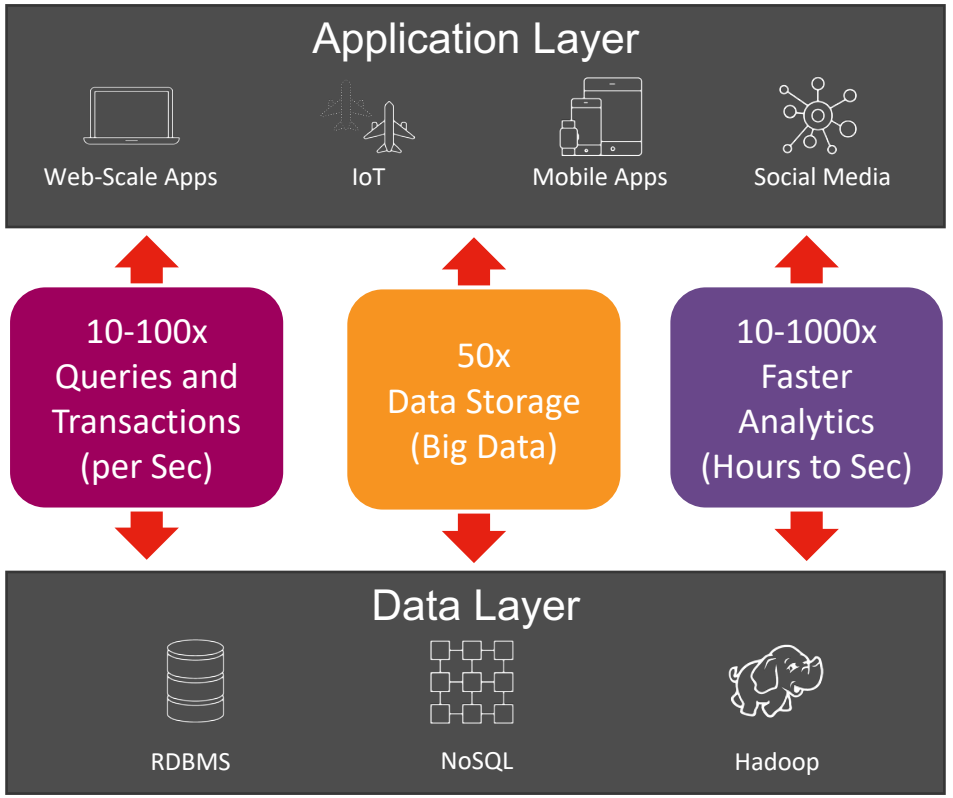
How-to for real-time streaming and analytics at scale with Apache Kafka and Apache Ignite

Viktor Gamov, Confluent, @gamussa

Denis Magda, GridGain, @denismagda

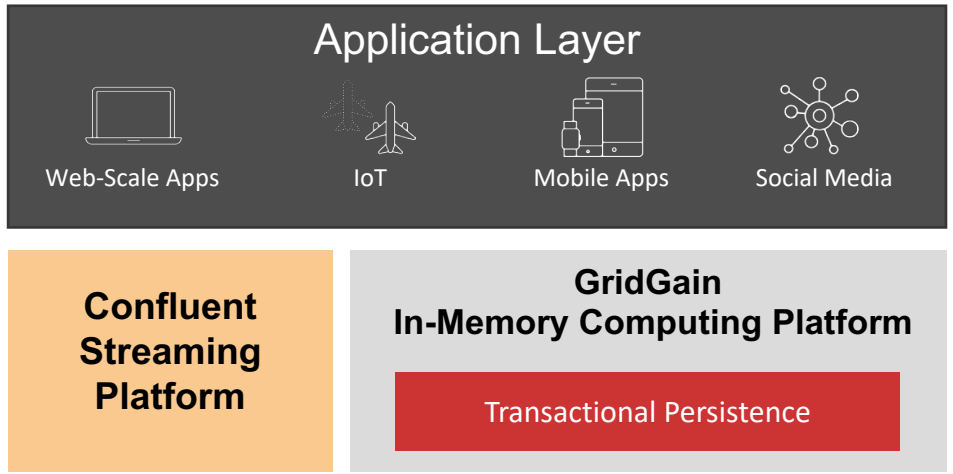
Digital Transformations Challenges

- 10-100x more queries and transactions
- 50x as much data today as a decade ago
- Overnight analytics becomes real-time

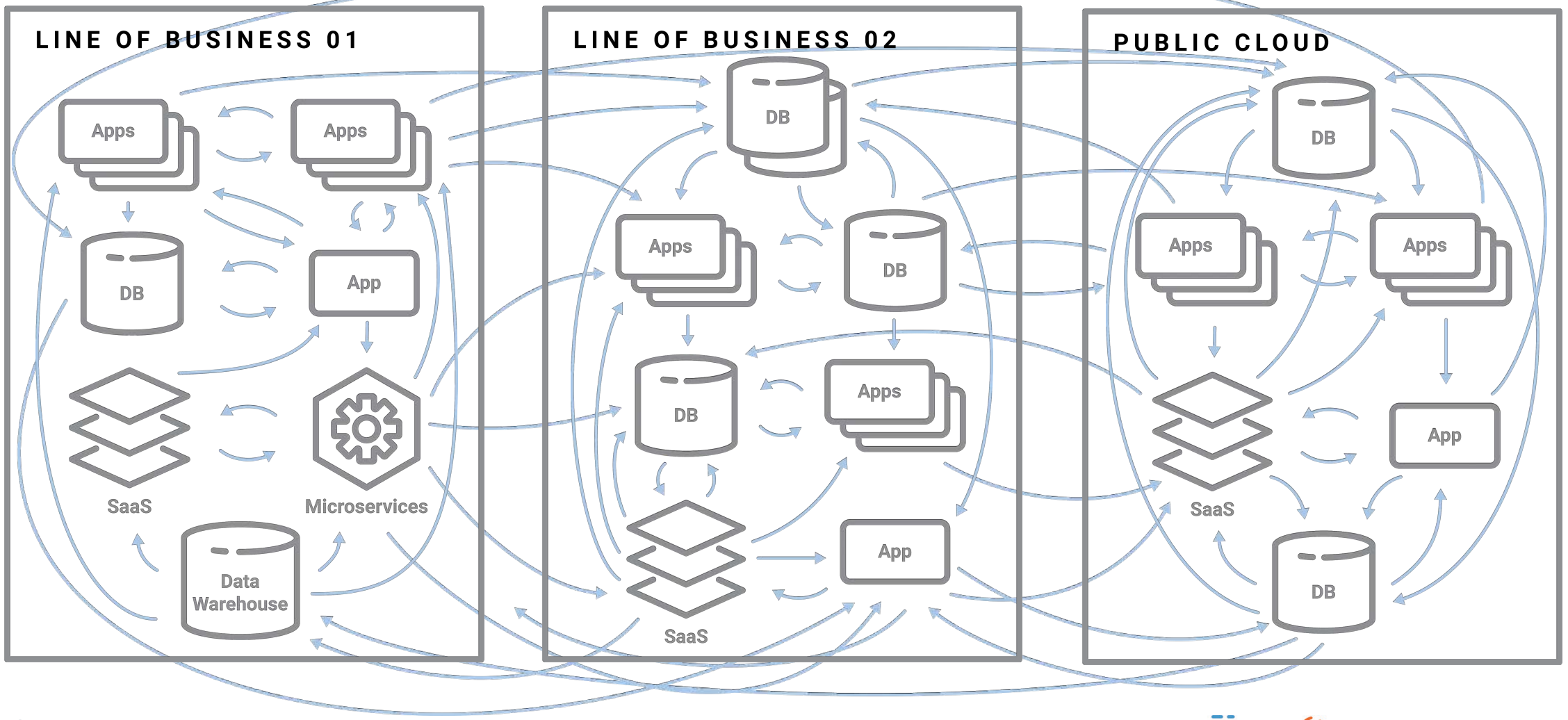


In-Memory Computing and Real-Time Streaming To Solve the Challenges

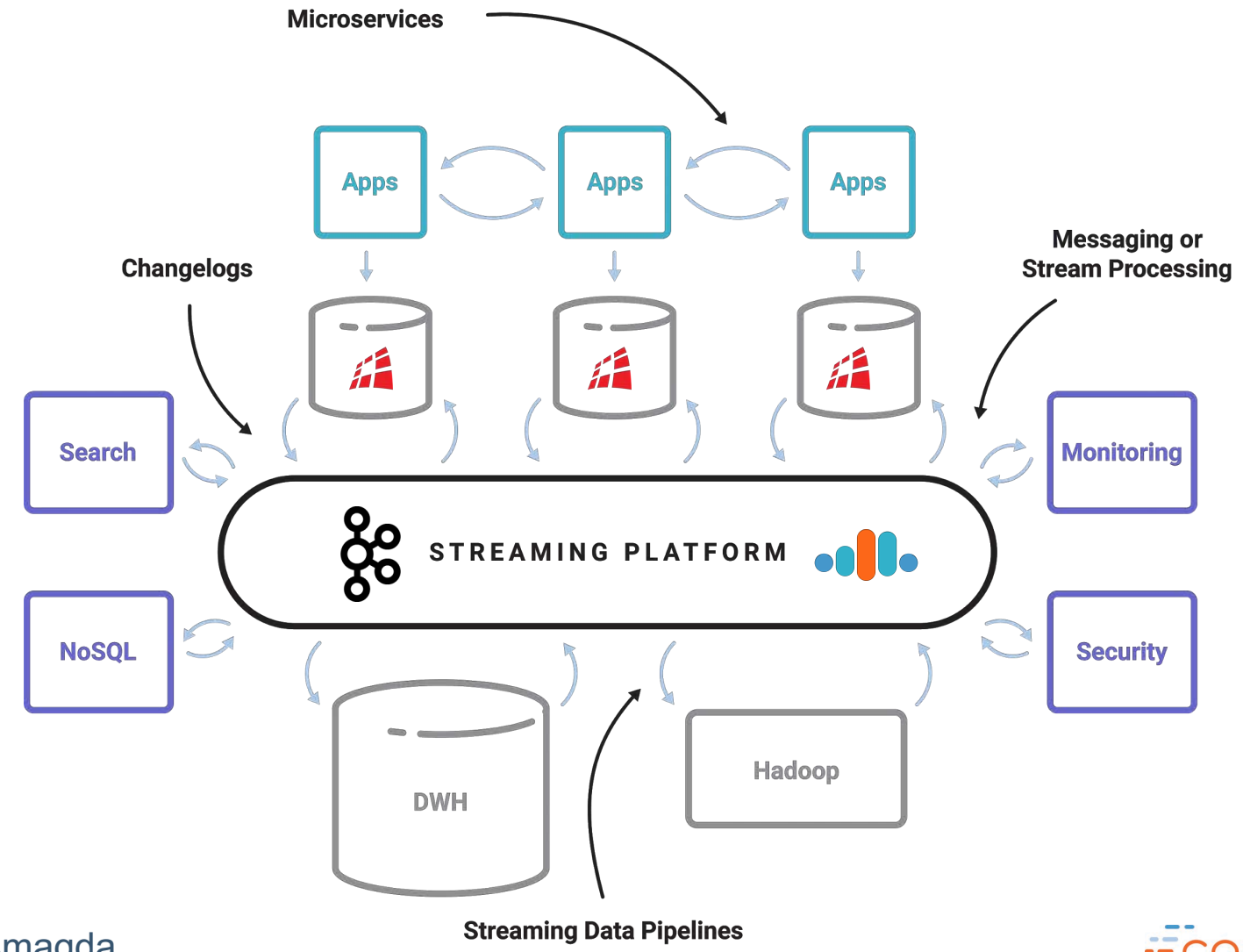
- Performance Increases 10x to 1,000x
- Act faster by analyzing streams of data
- Scalability up to petabytes of data



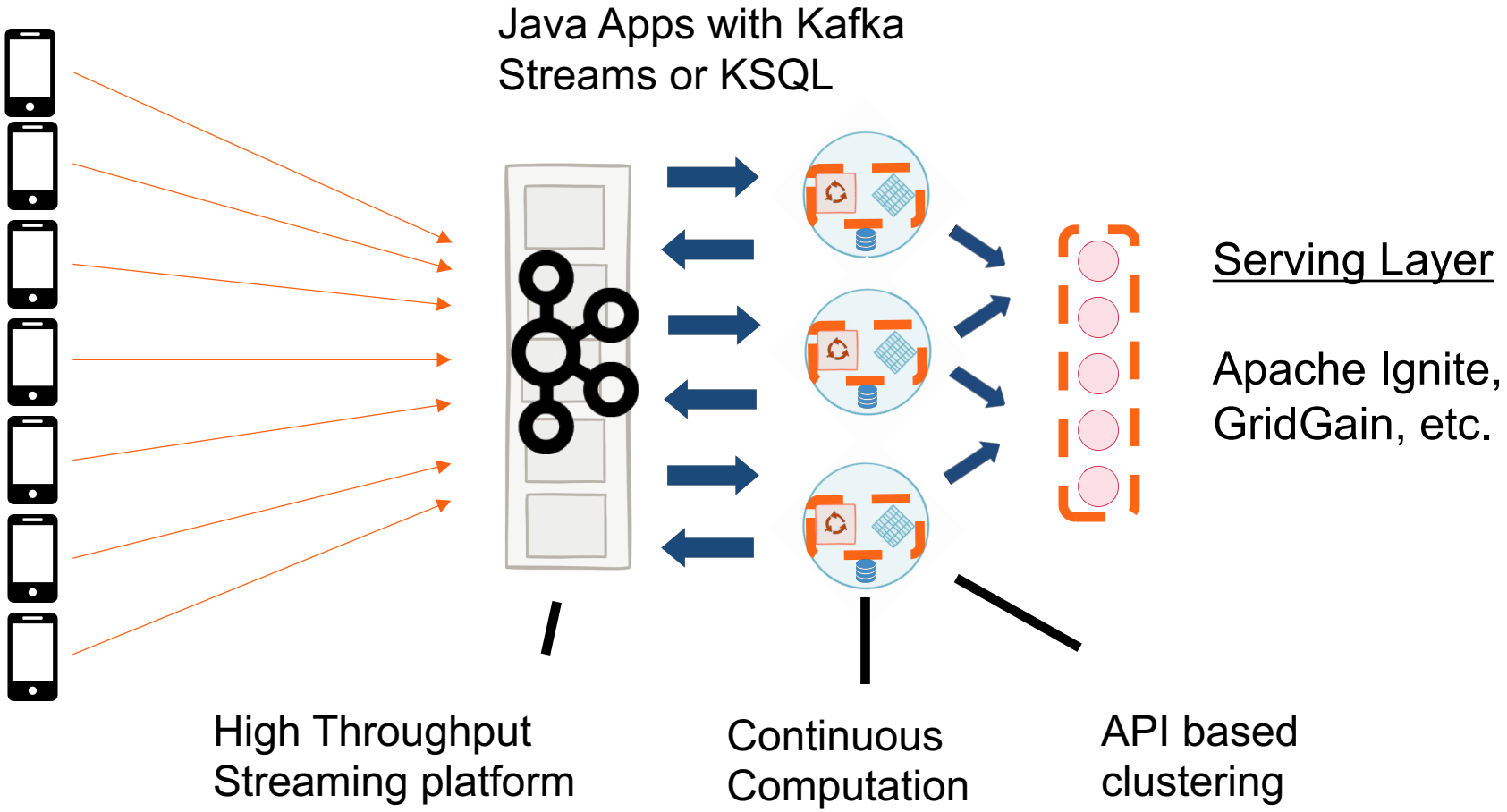
Pre-Streaming Era

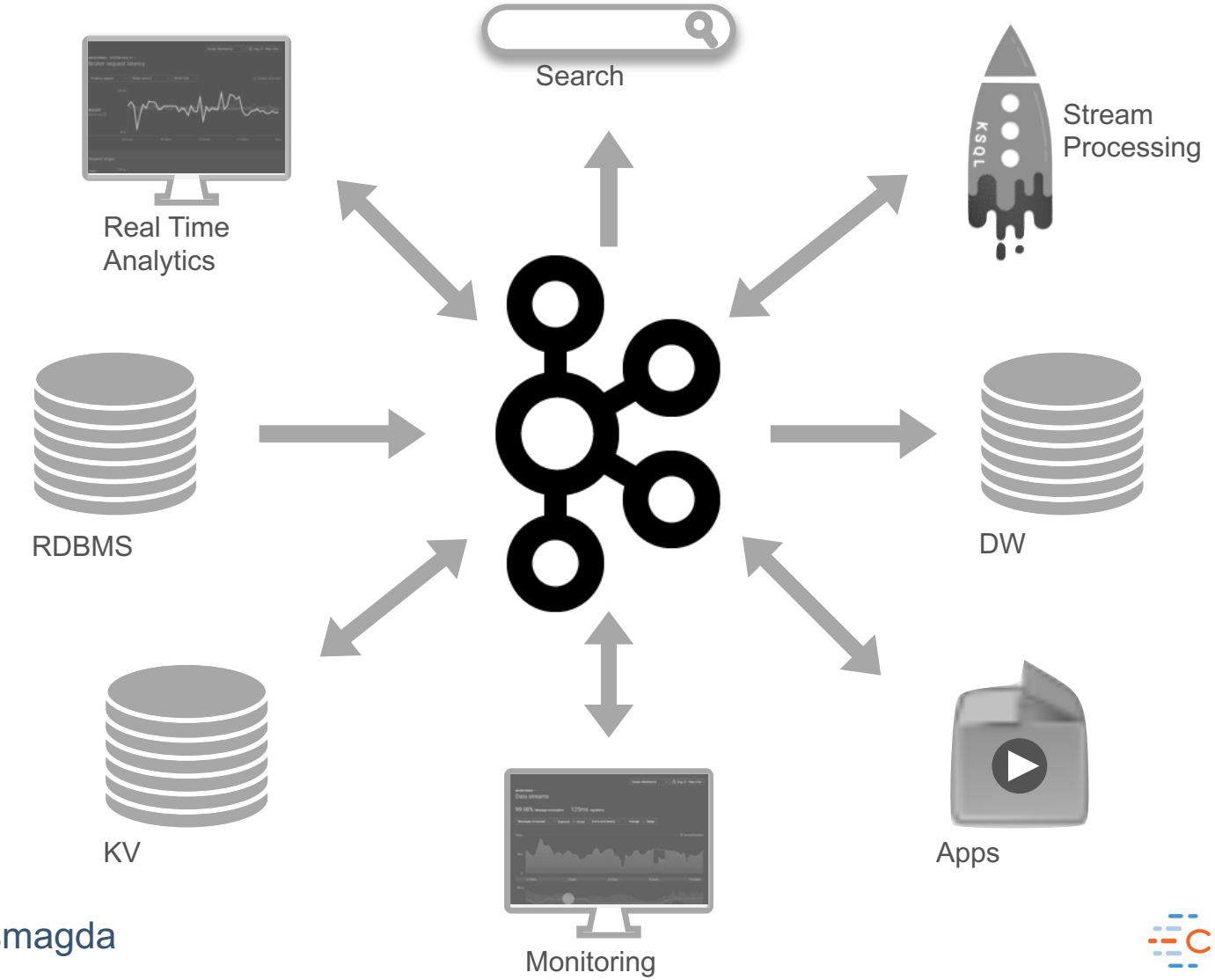


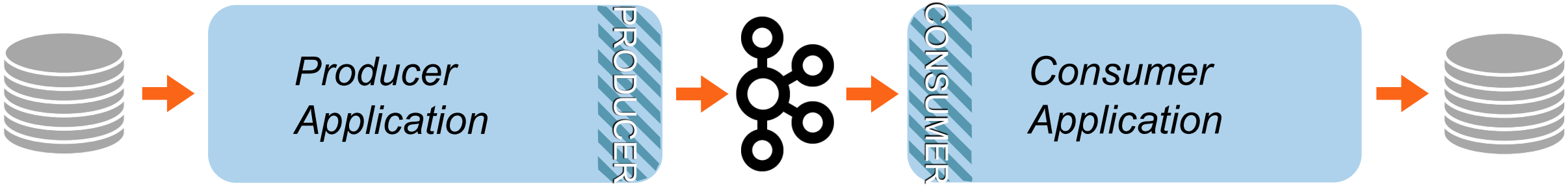
Streaming-First Workd



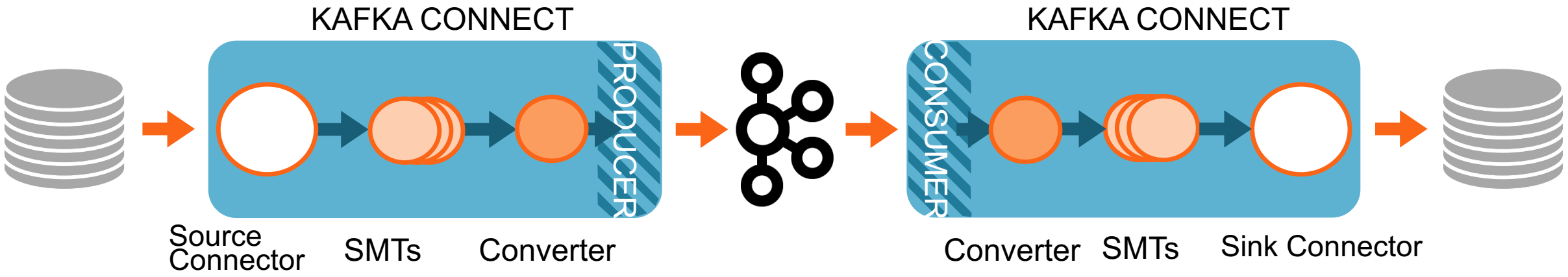
Origins in Streams Processing




Search
Real Time
Analytics
Stream
Processing
KSQL
RDBMS
DW
KV
Apps
Monitoring



- Where to restart ?
- How to scale and parallelize ?
- What metrics to capture ?
- How to handle failure & retries ?
- How to properly use the producer / consumer API ?

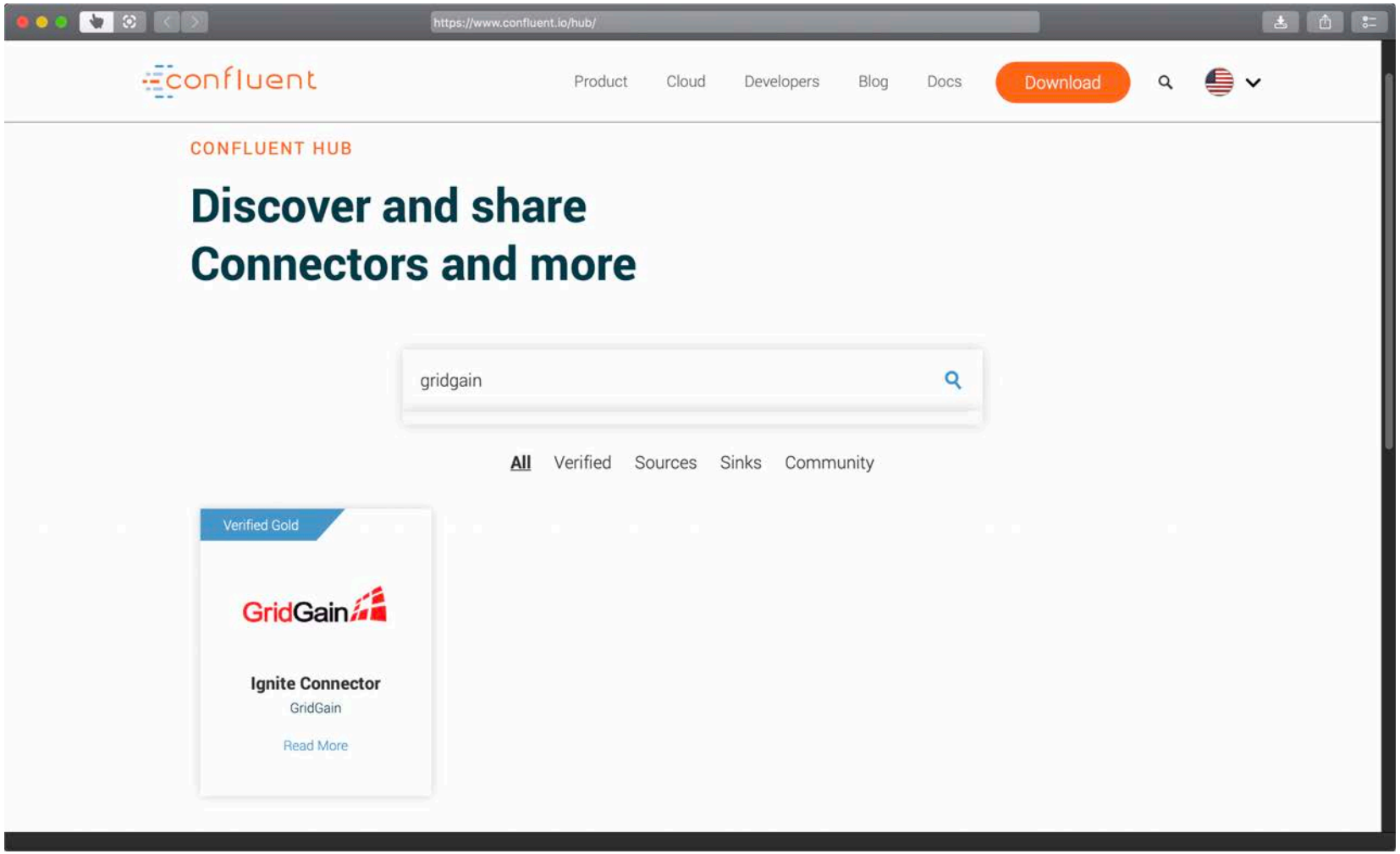


- Offset management
- Elastic scalability
- Parallelization

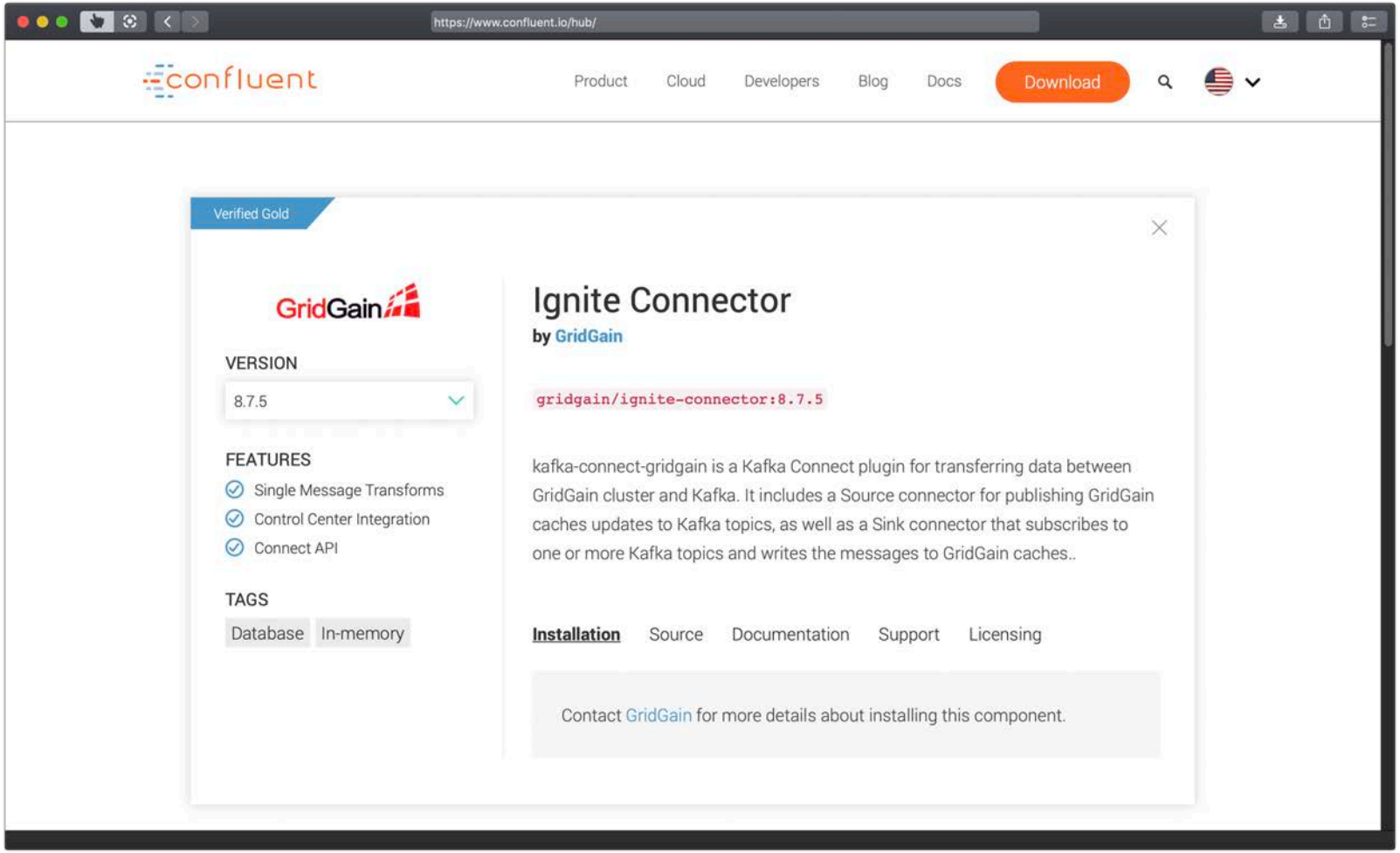
- Task distribution
- Metrics
- Failure & retries

- Configuration management
- REST API
- Schemas & data types

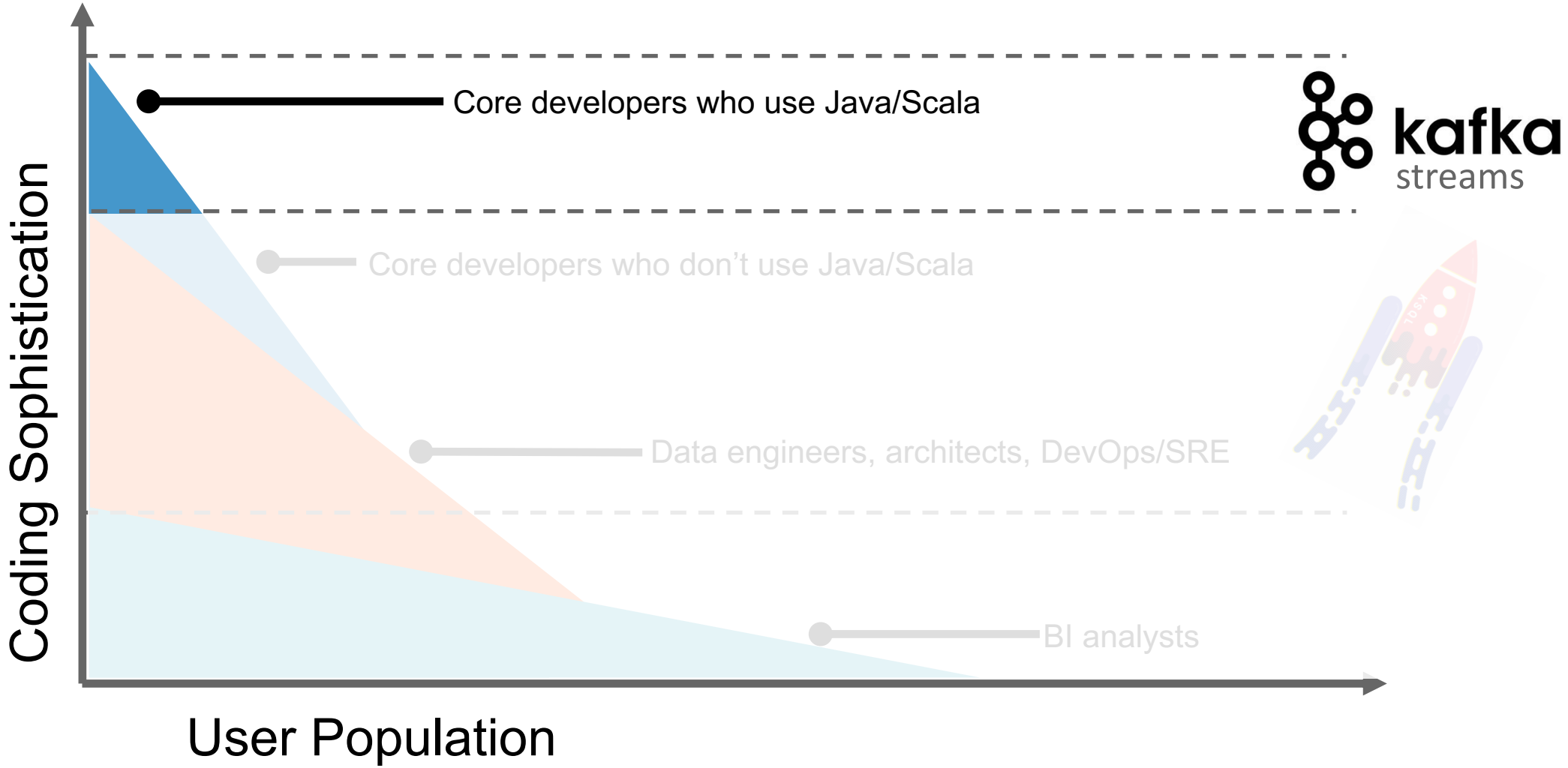
✓ Discover connectors, SMTs, and converters



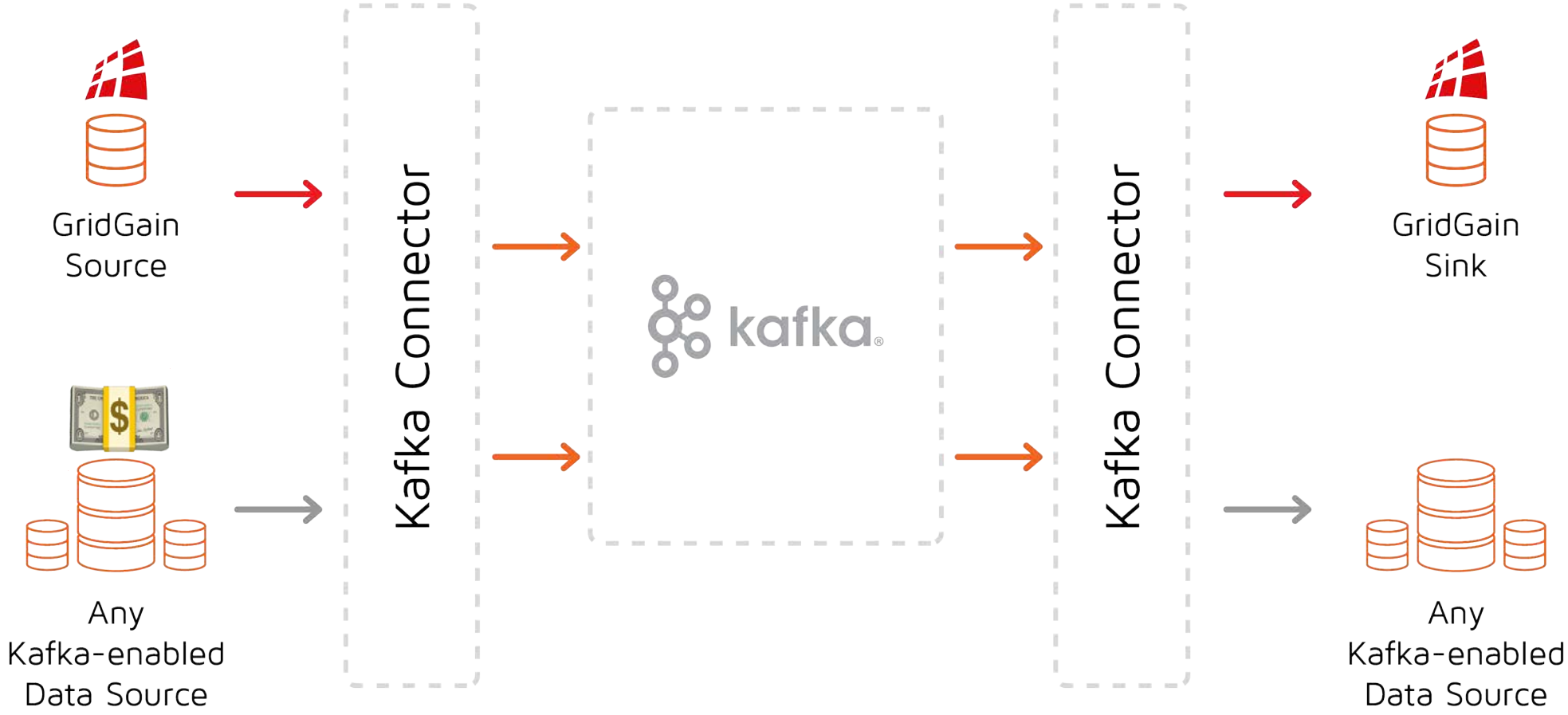
- ✓ Discover connectors, SMTs, and converters
- ✓ Descriptions, licensing, support, and more



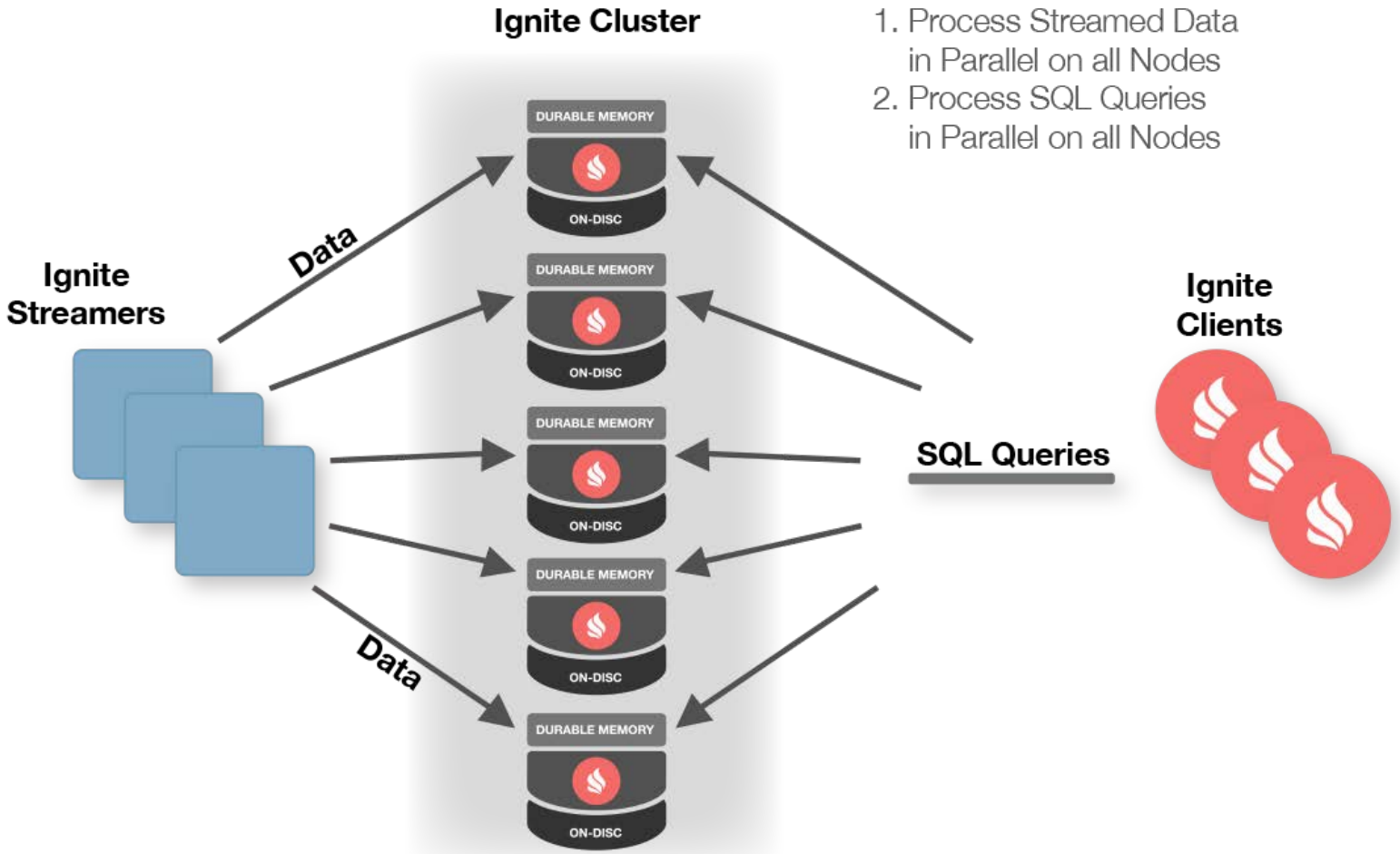
Lower the Bar to Enter the World



GridGain and Kafka Connect

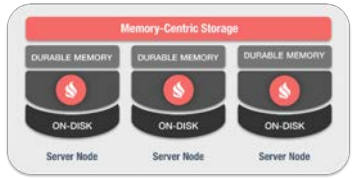


GridGain: Real-time Streaming and Analytics



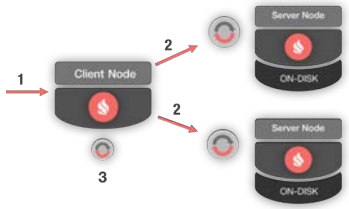
Essential GridGain APIs

Distributed memory-centric storage



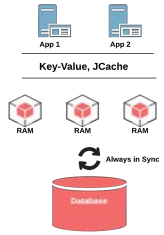
Combines the performance and scale of in-memory computing together with the disk durability and strong consistency in one system

Co-located Computations



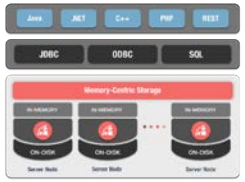
Brings the computations to the servers where the data actually resides, eliminating need to move data over the network

Distributed Key-Value



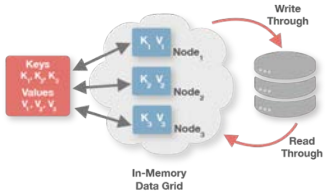
Read, write and transact with fast key-value APIs

Distributed SQL



Horizontally, fault-tolerant distributed SQL database that treats memory and disk as active storage tiers

ACID Transactions

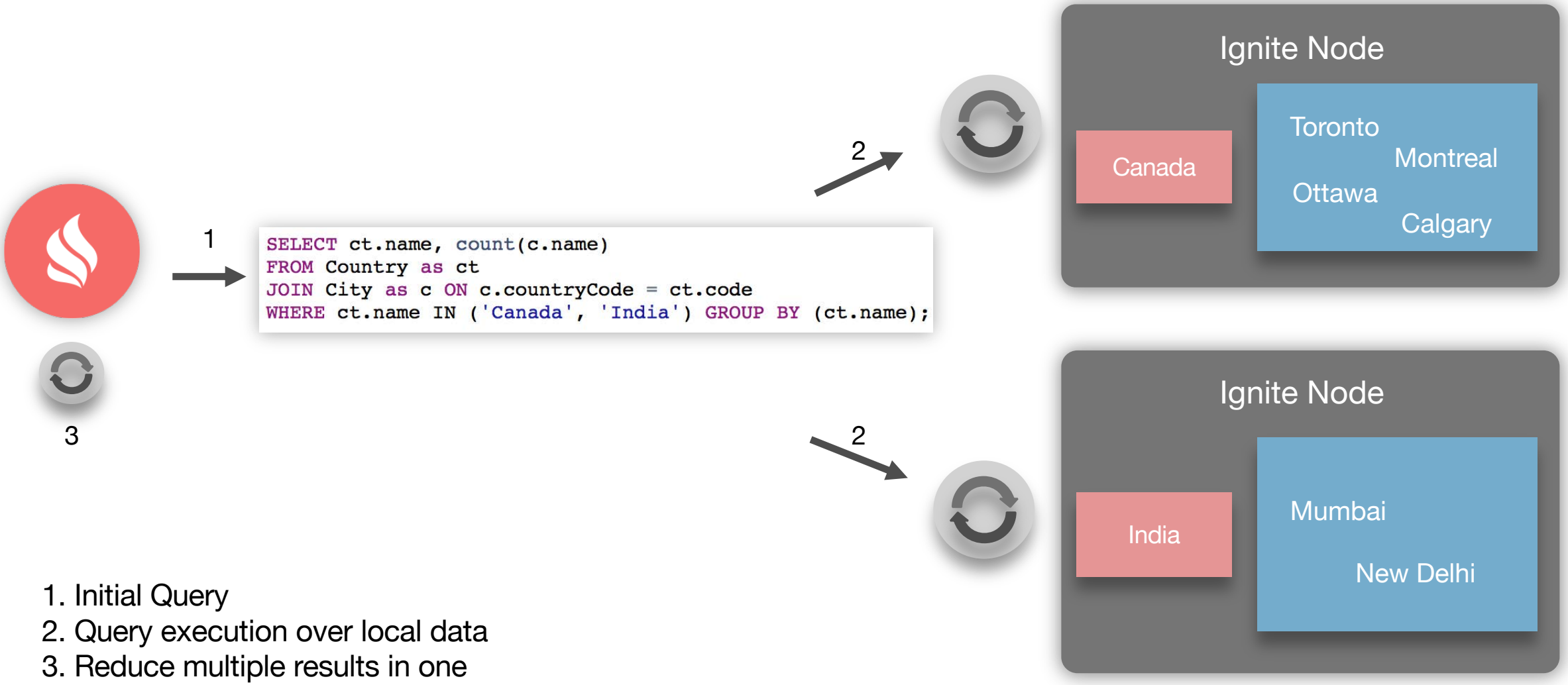


Supports distributed ACID transactions for key-value as well as SQL operations

Machine and Deep Learning

Set of simple, scalable and efficient tools that allow building predictive machine learning models without costly data transfers (ETL)

GridGain SQL For Real-Time Analytics

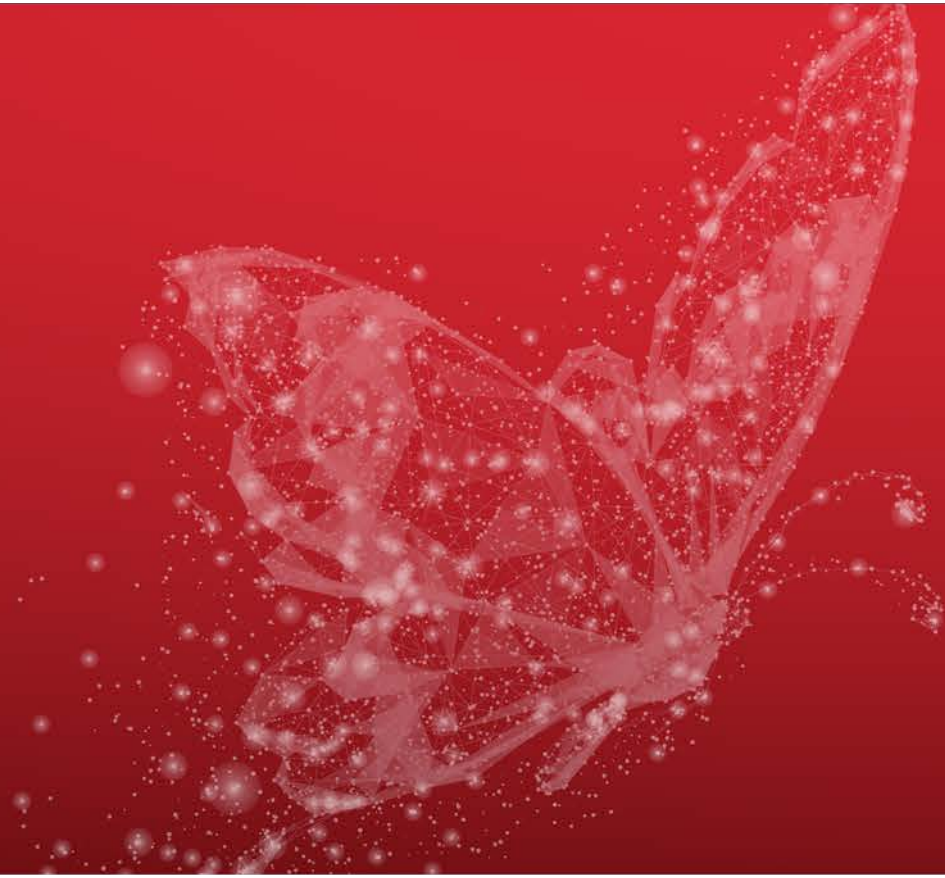


1. Initial Query
2. Query execution over local data
3. Reduce multiple results in one

Demo

confluent GridGain

Q&A



confluent GridGain