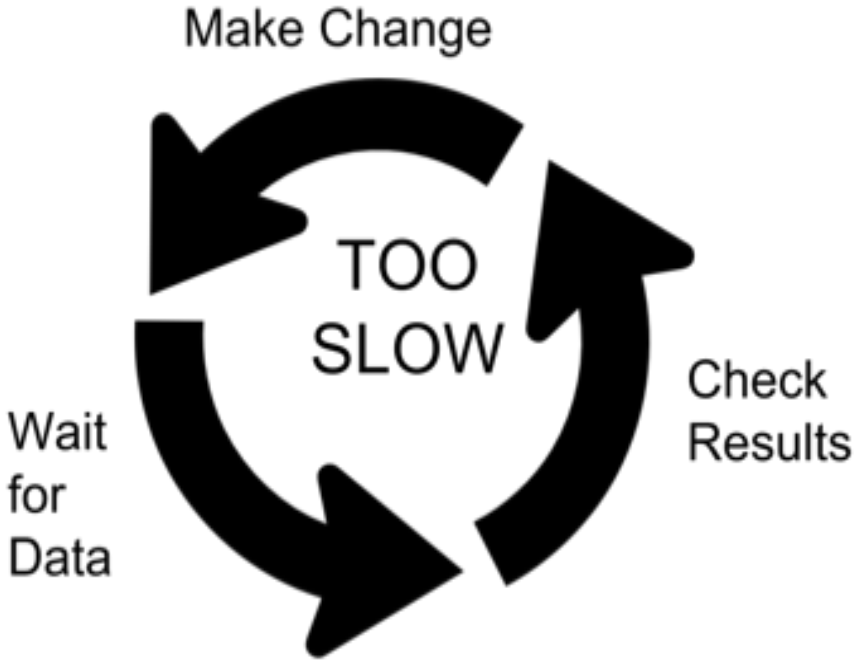
A Real Time Data Query Engine

Akshai Sarma, Nathan Speidel, Michael Natkovich

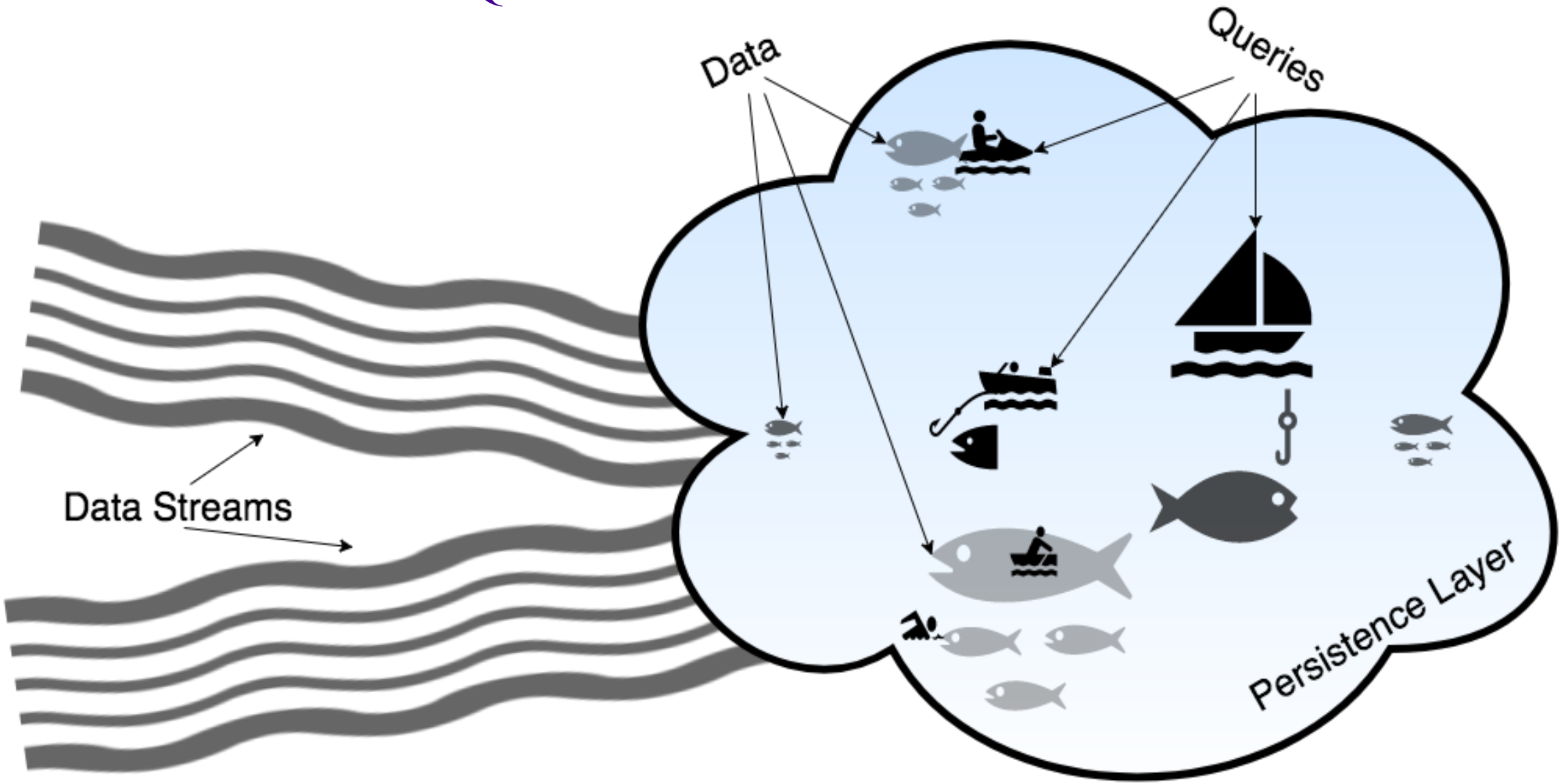
MOTIVATION: INSTRUMENTATION & THE CYCLE OF SADNESS

- Code on web pages/apps to capture usage and behavior
- Drives all data applications
- But validation is unbearably slow
  - Needs to be seconds not minutes/hours
  - Needs to be easy to query
  - Needs programmatic access



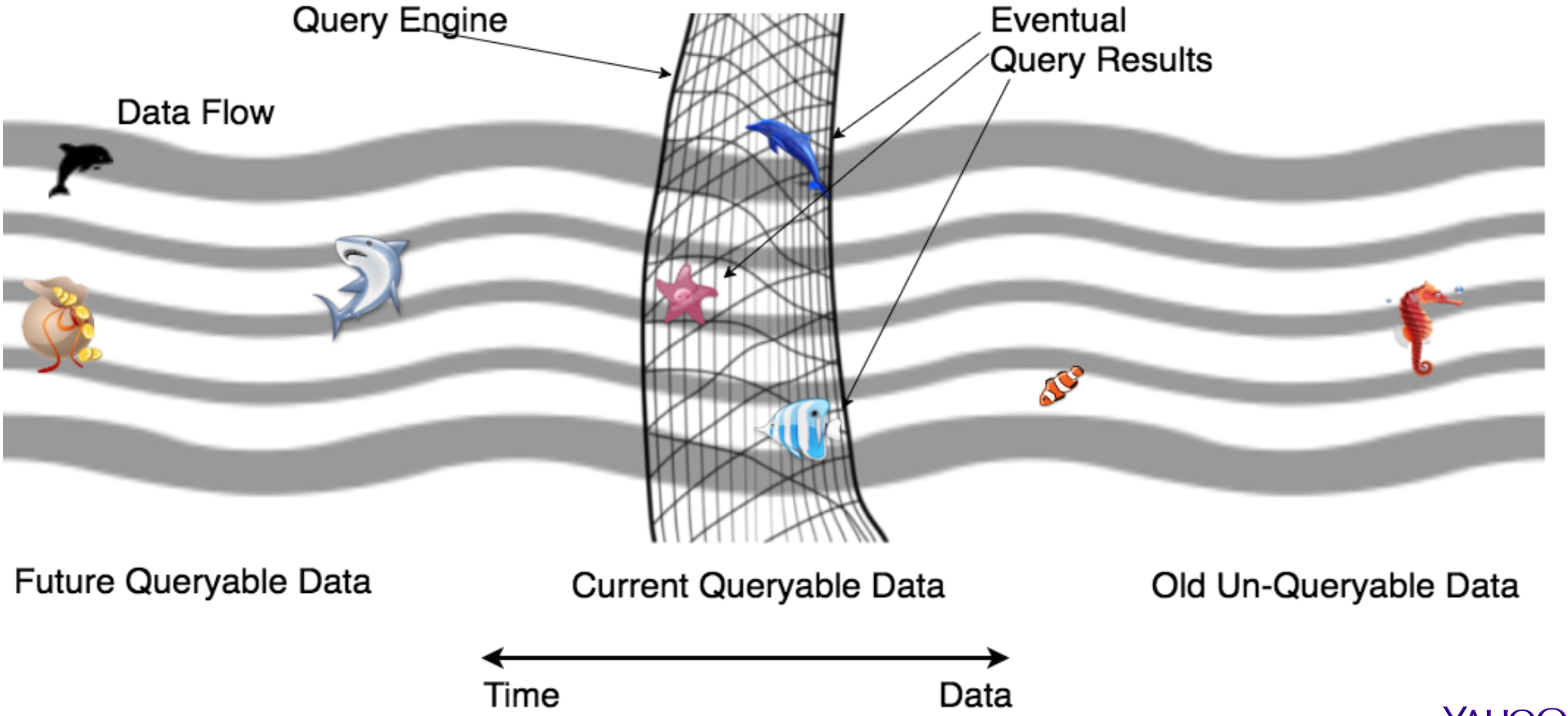
YAHOO!

A TYPICAL WAY OF QUERYING



YAHOO!

ATYPICAL WAY OF QUERYING



YAHOO!

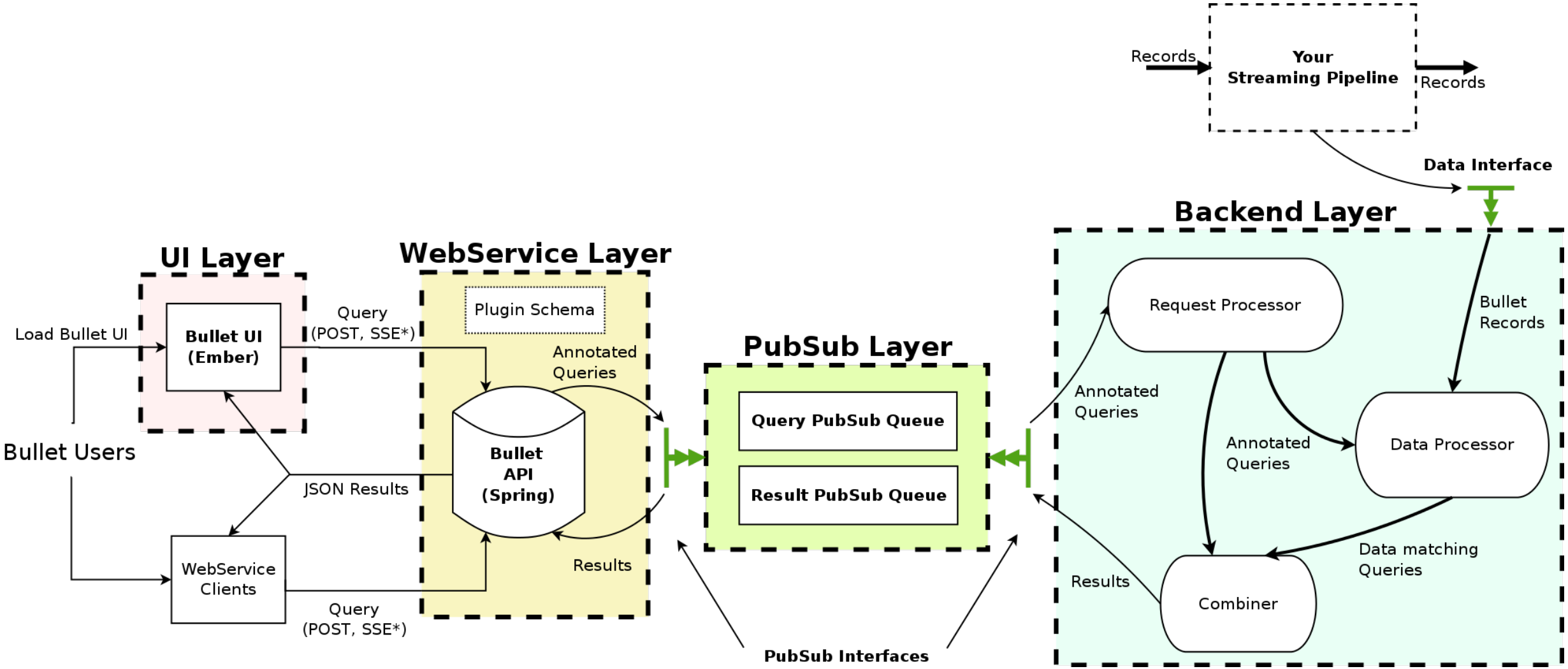
BULLET: LOOK-FORWARD QUERYING

- No persistence: light, fast and scalable!
- UI and API for ad-hoc or programmatic usage
- Pluggable with your streaming data
- Platform agnostic
- Queries run for time windows and can:
  - Filter data: Equals, Regex, And, Or and more
  - Get Raw or Aggregate data: Group By, Count Distinct, Top K, and more

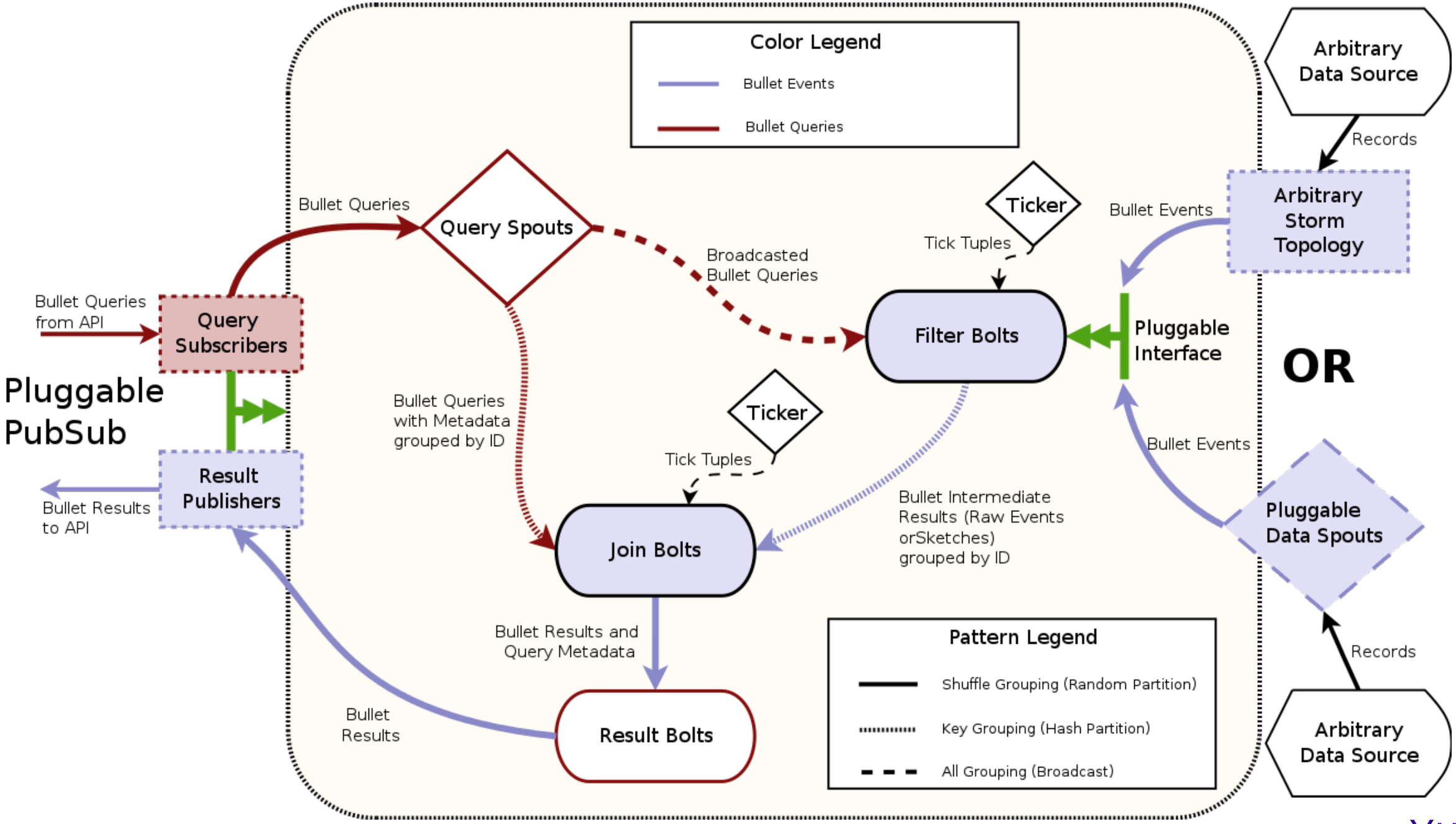
DATASKETCHES

- Sketches are a class of stochastic streaming algorithms
- Provides approximate results (if data is too large)
- Provable error bounds
- Fixed memory footprint
- Mergeable, allowing for parallel processing

ARCHITECTURE



YAHOO!


Color Legend
Bullet Events
Bullet Queries
Arbitrary Data Source
Records
Arbitrary Storm Topology
Bullet Events
Ticker
Tick Tuples
Bullet Queries
Query Spouts
Broadcasted Bullet Queries
Bullet Queries from API
Query Subscribers
Filter Bolts
Pluggable Interface
OR
Pluggable PubSub
Bullet Queries with Metadata grouped by ID
Ticker
Tick Tuples
Result Publishers
Bullet Results to API
Bullet Events
Pluggable Data Spouts
Bullet Intermediate Results (Raw Events orSketches) grouped by ID
Join Bolts
Bullet Results and Query Metadata
Records
Arbitrary Data Source
Bullet Results
Result Bolts
Pattern Legend
Shuffle Grouping (Random Partition)
Key Grouping (Hash Partition)
All Grouping (Broadcast)

YAHOO!

FULLY OPEN-SOURCED

- We are on GitHub - contributions, ideas, feedback welcome!

Component	Repo
Storm	https://github.com/yahoo/bullet-storm
Core	https://github.com/yahoo/bullet-core
API	https://github.com/yahoo/bullet-service
UI	https://github.com/yahoo/bullet-ui
Record	https://github.com/yahoo/bullet-record
Kafka	https://github.com/yahoo/bullet-kafka

LINKS

- Contact us
  - Developers : bullet-dev@googlegroups.com
  - Users : bullet-users@googlegroups.com
- For all Documentation: Usage, Performance, Quick Start, Examples
  - https://yahoo.github.io/bullet-docs
- DataSketches: https://datasketches.github.io

YAHOO!