



# Hands on Drill

Time to Query Complex Data Instantly on Hadoop

```
{ "name" : "Aditya Kishore", "github" : "adityakishore", "twitter" : "@adiore" }
```

January 29, 2015 – Big Data Everywhere, Seattle





# Aditya Kishore

Software Engineer – MapR Technologies

Committer & PMC Member – Apache Drill Project








# Time to Query Complex Data Instantly on Hadoop

## Leading the No-schema Revolution








# AGENDA

-  What is Drill
-  Installing & configuring Drill
-  Introducing Data Sets for the Demo
-  Demo
-  What's Next?



# AGENDA

-  What is Drill
-  Installing & configuring Drill
-  Introducing Data Sets for the Demo
-  Demo
-  What's Next?










# APACHE DRILL

- Distributed query engine
- Dynamic schema
- Scale-out
- Low latency
- Extensible
- Industry-standard APIs: ANSI SQL, ODBC/JDBC, REST



# AGENDA

-  What is Drill
-  Installing & configuring Drill
-  Introducing Data Sets for the Demo
-  Demo
-  What's Next?



# Running Drill Takes 10 minutes

DOWNLOAD

• <https://drill.apache.org/download/>

EXTRACT

• 

```
$ tar xf apache-drill-0.7.0.tar.gz  
$ cd apache-drill-0.7.0
```

RUN

• 

```
$ bin/sqlline -u jdbc:drill:zk=local
```

QUERY

In SQL

• 

```
> SELECT full_name, position_title, salary  
FROM cp.`employee.json`  
LIMIT 1;
```

RESULT






• 

```
+-----+-----+-----+  
| full_name | position_title | salary |  
+-----+-----+-----+  
| Sheri Nowmer | President | 80000.0 |  
+-----+-----+-----+  
1 row selected (0.417 seconds)
```





# AGENDA

-  What is Drill
-  Installing & configuring Drill
-  Introducing Data Sets for the Demo
-  Demo
-  What's Next?





TPC-H



YELP

# Datasets For the Demo



Name	Kind
clicks.json	JSON
customers.tbl	comm...values
▼ yelp	Folder
business.json	JSON
checkin.json	JSON
review.json	JSON
tip.json	JSON
user.json	JSON






	A	B	C	D	E	F	G	
1	c_custkey	c_name	c_address	c_nationkey	c_phone	c_acctbal	c_mktsegment	
2	1	Customer#00	IVhziApeRb of	15	25-989-741-25	711.56	BUILDING	to the
3	2	Customer#00	XSTf4,NCwDV	13	23-768-687-36	121.65	AUTOMOBILE	l accou
4	3	Customer#00	MG9kdTD2WB	1	11-719-748-33	7498.12	AUTOMOBILE	depos
5	4	Customer#00	XxVSJsLAGtn	4	14-128-190-59	2866.83	MACHINERY	reque
6	5	Customer#00	KvpyuHCplrB8	3	13-750-942-63	794.47	HOUSEHOLD	n acco
7	6	Customer#00	sKZz0CsnMD7	20	30-114-968-49	7638.57	AUTOMOBILE	tions. e
8	7	Customer#00	TcGe5gaZNgV	18	28-190-982-97	9561.95	AUTOMOBILE	ainst th
9	8	Customer#00	IOB10bB0Aymi	17	27-147-574-93	6819.74	BUILDING	among
10	9	Customer#00	xKiAFTjUsCux	8	18-338-906-36	8324.07	FURNITURE	r theoc
11	10	Customer#00	6LrEaV6KR6PL	5	15-741-346-98	2753.54	HOUSEHOLD	es regu
12	11	Customer#00	PkWS 3HIXqw	23	33-464-151-34	-272.6	BUILDING	ckages
13	12	Customer#00	9PWKuhzT4Zr	13	23-791-276-12	3396.49	HOUSEHOLD	to the

```
{
  "votes": {"funny": 0, "useful": 2, "cool": 1},
  "user_id": "Xqd0DzHaiyRqVH3WRG7hzig",
  "review_id": "15SdjuK7DmYqUAj6rjGowg",
  "stars": 5,
  "date": "2007-05-17",
  "text": "dr. goldberg offers everything ...",
  "type": "review",
  "business_id": "vcNAWiLM4dR7D2nwwJ7nCA"
}
```

**APACHE HBASE**  orders



# AGENDA

-  What is Drill
-  Installing & configuring Drill
-  Introducing Data Sets for the Demo
-  Demo
-  What's Next?



DEMO

# business.json (1)

```
{
  "business_id": "4bEj0yTaDG24SY5TxsaUNQ",
  "full_address": "3655 Las Vegas Blvd S\nThe Strip\nLas Vegas, NV 89109",
  "hours": {
    "Monday": {"close": "23:00", "open": "07:00"},
    "Tuesday": {"close": "23:00", "open": "07:00"},
    "Friday": {"close": "00:00", "open": "07:00"},
    "Wednesday": {"close": "23:00", "open": "07:00"},
    "Thursday": {"close": "23:00", "open": "07:00"},
    "Sunday": {"close": "23:00", "open": "07:00"},
    "Saturday": {"close": "00:00", "open": "07:00"}
  },
  "open": true,
  "categories": ["Breakfast & Brunch", "Steakhouses", "French", "Restaurants"],
  "city": "Las Vegas",
  "review_count": 4084,
  "name": "Mon Ami Gabi",
  "neighborhoods": ["The Strip"],
  "longitude": -115.172588519464,
```



## business.json (2)

```
"state": "NV",
"stars": 4.0,
"attributes": {
  "Alcohol": "full_bar",
  "Noise Level": "average",
  "Has TV": false,
  "Attire": "casual",
  "Ambience": {
    "romantic": true,
    "intimate": false,
    "touristy": false,
    "hipster": false,
    "classy": true,
    "trendy": false,
    "casual": false
  },
  "Good For": {"dessert": false, "latenight": false, "lunch": false,
    "dinner": true, "breakfast": false, "brunch": false},
}
```



# ● AGENDA

- What is Drill
- Installing & configuring Drill
- Introducing Data Sets for the Demo
- Use Cases
- What's Next?





# Next Steps for Drill

Security &  
Access  
Control

1

JSON in  
Any Shape  
or Form

2

Improved  
Multi  
Tenancy

3

New Data  
Sources

4



# Participate

- Learn: <http://drill.apache.org/>
- Download: <http://drill.apache.org/download/>
- Ask Questions: [user@drill.apache.org](mailto:user@drill.apache.org)
- Engage on Twitter: [@ApacheDrill](https://twitter.com/ApacheDrill)



Q&A

# Thank You

@mapr



maprtech

mapr-technologies



MapRTechnologies

aditya@mapr.com  
adi@apache.org



maprtech

