Apache Drill: interactive query and analysis on large-scale datasets

Michael Hausenblas, Chief Data Engineer EMEA, MapR
NoSQL matters Training Day, 2013-04-25
Agenda

• Introduction round (15min)
• Overview (45min)
• Interactive session (~3h)
  – Use Cases
  – Architecture
  – Hands-on
  – Wrap-up and mini-presentations
INTRODUCTION ROUND
Hi - I'm Michael -
I work as Chief Data Engineer at MapR in Europe
and contribute to Apache Drill

Get in touch:
@MichaelHausenblas
via Twitter

MichaelHausenblas@gmail.com
via mail

Right: Now, let's drill into Today's topic -
(pun intended)

Kudos to http://cmx.io/
... and who are you?
OVERVIEW
Which workloads do you encounter in your environment?
Batch processing

... for recurring tasks such as large-scale data mining, aggregation, ETL offloading, etc.
OLTP

... for example user-facing eCommerce transactions, real-time messaging at scale (FB), etc.
Stream processing

... in order to handle stream sources such as social media feeds or sensor data (mobile phones, RFID, weather stations, etc.)
... retrieval of items from semi-structured data formats (XML, JSON, etc.), documents (plain text, etc.) and datastores (MongoDB, CouchDB, etc.)
But what about interactive, ad-hoc query at scale?
Interactive Query (?)

- Hive
- Apache Drill
- Teradata Aster
- Greenplum
- Oracle MySQL
- Impala
- Citusdata
- HaAdapt

low-latency
Today’s Solutions

• RDBMS-focused
  – ETL data from MongoDB and Hadoop
  – Query data using SQL

• MapReduce-focused
  – ETL from RDBMS and MongoDB
  – Use Hive, etc.
And now for something completely different …
Google’s Dremel

Dremel is a scalable, interactive ad-hoc query system for analysis of read-only nested data. By combining multi-level execution trees and columnar data layout, it is capable of running aggregation queries over trillion-row tables in seconds. The system scales to thousands of CPUs and petabytes of data, and has thousands of users at Google.

…”

http://research.google.com/pubs/pub36632.html

Google’s Dremel

multi-level execution trees

columnar data layout
Google’s Dremel

nested data + schema

mapping nested data to tables

column-striped representation
Google’s Dremel

experiments:
datasets & query performance
Back to Apache Drill ...
Apache Drill—key facts

• Inspired by Google’s **Dremel**
• Standard **SQL 2003** support
• Plug-able **data sources**
• **Nested data** is a first-class citizen
• **Schema is optional**
• **Community** driven, **open**, 100’s involved
High-level Architecture
INTERACTIVE SESSION
Use Case I

- Jane, a marketing analyst
- Determine target segments
- Data from different sources
Use Case II

- Logistics – supplier status
- Queries
  - How many shipments from supplier X?
  - How many shipments in region Y?

<table>
<thead>
<tr>
<th>SUPPLIER_ID</th>
<th>NAME</th>
<th>REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>ACME Corp</td>
<td>US</td>
</tr>
<tr>
<td>GAL</td>
<td>GotALot Inc</td>
<td>US</td>
</tr>
<tr>
<td>BAP</td>
<td>Bits and Pieces Ltd</td>
<td>Europe</td>
</tr>
<tr>
<td>ZUP</td>
<td>Zu Pli</td>
<td>Asia</td>
</tr>
</tbody>
</table>

```json
{
  "shipment": 100123,
  "supplier": "ACM",
  "timestamp": "2013-02-01",
  "description": "first delivery today"
}
{
  "shipment": 100124,
  "supplier": "BAP",
  "timestamp": "2013-02-02",
  "description": "hope you enjoy it"
}
...
Your Use Case
Requirements

• Support for different data sources
• Support for different query interfaces
• Low-latency/real-time
• Ad-hoc queries
• Scalable, reliable
High-level Architecture

• Each node: **Drillbit** - maximize data locality
• Co-ordination, query planning, execution, etc, are **distributed**
• By default Drillbits hold all roles
• Any node can act as endpoint for a query
High-level Architecture

- **Zookeeper** for ephemeral cluster membership info
- **Distributed cache** (Hazelcast) for metadata, locality information, etc.
High-level Architecture

- **Originating Drillbit** acts as foreman, manages query execution, scheduling, locality information, etc.
- Streaming data **communication** avoiding SerDe
Principled Query Execution

Source Query → Parser → Logical Plan → Optimizer → Physical Plan → Execution

SQL 2003
DrQL
MongoQL
DSL

query: [
  {
    @id: "log",
    op: "sequence",
    do: [
      {
        op: "scan",
        source: "logs"
      },
      {
        op: "filter",
        condition: "x > 3"
      }
    ]
  }
]

scanner API
Logical Plan example

Drillbit Modules

RPC Endpoint

Logical Plan
- SQL
- HiveQL
- Pig
- Mongo

Optimizer

Physical Plan
- Scheduler
- Foreman
- Operators

Storage Engine Interface
- DFS Engine
- HBase Engine

Distributed Cache

Parser

APACHE DRILL

MAPR Technologies
Key Features

- Full SQL 2003
- Nested data
- Optional schema
- Extensibility points
Full SQL – ANSI SQL 2003

• SQL-like is often not enough
• Integration with existing tools
  – Datameer, Tableau, Excel, SAP Crystal Reports
  – Use standard ODBC/JDBC driver
Nested Data

• Nested data becoming prevalent
  – JSON/BSON, XML, ProtoBuf, Avro
  – Some data sources support it natively (MongoDB, etc.)

• Flattening nested data is error-prone

• Extension to ANSI SQL 2003
Optional Schema

• Many data sources don’t have rigid schemas
  – Schema changes rapidly
  – Different schema per record (e.g. HBase)
• Supports queries against unknown schema
• User can define schema or via discovery

<table>
<thead>
<tr>
<th>Row Key</th>
<th>CF contents</th>
<th>CF anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;com.cnn.www&quot;</td>
<td>contents:html = &quot;&lt;html&gt;...&quot;</td>
<td>anchor:my.look.ca = &quot;CNN.com&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>anchor:cnnsi.com = &quot;CNN&quot;</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Extensibility Points

- Source query $\rightarrow$ parser API
- Custom operators, UDF $\rightarrow$ logical plan
- Serving tree, CF, topology $\rightarrow$ physical plan/optimizer
- Data sources & formats $\rightarrow$ scanner API
... and Hadoop?

• HDFS can be a data source

• Complementary use cases*
  
• ... use Apache Drill
    – Find record with specified condition
    – Aggregation under dynamic conditions

• ... use MapReduce
  – Data mining with multiple iterations
  – ETL

* https://cloud.google.com/files/BigQueryTechnicalWP.pdf
Hands-on

data source: donuts.json

query: 

```json
{ 
  "id": "0001",
  "type": "donut",
  "ppu": 0.55,
  "batters": 
  [ 
    "batter": 
      [ 
        "id": "1001", "type": "Regular" 
      ],
      [ "id": "1002", "type": "Chocolate" ]
  ]
}
```

logical plan: simple_plan.json

result: out.json

https://cwiki.apache.org/confluence/display/DRILL/Demo+HowTo
Implement Your Usage Scenario
Status

• Heavy development by multiple organizations

• Available
  – Logical plan (ADSP)
  – Reference interpreter
  – Basic SQL parser
  – Basic demo
Status

April 2013

• Extend SQL support
• Physical plan
• In-memory compressed data interfaces
• Distributed execution
• HBase and MySQL storage engine
Contributing

• Learn where and how to contribute
  https://cwiki.apache.org/confluence/display/DRILL/Contributing

• Jira, Git, Apache build and test tools

• Preparing for dependencies
  – Hazelcast
  – Netflix Curator
Contributing

General contributions appreciated:

• Test data & test queries

• Use case scenarios (textual desc./SQL queries)

• Documentation
Contributing

- Dremel-inspired columnar format
  - Twitter’s Parquet
  - Hive’s ORC file

- Integration with Hive metastore (?)

- **DRILL-13** Storage Engine: Define Java Interface

- **DRILL-15** Build HBase storage engine implementation
Contributing

• **DRILL-48** RPC interface for query submission and physical plan execution

• **DRILL-53** Setup cluster configuration and membership mgmt system

• Further schedule
  – Alpha Q2
  – Beta Q3
Kudos to ...

• Julian Hyde, Pentaho
• Lisen Mu
• Tim Chen, Microsoft
• Chris Merrick, RJMetrics
• David Alves, UT Austin
• Sree Vaadi, SSS/NGData
• Jacques Nadeau, MapR
• Ted Dunning, MapR
Wrap-up & mini-presentations
Engage!

- Follow @ApacheDrill on Twitter

- Sign up at mailing lists (user | dev)
  http://incubator.apache.org/drill/mailing-lists.html

- Standing G+ hangouts every Tuesday at 18:00 CET

- Keep an eye on http://drill-user.org/