
druid

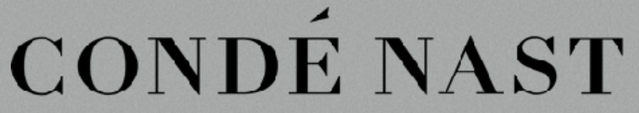
Plan

- Druid concepts
- Druid internals
- Druid practice
- Superset

Druid is a high-performance,
column-oriented,
distributed data store.

- Interactive Queries: sub-second ad-hoc queries to group, filter, and aggregate data
- Real-time Streams: explore events immediately after they occur
- Horizontally Scalable: existing Druid clusters have scaled to petabytes of data and trillions of events, ingesting millions of events every second
- Visualization: Pivot or Superset

Powered by Druid



Druid is for you if…

- You are building an application that requires fast aggregations and OLAP queries
- You want to do real-time analysis
- You have lots of data (trillions of events, petabytes of data)
- You need a data store that is always available with no single point of failure

Property	OLTP	OLAP
Main read pattern	Small number of records per query, fetched by key	Aggregate over large number of records
Main write pattern	Random-access, low-latency writes from user input	Bulk import (ETL) or event stream
Primary used by	End user/customer, via web application	Internal analyst, for decision support
What data represents	Latest state of data (current point in time)	History of events that happened over time
Dataset size	Gigabytes to terabytes	Terabytes to petabytes

Druid Concepts

The Data

The Data

timestamp	publisher	advertiser	gender	country	click	price
2011-01-01T01:01:35Z	bieberfever.com	google.com	Male	USA	0	0.65
2011-01-01T01:03:63Z	bieberfever.com	google.com	Male	USA	0	0.62
2011-01-01T01:04:51Z	bieberfever.com	google.com	Male	USA	1	0.45
2011-01-01T01:00:00Z	ultratrimfast.com	google.com	Female	UK	0	0.87
2011-01-01T02:00:00Z	ultratrimfast.com	google.com	Female	UK	0	0.99
2011-01-01T02:00:00Z	ultratrimfast.com	google.com	Female	UK	1	1.53

Roll-up

Roll-up

timestamp	publisher	advertiser	gender	country	click	price
2011-01-01T01:01:35Z	bieberfever.com	google.com	Male	USA	0	0.65
2011-01-01T01:03:63Z	bieberfever.com	google.com	Male	USA	0	0.62
2011-01-01T01:04:51Z	bieberfever.com	google.com	Male	USA	1	0.45
2011-01-01T01:00:00Z	ultratrimfast.com	google.com	Female	UK	0	0.87
2011-01-01T02:00:00Z	ultratrimfast.com	google.com	Female	UK	0	0.99
2011-01-01T02:00:00Z	ultratrimfast.com	google.com	Female	UK	1	1.53

GROUP BY timestamp, publisher, advertiser, gender, country
:: impressions = COUNT(1), clicks = SUM(click), revenue = SUM(price)

timestamp	publisher	advertiser	gender	country	impressions	clicks	revenue
2011-01-01T01:00:00Z	ultratrimfast.com	google.com	Male	USA	1800	25	15.70
2011-01-01T01:00:00Z	bieberfever.com	google.com	Male	USA	2912	42	29.18
2011-01-01T02:00:00Z	ultratrimfast.com	google.com	Male	UK	1953	17	17.31
2011-01-01T02:00:00Z	bieberfever.com	google.com	Male	UK	3194	170	34.01

Sharding the Data

Sharding the Data

Segment `sampleData_2011-01-01T01:00:00:00Z_2011-01-01T02:00:00:00Z_v1_0` contains

```
2011-01-01T01:00:00Z  ultratrimfast.com  google.com  Male  USA  1800  25  15.70
2011-01-01T01:00:00Z  bieberfever.com    google.com  Male  USA  2912  42  29.18
```

Segment `sampleData_2011-01-01T02:00:00:00Z_2011-01-01T03:00:00:00Z_v1_0` contains

```
2011-01-01T02:00:00Z  ultratrimfast.com  google.com  Male  UK  1953  17  17.31
2011-01-01T02:00:00Z  bieberfever.com    google.com  Male  UK  3194  170  34.01
```

Loading the Data

Loading the Data

- real-time/batch
- at-least-once guarantee for real-time
- immutable snapshots of data
- column store
- compression
- indices for columns

Querying the Data

Querying the Data

- JSON over HTTP
- single table queries
- no joins
- denormalized data

The Druid Cluster

The Druid Cluster

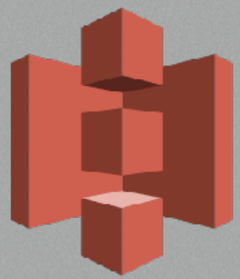
- Historical Nodes
- Broker Nodes
- Coordinator Nodes
- Overlord Nodes
- MiddleManager Nodes

} aka Real-time Nodes

External Dependencies

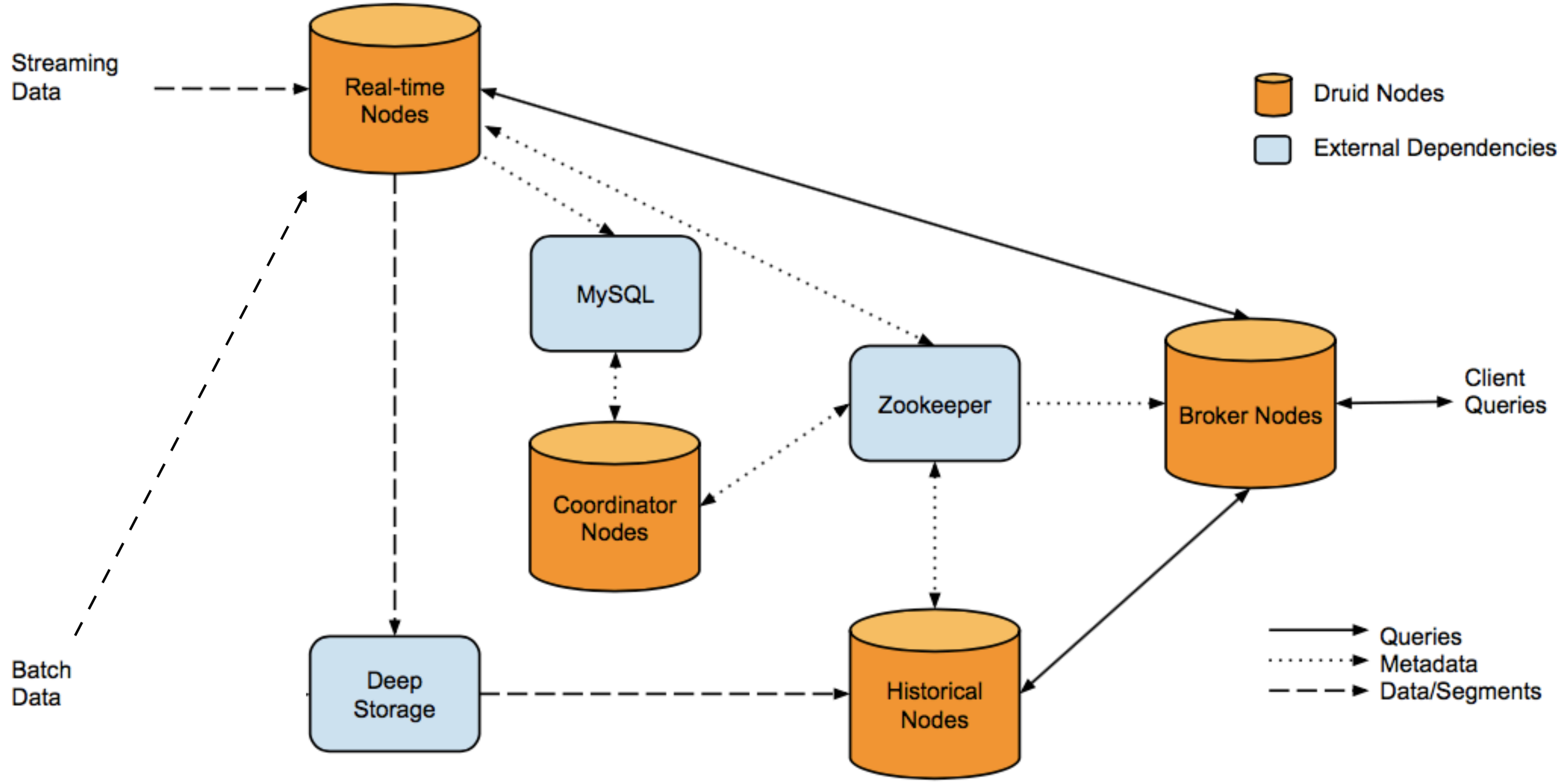
External Dependencies

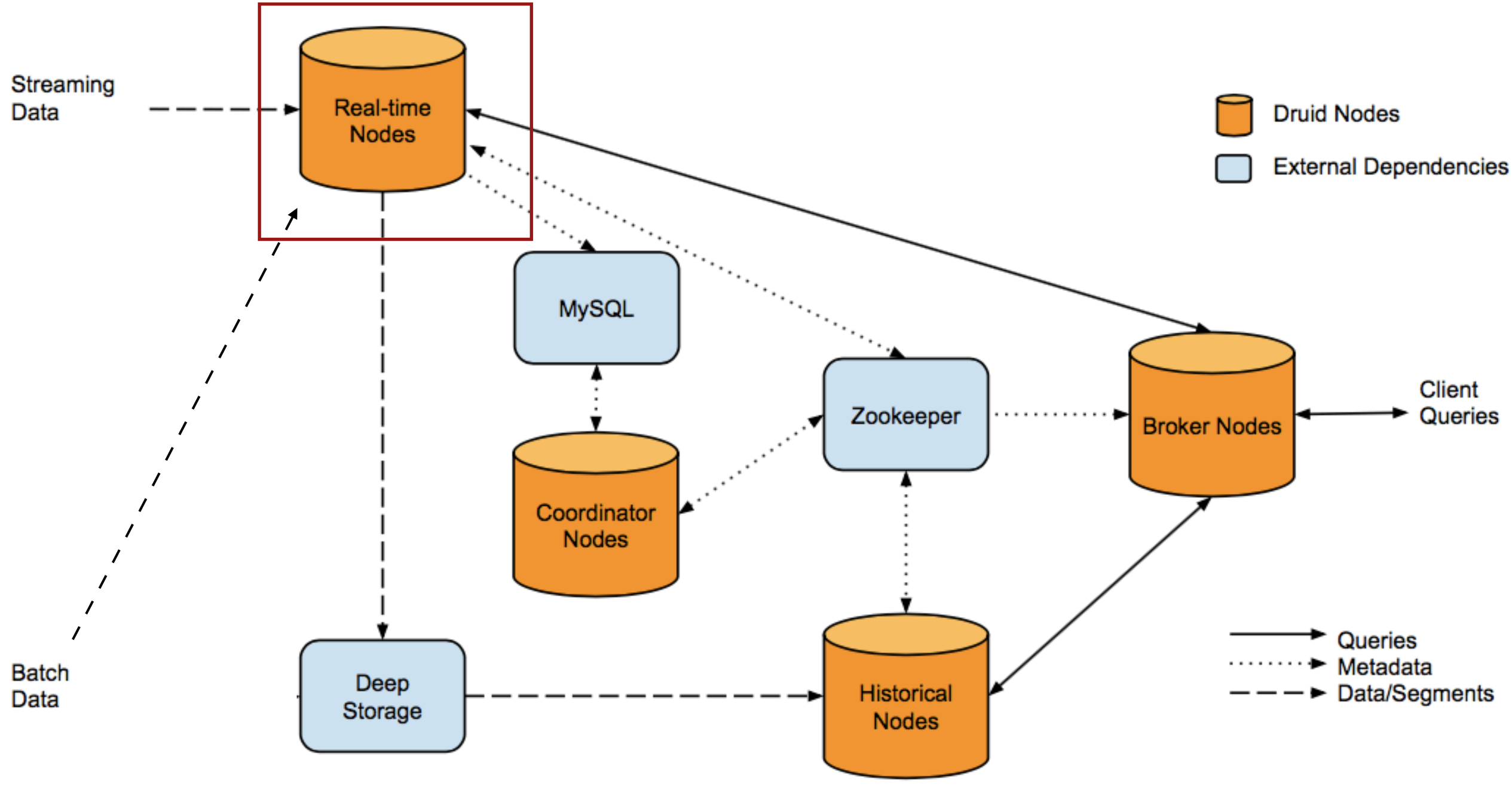
- Zookeeper
- Metadata Storage: Derby, MySQL, PostgreSQL
- Deep Storage: local, HDFS, S3

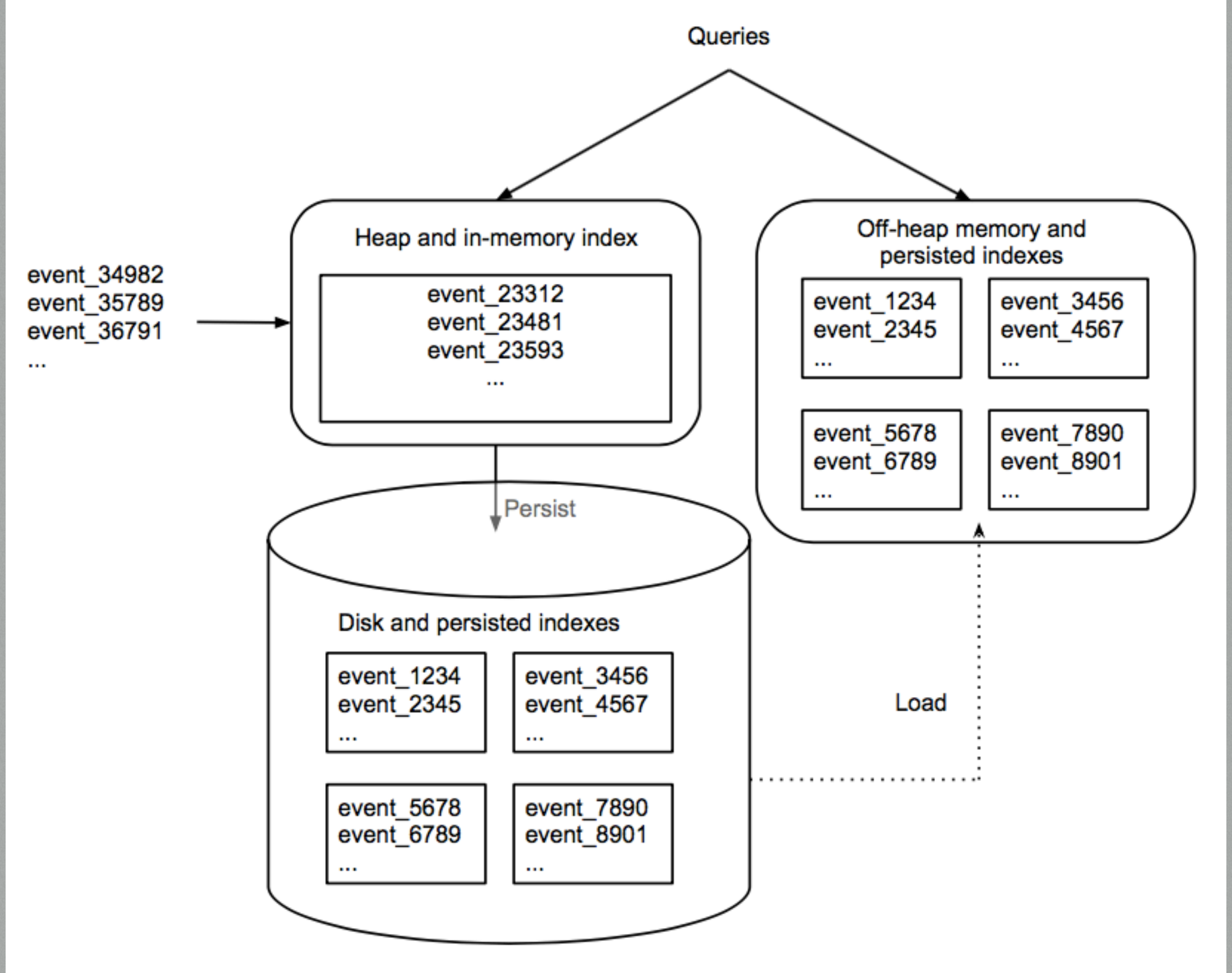


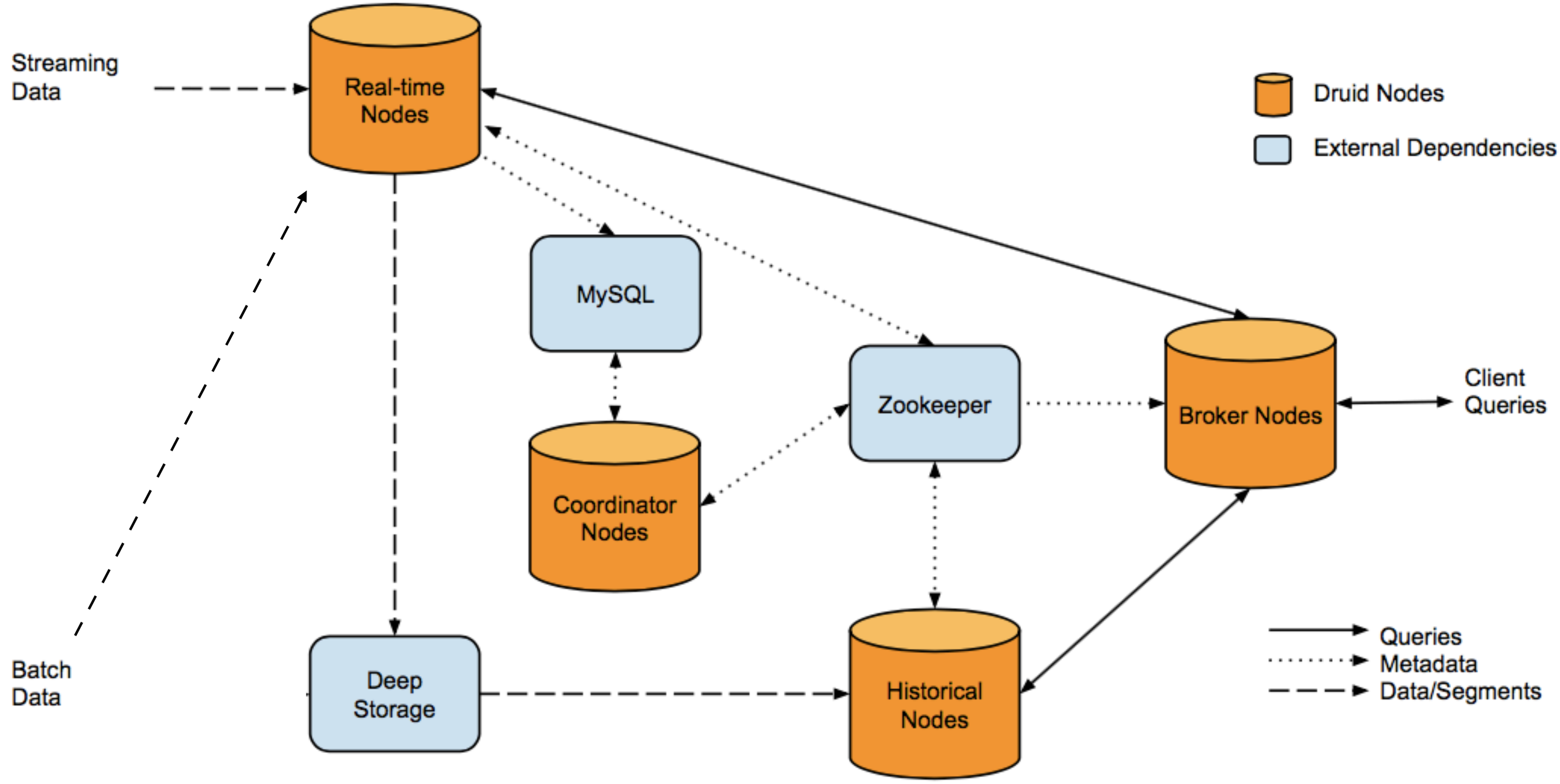
Druid Internals

Architecture

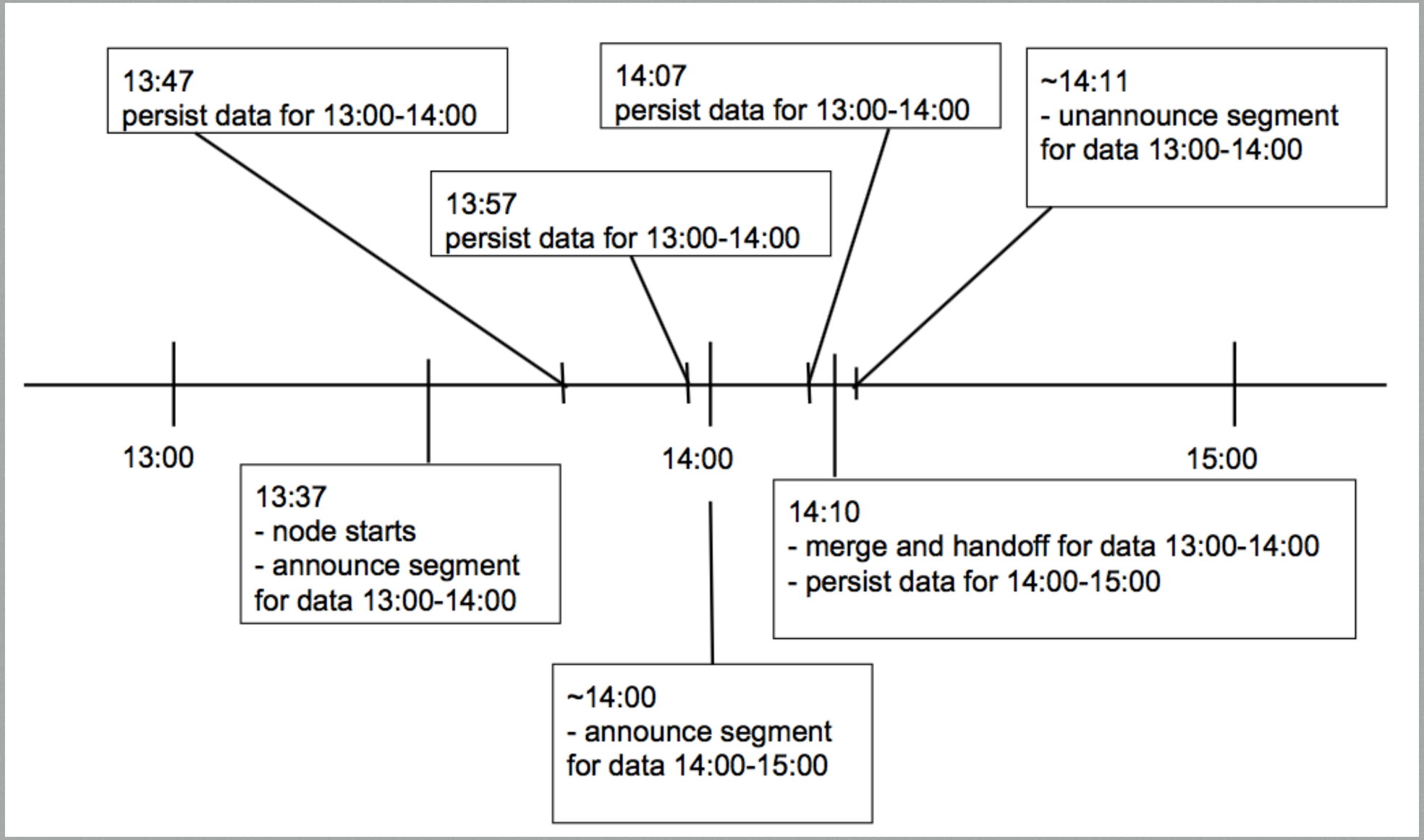

Streaming Data
Real-time Nodes
Druid Nodes
External Dependencies
MySQL
Zookeeper
Broker Nodes
Client Queries
Coordinator Nodes
Batch Data
Deep Storage
Historical Nodes
Queries
Metadata
Data/Segments

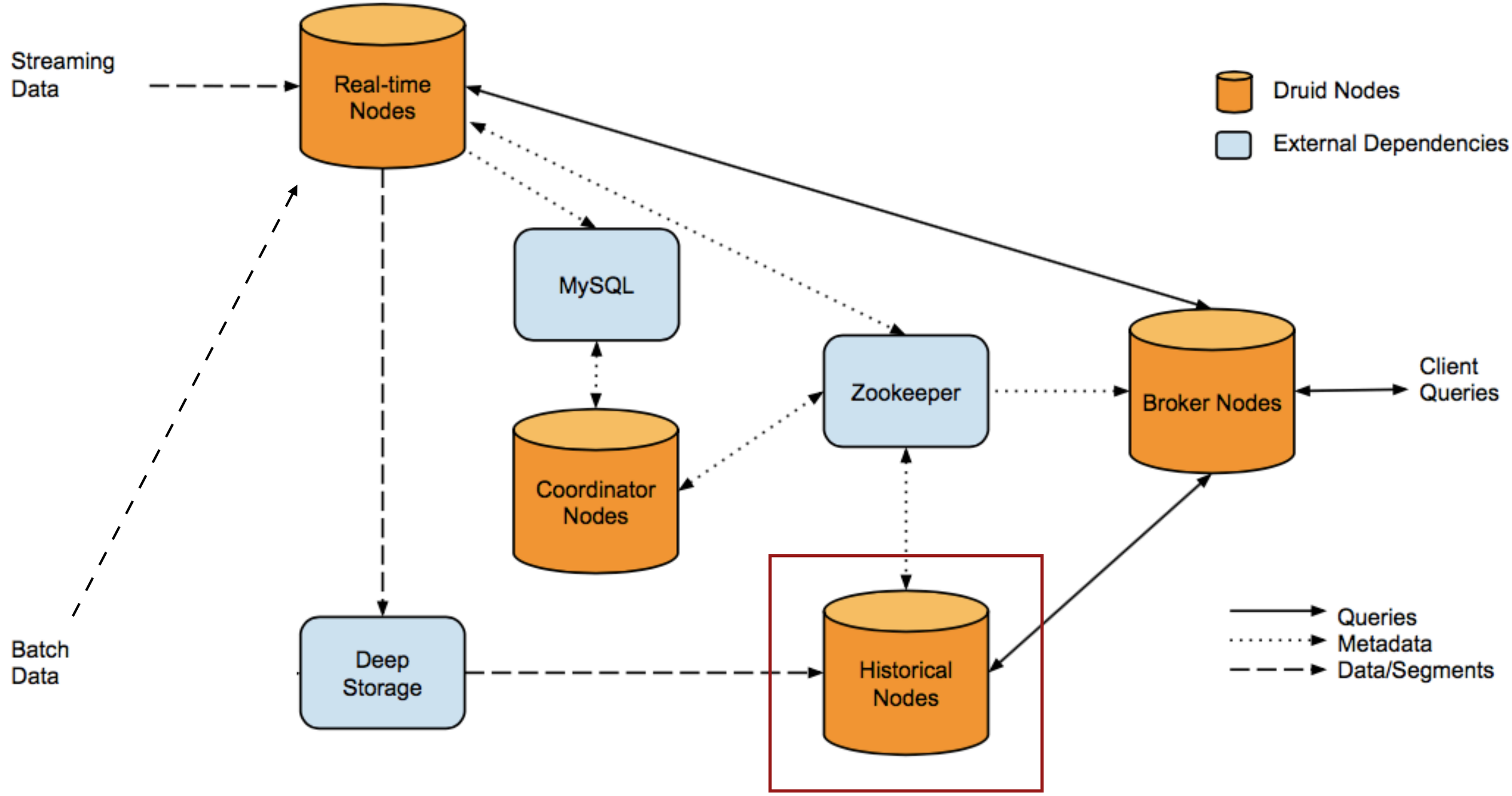

Streaming Data
Real-time Nodes
MySQL
Zookeeper
Broker Nodes
Client Queries
Coordinator Nodes
Batch Data
Deep Storage
Historical Nodes
Druid Nodes
External Dependencies
Queries
Metadata
Data/Segments

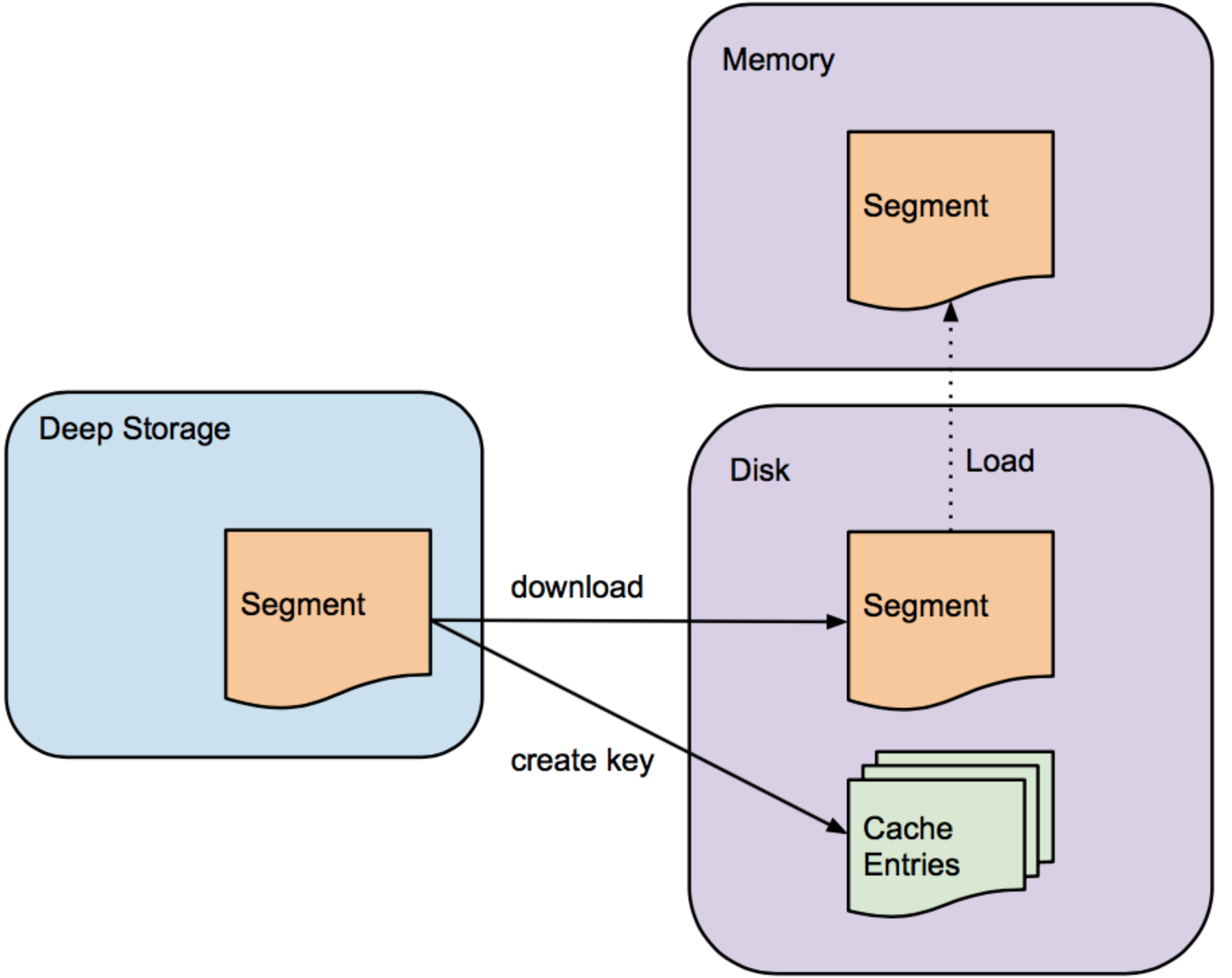

Queries
event_34982
event_35789
event_36791
...
Heap and in-memory index
event_23312
event_23481
event_23593
...
Off-heap memory and
persisted indexes
event_1234
event_2345
...
event_3456
event_4567
...
event_5678
event_6789
...
event_7890
event_8901
...
Persist
Disk and persisted indexes
event_1234
event_2345
...
event_3456
event_4567
...
event_5678
event_6789
...
event_7890
event_8901
...
Load

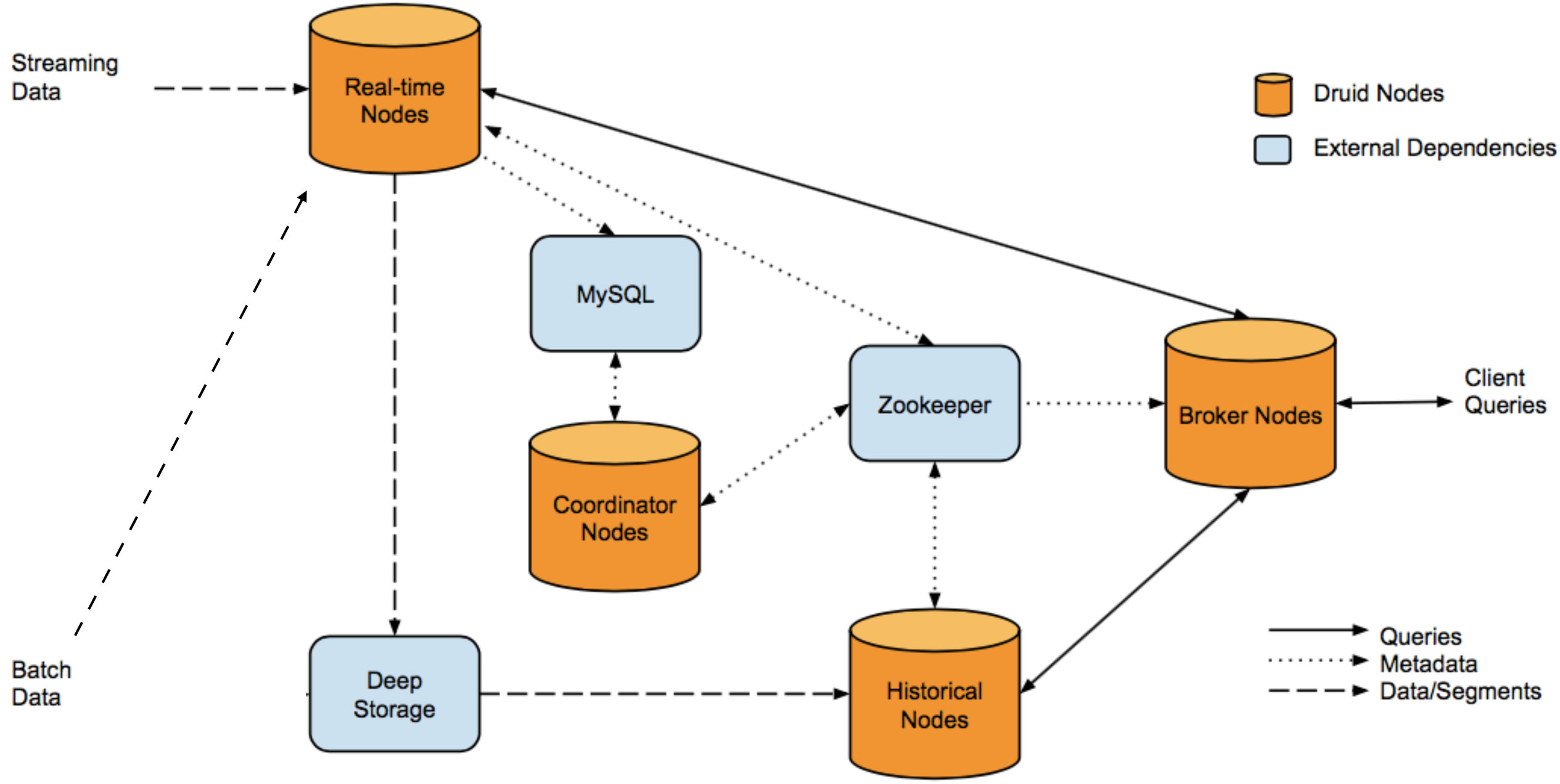

Streaming Data
Real-time Nodes
Druid Nodes
External Dependencies
MySQL
Zookeeper
Broker Nodes
Client Queries
Coordinator Nodes
Batch Data
Deep Storage
Historical Nodes
Queries
Metadata
Data/Segments

Life of segment




Streaming Data
Real-time Nodes
Druid Nodes
External Dependencies
MySQL
Zookeeper
Broker Nodes
Client Queries
Coordinator Nodes
Batch Data
Deep Storage
Historical Nodes
Queries
Metadata
Data/Segments


Memory
Segment
Deep Storage
Segment
download
Disk
Load
Segment
create key
Cache
Entries


Streaming Data
Real-time Nodes
Druid Nodes
External Dependencies
MySQL
Zookeeper
Broker Nodes
Client Queries
Coordinator Nodes
Batch Data
Deep Storage
Historical Nodes
Queries
Metadata
Data/Segments

Storage format

Storage format

- segment files
- partitioned by time
- recommended size 300mb-700mb
- …or 5 million rows
- columnar

Storage format

Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke$ha	Xeno	Male	Taiyuan	3194	170

Table 1: Sample Druid data for edits that have occurred on Wikipedia.

Storage format

Timestamp

Dimensions

Metrics

Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke$ha	Xeno	Male	Taiyuan	3194	170

Table 1: Sample Druid data for edits that have occurred on Wikipedia.

Storage format

```
1: Dictionary that encodes column values
  {
    "Justin Bieber": 0,
    "Ke$ha":         1
  }

2: Column data
  [0,
   0,
   1,
   1]

3: Bitmaps - one for each unique value of the column
  value="Justin Bieber": [1,1,0,0]
  value="Ke$ha":         [0,0,1,1]
```

Storage format

```
1: Dictionary that encodes column values
  {
    "Justin Bieber": 0,
    "Ke$ha":         1
  }

2: Column data
  [0,
   0,
   1,
   1]

3: Bitmaps - one for each unique value of the column
  value="Justin Bieber": [1,1,0,0]
  value="Ke$ha":         [0,0,1,1]
```

for encoding

for group by

for filtering

Practice

Practice

- Start Druid
- Upload batch
- Query
- Superset
- Ingest from Kafka

Thank you!