China · Beijing

KV Storage Online Production Application in PayPal Risk

Bruce Li

Paypal Risk Tech Infra

Agenda

1、Paypal Risk Data Challenges

2、Considerations on Risk Storage

3、Generic KV Storage Data Access Solution

4、Experiences Learnt from Aerospike Migration

5、Q&A

Challenges of Paypal Risk Data Access

Business Requirements:

- Provide sub-second level high quality risk decision service
- Support rapid business growth
- Support flexible & fast-evolving data schema changes
- Risk decisions should be offline simulatable

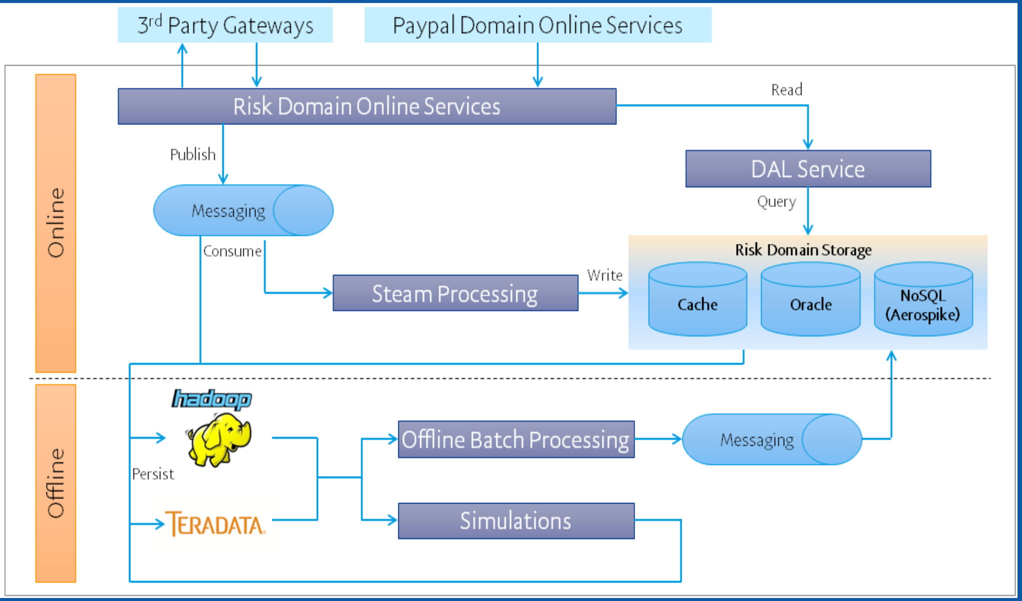
Tech Challenges:

- Low latency & large parallel data loading
- 10X scalability and high availability
- The same schema supports both online & offline

Risk Online Data & Flow

Major Risk Data Set Types

- Real-time events
- Real-time computed data
- Near-real-time computed data
- Offline computed data
- Static data
- Others



Online Risk Data Storage Requirements

	Events	Offline Computed Data	Real-time Computed Data	Static Data
Is KV use case	Storage is not, but event cache can be	Yes	Yes	Not natural KV, but convertible
Join needed	Limited	No	No	Very limited
Row size	Medium-Large	Small-Medium	Medium-Large	Small-Medium
Raw data size	~X00 TB	~X0 TB	~X0 TB	~X0 GB
Column based	Yes	Yes	Yes	Yes
Secondary index needed	No	No	No	No
Need server-side compute on write	Yes (for cache)	No	Yes	No
Need server-side compute on read	Yes	No	Yes	No
Recommended storage	DB + KV cache	KV storage	KV storage/cache	Embedded DB

Agenda

1、Paypal Risk Data Challenges
2、Considerations on Risk Storage
3、Generic KV Storage Data Access Solution
4、Experiences Learnt from Aerospike Migration
5、Q&A

KV Storage Selection

NoSQL Feature	AeroSpike	Couchbase	Cassandra	MongoDB
SSD Support?	Yes	Yes	Yes	Yes
Type	Key/Value (& Columnar)	Document (JSON)	Columnar	Document (BSON)
User Defined Functions (UDF) or Scripting?	Yes (Lua)	No	Yes (Java)	Yes (JavaScript)
Partial Read/Write Support?	Yes	No	Yes	No
Server-side Computation?	Yes	No	Yes	Yes
Complex data type Support?	Yes	Yes (JSON)	Yes	Yes (JSON)
Secondary Indexes?	Yes	Yes	Yes	Yes
TTL expiration?	Yes	Yes	Yes	Yes
Cross Data Center Replication?	Yes	Yes	Yes	No
Range Queries?	Yes (not mature enough)	No	Yes	No
Automatic failover & rebalancing?	Yes	Yes	Yes	Yes
Auto-Sharding	Yes	Yes	Yes	Yes (not mature enough)
Tunable Consistency levels?	Yes	No	Yes	Yes
SQL Like Interface?	Yes	No	Yes	No
Aggregation Query Support?	Yes	No	Yes	No
Share-nothing Architecture?	Yes	Yes	Yes	No (Master/slave)
Cross Datacenter Replication (XDCR/XDR)?	Yes	Yes	Yes	Yes
Performance / Scalability Concerns	N/A - Green Light!	N/A	GC Pauses	Master/slave SPOF

KV Storage & Data Access

KV Storage

- Generic KV storage abstraction
- Won't be locked in by a specific solution
- Separate operations and business

Data Access

- Flexible mapping from data set to KV
- Data compaction
- Metadata driven, support fast-evolving schema changes
- Very high throughput but low latency

Data Access Layered Abstraction

Data Set

- Multi-columns but only one key

Data Access

- Generic async/sync API，e.g. get/put/compute/batch/metadata API
- Data set to KV mapping
- Data set name, column name dictionary

KV Storage/Cache

- Sharding, node add/remove, XDR, data migration…

Agenda

1、Paypal Risk Data Challenges

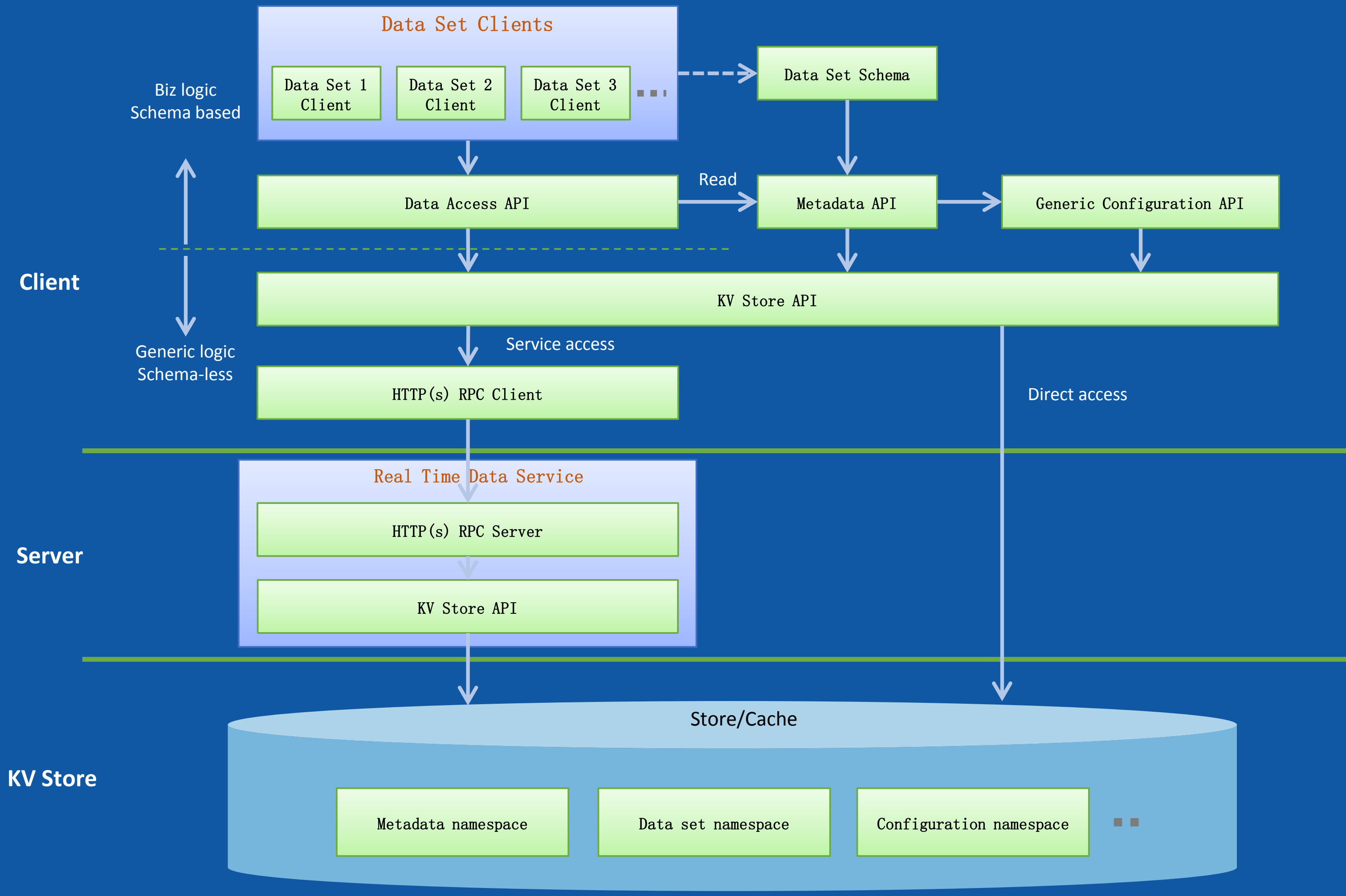
2、Considerations on Risk Storage

3、Generic KV Storage Data Access Solution

4、Experiences Learnt from Aerospike Migration

5、Q&A

Generic KV Data Access Design



Generic KV Data Access API

Read API

- Get, exists

Write API

- Insert, update, upsert, delete, CAS based update/insert/…

Server-side Compute API

- Increase/decrease, list/map/set operations, user defined function（UDF）

Batch API

- Read/write/compute/…

KV Data Access Service Considerations

High Throughput

- Customized async RPC based on Netty, lock-free implementation
- Async storage client, always use batch when possible

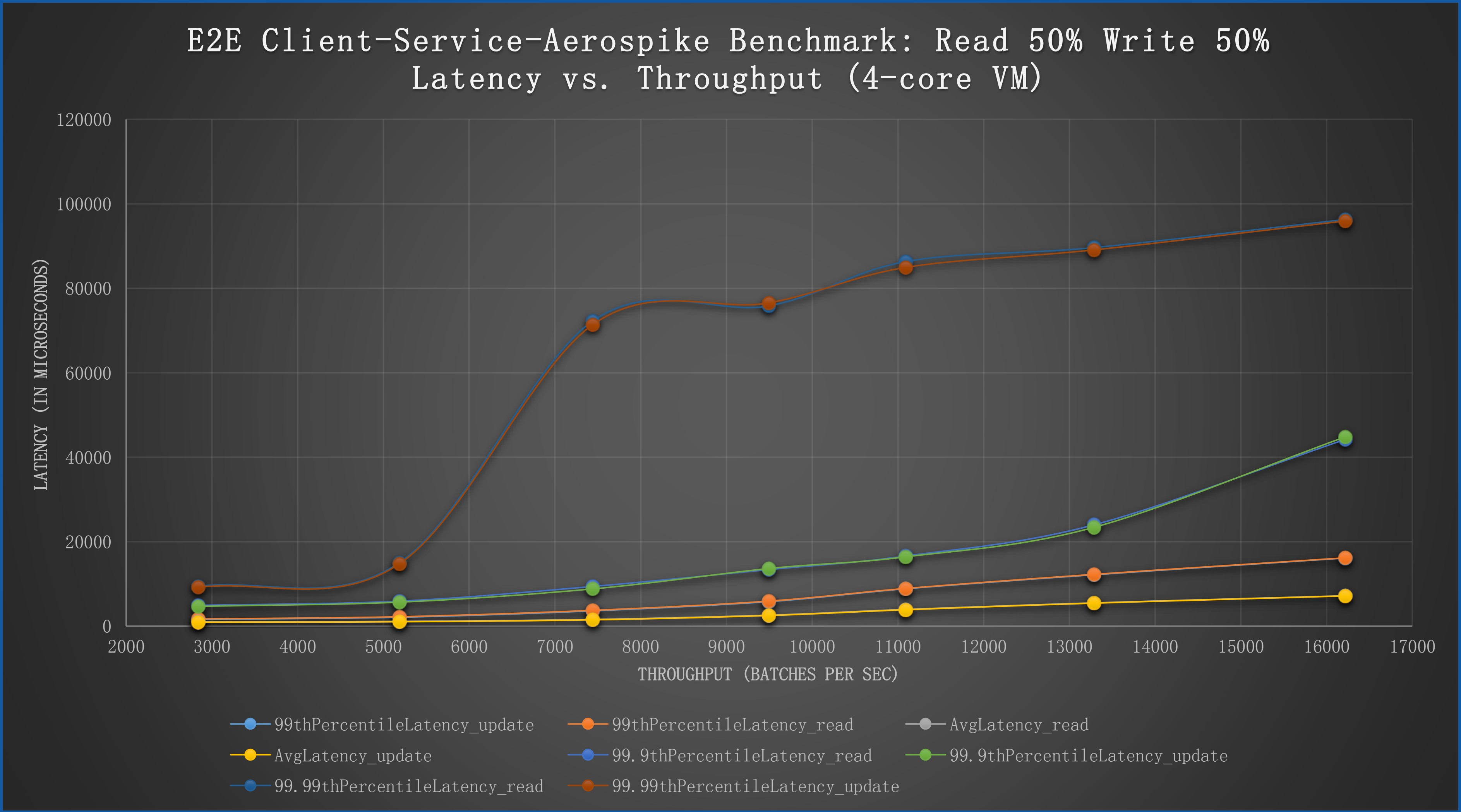
Low Latency

- De/Serialization on direct NIO buffers
- Avoid unnecessary object creation/memory copy, GC optimizations
- Release unused objects as earlier as possible
- Workaround HTTP/1.1 limits, or embrace HTTP/2

Isolation

- Cluster vs. node level
- Connection level
- JVM level

Asynchronous Data Access Service Benchmark



Agenda

1、Paypal Risk Data Challenges

2、Considerations on Risk Storage

3、Generic KV Storage Data Access Solution

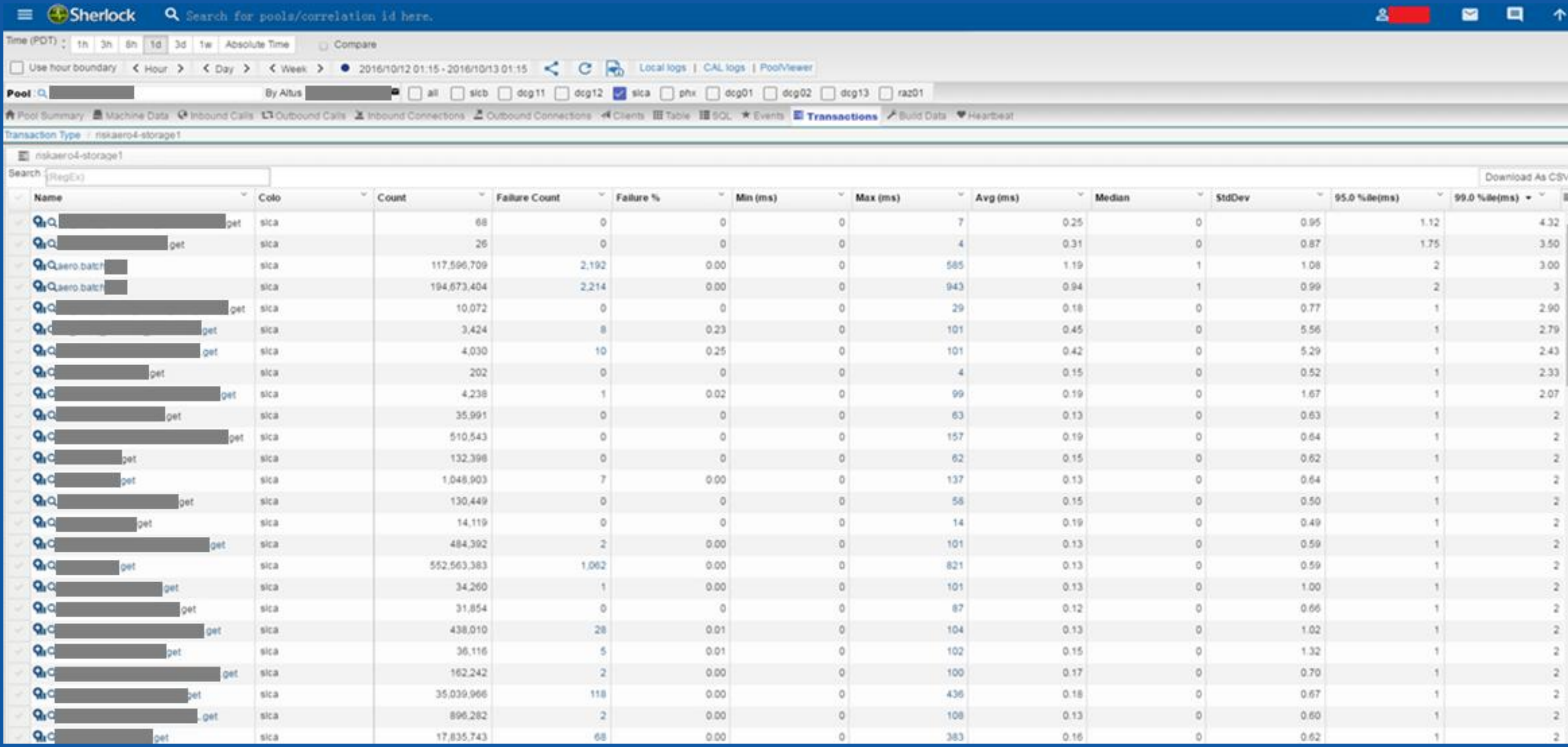
4、Experiences Learnt from Aerospike Migration

5、Q&A

Why Aerospike

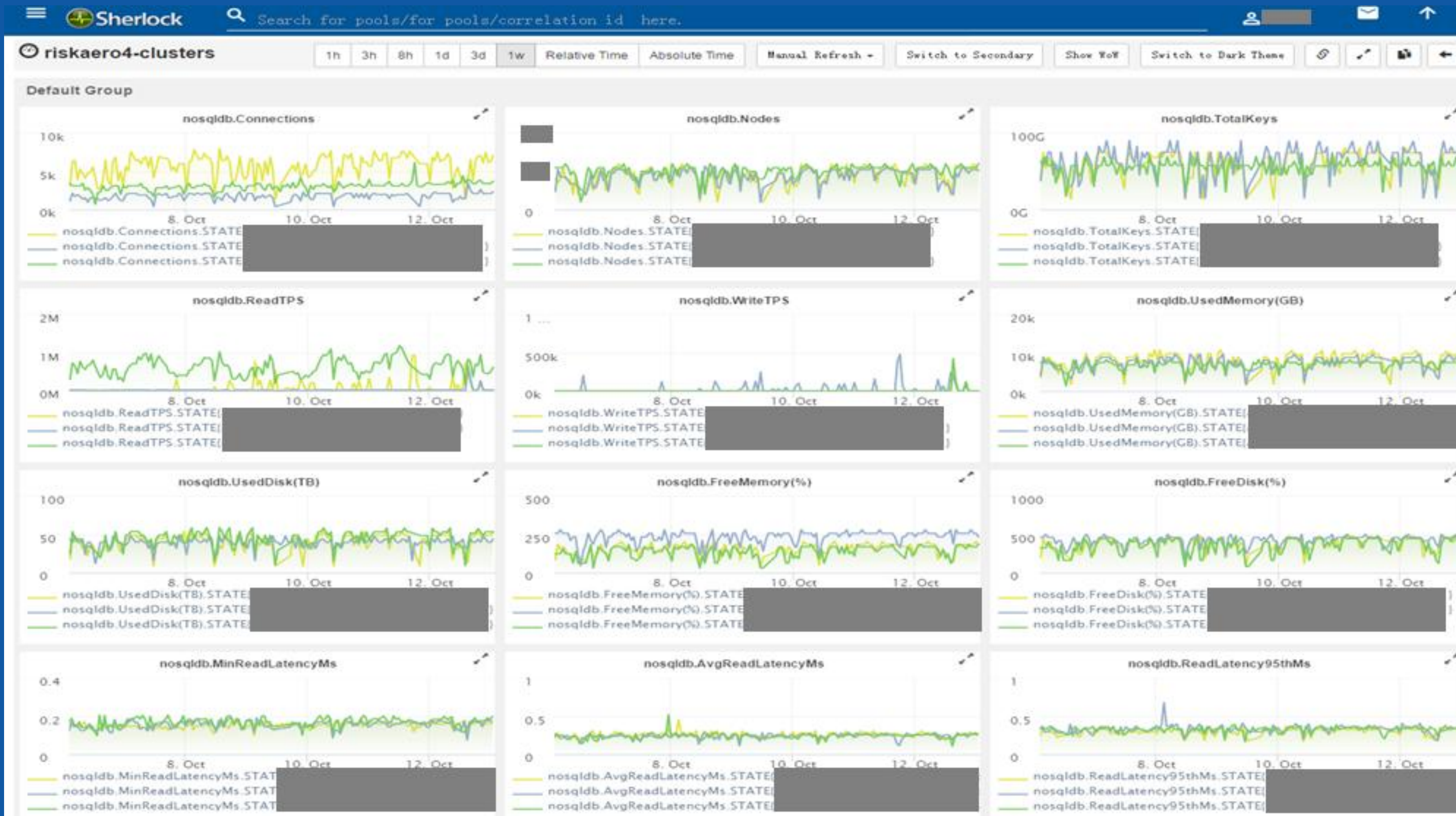
- Share-nothing architecture
- Best performance among the other options
- User defined function support
- Native XDR support

Aerospike Online Performance



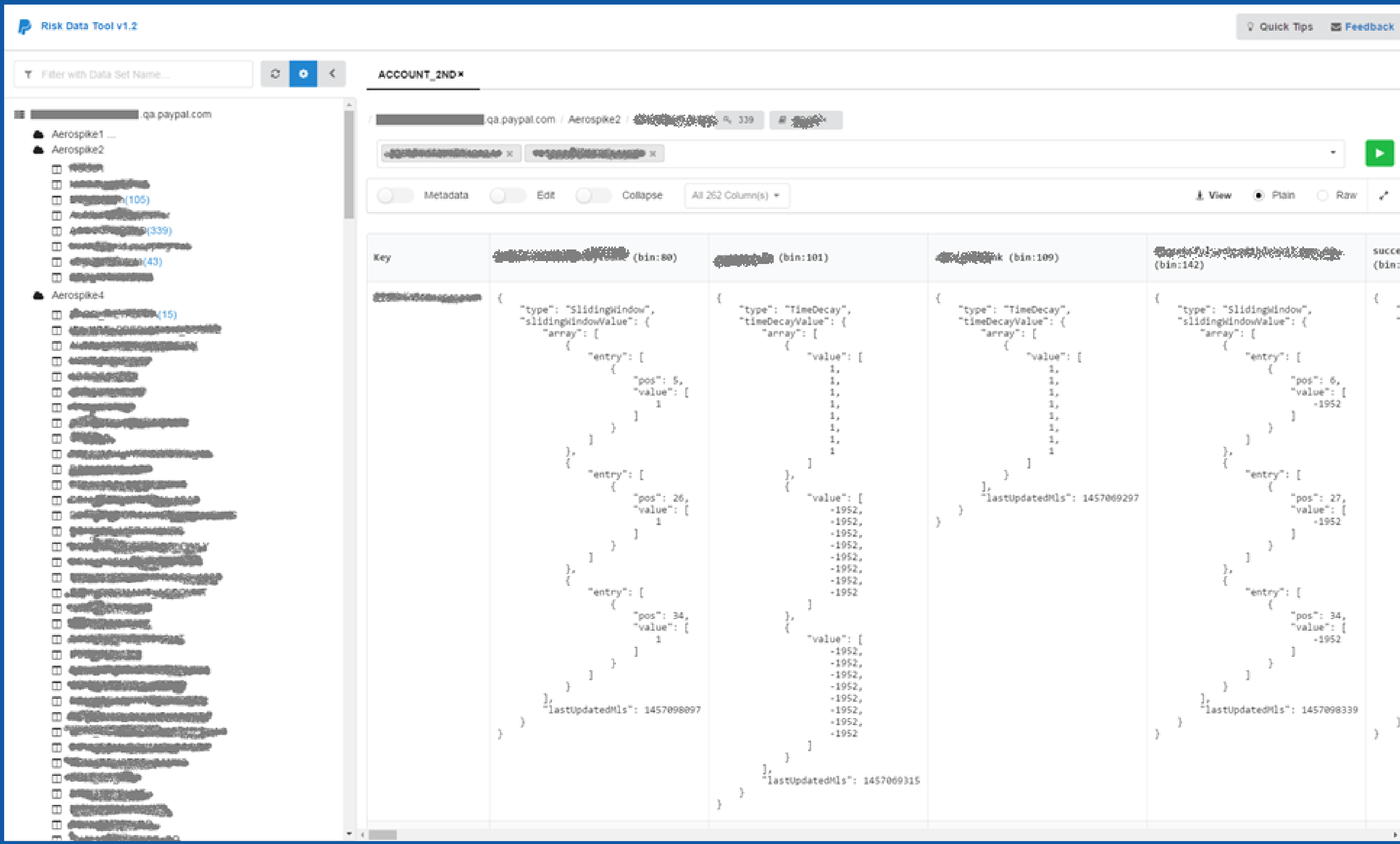
Name	Colo	Count	Failure Count	Failure %	Min (ms)	Max (ms)	Avg (ms)	Median	StdDev	95.0 %ile(ms)	99.0 %ile(ms)
[redacted]get	slca	68	0	0	0	7	0.25	0	0.95	1.12	4.32
[redacted]get	slca	26	0	0	0	4	0.31	0	0.87	1.75	3.50
aero.batch[redacted]	slca	117,596,709	2,192	0.00	0	585	1.19	1	1.08	2	3.00
aero.batch[redacted]	slca	194,673,404	2,214	0.00	0	943	0.94	1	0.99	2	3
[redacted]get	slca	10,072	0	0	0	29	0.18	0	0.77	1	2.90
[redacted]get	slca	3,424	8	0.23	0	101	0.45	0	5.56	1	2.79
[redacted]get	slca	4,030	10	0.25	0	101	0.42	0	5.29	1	2.43
[redacted]get	slca	202	0	0	0	4	0.15	0	0.52	1	2.33
[redacted]get	slca	4,238	1	0.02	0	99	0.19	0	1.67	1	2.07
[redacted]get	slca	35,991	0	0	0	63	0.13	0	0.63	1	2
[redacted]get	slca	510,543	0	0	0	157	0.19	0	0.64	1	2
[redacted]get	slca	132,398	0	0	0	62	0.15	0	0.62	1	2
[redacted]get	slca	1,048,903	7	0.00	0	137	0.13	0	0.64	1	2
[redacted]get	slca	130,449	0	0	0	58	0.15	0	0.50	1	2
[redacted]get	slca	14,119	0	0	0	14	0.19	0	0.49	1	2
[redacted]get	slca	484,392	2	0.00	0	101	0.13	0	0.59	1	2
[redacted]get	slca	552,563,383	1,062	0.00	0	821	0.13	0	0.59	1	2
[redacted]get	slca	34,260	1	0.00	0	101	0.13	0	1.00	1	2
[redacted]get	slca	31,854	0	0	0	87	0.12	0	0.66	1	2
[redacted]get	slca	438,010	28	0.01	0	104	0.13	0	1.02	1	2
[redacted]get	slca	36,116	5	0.01	0	102	0.15	0	1.32	1	2
[redacted]get	slca	162,242	2	0.00	0	100	0.17	0	0.70	1	2
[redacted]get	slca	35,039,966	118	0.00	0	436	0.18	0	0.67	1	2
[redacted]get	slca	896,282	2	0.00	0	108	0.13	0	0.60	1	2
[redacted]get	slca	17,835,743	68	0.00	0	383	0.16	0	0.62	1	2

Aerospike Online Performance – Cont.



ATB Improvement with Aerospike Adoption

Web Data Query Tool



Experiences Learnt from Aerospike Migration

- Control the size for a single Aerospike cluster
- Balance the data density for each node
- UDF should be simple & fast enough
- Leverage asynchronous client
- Provide different tools for easier user adoption
- Build benchmark tool & automate at the beginning

Agenda

1、Paypal Risk Data Challenges
2、Considerations on Risk Storage
3、Generic KV Storage Data Access Solution
4、Experiences Learnt from Aerospike Migration
5、Q&A

THANKS

ArchSummit 全球架构师峰会 2016

[北京站]

主办方 Geekbang 极客邦科技 InfoQ