

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**
- **Scalabilità**
- **Nessun linguaggio standard**
- **BASE properties**
- **Sistemi distribuiti**

RDBMS

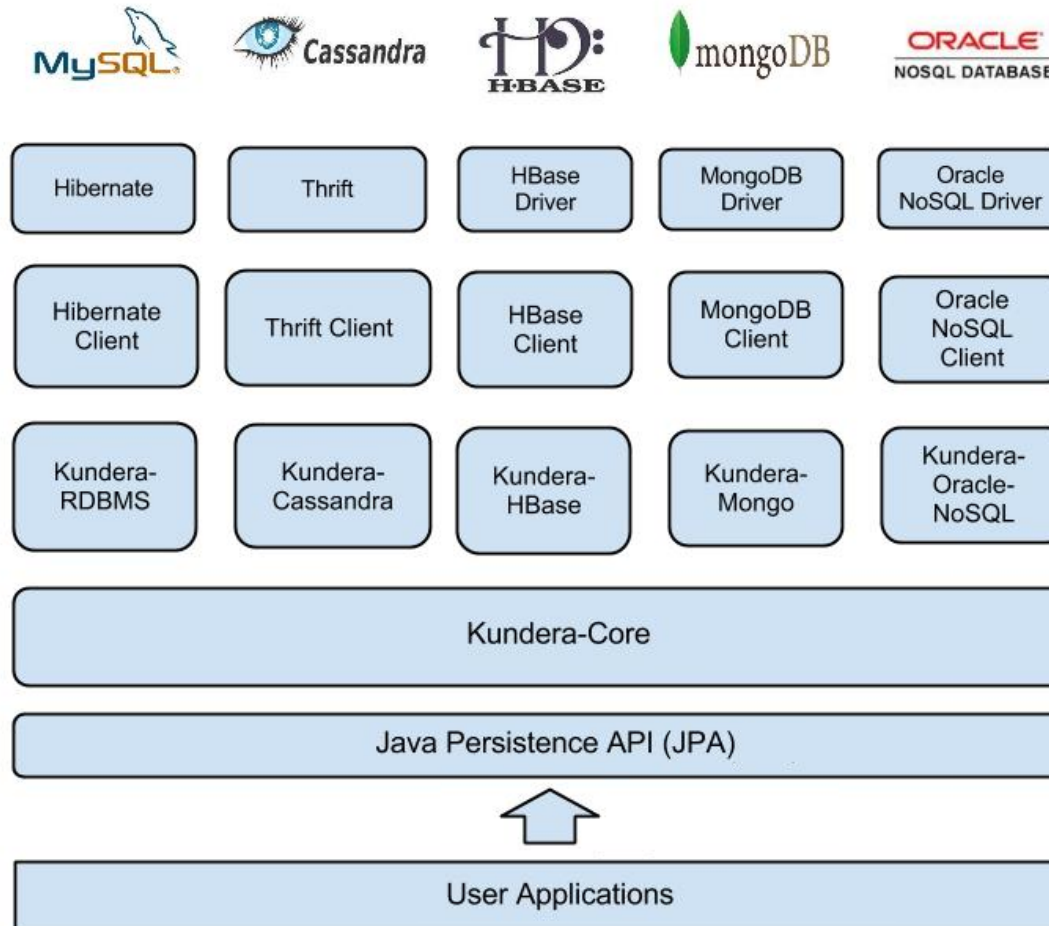
- **Dati benstrutturati**
- **Poca scalabilita**
- **SQL**
- **ACID transaction**
- **Centralizzati**

Common language approaches

Great content

Kundera

A JPA 2.1 compliant ORM Library for NoSQL databases



Cloud Platform Independent Model

content

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 data migration tra database NoSQL attraverso Hegira**

Data migration

Great content

Perché Kundera

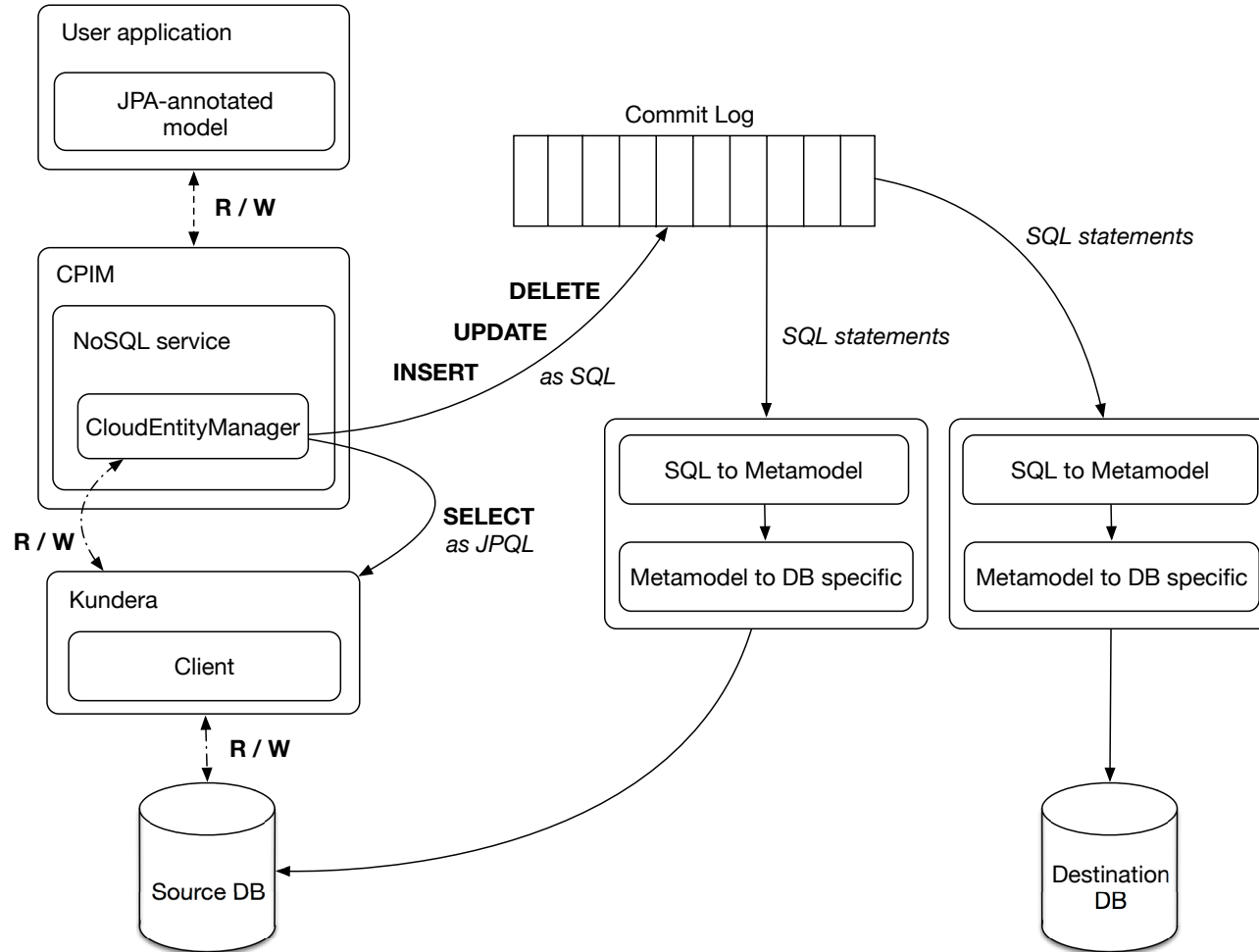
- **Open source**
- **Modulare, sviluppata con l'obiettivo di essere estendibile**
- **Ployglot persistency**
- **Nel campo dal 2010 con una community attiva**
- **Utilizzato in production**
- **Supporto a molti diversi database NoSQL**

Integrazione Kundera in CPIM

Great content

Supporto a Hegira

Servizio di migratione, attraverso Hegira, in modo trasparente all'utente della libreria CPIM



Title

Great content

Title

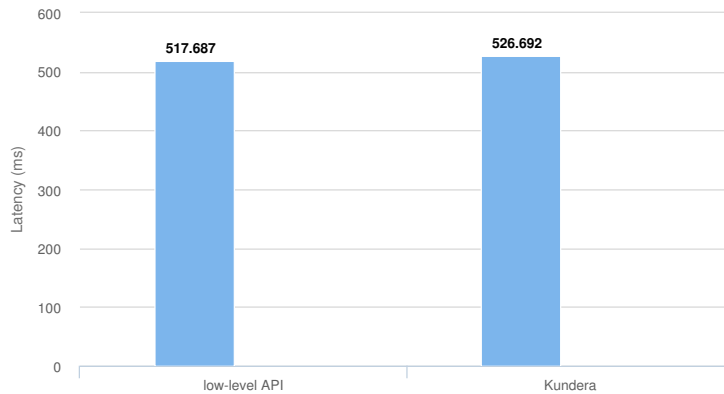
Great content

Performance

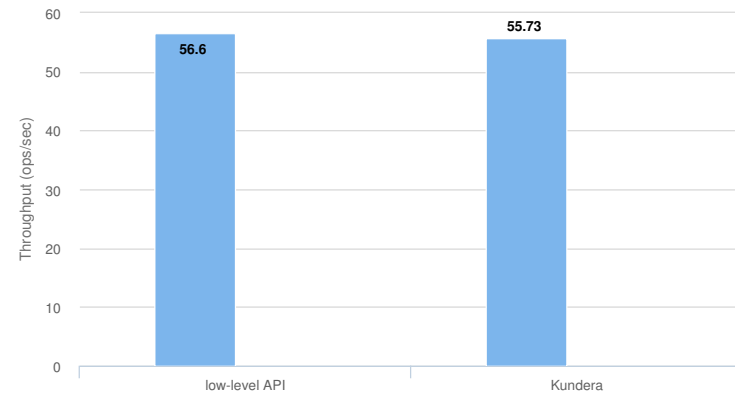
**Test di performance utilizzando YCSB (Yahoo Cloud Serving Benchmark)
sui client di Kundera sviluppati**

Azure Tables results

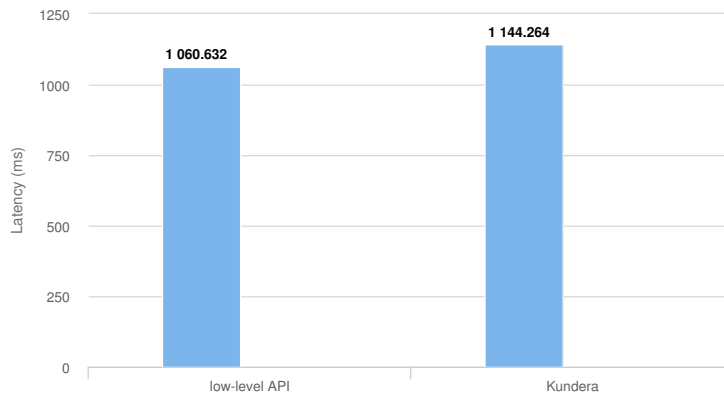
Read latency



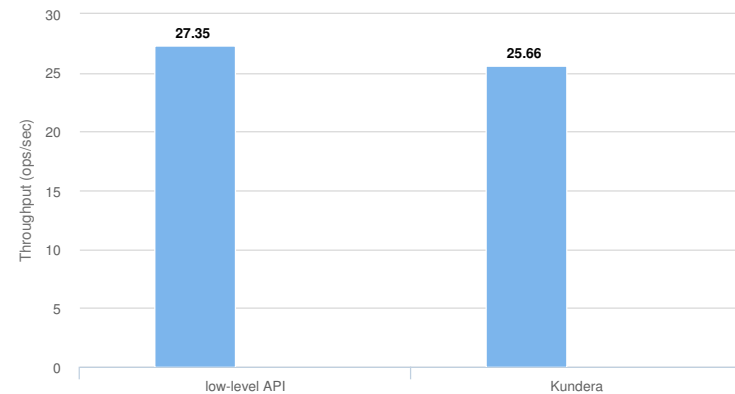
Read throughput



Write latency

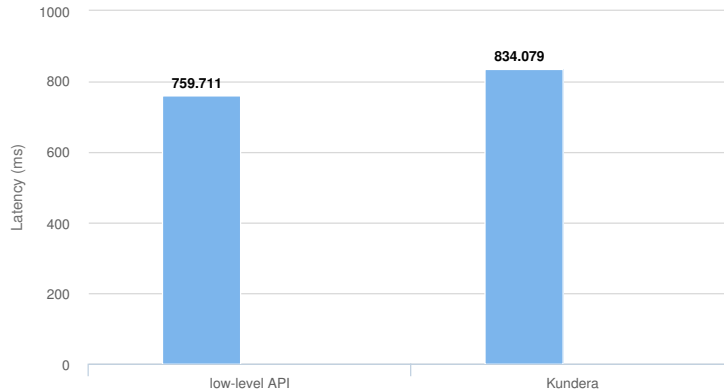


Write throughput

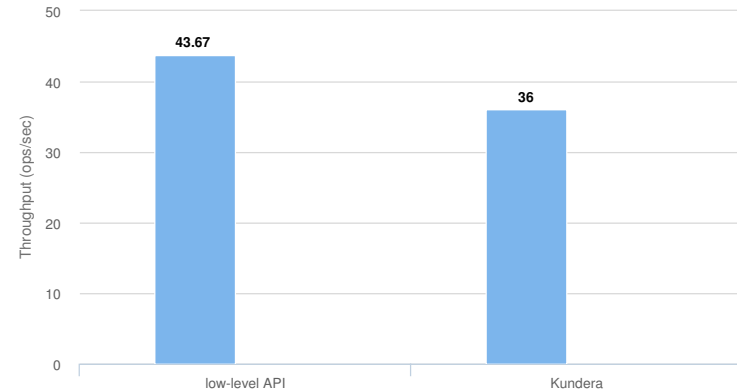


GAE Datastore results

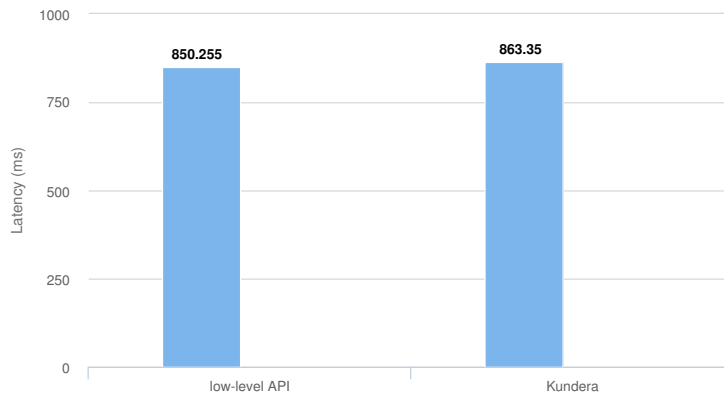
Read latency



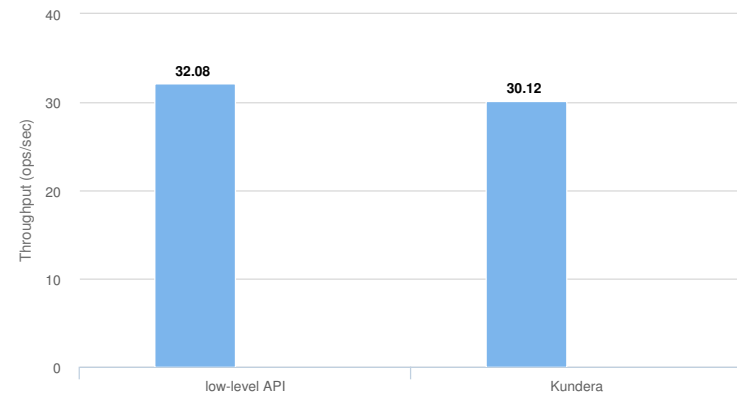
Read throughput



Write latency



Write throughput



Conclusioni

content

Ultima slide