



Leveraging UBI to enhance Search Relevance



Aswath Srinivasan
aswathns | opensearch.org
September 30, 2024



How is OpenSearch doing?

OpenSearch is 3 years old!



OpenSearch Project

OpenSearch is an Apache 2.0-licensed search and analytics suite comprising OpenSearch, OpenSearch Dashboards, and a suite of plugins providing advanced anomaly detection, alerting, observability, and security analytics



>700MM

OpenSearch project downloads since launch



Top 4 search engine

DB-Engines ranking



80+

Partners and growing



122 GitHub repos

1000 Contributors and 43 non-AWS maintainers



Multiple service providers

AWS, OCI, Aiven-Azure, Bonsai-GCP, AliCloud

Did you see the announcement?



Agenda

The Why

The What – What does UBI provide?

The How – UBI with OpenSearch & beyond

The Art of the Possibilities

The Why

Why **User Behavior** Insights?

UBI's Goal

Why not a off the shelf Click Metrics tool?

The virtuous circle of search relevance

Search processing

- Pipeline processor
 - Modular query processing
 - Query rewriter
 - Modular result processing
 - Relevance integrations
 - Personalization integrations
- Integrated semantic search
- Combining semantic and lexical (BM25)
- Sparse neural

Measurement and analysis

- End-to-end search behavior (UBI)
- Online evaluation tools
- A/B testing
- Creation of judgement sets
- Offline evaluation tools
- Metrics

Search tuning

- Search configurations
- Regression evaluation
- Manual tuning, Bayesian optimization
- Semantic model fine-tuning
- Learning to Rank

The What

What does UBI Provide?

UBI Priorities




Define standardized Schema



Deliver client-side instrumentation



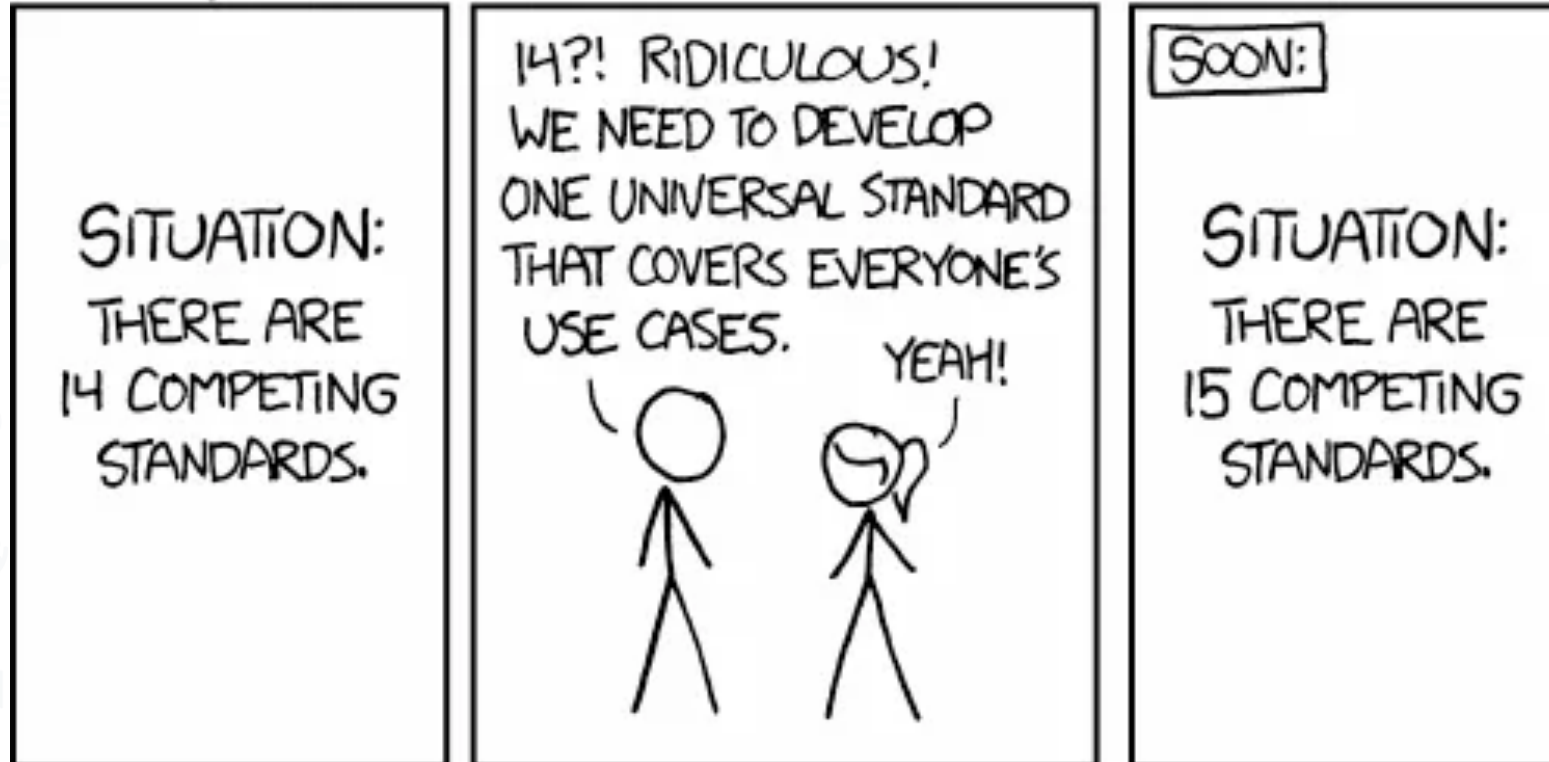
Collect end-to-end search behavior events



Build Dashboard to Visualize, Investigate, and tune

Standardizing the Schema

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



UBI Index schemas

Index

Queries

Index

Events

UBI Query Index

```
1 {
2   "ubi_queries": { Schema
3     "aliases": {},
4     "mappings": {
5       "dynamic": "false",
6       "properties": {
7         "client_id": {
8           "type": "keyword",
9           "ignore_above": 100
10        },
11        "query": {
12          "type": "text"
13        },
14        "query_attributes": {
15          "type": "flat_object"
16        },
17        "query_id": {
18          "type": "keyword",
19          "ignore_above": 100
20        },
21        "query_response_hit_ids": {
22          "type": "keyword"
23        },
24        "query_response_id": {
25          "type": "keyword",
26          "ignore_above": 100
27        },
28        "timestamp": {
29          "type": "date"
30        },
31        "user_query": {
32          "type": "text",
33          "fields": {
34            "keyword": {
35              "type": "keyword",
36              "ignore_above": 256
37            }
38          }
39        }
40      }
41    }
42  }
```

```
855 {
856   "_index": "ubi_queries",
857   "_id": "255",
858   "_score": 1,
859   "_source": {
860     "query_response_id": "121486af-c841-473a-b4c1-6780c04f7663",
861     "user_query": "tablet",
862     "query_id": "c4d15334-4621-456b-9ee9-b437c6ad5f78",
863     "query_response_object_ids": [
864       "0085896679318",
865       "0085896679196",
866       "0808736521253",
867       "0882780564893",
868       "0882861340606",
869       "0065030888028",
870       "4905524608922",
871       "4905524608915",
872       "0092636344559",
873       "5711724162510",
874       "0793442009532",
875       "0085896387008",
876       "0609585191624",
877       "0085896625711",
878       "0085896679332",
879       "0085896679202",
880       "0085896679301",
881       "0085896973201",
882       "0793442009181",
883       "0065030880244",
884       "0085896679325",
885       "0193638288510",
886       "8806088952321",
887       "0193386748021",
888       "5711724010613"
889     ],
890     "query": ""{"size":25,"query":{"query_string":{"query":"(title:\"tablet\"
891       ,\"max_determinized_states\":10000,\"enable_position_increments\":true,\"fuzz
892       :false,\"auto_generate_synonyms_phrase_query\":false,\"fuzzy_transpositions
893       :\"e1e49e1-066a-4251-b70e-3bee1dbec51f\",\"object_id_field\":\"primary_ean\",
894       \"query_attributes\": {
895         \"application\": \"ubi-demo\"
896       }
897     }
898   }
```

UBI Events Index

```
1- [
2-   "ubi_events": {
3-     "aliases": {},
4-     "mappings": {
5-       "properties": {
6-         "action_name": {
7-           "type": "keyword",
8-           "ignore_above": 100
9-         },
10-        "application": {
11-          "type": "keyword",
12-          "ignore_above": 256
13-        },
14-        "client_id": {
15-          "type": "keyword",
16-          "ignore_above": 100
17-        },
18-        "event_attributes": { },
387-        "message": {
388-          "type": "keyword",
389-          "ignore_above": 1024
390-        },
391-        "message_type": {
392-          "type": "keyword",
393-          "ignore_above": 100
394-        },
395-        "query_id": {
396-          "type": "text",
397-          "fields": {
398-            "keyword": {
399-              "type": "keyword",
400-              "ignore_above": 256
401-            }
402-          }
403-        },
404-        "timestamp": {
405-          "type": "date",
406-          "format": "rfc3339_lenient|epoch_millis",
407-          "ignore_malformed": true
408-        }
409-      }
410-    }
  ],
```

Schema

```
17- {
18-   "_index": "ubi_events",
19-   "_id": "258",
20-   "_score": null,
21-   "_source": {
22-     "action_name": "item_click",
23-     "client_id": "eb1e49e1-066a-4251-b70e-3bee1dbec51f",
24-     "query_id": "c4d15334-4621-456b-9ee9-b437c6ad5f78",
25-     "timestamp": 1717527764188,
26-     "message_type": "INFO",
27-     "event_attributes": { },
67-     "message": "crece iniciativas escalables",
68-     "application": "ubi-demo"
69-   },
```

Sample document

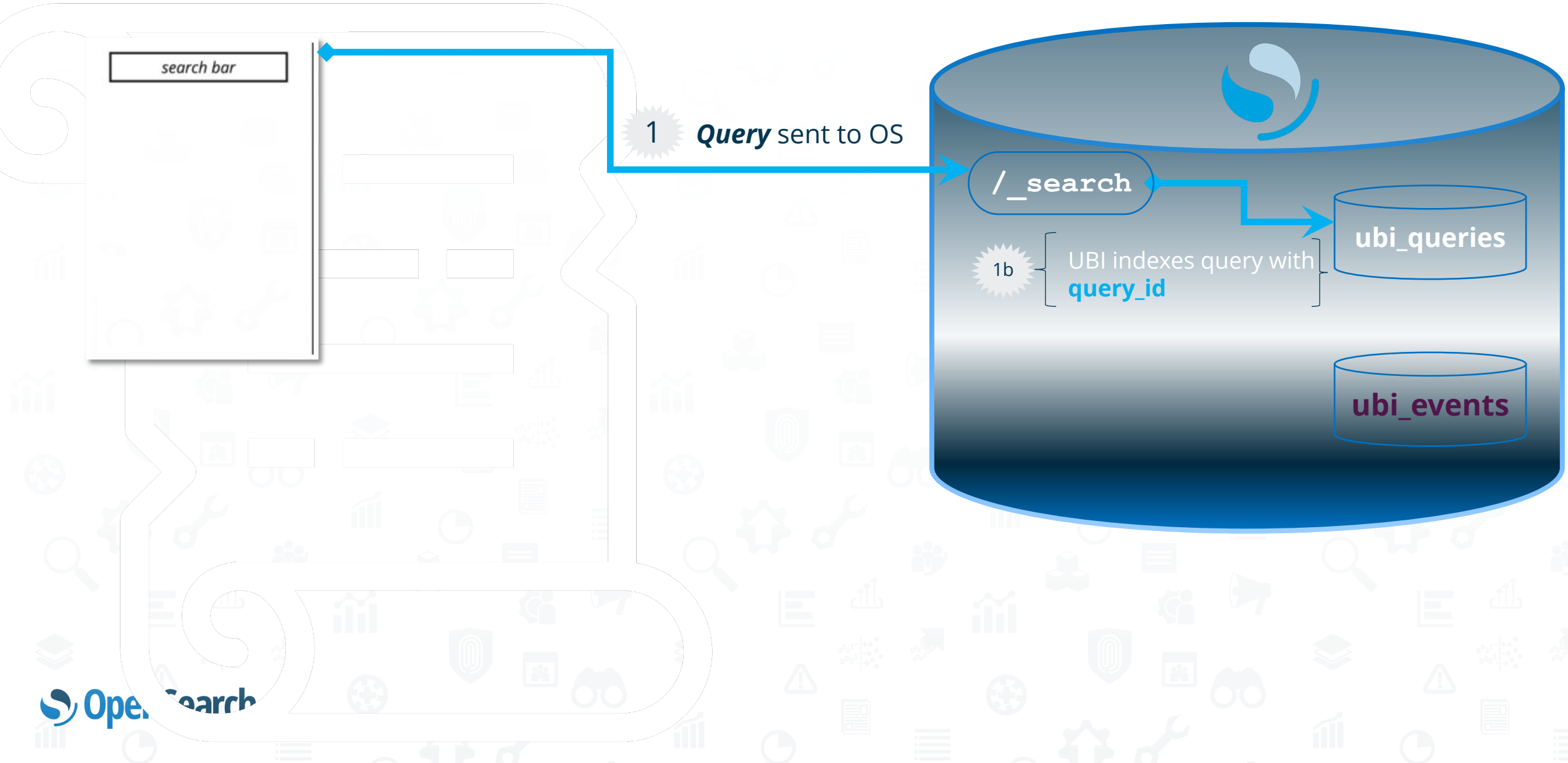
```
curl -s -X POST http://localhost:9200/ubi_events/_doc/ -H "Content-Type: application/json" -d'
{
  "action_name": "page_exit",
  "user_id": "1821196507152684",
  "query_id": "00112233-4455-6677-8899-aabccddeeff",
  "session_id": "c3d22be7-6bdc-4250-91e1-fc8a92a9b1f9",
  "page_id": "/docs/latest/",
  "timestamp": "2024-05-16T12:34:56.789Z",
  "message_type": "INFO",
  "message": "On page /docs/latest/ for 3.35 seconds",
  "event_attributes": {
    "position": {},
    "object": {
      "idleTimeoutMs": 5000,
      "currentIdleTimeMs": 250,
      "checkIdleStateRateMs": 250,
      "isUserCurrentlyOnPage": true,
      "isUserCurrentlyIdle": false,
      "currentPageName": "http://localhost:4000/docs/latest/",
      "timeElapsedCallbacks": [],
      "userLeftCallbacks": [],
      "userReturnCallbacks": [],
      "visibilityChangeEventName": "visibilitychange",
      "hiddenPropName": "hidden"
    }
  }
}'
```

cURL POST

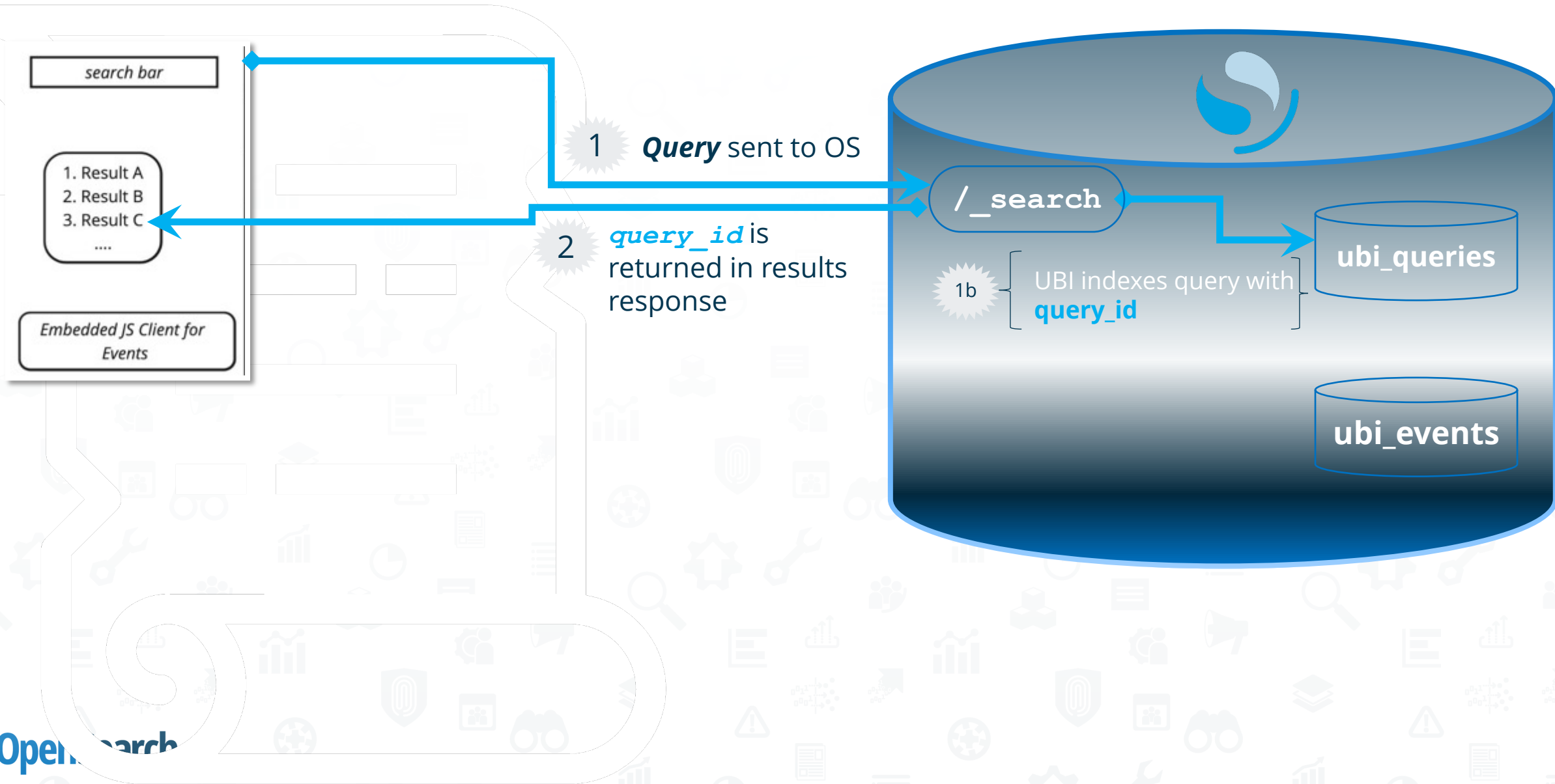
UBI with OpenSearch

... & beyond

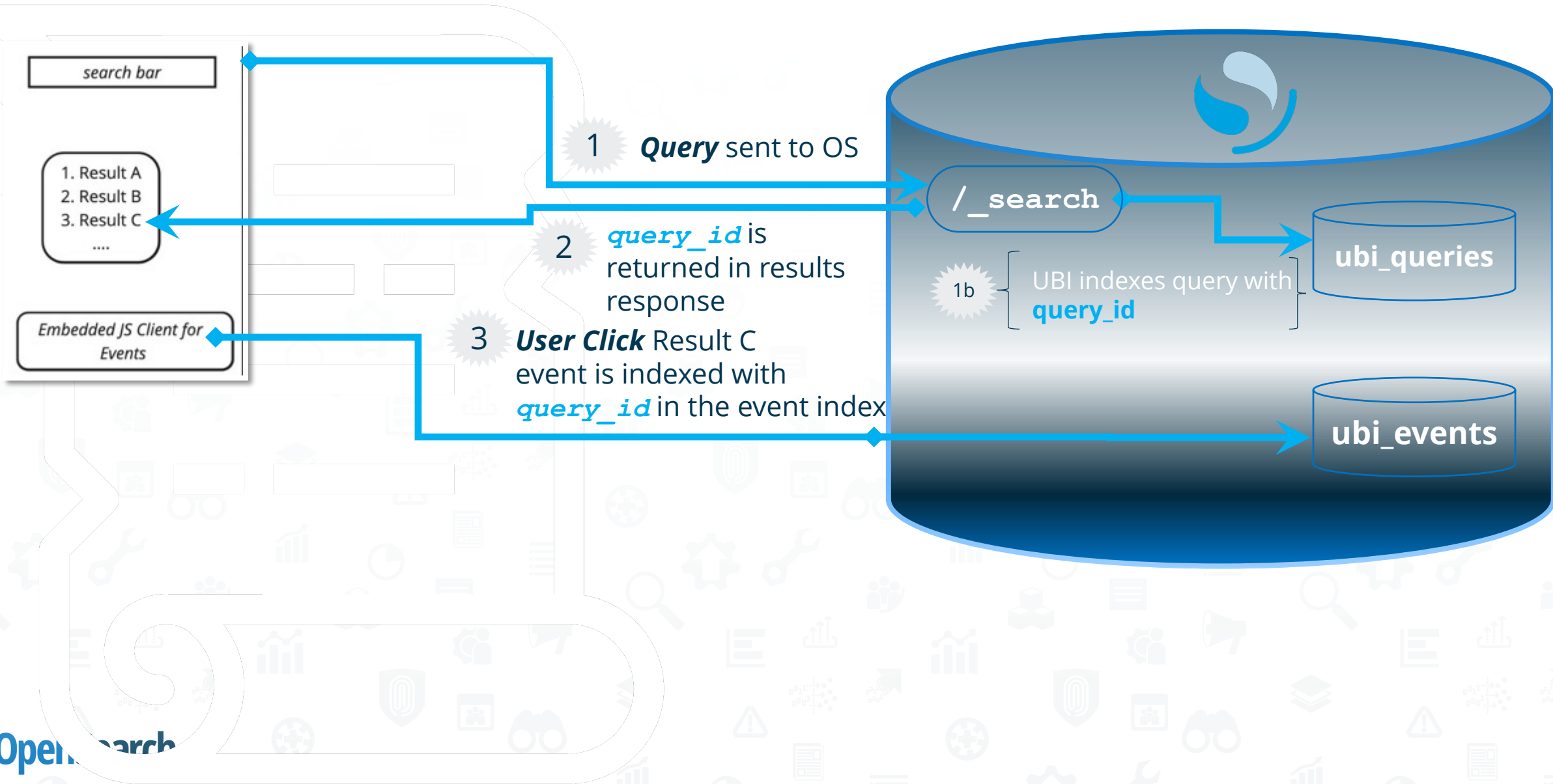
UBI data flow for OpenSearch



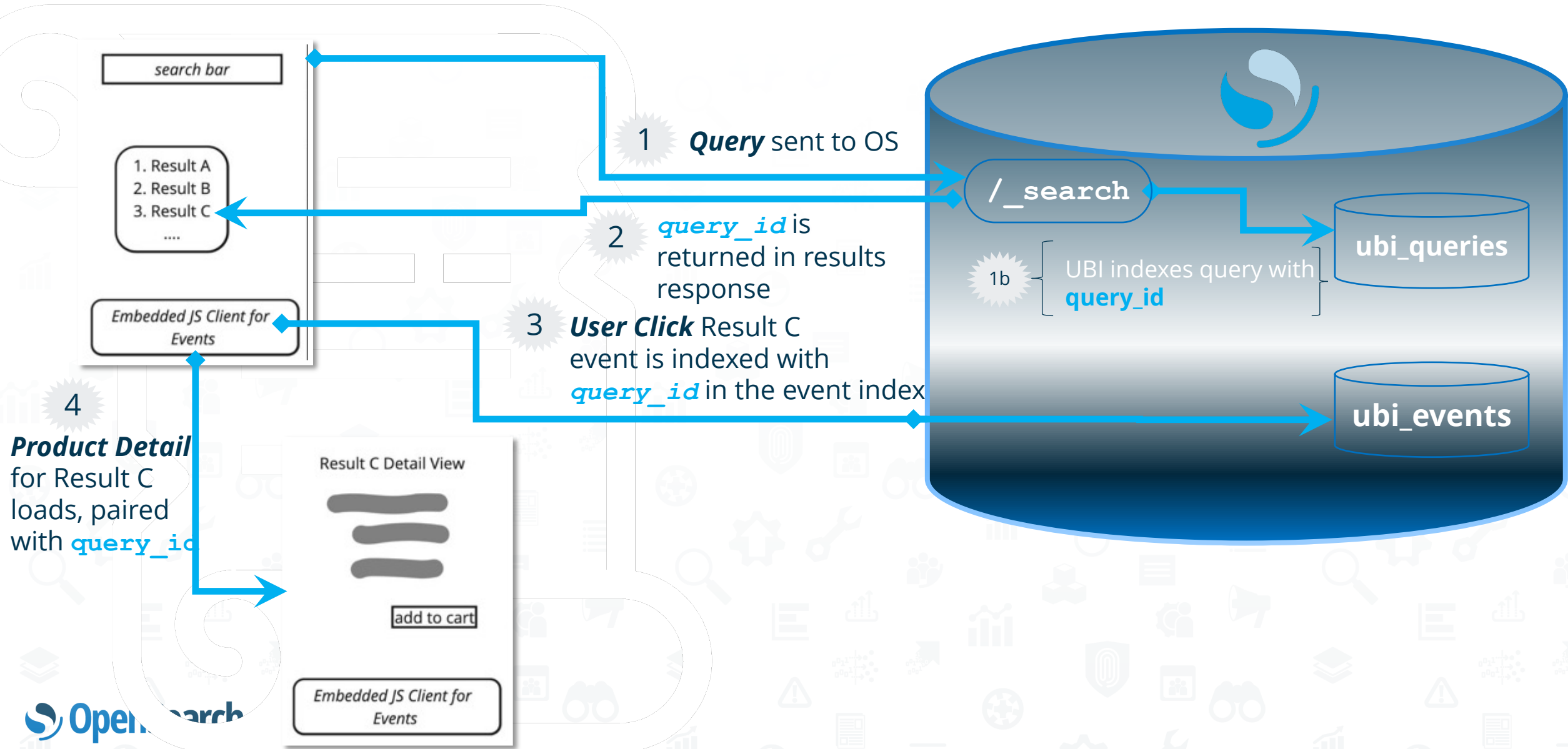
UBI data flow for OpenSearch



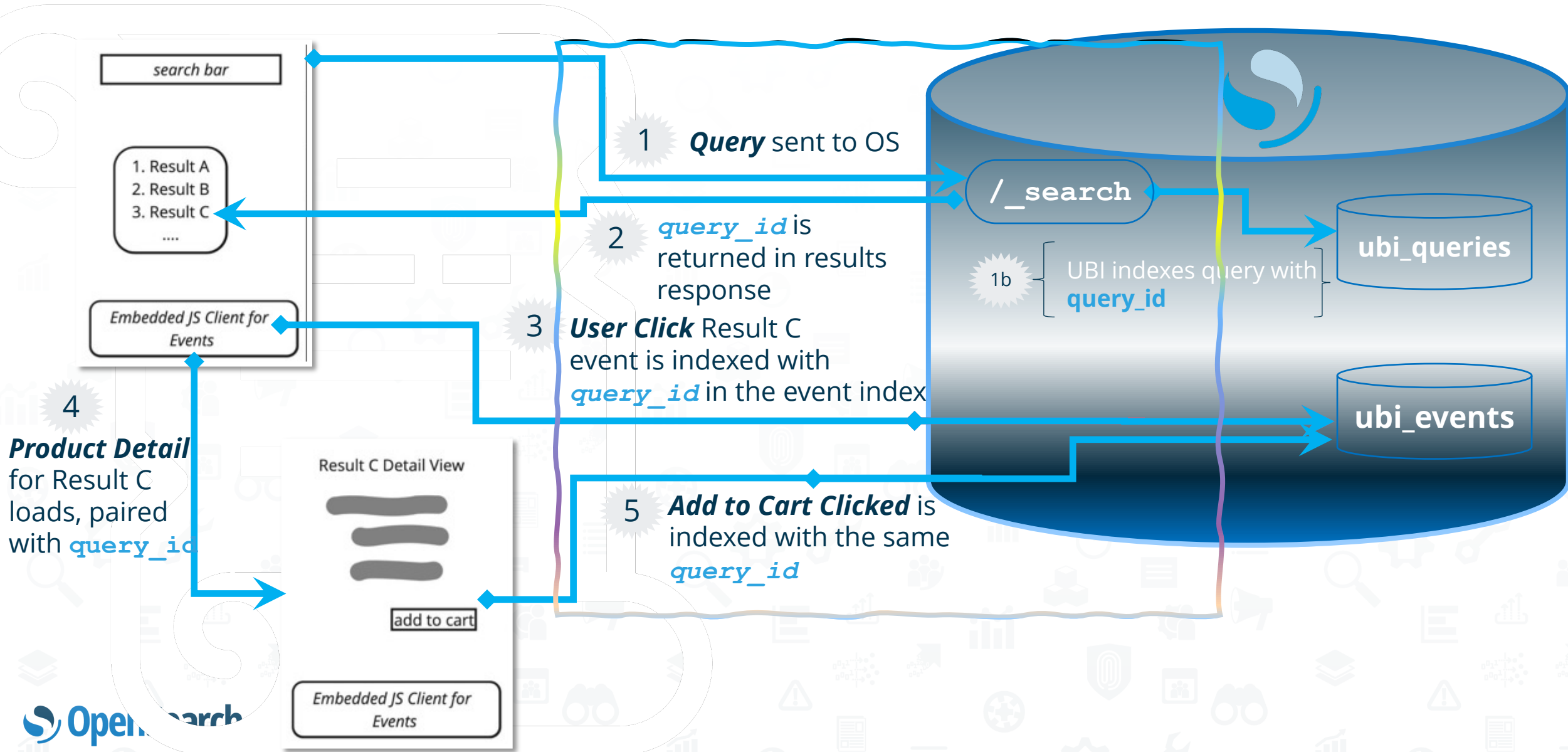
UBI data flow for OpenSearch



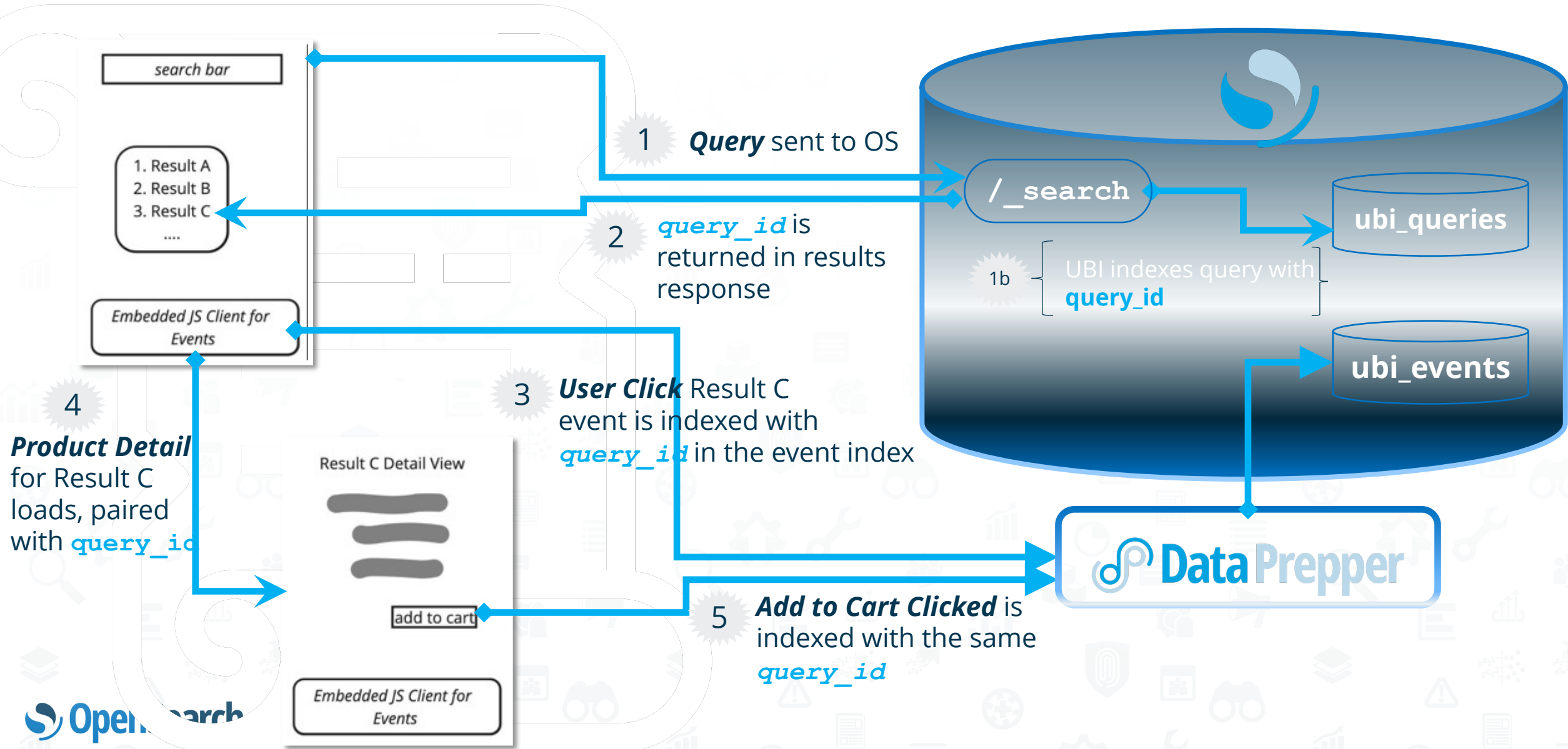
UBI data flow for OpenSearch



UBI data flow for OpenSearch



UBI data flow for OpenSearch



...and Beyond!

with a quick demo



Chorus for OpenSearch

The Art of the Possibilities

Secondary re-Ranker

Learning to Rank
Learning to Boost
Amazon Personalize
MetaRank

Weightage for hybrid search

Common Denominator

Judgement list

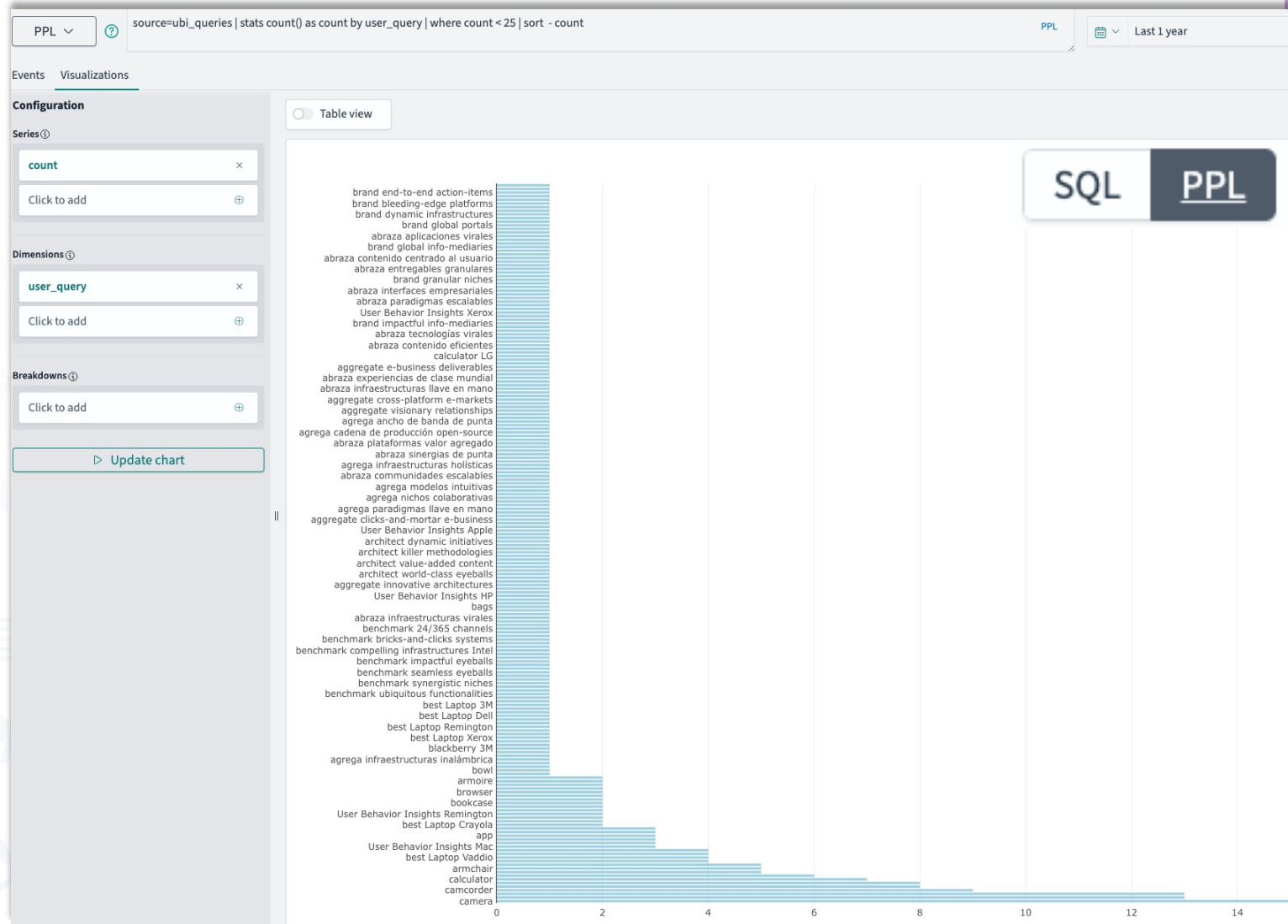
UBI based Judgement list

```
1  SELECT
2     q.user_query,
3     q.query_id,
4     e.query_response_hit_ids AS document_id,
5     CASE
6         WHEN COUNT(DISTINCT e.action_name) = 3 THEN 3
7         WHEN COUNT(DISTINCT e.action_name) = 2 THEN 2
8         WHEN COUNT(DISTINCT e.action_name) = 1 THEN 1
9         ELSE 0
10    END AS grade
11 FROM
12     ubi_queries q
13 JOIN
14     ubi_events e ON q.query_id = e.query_id
15 WHERE
16     e.action_name IN ('item_click', 'add_to_cart', 'purchase')
17 GROUP BY
18     q.user_query, q.query_id, e.query_response_hit_ids
19 ORDER BY
20     grade DESC, q.user_query ASC
21 LIMIT 100
```

24	+				
25		user_query	query_id	document_id	grade
26	+				
27		laptop	query_123456789	doc_987654321	3
28		laptop	query_123456789	doc_246813579	2
29		laptop	query_123456789	doc_135792468	1
30		laptop	query_123456789	doc_369258147	0
31		laptop	query_123456789	doc_159753486	0
32		smartphone	query_987654321	doc_123456789	3
33		smartphone	query_987654321	doc_369258147	2
34		smartphone	query_987654321	doc_246813579	0
35		smartphone	query_987654321	doc_135792468	0
36		headphones	query_246813579	doc_135792468	2
37		headphones	query_246813579	doc_147258369	1
38		headphones	query_246813579	doc_987654321	0
39		headphones	query_246813579	doc_369258147	0
40		smartwatch	query_135792468	doc_246813579	3
41		smartwatch	query_135792468	doc_987654321	1
42		smartwatch	query_135792468	doc_123456789	0
43		smartwatch	query_135792468	doc_369258147	0
44		tablet	query_369258147	doc_369258147	2
45		tablet	query_369258147	doc_123456789	1
46		tablet	query_369258147	doc_246813579	0
47		tablet	query_369258147	doc_135792468	0
48	+				

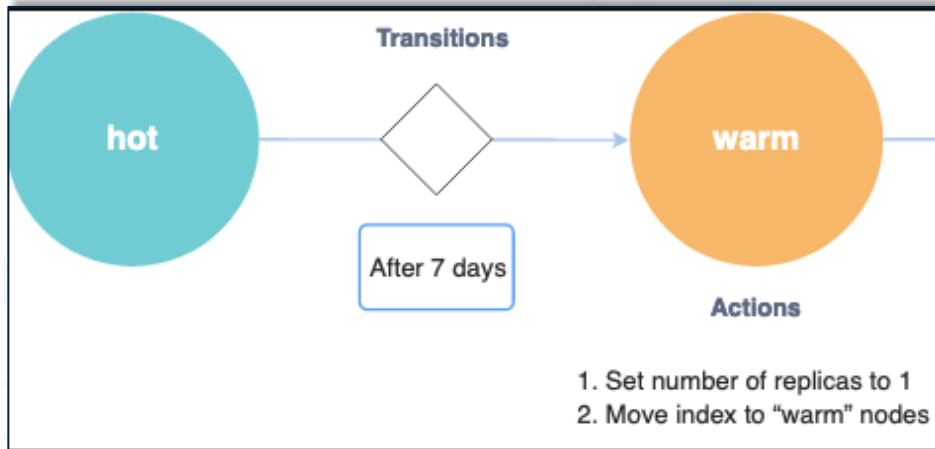
Long Tail queries

```
SQL PPL
1 SELECT user_query, COUNT(*) AS count
2 FROM ubi_queries
3 GROUP BY user_query
4 HAVING COUNT(*) < 50
5 ORDER BY count ASC;
```



Archive UBI data – Save Cost

Warm Nodes



Rollups

SQL PPL [SQL documentation](#)

```
1 SELECT
2 application_terms,
3 action_name_terms,
4 message_terms AS search_terms,
5 message_type_terms,
6 transform._doc_count AS doc_count
7 FROM
8 ubi_events_transform
9 ORDER BY
10 action_name_terms DESC, doc_count DESC;
```

Index	docs.count	pri.store.size
ubi_events	26828	13.3mb
ubi_events_transform	9636	826.2kb

Run Clear Explain

Results

Output [Full screen view](#)

(200) [Download](#)

application_terms	action_name_terms	search_terms	message_type_terms	doc_count
ubi-demo	view_search_results	reinventa relaciones personalizado	INFO	8
ubi-demo	view_search_results	facilita sinergias open-source	INFO	7
ubi-demo	view_search_results	transform back-end experiences	INFO	7
ubi-demo	view_search_results	visualiza esquemas virtuales	INFO	7
ubi-demo	view_search_results	enable value-added solutions	INFO	6
ubi-demo	view_search_results	escala plataformas centrado al usuario	INFO	6
ubi-demo	view_search_results	optimiza arquitecturas colaborativas HP	INFO	6
ubi-demo	view_search_results	re-intermediate web-enabled action-items	INFO	6
ubi-demo	view_search_results	reinventa mercados en tiempo real	INFO	6
ubi-demo	view_search_results	sintetiza soluciones open-source	INFO	6

Rows per page: 10 < 1 2 3 4 5 ... 20 >

Insights via SQL

Slice & Dice UBI data

```
%sql
```

```
select
  user_query, count(0) Total
from ubi_queries
group by user_query
order by Total desc
```

Specify the input language on the first line using %[language type]. Supported languages include

Refresh

✓ Last successful run 09/11/2024 03:49 PM.

```
select user_query, count(0) Total from ubi_queries group by user_query order
```

Columns Density Sort fields Full screen

user_query	count(0)
User Behavior Insights	
best Laptop	55
camera	
briefcase	13
backpack	13
camcorder	9
box	8
cabinet	8
calculator	7
bed	6

Anomaly Detection & Alerting

Detect **Search Quality issues** – Usual drop in CTR, Conversion Rate

Detect **Trending Searches**

Monitoring **Zero-Results Searches**

A/B testing

- Click Through Rate
- Add to Cart Rate
- Position of Clicked Results
- Search Refinement Rate
- Filter/Facet usage Rate
- Zero Results Rate
- Search Abandonment Rate
- Product Return Rate

Wrap

Need for tailor made user behavior metrics for Search

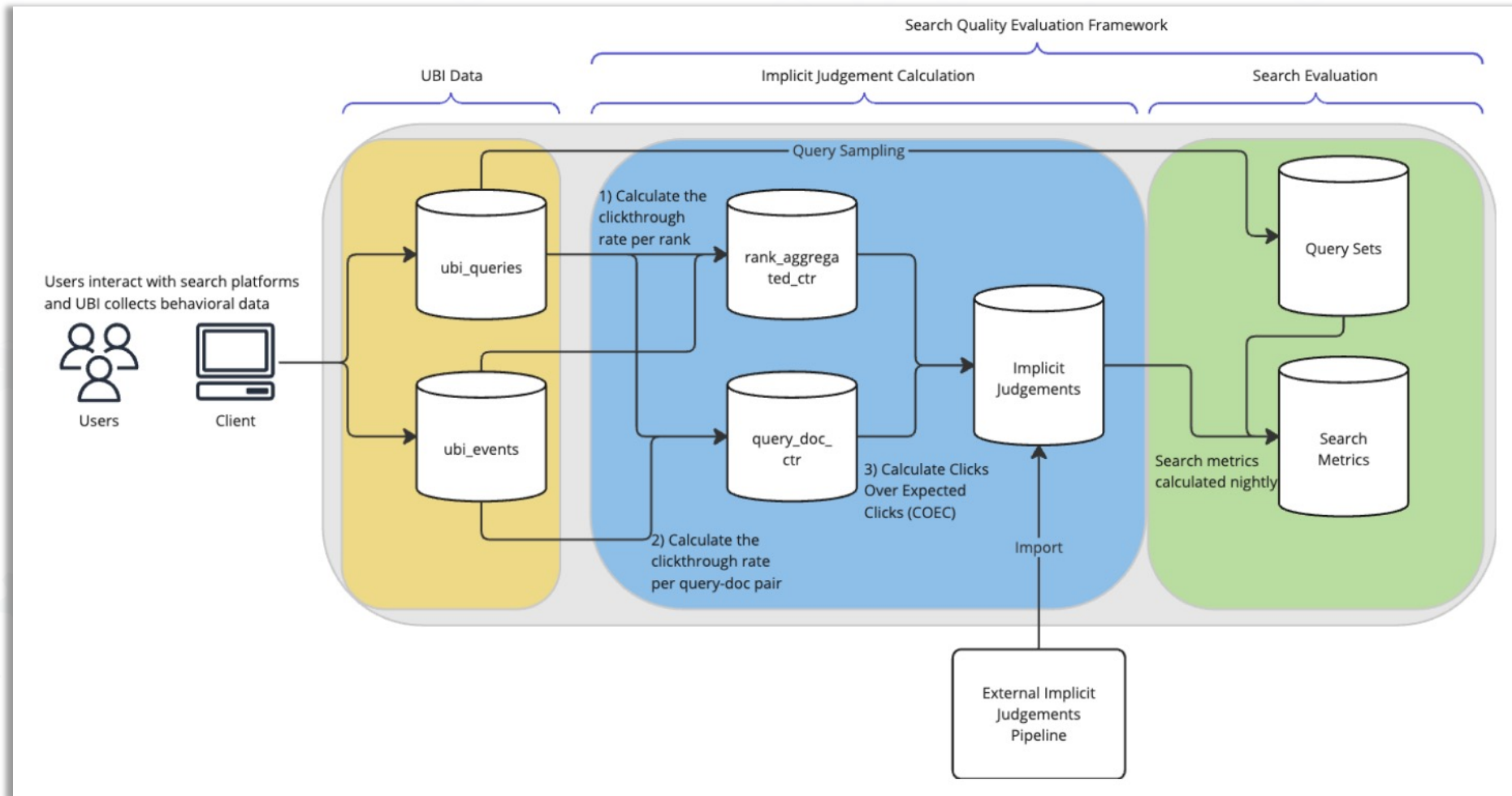
Standardized Schema for UBI

UBI for OpenSearch & Beyond

Leverage UBI data to improve Search

[RFC] OS Search Quality Evaluation Framework

#15354



RFC 15354



Thank you!
Q&A



Aswath Srinivasan

