

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

Tecnologie per la gestione dei dati

NoSQL

- Dati non strutturati
- Grande volume di dati
- Horizontal scaling
- Nessun linguaggio standard
- BASE properties

RDBMS

- Dati benstrutturati
- Vertical scaling
- SQL
- ACID transaction

Approcci per un linguaggio standard

Meta-model

- Apache MetaModel
- SOS platform

SQLification

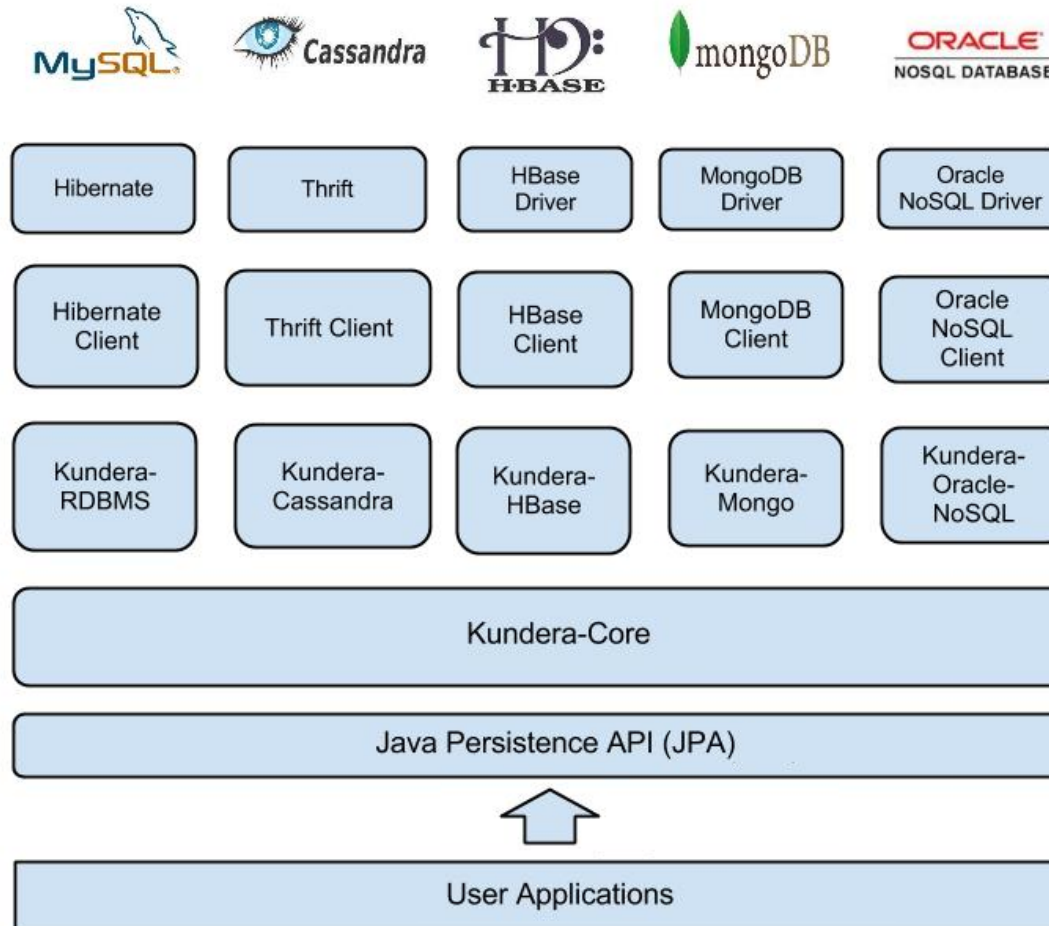
- Apache Phoenix
- UnQL

ORM

- Kundera
- PlayORM
- Spring-data
- Apache Gora

Kundera

A JPA 2.1 compliant ORM Library for NoSQL databases

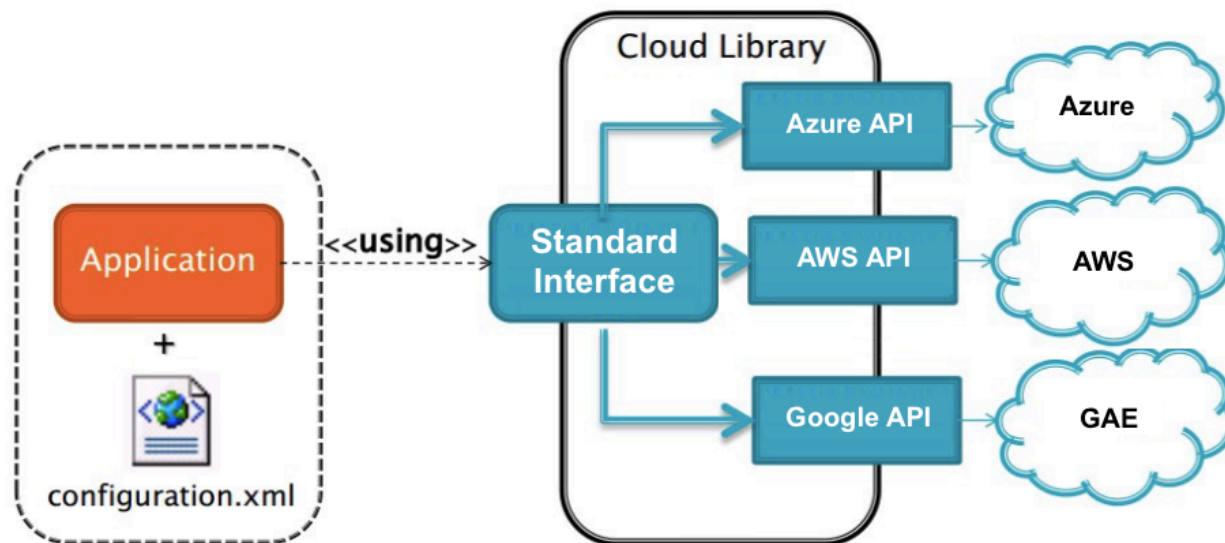


Cloud Platform Independent Model

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

Many supported services:

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

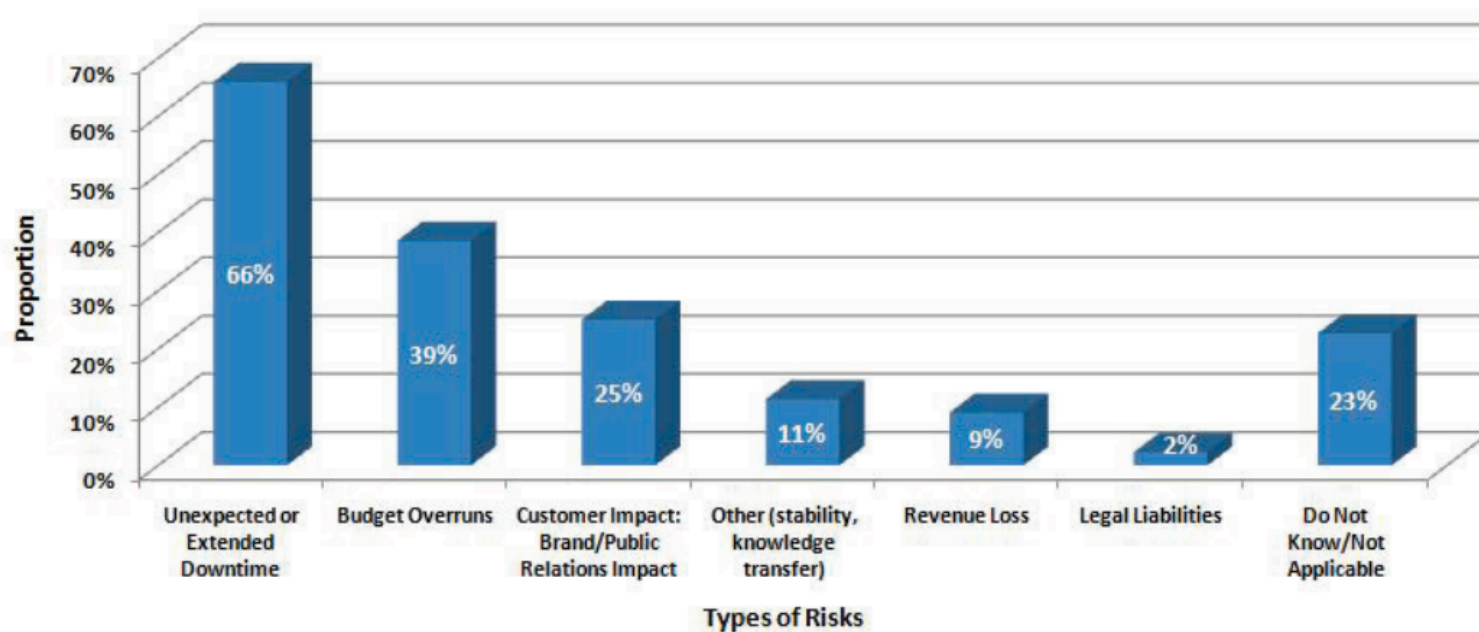


Obbiettivi del lavoro

- Integrazione di Kundera in CPIM
 - estensione del numero di NoSQL supportati
 - fix dei problemi del servizio NoSQL
- Contribuire a Kundera come progetto open-source:
 - sviluppo di un client per GAE Datastore
 - sviluppo di un client per Azure Tables
- Supportare migrazione dei dati tra database NoSQL attraverso 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

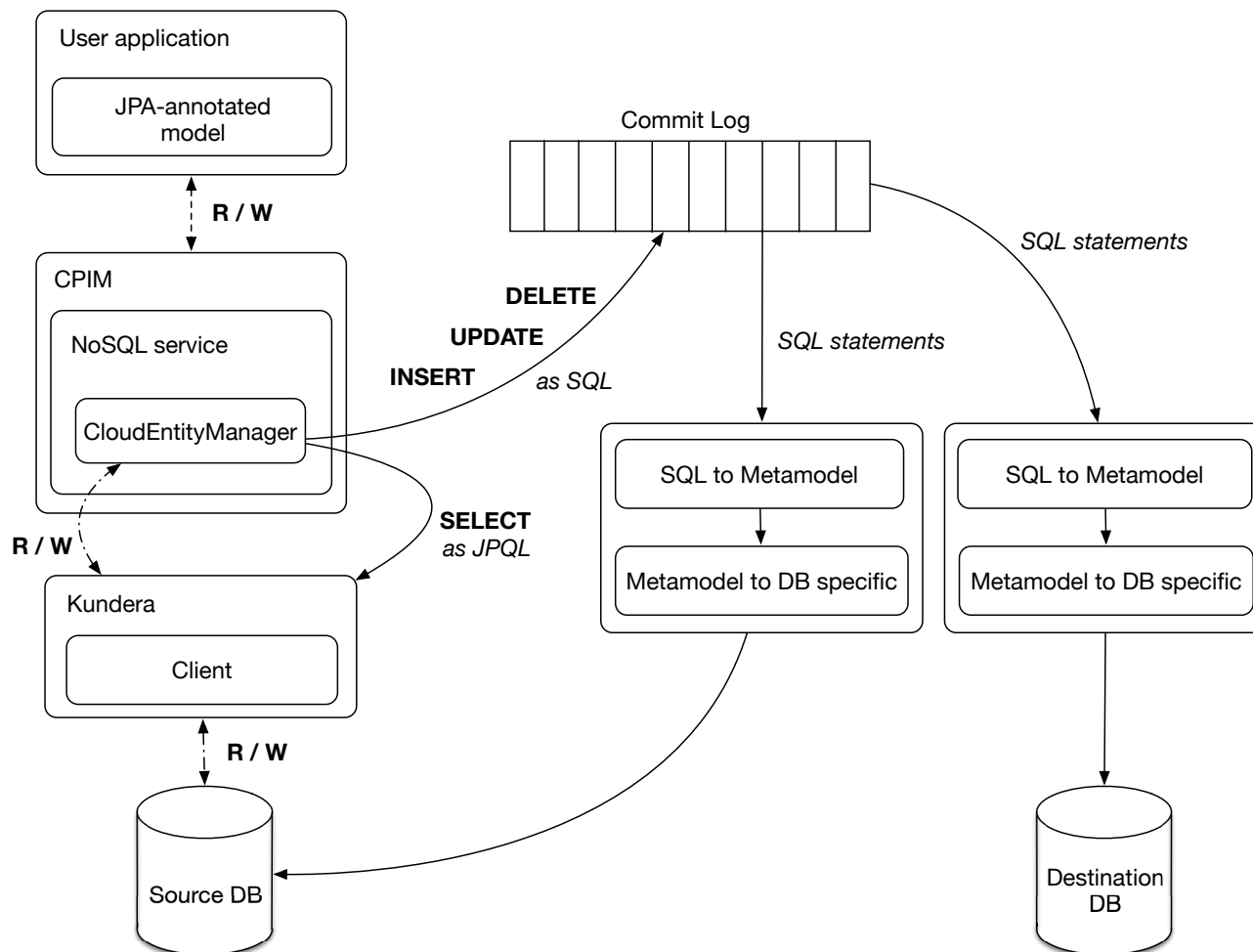


Perché Kundera

- Open source
- Modulare, sviluppata con l'obiettivo di essere estendibile
- Ployglot persistency
- Nel campo dal 2010 con una community attiva
- Utilizzato in production
- Supporto a molti diversi database NoSQL

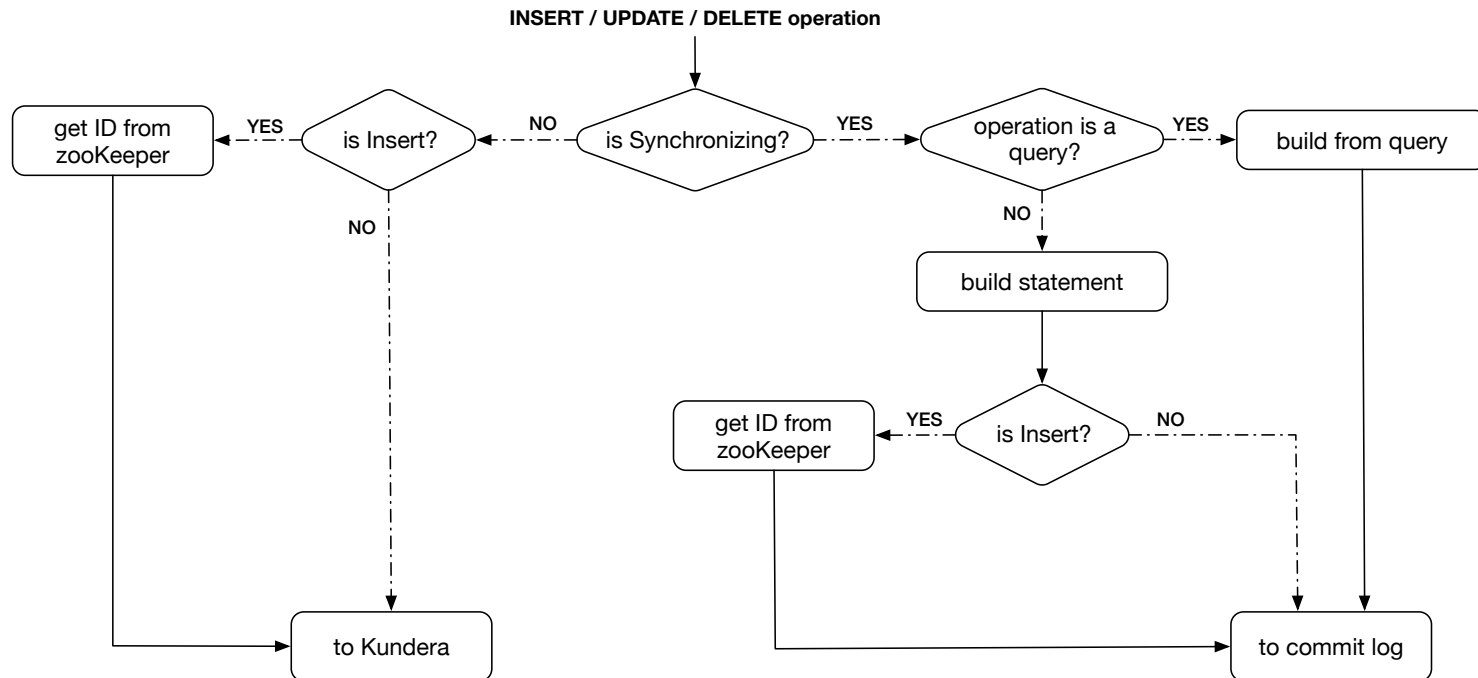
Supporto a Hegira

Servizio di migrazione in modo trasparente all'utente della libreria CPIM



Supporto a Hegira

1. Intercattare le operazioni dell'utente (DMQ)
2. Tradurle in statements SQL
3. Inviarle al commit-log di Hegira



Performance

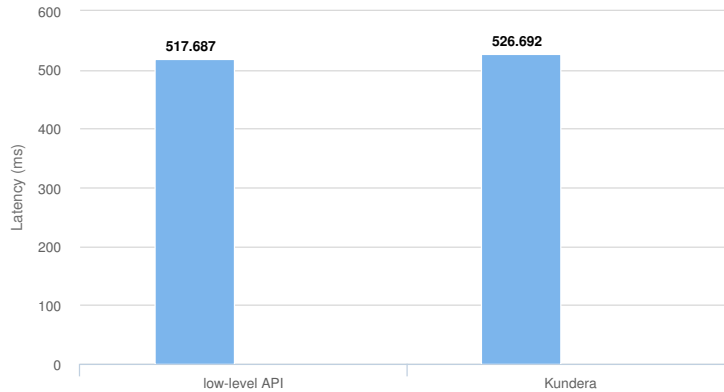
Test di performance dei client di Kundera sviluppati utilizzando YCSB (Yahoo Cloud Serving Benchmark).

- Adapter YCSB per operazioni attraverso Kundera
- Adapter YCSB per operazioni attraverso low-level API

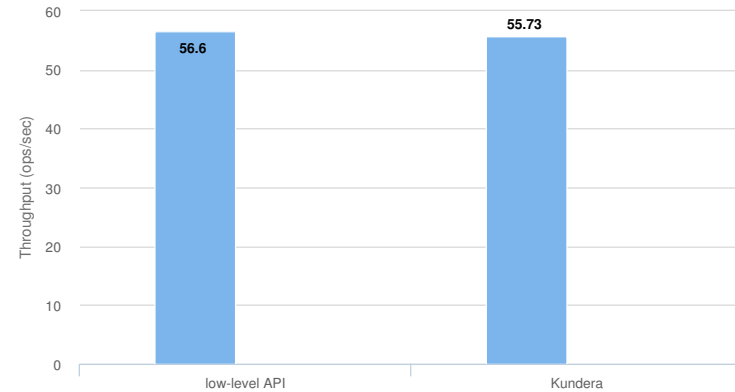
Workload da 100.000 operazioni in due fasi (write/read) su istanze remote di Google Datastore e Azure Tables.

Azure Tables results

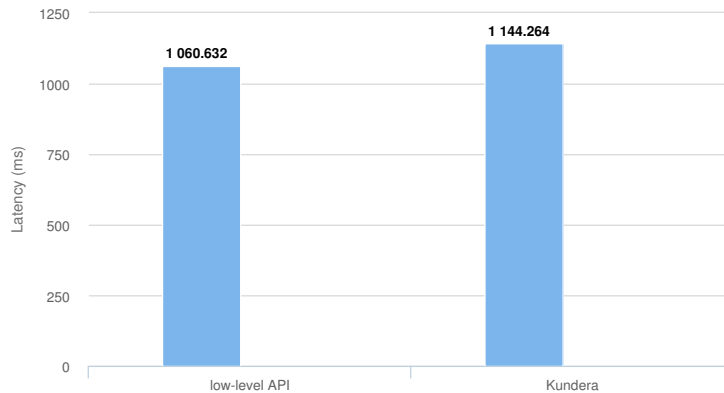
Read latency



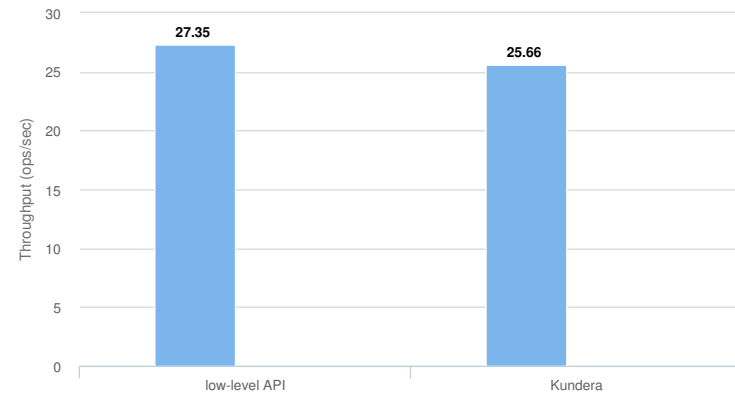
Read throughput



Write latency

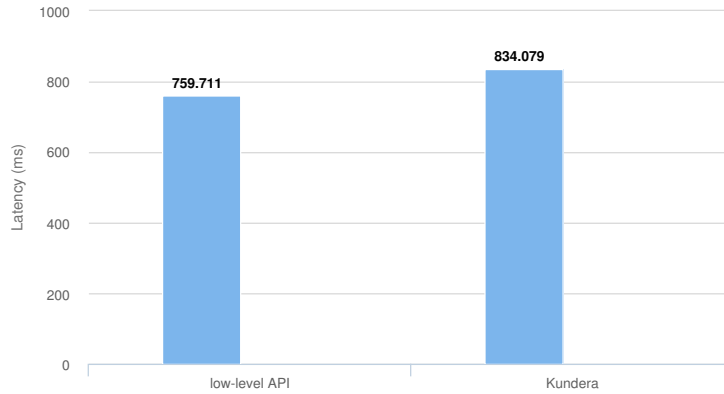


Write throughput

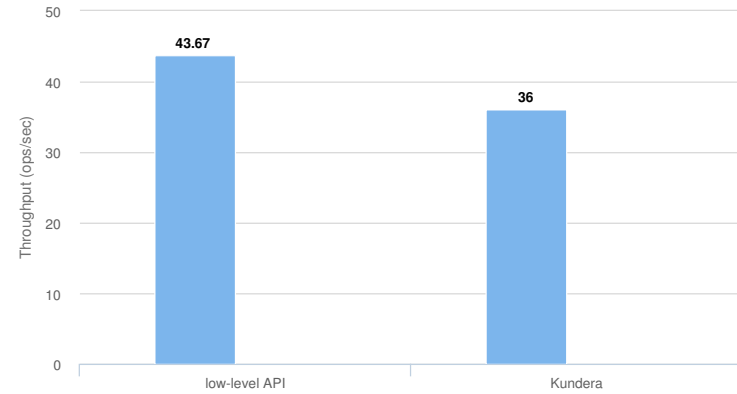


GAE Datastore results

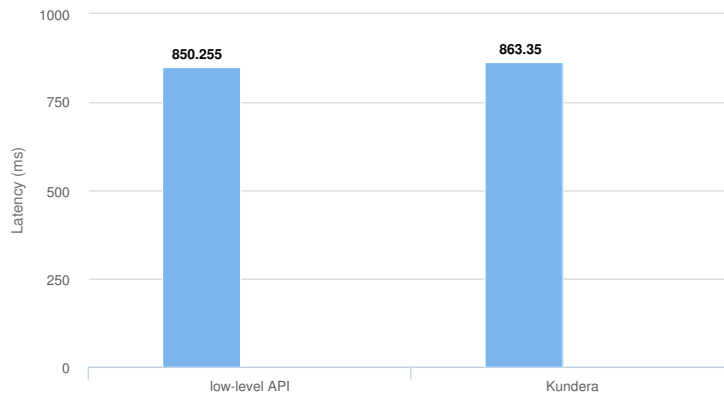
Read latency



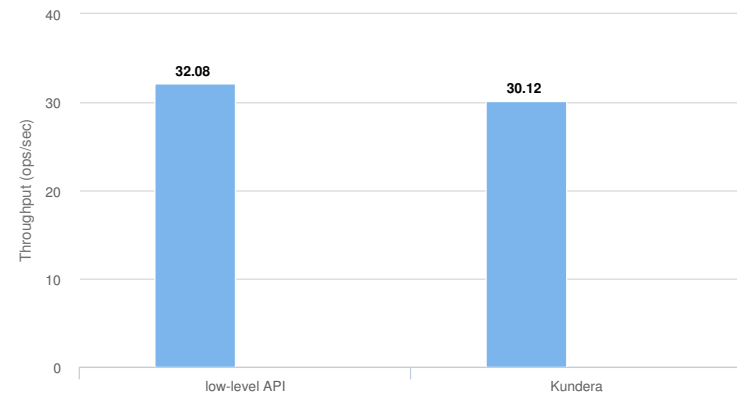
Read throughput



Write latency



Write throughput



Conclusioni

Contributi:

- Nuovi client per Kundera per il supporto di Google Datastore e Azure Tables
- Integrazione di Hegira all'interno della libreria CPIM

Sviluppi futuri:

- Estendere la libreria CPIM per supportare nuovi cloud e/o nuovi servizi
- Creare nuove estensioni di Kundera per ampliare i database NoSQL supportati