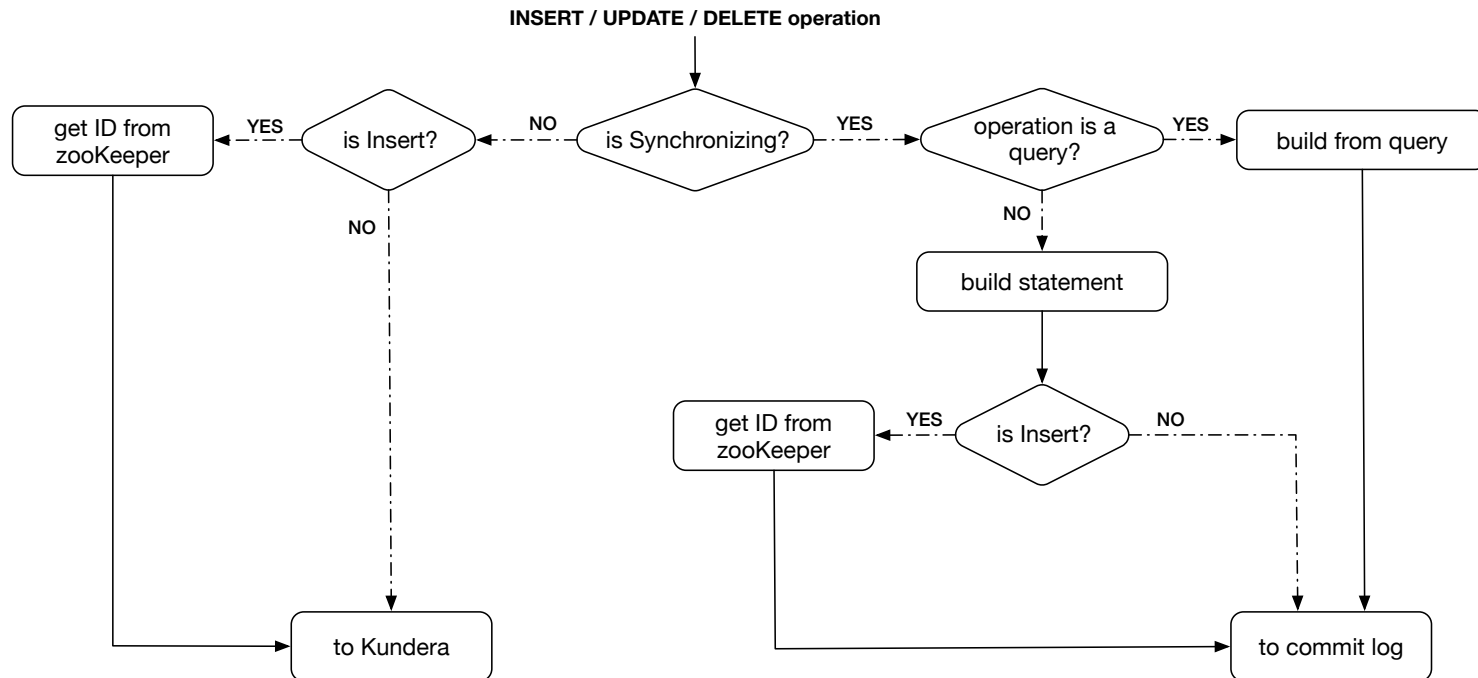
Hegira support (1)

Migration service transparent to the user of the CPIM library



Hegira support (2)

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



Performance

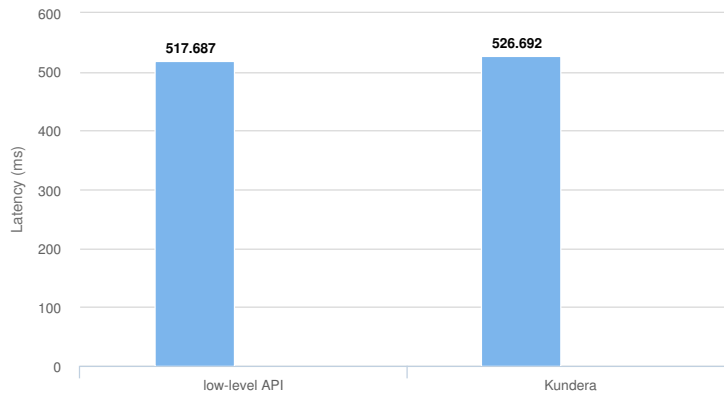
Performance test of the developed Kundera clients by using YCSB (Yahoo Cloud Serving Benchmark).

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

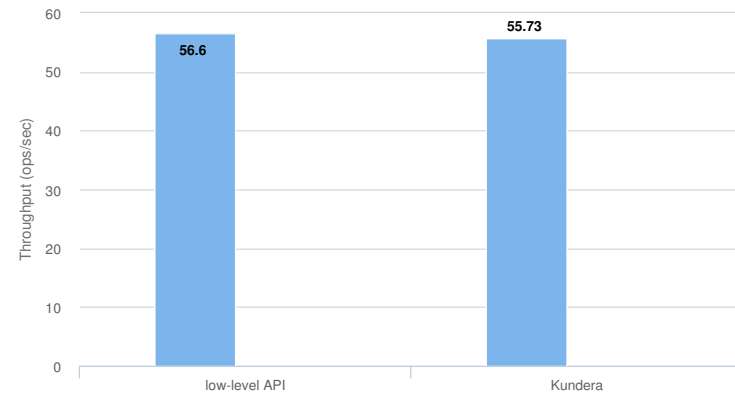
Workload of 100.000 operations, splitted in two phases (write/read) on remote instances of Google Datastore and Azure Tables.

Azure Tables results

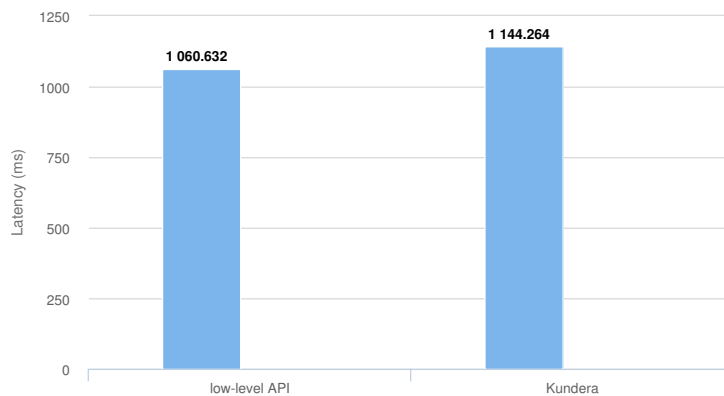
Read latency



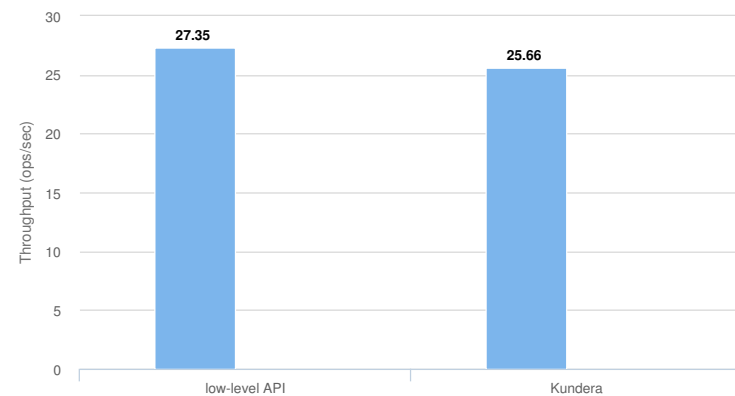
Read throughput



Write latency

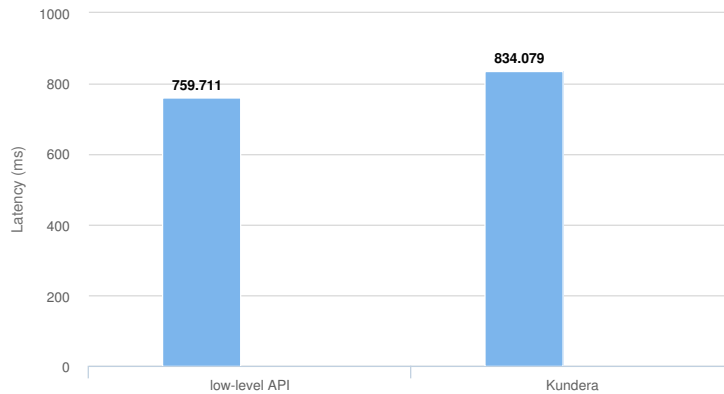


Write throughput

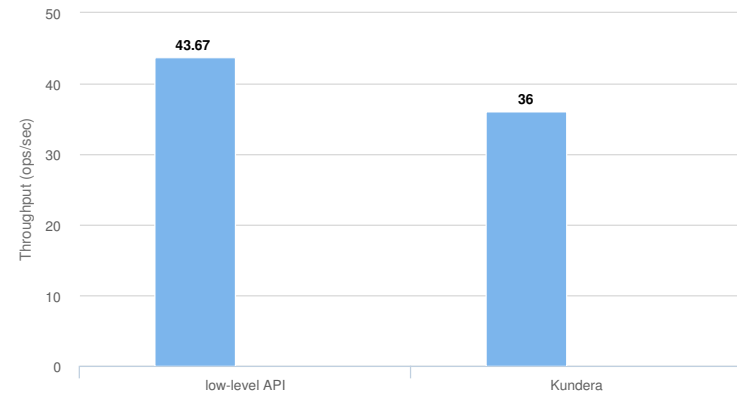


GAE Datastore results

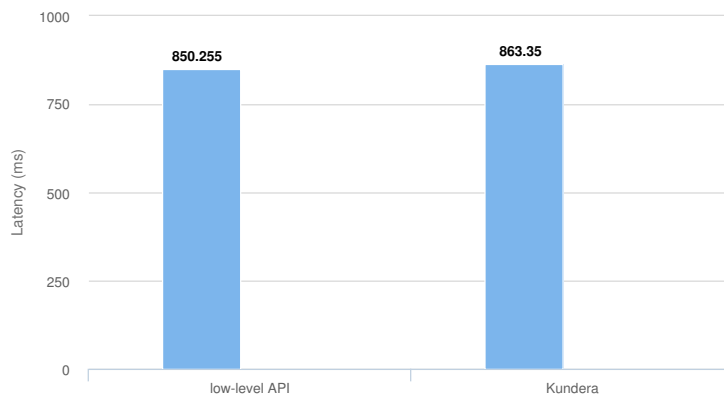
Read latency



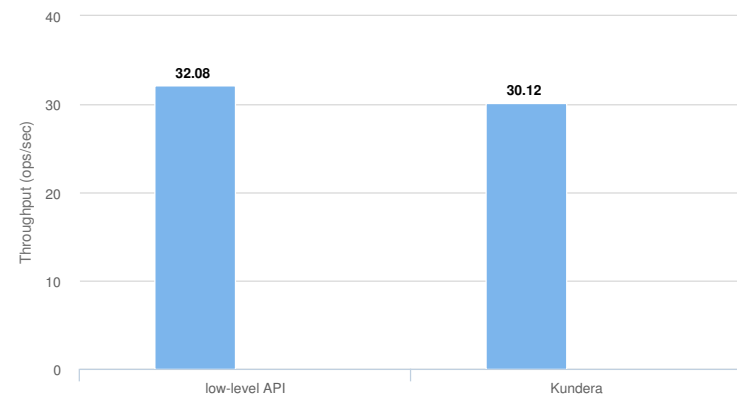
Read throughput



Write latency



Write throughput



Conclusions

Contributions:

- New clients for Kundera 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