

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THỐNG THÔNG TIN



BÁO CÁO SEMINAR CHƯƠNG 3
DỮ LIỆU LỚN – IS405.P11.HTCL
APACHE DRILL

GVHD: ThS. Nguyễn Hồ Duy Tri

Trương Vĩnh Thuận - 21522653

Hoàng Quốc Việt - 21522790

Phùng Thiên Phúc - 21521297

HO CHI MINH CITY, 2024

Mục lục

1. Thông tin chung	3
2. Ưu/khuyết điểm, những đặc điểm khiến sản phẩm nổi bật hơn so với các sản phẩm khác cùng loại	3
2.1 Ưu điểm của Apache Drill	3
2.2 Khuyết điểm của Apache Drill	4
❖ Đặc điểm nổi bật so với các sản phẩm khác.....	5
❖ So sánh với các sản phẩm khác cùng loại.....	6
3. Ứng dụng cụ thể.....	6
4. Triển khai Apache Drill.....	8
4.1 Embedded mode.....	8
4.2 Distributed mode.....	15

1. Thông tin chung

Drill là một công cụ truy vấn SQL mã nguồn mở của Apache, được thiết kế để phân tích dữ liệu lớn một cách hiệu quả. Nó hỗ trợ phân tích dữ liệu bán cấu trúc và phát triển nhanh từ các ứng dụng Big Data hiện đại, đồng thời sử dụng ngôn ngữ truy vấn ANSI SQL quen thuộc. Drill có thể tích hợp dễ dàng với Apache Hive và Apache HBase.

Các tính năng chính của Apache Drill bao gồm:

- **Truy vấn SQL độ trễ thấp:** Tối ưu hóa hiệu suất truy vấn.
- **Truy vấn động trên dữ liệu tự mô tả:** Hỗ trợ các định dạng như JSON, Parquet, văn bản và bảng HBase mà không cần định nghĩa siêu dữ liệu trong metastore của Hive.
- **Hỗ trợ ANSI SQL:** Cung cấp ngôn ngữ truy vấn tiêu chuẩn.
- **Hỗ trợ dữ liệu lồng ghép:** Xử lý cấu trúc dữ liệu phức tạp.
- **Tích hợp với Apache Hive:** Truy vấn trên bảng và chế độ xem Hive, hỗ trợ tất cả định dạng tệp Hive và UDF của Hive.
- **Tích hợp công cụ BI/SQL:** Sử dụng trình điều khiển JDBC/ODBC tiêu chuẩn.

<https://drill.apache.org/docs/drill-introduction/>

2. Ưu/khuyết điểm, những đặc điểm khiến sản phẩm nổi bật hơn so với các sản phẩm khác cùng loại

2.1 Ưu điểm của Apache Drill

1. Hỗ trợ truy vấn dữ liệu phi cấu trúc và bán cấu trúc:

- Apache Drill nổi bật nhờ khả năng truy vấn trực tiếp dữ liệu từ nhiều định dạng khác nhau như **JSON**, **Parquet**, **CSV**, **HBase**, **MongoDB**, **HDFS** mà không cần chuyển đổi dữ liệu sang dạng quan hệ hoặc qua bước ETL phức tạp.
- Drill sử dụng cơ chế **schema-free** (không cần lược đồ cố định) nên có thể xử lý dữ liệu linh hoạt từ các định dạng bán cấu trúc như JSON, giúp tiết kiệm thời gian và công sức khi làm việc với các tập dữ liệu lớn và không đồng nhất.

2. Tính linh hoạt trong việc kết nối với nhiều hệ thống lưu trữ:

- Drill có thể kết nối với các hệ thống lưu trữ đa dạng như **HDFS**, **Amazon S3**, **Google Cloud Storage**, **NoSQL (MongoDB, HBase)**, và các cơ sở dữ liệu quan hệ. Điều này giúp người dùng có thể phân tích dữ liệu từ nhiều nguồn mà không cần phải tích hợp qua các công cụ khác.
- Drill còn hỗ trợ truy vấn qua **ODBC**, **JDBC** và **REST API**, dễ dàng tích hợp vào các ứng dụng hiện có.

3. Dễ sử dụng với SQL quen thuộc:

- Người dùng có thể sử dụng **SQL** quen thuộc để truy vấn dữ liệu từ các nguồn khác nhau mà không cần học một ngôn ngữ mới. Drill hỗ trợ hầu hết các cú pháp SQL chuẩn (ANSI SQL).
- Người dùng có thể thực hiện truy vấn dữ liệu ngay cả khi không có kiến thức sâu về các cấu trúc dữ liệu phức tạp, nhờ việc Drill tự động phát hiện lược đồ.

4. Khả năng mở rộng và phân tán:

- Apache Drill được thiết kế để chạy trên các cụm máy (cluster), giúp xử lý dữ liệu lớn với quy mô từ hàng terabyte đến petabyte. Drill có thể chia nhỏ các truy vấn thành các phân đoạn (fragments) và thực hiện phân tán trên nhiều node, giúp tăng tốc độ xử lý.
- Drill dễ dàng mở rộng khi thêm nhiều node vào cluster, mà không ảnh hưởng đến hiệu năng.

5. Không cần ETL và khả năng thời gian thực:

- Drill cho phép truy vấn dữ liệu trực tiếp mà không cần các bước **ETL** (Extract, Transform, Load) truyền thống. Điều này giúp phân tích dữ liệu gần như trong thời gian thực, hỗ trợ các quyết định kinh doanh nhanh chóng.
- Khả năng thời gian thực là một lợi thế lớn đối với những hệ thống yêu cầu thông tin ngay lập tức, chẳng hạn như hệ thống quảng cáo hay dữ liệu cảm biến.

6. Cộng đồng và phát triển nguồn mở:

- Apache Drill là một dự án mã nguồn mở thuộc Apache Foundation, có một cộng đồng lớn và đang phát triển mạnh mẽ. Điều này đảm bảo rằng Drill luôn được cải tiến và hỗ trợ tốt từ cộng đồng cũng như từ các doanh nghiệp sử dụng.

2.2 Khuyết điểm của Apache Drill

1. Hiệu năng có thể không ổn định với các tập dữ liệu rất lớn:

- Mặc dù Drill có thể xử lý tốt dữ liệu phi cấu trúc, nhưng trong một số trường hợp với các tập dữ liệu rất lớn và phức tạp, hiệu năng của Drill có thể không ổn định hoặc không tối ưu như các công cụ chuyên biệt khác như **Presto** hay **Hive**.
- Drill có thể tiêu tốn nhiều bộ nhớ và tài nguyên khi xử lý các truy vấn phức tạp với các tập dữ liệu khổng lồ.

2. Chưa hỗ trợ tốt các tác vụ phân tích chuyên sâu:

- Drill thiên về khả năng truy vấn dữ liệu từ nhiều nguồn và định dạng khác nhau, nhưng chưa thực sự mạnh trong các tác vụ phân tích chuyên sâu như **machine learning** hay các tác vụ thống kê phức tạp. Với các yêu cầu phức tạp hơn, người dùng cần tích hợp với các công cụ khác (như **Spark** hay **Flink**).

3. Yêu cầu cấu hình và quản lý cụm (cluster) phức tạp:

- Việc cài đặt và quản lý Drill trong môi trường phân tán (cluster) yêu cầu phải hiểu rõ về cấu hình Zookeeper, các tham số điều chỉnh phân mảnh, bộ nhớ. Với những người không có kinh nghiệm về hệ thống phân tán, có thể gặp khó khăn trong việc tối ưu hóa và bảo trì hệ thống.

4. Giới hạn trong hỗ trợ ACID và bảo mật:

- Drill không hỗ trợ tốt các tính năng **ACID** (Atomicity, Consistency, Isolation, Durability), điều này có thể là trở ngại trong các hệ thống yêu cầu giao dịch dữ liệu mạnh mẽ.
- Mặc dù Drill cung cấp các cơ chế bảo mật như **authentication** và **authorization**, nhưng tính năng này vẫn hạn chế so với các giải pháp dữ liệu lớn khác như Hive hoặc Impala, đặc biệt trong các môi trường yêu cầu bảo mật cao.

❖ Đặc điểm nổi bật so với các sản phẩm khác

1. Hỗ trợ schema-free:

- Không như **Hive** hay **Impala** (yêu cầu lược đồ cố định), Drill nổi bật với khả năng truy vấn **schema-free**, tức là không cần phải thiết lập lược đồ trước khi truy vấn. Điều này giúp nó dễ dàng làm việc với các dữ liệu phi cấu trúc và bán cấu trúc như JSON và Parquet.

2. Truy vấn nhiều nguồn dữ liệu từ một hệ thống:

- Drill hỗ trợ truy vấn từ nhiều hệ thống lưu trữ khác nhau trong cùng một lần chạy truy vấn SQL mà không cần qua bước trung gian. Điều này giúp dễ dàng kết hợp dữ liệu từ **HDFS, NoSQL, RDBMS**, và các dịch vụ đám mây (cloud storage) trong cùng một truy vấn, mà không phải chuyển đổi dữ liệu.

3. Cộng đồng phát triển mạnh mẽ:

- Là một dự án thuộc **Apache Foundation**, Apache Drill nhận được sự đóng góp mạnh mẽ từ cộng đồng nguồn mở, đảm bảo rằng các tính năng và cải tiến mới liên tục được phát triển, đồng thời có nhiều tài liệu và hướng dẫn hỗ trợ.

4. Khả năng xử lý dữ liệu bán cấu trúc tốt hơn:

- So với các công cụ truy vấn phân tán khác như **Presto, Hive**, hoặc **Impala**, Drill có lợi thế trong việc xử lý dữ liệu bán cấu trúc, nhờ vào việc hỗ trợ JSON và các định dạng tương tự mà không cần phải thiết lập cấu trúc trước đó.

❖ So sánh với các sản phẩm khác cùng loại

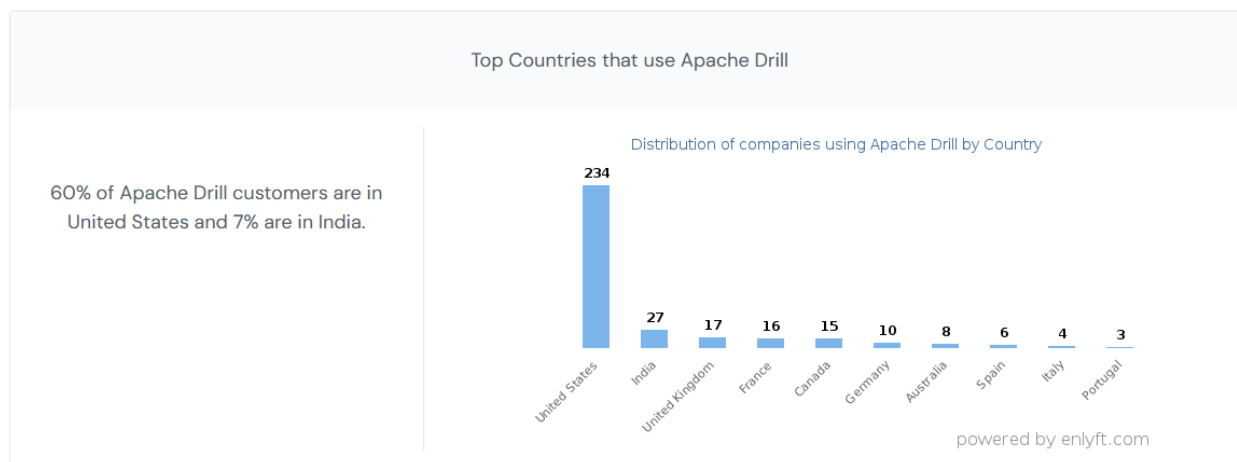
- **Apache Hive:** Mạnh trong xử lý dữ liệu dạng bảng và hỗ trợ tốt cho các tác vụ ETL, nhưng yêu cầu định nghĩa lược đồ trước khi truy vấn và không mạnh về dữ liệu phi cấu trúc.
- **Presto:** Hiệu năng truy vấn nhanh hơn trong các tác vụ phân tích trên tập dữ liệu rất lớn, nhưng không hỗ trợ tốt dữ liệu phi cấu trúc như Drill.
- **Apache Impala:** Tốt cho phân tích dữ liệu bảng trên HDFS, nhưng không linh hoạt với dữ liệu phi cấu trúc.

<https://drill.apache.org/docs/why-drill/>

3. Ứng dụng cụ thể

Apache Drill được sử dụng bởi gần 400 công ty và doanh nghiệp toàn cầu. Phần lớn các công ty này có trụ sở ở nước Mỹ và thuộc ngành Công nghệ Thông tin hoặc Dịch vụ. Ứng dụng được bởi các công ty lớn có hơn 10 nghìn nhân viên và tổng lợi nhuận trên nghìn triệu đô

Top Countries that use Apache Drill



Các công ty sử dụng Apache Drill trong công việc thường ngày.

•Intuit

Intuit là công ty cung cấp phần mềm tài chính sử dụng Apache Drill để phân tích và tạo báo cáo từ các chủ đề về luồng nhấp chuột của Kafka. Ưu điểm của việc sử dụng Apache Drill là loại kết hợp của nó cung cấp sự hỗ trợ cho luồng nhấp chuột cũng là dữ liệu định dạng JSON.

•Ericsson

Ericsson là một công ty mạng và viễn thông đa quốc gia sử dụng Apache Drill để phân tích việc sử dụng các thiết bị mạng không dây của họ bằng cách chạy truy vấn SQL trên số lượng bộ dữ liệu lớn.

•RedBus

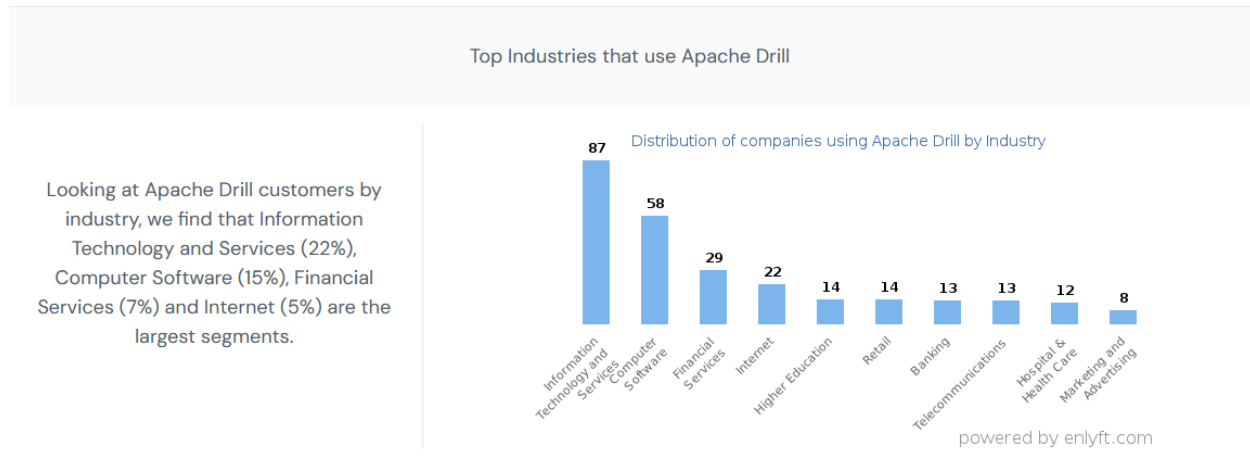
RedBus là nền tảng bán vé BUS trực tuyến ở Ấn Độ sử dụng Apache Drill để thực hiện thao tác ETL (trích xuất, chuyển đổi và tải). Nó sử dụng Kafka, Amazon S3 và MongoDB cho kho dữ liệu.

•Ant Financial

Ant Financial là nền tảng thanh toán kỹ thuật số lớn nhất Trung Quốc mà Apache Drill để thực hiện thao tác truy vấn Adhoc cũng sử dụng máy khoan để thực hiện phân tích dữ liệu dịch vụ nhằm hiểu rõ hơn để đưa ra quyết định kinh doanh.

• TransUnion

TransUnion là cơ quan báo cáo tín dụng tiêu dùng sử dụng Apache Drill để thực hiện báo cáo và vận hành BI.



Các ngành công nghiệp đi đầu trong việc sử dụng Apache Drill

4. Triển khai Apache Drill

4.1 Embedded mode

```
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.2# ./bin/drill-embedded
Apache Drill 1.21.2
"Keep your data close, but your Drillbits closer."
apache drill> 
```


Kết quả:

The screenshot displays the Apache Drill web interface in a browser window. The address bar shows 'localhost:8047'. The navigation bar includes links for 'Query', 'Profiles', 'Storage', 'Metrics', 'Threads', 'Logs', 'Options', and 'Documentation'. The main content area is titled 'Drillbits 1' and contains a table with the following data:

#	Address	Heap Memory Usage	Direct Memory Usage	CPU Usage	Avg Sys Load	User Port	Control Port	Data Port	Version	Status	Uptime	Shutdown
1	thuantv-VMware-Virtual-Platform Current	0.51GB (13% of 4GB)	0GB (0% of 0GB)	0.30%	0.20	31010	31011	31012	1.21.2	ONLINE	1m 51s	

Below the table, there are two sections: 'Encryption' and 'Query Throttling'. The 'Encryption' section shows 'Client to Bit Encryption' and 'Bit to Bit Encryption' both set to 'Disabled'. The 'Query Throttling' section shows 'Queue Status' set to 'Disabled'.

❖ Tổng quan về dataset (Yelp Business)

```
{  
  // string, 22 character unique string business id  
  "business_id": "tnhfDv5Il8EaGSXZGiuQGg",  
  
  // string, the business's name  
  "name": "Garaje",  
  
  // string, the full address of the business  
  "address": "475 3rd St",  
  
  // string, the city  
  "city": "San Francisco",  
  
  // string, 2 character state code, if applicable  
  "state": "CA",  
  
  // string, the postal code  
  "postal code": "94107",  
  
  // float, latitude  
  "latitude": 37.7817529521,  
  
  // float, longitude  
  "longitude": -122.39612197,  
  
  // float, star rating, rounded to half-stars  
  "stars": 4.5,  
  
  // integer, number of reviews  
  "review_count": 1198,
```

```

// integer, 0 or 1 for closed or open, respectively
"is_open": 1,

// object, business attributes to values. note: some attribute values m
"attributes": {
  "RestaurantsTakeOut": true,
  "BusinessParking": {
    "garage": false,
    "street": true,
    "validated": false,
    "lot": false,
    "valet": false
  },
},

// an array of strings of business categories
"categories": [
  "Mexican",
  "Burgers",
  "Gastropubs"
],

// an object of key day to value hours, hours are using a 24hr clock
"hours": {
  "Monday": "10:00-21:00",
  "Tuesday": "10:00-21:00",
  "Friday": "10:00-21:00",
  "Wednesday": "10:00-21:00",
  "Thursday": "10:00-21:00",
  "Sunday": "11:00-18:00",
  "Saturday": "10:00-21:00"
}
}

```

Truy vấn 1: Tổng quan dữ liệu business

```

SELECT *
FROM dfs.`/home/thuantv/yelp_dataset/yelp_academic_dataset_business.json`
LIMIT 15;

```

Apache Drill Query Profiles Storage Metrics Threads Logs Options Documentation

Query Profile: 18e0f5e-e143-3d17-7eac-3649cd9c14c **COMPLETED**

Column visibility Show 10 entries

Search:

business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attributes
Pns2I4eNsfO8kk83dixA6A	Abby Rappoport, LAC, CMQ	1616 Chapala St, Ste 2	Santa Barbara	CA	93101	34.4266787	-119.7111968	5.0	7	0	("ByAppointmentC
mpf3x-BjtTdTEA3yCzIAYPw	The UPS Store	87 Grasso Plaza Shopping Center	Alfton	MO	63123	38.551126	-90.335695	3.0	15	1	("BusinessAccept
tUFWirkKI_TansVWINQQ	Target	5255 E Broadway Blvd	Tucson	AZ	85711	32.223236	-110.880452	3.5	22	0	("ByAppointmentC
MTSW4McQd7CbVtyjqoe9mw	St Honore Pastries	935 Race St	Philadelphia	PA	19107	39.9555052	-75.1555641	4.0	80	1	("ByAppointmentC
mWMc6_wTdEOEUBKIGXDVIA	Perkiomen Valley Brewery	101 Walnut St	Green Lane	PA	18054	40.3381827	-75.4716585	4.5	13	1	("BusinessAccept
CF33F8-E6oudUQ46HnavjQ	Sonic Drive-In	615 S Main St	Ashland City	TN	37015	36.269593	-87.058943	2.0	6	1	("ByAppointmentC
n_0UpQx1hsNbnPUSlodU8w	Famous Footwear	8522 Eager Road, Dierbergs Brentwood Point	Brentwood	MO	63144	38.627695	-90.340465	2.5	13	1	("BusinessAccept
qkRM_2X51yqxk3btwAQIg	Temple Beth-El	400 Pasadena Ave S	St. Petersburg	FL	33707	27.76659	-82.732983	3.5	5	1	()
k0hIBqXX-Bt0vT1op7Jr1w	Tsevi's Pub And Grill	8025 Mackenzie Rd	Alfton	MO	63123	38.5651648	-90.3210868	3.0	19	0	("BusinessAccept
bBDEgkFA1Oxb9Le7BZUQ	Sonic Drive-In	2312 Dickerson Pike	Nashville	TN	37207	36.2081024	-86.7681696	1.5	10	1	("ByAppointmentC

Showing 1 to 10 of 15 entries

Previous 1 2 Next

Truy vấn 2: Tổng số lượt đánh giá của dataset

SELECT sum(review_count) as totalreviews

FROM dfs.`/home/thuantv/yelp_dataset/yelp_academic_dataset_business.json`;

Apache Drill Query Profiles Storage Metrics Threads Logs Options Documentation

Query Profile: 18e60e17-d3fe-6122-eece-1ef3621b9aeb **COMPLETED**

Column visibility Show 10 entries

Search:

totalreviews
6745508

Truy vấn 3: Top 10 các bang và thành phố có số review cao nhất

SELECT state, city, count(*) totalreviews

FROM dfs.`/home/thuantv/yelp_dataset/yelp_academic_dataset_business.json`

GROUP BY state, city ORDER BY count(*) DESC LIMIT 10;

Apache Drill Query Profiles Storage Metrics Threads Logs Options Documentation

Query Profile: 18e60cf3-f797-6d48-aae1-bbc78365344c **COMPLETED**

Delimiter: , Export

Column visibility Show 10 entries Search:

state	city	totalreviews
PA	Philadelphia	14567
AZ	Tucson	9249
FL	Tampa	9048
IN	Indianapolis	7540
TN	Nashville	6968
LA	New Orleans	6208
NV	Reno	5932
AB	Edmonton	5054
MO	Saint Louis	4827
CA	Santa Barbara	3829

Truy vấn 4: Thời gian mở cửa, đóng cửa của các doanh nghiệp vào thứ 7

```
SELECT b.name, b.hours.`Saturday`
```

```
FROM dfs.`/home/thuantv/yelp_dataset/yelp_academic_dataset_business.json` b
LIMIT 10;
```

Apache Drill Query Profiles Storage Metrics Threads Logs Options Documentation

Query Profile: 18e60a97-4ebd-3666-fc48-119ec091135e **COMPLETED**

Delimiter: ,

Export

Column visibility Show 10 entries Search:

name	EXPRS1
Abby Rappoport, LAC, CMQ	null
The UPS Store	8:0-14:0
Target	8:0-23:0
St Honore Pastries	7:0-21:0
Perkiomen Valley Brewery	12:0-22:0
Sonic Drive-In	9:0-22:0
Famous Footwear	10:0-18:0
Temple Beth-El	null
Tsevi's Pub And Grill	null
Sonic Drive-In	6:0-17:0

Truy vấn 5: Xem các tiện nghi của doanh nghiệp

```
alter system set `store.json.all_text_mode` = true;
```

Apache Drill Query Profiles Storage Metrics Threads Logs Options Documentation

Query Profile: 18e60a45-cf91-0e21-9fcd-e850d65988b6 **COMPLETED**

Delimiter: ,

Export

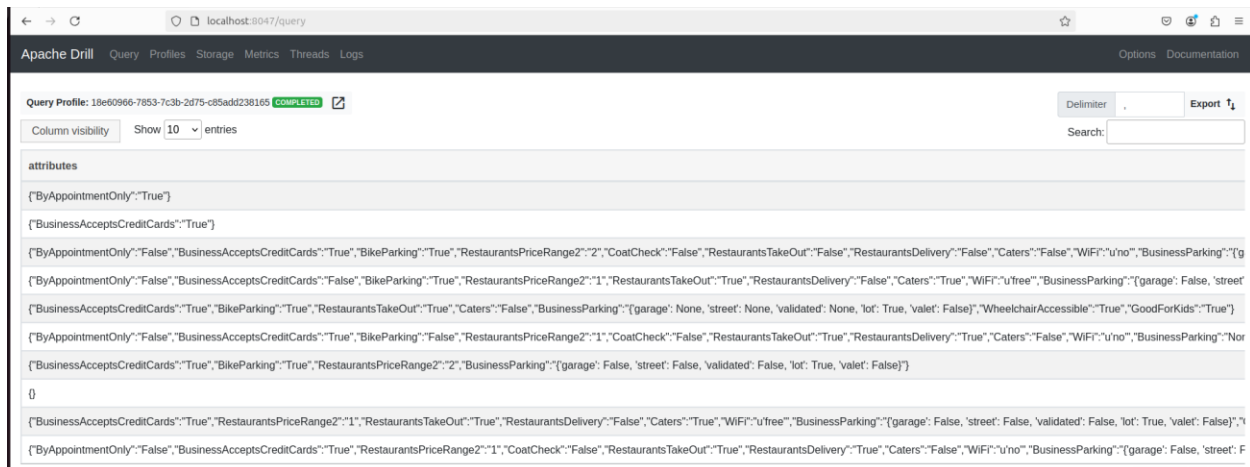
Column visibility Show 10 entries Search:

ok	summary
true	store.json.all_text_mode updated.

Showing 1 to 1 of 1 entries Previous 1 Next

SELECT attributes

```
FROM dfs.`/home/thuantv/yelp_dataset/yelp_academic_dataset_business.json`  
LIMIT 10;
```



4.2 Distributed mode

❖ Cấu hình Zookeeper (../apache-zookeeper-3.5.10-bin/conf/zoo.cfg):



- **tickTime=2000:** Thiết lập khoảng thời gian cơ bản giữa các lần gửi tín hiệu giữa các server trong cluster Zookeeper (tính bằng mili giây).
- **dataDir=/home/thuantv/apache-zookeeper-3.5.10-bin/data:** Thư mục nơi Zookeeper lưu trữ dữ liệu trạng thái và nhật ký giao dịch.
- **clientPort=2182:** Cổng mà Zookeeper sẽ lắng nghe kết nối từ client.
- **initLimit=5:** Giới hạn số tick mà Zookeeper cho phép một follower hoàn tất quá trình khởi tạo kết nối với leader.
- **syncLimit=2:** Giới hạn số tick mà follower có thể không đồng bộ với leader trước khi bị coi là không khả dụng.
- **server.1=127.0.0.1:2888:3888:** Định nghĩa một server trong cluster. Ở đây, server với ID 1 chạy trên IP 127.0.0.1, cổng 2888 (kết nối giữa các server) và cổng 3888 (dùng để bầu leader).

❖ Cấu hình Drillbit1 (../apache-drill-1.21.2/conf/drill-override.conf)

```
drill.exec: {
  cluster-id: "drill-cluster",
  zk.connect: "127.0.0.1:2182",
  drillbit: {
    port: 31010,
    http: {
      port: 8047
    }
  },
  exec: {
    allow_loopback_address_binding: true
  }
}
```

- **cluster-id: "drill-cluster"**: Đặt tên cho cluster của Apache Drill. Các Drillbits trong cùng một cluster sẽ chia sẻ cùng ID này để làm việc chung.
- **zk.connect: "127.0.0.1:2182"**: Địa chỉ kết nối với Zookeeper, ở đây Drill kết nối với Zookeeper tại địa chỉ 127.0.0.1 và cổng 2182.
- **drillbit.port: 31010**: Cổng mà Drillbit lắng nghe để xử lý các truy vấn và giao tiếp với các Drillbit khác.
- **drillbit.http.port: 8047**: Cổng cho giao diện quản lý web của Drill (Web UI) để người dùng truy cập và quản lý hệ thống.
- **exec.allow_loopback_address_binding: true**: Cho phép Drillbit được phép kết nối với địa chỉ loopback (localhost), điều này hữu ích trong môi trường cài đặt đơn giản trên một máy.

❖ **Cấu hình Drillbit2 (../apache-drill-1.21.1/conf/drill-override.conf)**


```
drill.exec: {  
  cluster-id: "drill-cluster",  
  zk.connect: "127.0.0.1:2182",  
  drillbit: {  
    port: 31011,  
    http: {  
      port: 8048  
    }  
  },  
  exec: {  
    allow_loopback_address_binding: true  
  }  
}
```

- **cluster-id: "drill-cluster"**: Đặt tên cho cluster của Apache Drill. Các Drillbits trong cùng một cluster sẽ chia sẻ cùng ID này để làm việc chung.
- **zk.connect: "127.0.0.1:2182"**: Địa chỉ kết nối với Zookeeper, ở đây Drill kết nối với Zookeeper tại địa chỉ 127.0.0.1 và cổng 2182.
- **drillbit.port: 31011**: Cổng mà Drillbit lắng nghe để xử lý các truy vấn và giao tiếp với các Drillbit khác.
- **drillbit.http.port: 8048**: Cổng cho giao diện quản lý web của Drill (Web UI) để người dùng truy cập và quản lý hệ thống.
- **exec.allow_loopback_address_binding: true**: Cho phép Drillbit được phép kết nối với địa chỉ loopback (localhost), điều này hữu ích trong môi trường cài đặt đơn giản trên một máy.

Start zookeeper

```
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-zookeeper-3.5.10-bin#  
./bin/zkServer.sh start  
ZooKeeper JMX enabled by default  
Using config: /home/thuantv/apache-zookeeper-3.5.10-bin/bin/../conf/zoo.cfg  
Starting zookeeper ... STARTED  
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-zookeeper-3.5.10-bin#  
./bin/zkServer.sh status  
ZooKeeper JMX enabled by default  
Using config: /home/thuantv/apache-zookeeper-3.5.10-bin/bin/../conf/zoo.cfg  
Client port found: 2182. Client address: localhost. Client SSL: false.  
Mode: standalone  
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-zookeeper-3.5.10-bin#
```

Start Drillbit:

```
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.2# ./bin/drillbit.sh start  
Starting drillbit, logging to /home/thuantv/apache-drill-1.21.2/log/drillbit.out  
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.2# ./bin/drillbit.sh status  
drillbit is running.  
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.2# ./bin/drillbit.sh status  
drillbit is running.
```

Kết quả:

The screenshot shows the Apache Drill web interface in a browser window. The address bar shows 'localhost:8047'. The interface has a top navigation bar with links: Apache Drill, Query, Profiles, Storage, Metrics, Threads, Logs, Options, and Documentation. Below this is a 'Drillbits' section with a count of 1. A table lists the Drillbit details:

#	Address	Heap Memory Usage	Direct Memory Usage	CPU Usage	Avg Sys Load	User Port	Control Port	Data Port	Version	Status	Uptime	Shutdown
1	thuantv-VMware-Virtual-Platform	0.22GB (5% of 4GB)	0GB (0% of 0GB)	0.27%	0.11	31010	31011	31012	1.21.2	ONLINE	3m 38s	

Below the table, there are two sections: 'Encryption' and 'Query Throttling'. Under 'Encryption', 'Client to Bit Encryption' and 'Bit to Bit Encryption' are both 'Disabled'. Under 'Query Throttling', 'Queue Status' is 'Disabled'.

Lỗi khi khởi chạy Drillbit thứ 2

(`drillbit.pid` file is present but drillbit is not running`)

```
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh start
Starting drillbit, logging to /home/thuantv/apache-drill-1.21.1/log/drillbit.out
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
drillbit is running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
drillbit is running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
drillbit is running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
drillbit is running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
drillbit is running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1# ./bin/drillbit.sh status
/home/thuantv/apache-drill-1.21.1/drillbit.pid file is present but drillbit is not running.
root@thuantv-VMware-Virtual-Platform:/home/thuantv/apache-drill-1.21.1#
```