AQUAQ Analytics

Common Design Principles for kdb+ Gateways

Experts in fast data solutions

for demanding environments

AQUAQ Analytics

- Established in 2011
- Headquarters in Belfast, N.Ireland
- Headcount of 160 staff
- 2016 US Subsidiary launched
- 2018 Singapore subsidiary launch
- 2020 Hong Kong subsidiary launch

What do we do?

Technology Consultancy Services

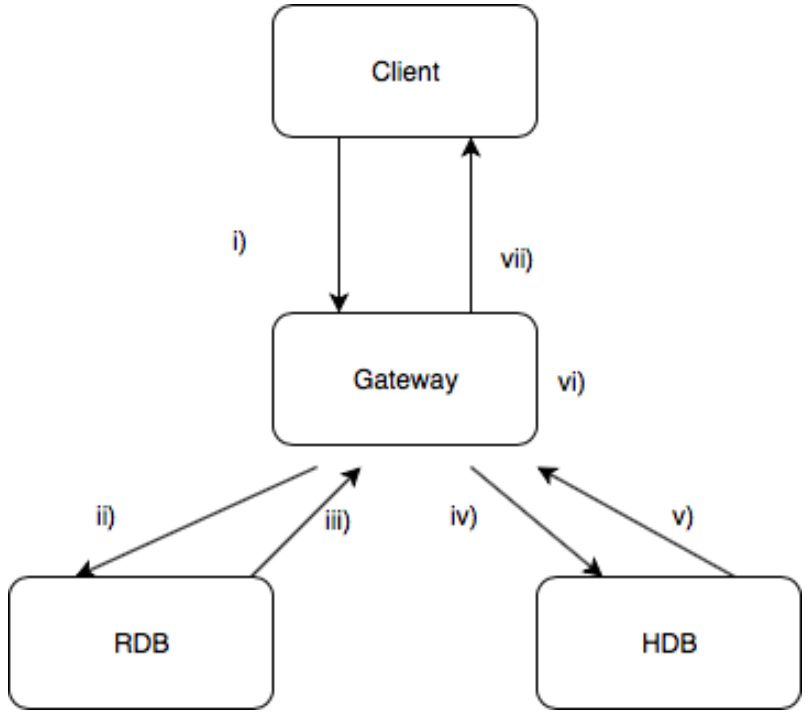
Altair Panopticon Professional Services

Remote (24/7) Support Centre of Excellence

AQUAQ Analytics

What is a gateway?

- Component for controlling process access, performing business logic and stitching datasets
- Normally includes a function to access the data and perform queries in the most efficient manner



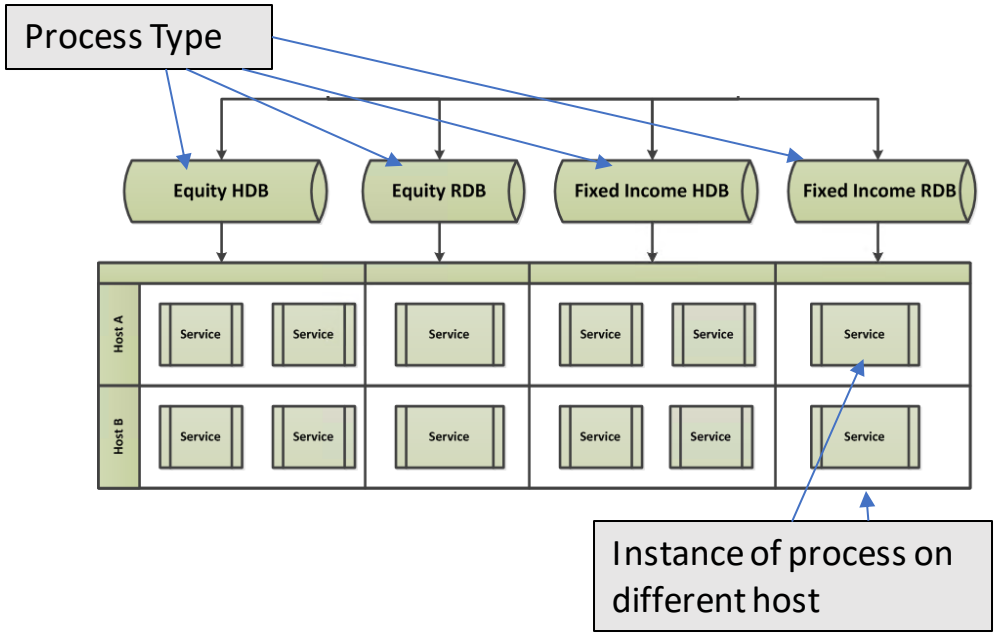
https://kx.com/blog/kdb-q-insights-deferred-response/

AQUAQ Analytics

Why are gateways necessary?

AQUAQ Analytics

Abstraction

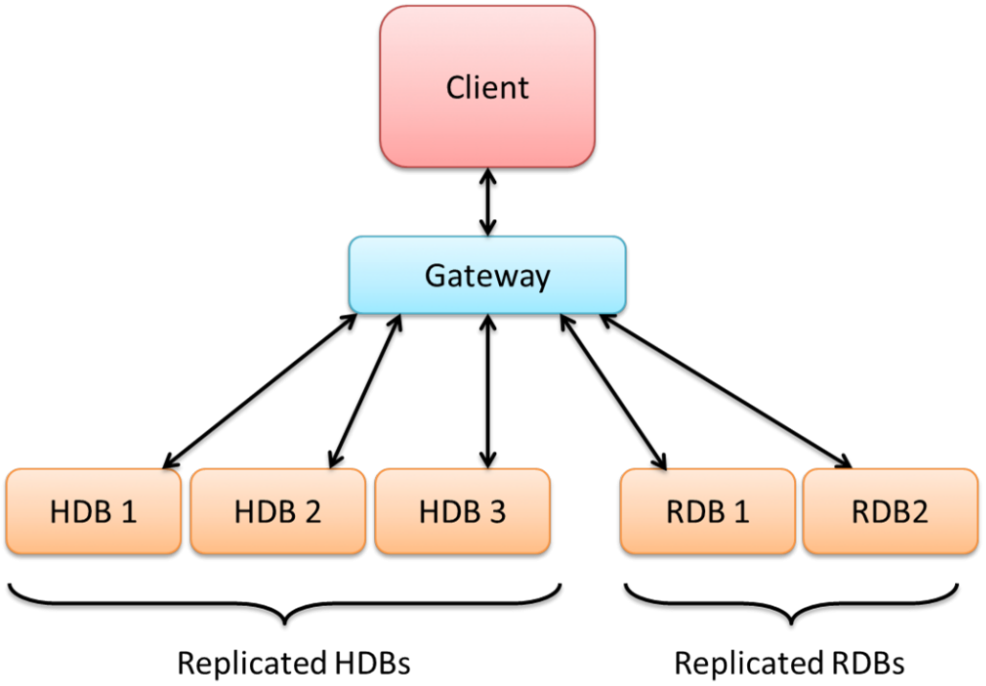


- Often complex process architecture
- Different hosts, different locations
- Mix of historical and real-time
- Potentially cross-asset
- Various methods for storing same-asset data
  - On-disk RDBs
  - Segmented HDBs
  - Streaming services

https://code.kx.com/q/wp/query-routing/#technical-overview

AQUAQ Analytics

Simplification



https://www.aquaq.co.uk/q/kdb-gateways/

- Single entry point
  - One function e.g. getData, not one gateway
- Dictionary of arguments describing data to fetch:
  - Data type: `trades or `quotes
  - Date range: `startTime`endTime
  - Symbols: `VOD`AAPL
  - Filters: (>;`qty;100f)
  - GroupBy: `exchange
  - Aggregations: (sum;`notional)
  - …

AQUAQ Analytics

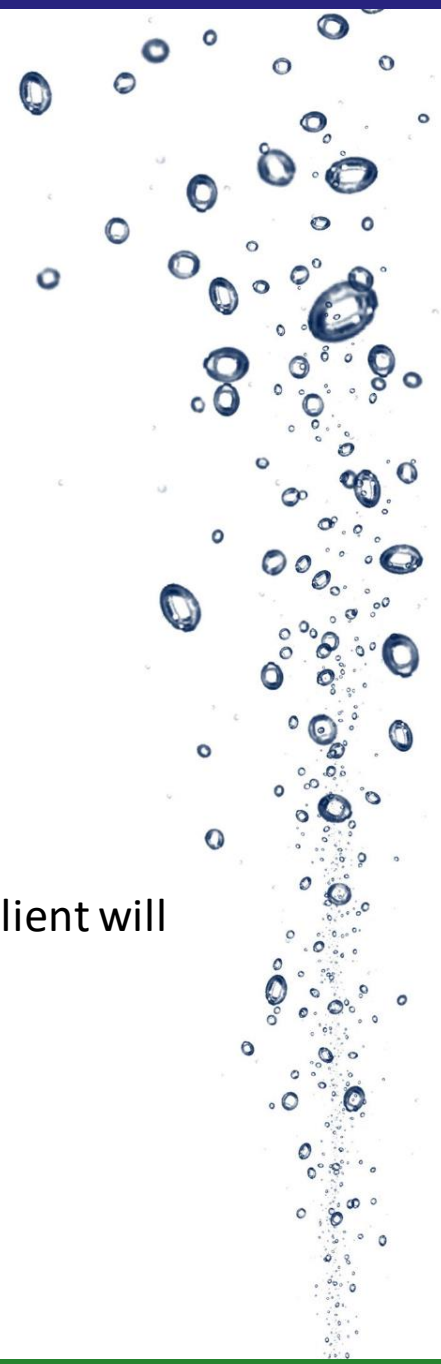
What do we need to consider?

1. Number of client queries, query types
2. Number of processes and sites
3. Requirement for data aggregation
4. Level of redundancy and failover

AQUAQ Analytics

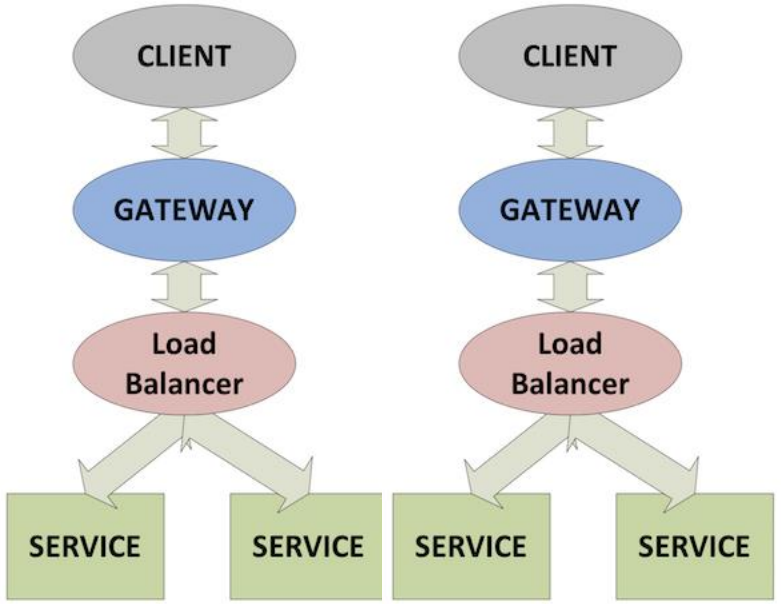
Tech #1: Load Balancing

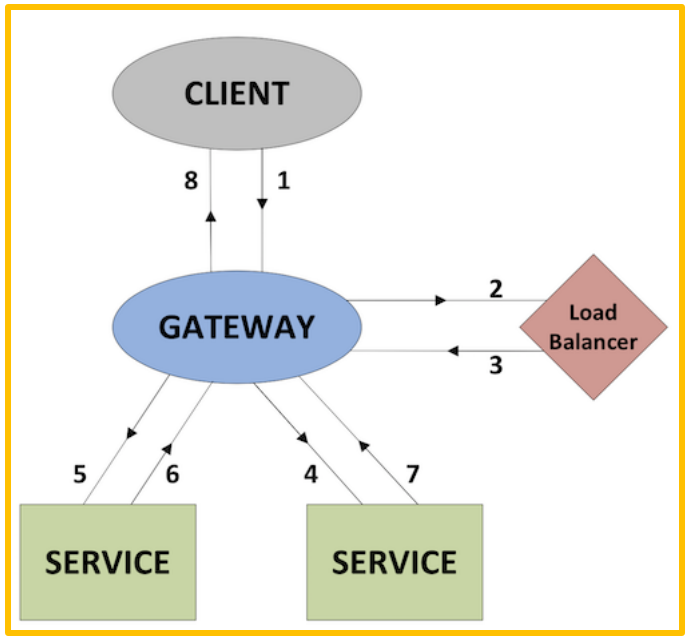
- Utilizing resources in the most efficient way possible
- Always send user queries via the “most available” route:
  - Random
  - Round-robin among available resources
  - Routed based on query type/user tiering
- 2 modes of operation:
  - As a pass-through (embedded load-balancer)
  - As a connection manager (separate load-balancer)
- Client queries must be asynchronous (or deferred sync), otherwise all operations beneath the client will also need to be synchronous



AQUAQ Analytics

Embedded vs. Separate Load Balancing

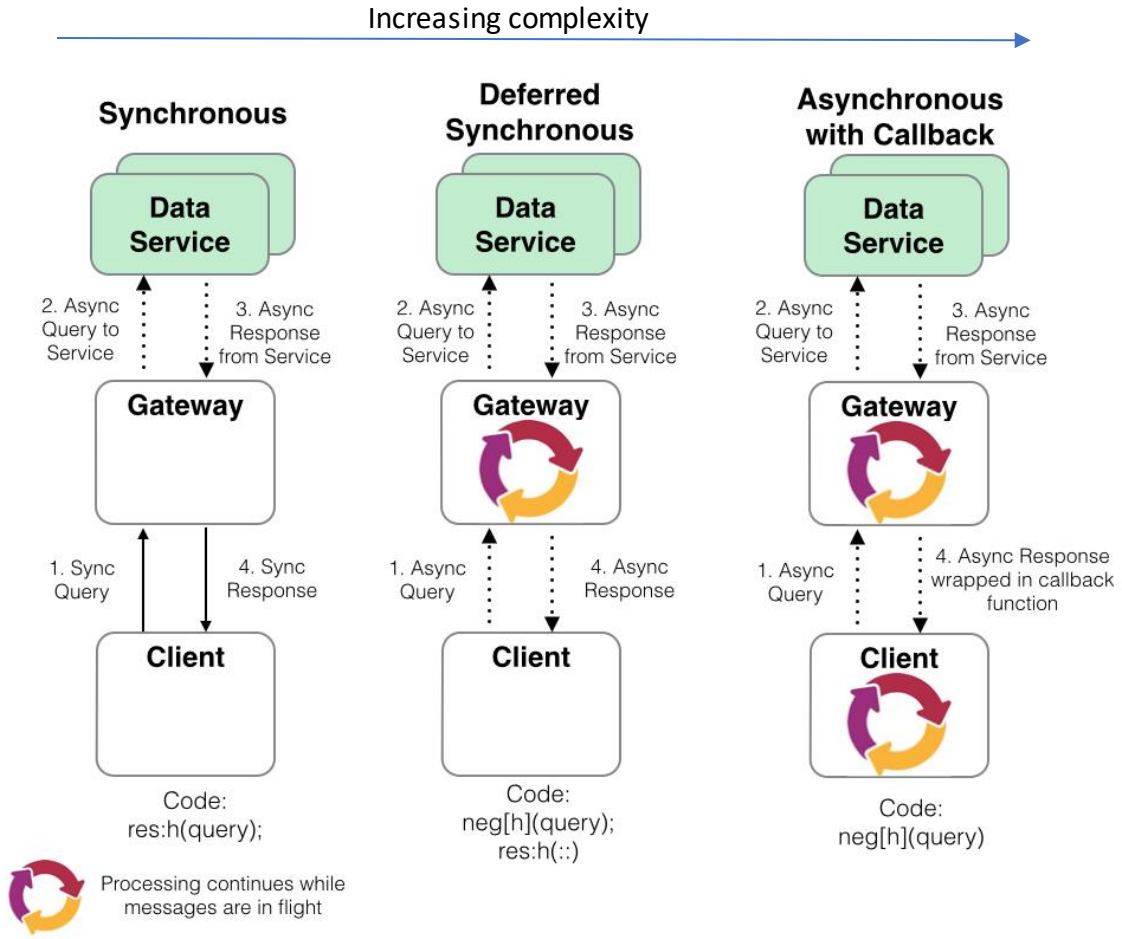




https://code.kx.com/q/wp/gateway-design/#gateway-design

AQUAQ Analytics

Tech #2: Interprocess Communication



- Synchronous client calls
  - Gateway, services are blocked
- Asynchronous client calls
  - Allows concurrent requests
  - More complex architecture
  - Callback on client side
- Deferred Synchronous
  - Client sends async and blocks
- Deferred response (v3.6):
  - Result explicitly returned to the user
  - Include -30!x in .z.pg handler
  - Includes error term in signature

https://www.aquaq.co.uk/q/kdb-gateways/

AQUAQ Analytics

Tech #3: Transport and Aggregation

- AIM: minimize the amount of data travelling over IPC
  - Aggregate at the process level where possible
  - For complex aggs across > 1 process, we need to pass data back to the gateway level
  - Location plays a part, larger overhead passing data from EMEA processes to a gateway sitting in Asia or North America
- Summary stats processes
  - Pre-aggregating metrics, grouped by the most (feasible) granular level
  - Example: sum notional by symbol, country, exchange allows us to “roll-up” by any subset of that grouping

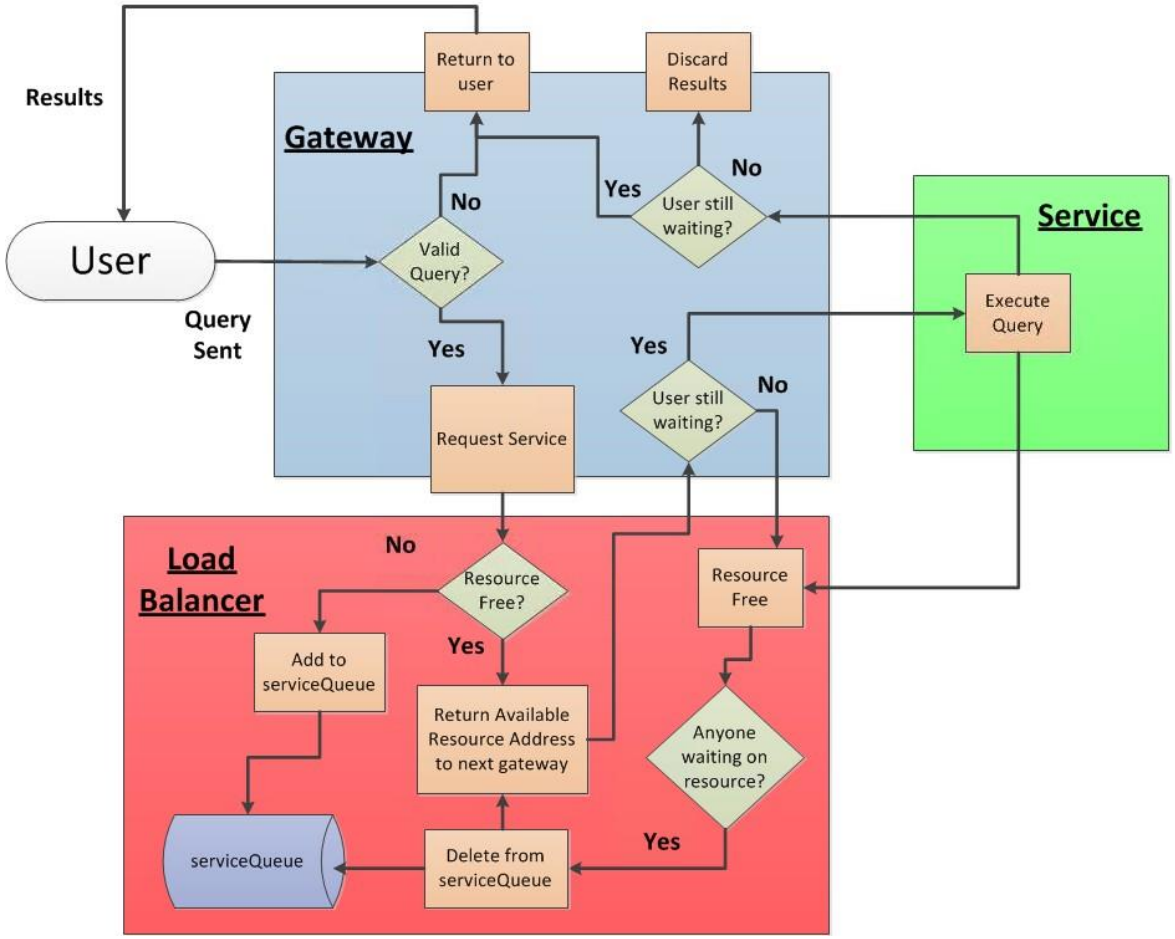
AQUAQ Analytics

Tech #4: Resilience + Availability

- Hot-hot setup vs. hot-cold
- Process replication
- Connection management
  - Dropped connections, dead processes
  - Handlers
  - Timeouts
- User prioritization
- Query categorization

AQUAQ Analytics

Example Implementation



https://code.kx.com/q/wp/query-routing/

AQUAQ Analytics

Additional Features

- Socket sharding of processes like HDBs (kdb+ v3.5)
  - Multiple processes running on the same port (the TCP socket is “sharded” or split)
  - Simple set-up, more direct access to processes, static load balancing
- Introduce a query manager to handle entitlements, basic checks
  - Allows us to have > 1 gateway, increasing scalability
  - Handles deferred response conversion
  - Replicate this query manager as necessary (one per region)
- Increase scalability of process layer
- Streaming gateways
- Parallelization + slaves

AQUAQ Analytics

Thanks!

Q+A

14th May: Grafana and kdb+

21st May: kdb+ 4.0

28th May: kdb+ in Containers