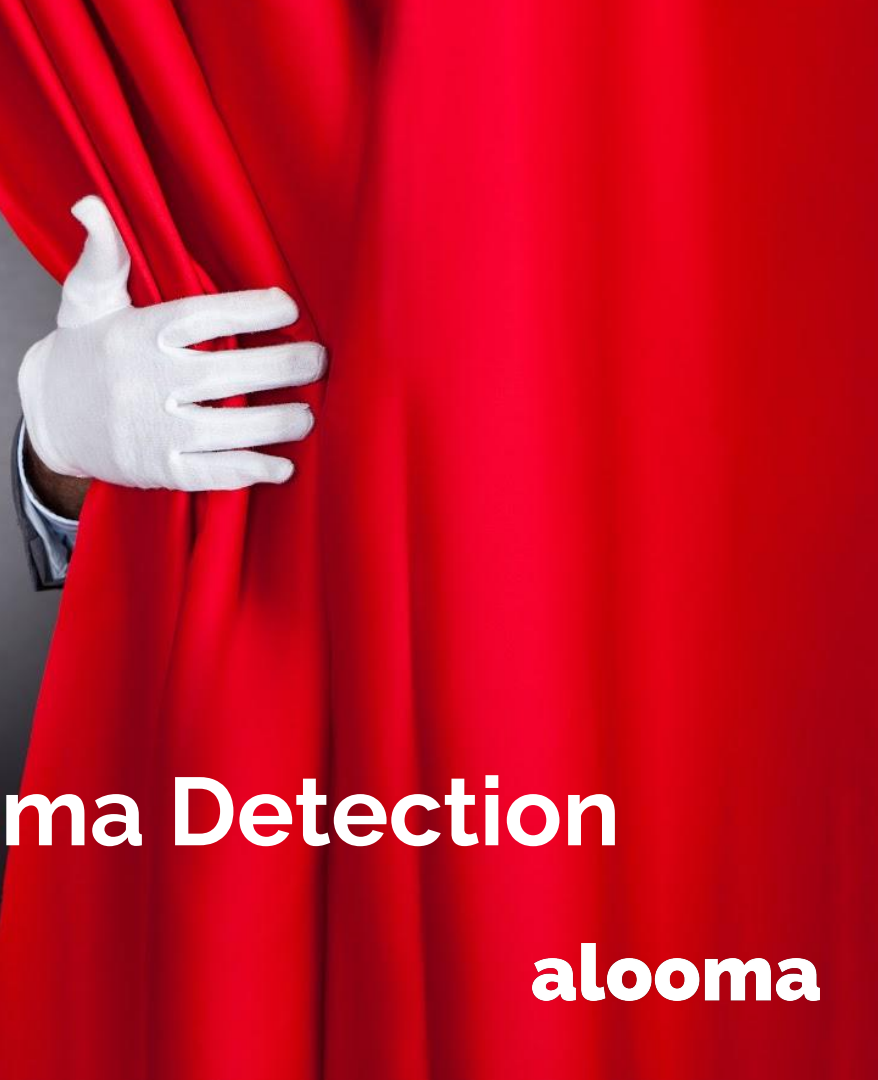
redislabs
home of redis

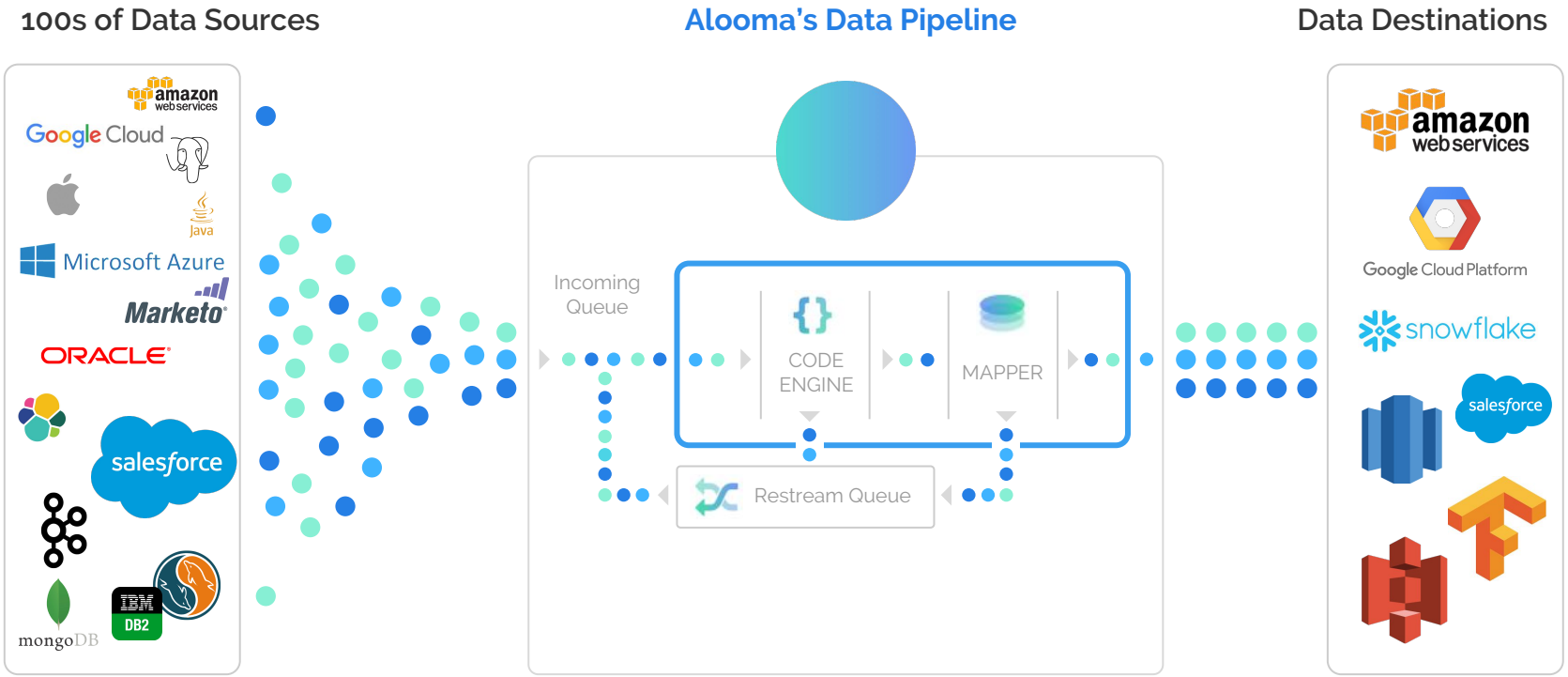
redisday
TLV 2018



Using Redis for Schema Detection

alooma

Alooma: Data Pipeline as a Service

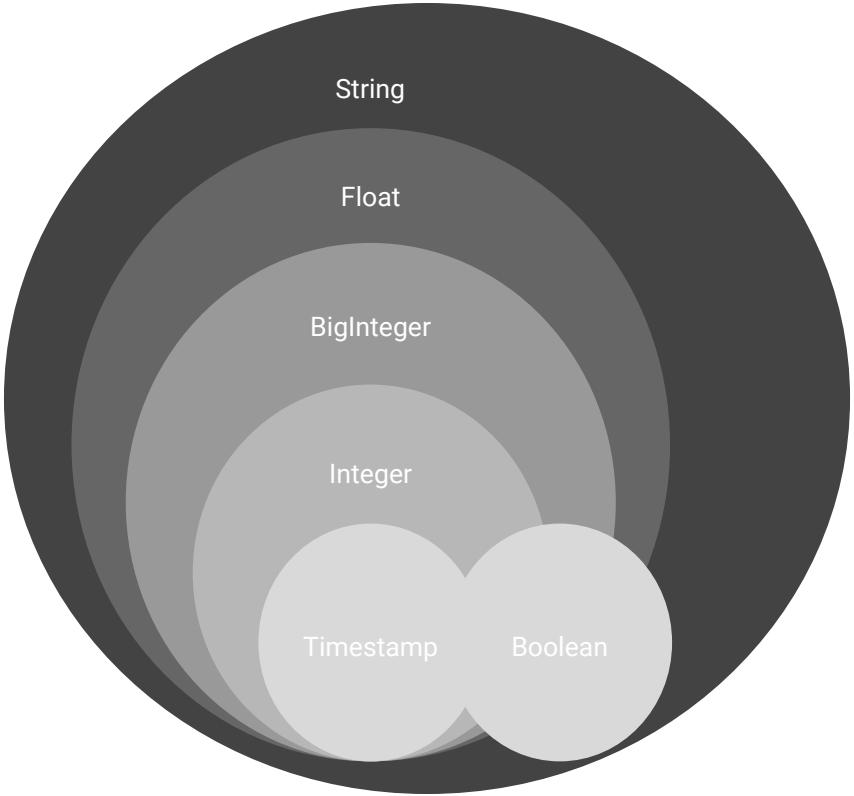


Challenge: Many data sources has no clear schema

- Apps & Agents
- WebHooks
- Csv lines

Naive (Bad) solution: Load data as text

Solution: Use samples to detect types



Solution: Use samples to detect types

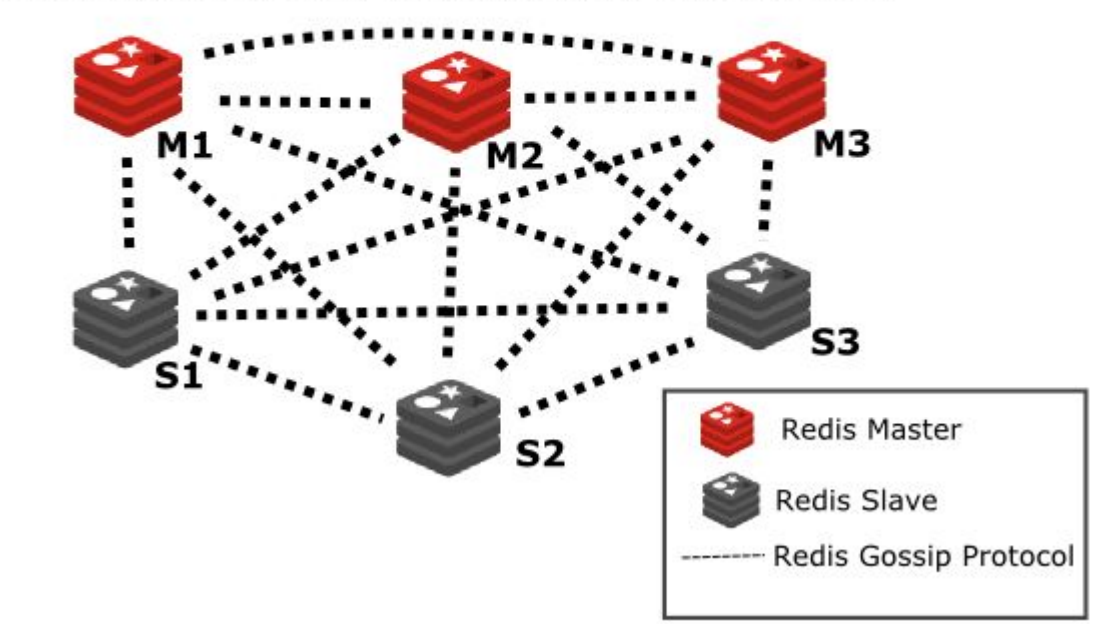
- Build statistics of types based on samples
- Choose strategy:
  - Choose type only if 100% of the samples suite it
  - Choose type if 95% of the samples suite it

Challenge: meet our performance requirements

- Super high-scale
- Real time event streams
- Distributed System

Solution: Redis cluster

- ✓ High Performance
- ✓ High Availability
- ✓ Key-value Store
- ✓ Support for Sub Keys



Solution: Redis cluster

- Update Statistics in real time
- Store statistics for each fields
- Update redis every micro batch

Thank You!
redisday
TLV 2018
redislabs
home of redis