DrillProposal

Drill

Abstract

Drill is a distributed system for interactive analysis of large-scale datasets, inspired by Google's Dremel.

Proposal

Drill is a distributed system for interactive analysis of large-scale datasets. Drill is similar to Google's Dremel, with the additional flexibility needed to support a broader range of query languages, data formats and data sources. It is designed to efficiently process nested data. It is a design goal to scale to 10,000 servers or more and to be able to process petabyes of data and trillions of records in seconds.

Background

Many organizations have the need to run data-intensive applications, including batch processing, stream processing and interactive analysis. In recent years open source systems have emerged to address the need for scalable batch processing (Apache Hadoop) and stream processing (Storm, Apache S4). In 2010 Google published a paper called "Dremel: Interactive Analysis of Web-Scale Datasets," describing a scalable system used internally for interactive analysis of nested data. No open source project has successfully replicated the capabilities of Dremel.

Rationale

There is a strong need in the market for low-latency interactive analysis of large-scale datasets, including nested data (eg, JSON, Avro, Protocol Buffers). This need was identified by Google and addressed internally with a system called Dremel.

In recent years open source systems have emerged to address the need for scalable batch processing (Apache Hadoop) and stream processing (Storm, Apache S4). Apache Hadoop, originally inspired by Google's internal MapReduce system, is used by thousands of organizations processing large-scale datasets. Apache Hadoop is designed to achieve very high throughput, but is not designed to achieve the sub-second latency needed for interactive data analysis and exploration. Drill, inspired by Google's internal Dremel system, is intended to address this need.

It is worth noting that, as explained by Google in the original paper, Dremel complements MapReduce-based computing. Dremel is not intended as a replacement for MapReduce and is often used in conjunction with it to analyze outputs of MapReduce pipelines or rapidly prototype larger computations. Indeed, Dremel and MapReduce are both used by thousands of Google employees.

Like Dremel, Drill supports a nested data model with data encoded in a number of formats such as JSON, Avro or Protocol Buffers. In many organizations nested data is the standard, so supporting a nested data model eliminates the need to normalize the data. With that said, flat data formats, such as CSV files, are naturally supported as a special case of nested data.

The Drill architecture consists of four key components/layers:

- Query languages: This layer is responsible for parsing the user's query and constructing an execution plan. The initial goal is to support the SQL-like language used by Dremel and Google BigQuery, which we call DrQL. However, Drill is designed to support other languages and programming models, such as the Mongo Query Language, Cascading or Plume.
- Low-latency distributed execution engine: This layer is responsible for executing the physical plan. It provides the scalability and fault tolerance needed to efficiently query petabytes of data on 10,000 servers. Drill's execution engine is based on research in distributed execution engines (eg, Dremel, Dryad, Hyracks, CIEL, Stratosphere) and columnar storage, and can be extended with additional operators and connectors.
- Nested data formats: This layer is responsible for supporting various data formats. The initial goal is to support the column-based format used by Dremel. Drill is designed to support schema-based formats such as Protocol Buffers/Dremel, Avro/AVRO-806/Trevni and CSV, and schema-less formats such as JSON, BSON or YAML. In addition, it is designed to support column-based formats such as Dremel, AVRO-806/Trevni and RCFile, and row-based formats such as Protocol Buffers, Avro, JSON, BSON and CSV. A particular distinction with Drill is that the execution engine is flexible enough to support column-based processing as well as row-based processing. This is important because column-based processing can be much more efficient when the data is stored in a column-based format, but many large data assets are stored in a row-based format that would require conversion before use.
- Scalable data sources: This layer is responsible for supporting various data sources. The initial focus is to leverage Hadoop as a data source.

It is worth noting that no open source project has successfully replicated the capabilities of Dremel, nor have any taken on the broader goals of flexibility (eg, pluggable query languages, data formats, data sources and execution engine operators/connectors) that are part of Drill.

Initial Goals

The initial goals for this project are to specify the detailed requirements and architecture, and then develop the initial implementation including the execution engine and DrQL. Like Apache Hadoop, which was built to support multiple storage systems (through the FileSystem API) and file formats (through the InputFormat OutputFormat APIs), Drill will be built to support multiple query languages, data formats and data sources. The initial implementation of Drill will support the DrQL and a column-based format similar to Dremel.

Current Status

Significant work has been completed to identify the initial requirements and define the overall system architecture. The next step is to implement the four components described in the Rationale section, and we intend to do that development as an Apache project.

Meritocracy

We plan to invest in supporting a meritocracy. We will discuss the requirements in an open forum. Several companies have already expressed interest in this project, and we intend to invite additional developers to participate. We will encourage and monitor community participation so that privileges can be extended to those that contribute. Also, Drill has an extensible/pluggable architecture that encourages developers to contribute various extensions, such as query languages, data formats, data sources and execution engine operators and connectors. While some companies will surely develop commercial extensions, we also anticipate that some companies and individuals will want to contribute such extensions back to the project, and we look forward to fostering a rich ecosystem of extensions.

Community

The need for a system for interactive analysis of large datasets in the open source is tremendous, so there is a potential for a very large community. We believe that Drill's extensible architecture will further encourage community participation. Also, related Apache projects (eg, Hadoop) have very large and active communities, and we expect that over time Drill will also attract a large community.

Core Developers

The developers on the initial committers list include experienced distributed systems engineers:

- Tomer Shiran has experience developing distributed execution engines. He developed Parallel DataSeries, a data-parallel version of the open source DataSeries system. He is also the author of Applying Idealized Lower-bound Runtime Models to Understand Inefficiencies in Data-intensive Computing (SIGMETRICS 2011). Tomer worked as a software developer and researcher at IBM Research, Microsoft and HP Labs, and is now at MapR Technologies. He has been active in the Hadoop community since 2009.
- Jason Frantz was at Clustrix, where he designed and developed the first scale-out SQL database based on MySQL. Jason developed the distributed query optimizer that powered Clustrix. He is now a software engineer and architect at MapR Technologies.
- Ted Dunning is a PMC member for Apache ZooKeeper and Apache Mahout, and has a history of over 30 years of contributions to open source. He is now at MapR Technologies. Ted has been very active in the Hadoop community since the project's early days.
- MC Srivas is the co-founder and CTO of MapR Technologies. While at Google he worked on Google's scalable search infrastructure. MC Srivas has been active in the Hadoop community since 2009.
- Chris Wensel is the founder and CEO of Concurrent. Prior to founding Concurrent, he developed Cascading, an Apache-licensed open source application framework enabling Java developers to quickly and easily develop robust Data Analytics and Data Management applications on Apache Hadoop. Chris has been involved in the Hadoop community since the project's early days.
- Keys Botzum was at IBM, where he worked on security and distributed systems, and is currently at MapR Technologies.
- Gera Shegalov was at Oracle, where he worked on networking, storage and database kernels, and is currently at MapR Technologies.
- Ryan Rawson is the VP Engineering of Drawn to Scale where he developed Spire, a real-time operational database for Hadoop. He is also a committer and PMC member for Apache HBase, and has a long history of contributions to open source. Ryan has been involved in the Hadoop community since the project's early days.

We realize that additional employer diversity is needed, and we will work aggressively to recruit developers from additional companies.

Alignment

The initial committers strongly believe that a system for interactive analysis of large-scale datasets will gain broader adoption as an open source, community driven project, where the community can contribute not only to the core components, but also to a growing collection of query languages and optimizers, data formats, data formats, and execution engine operators and connectors. Drill will integrate closely with Apache Hadoop. First, the data will live in Hadoop. That is, Drill will support Hadoop FileSystem implementations and HBase. Second, Hadoop-related data formats will be supported (eg, Apache Avro, RCFile). Third, MapReduce-based tools will be provided to produce column-based formats. Fourth, Drill tables can be registered in HCatalog. Finally, Hive is being considered as the basis of the DrQL implementation.

Known Risks

Orphaned Products

The contributors are leading vendors in this space, with significant open source experience, so the risk of being orphaned is relatively low. The project could be at risk if vendors decided to change their strategies in the market. In such an event, the current committers plan to continue working on the project on their own time, though the progress will likely be slower. We plan to mitigate this risk by recruiting additional committers.

Inexperience with Open Source

The initial committers include veteran Apache members (committers and PMC members) and other developers who have varying degrees of experience with open source projects. All have been involved with source code that has been released under an open source license, and several also have experience developing code with an open source development process.

Homogenous Developers

The initial committers are employed by a number of companies, including MapR Technologies, Concurrent and Drawn to Scale. We are committed to recruiting additional committers from other companies.

Reliance on Salaried Developers

It is expected that Drill development will occur on both salaried time and on volunteer time, after hours. The majority of initial committers are paid by their employer to contribute to this project. However, they are all passionate about the project, and we are confident that the project will continue even if no salaried developers contribute to the project. We are committed to recruiting additional committers including non-salaried developers.

Relationships with Other Apache Products

As mentioned in the Alignment section, Drill is closely integrated with Hadoop, Avro, Hive and HBase in a numerous ways. For example, Drill data lives inside a Hadoop environment (Drill operates on in situ data). We look forward to collaborating with those communities, as well as other Apache communities.

An Excessive Fascination with the Apache Brand

Drill solves a real problem that many organizations struggle with, and has been proven within Google to be of significant value. The architecture is based on academic and industry research. Our rationale for developing Drill as an Apache project is detailed in the Rationale section. We believe that the Apache brand and community process will help us attract more contributors to this project, and help establish ubiquitous APIs. In addition, establishing consensus among users and developers of a Dremel-like tool is a key requirement for success of the project.

Documentation

Drill is inspired by Google's Dremel. Google has published a paper highlighting Dremel's innovative nested column-based data format and execution engine.

Initial Source

The requirement and design documents are currently stored in MapR Technologies' source code repository. They will be checked in as part of the initial code dump. Check out the Drill slides.pdf.

Cryptography

Drill will eventually support encryption on the wire. This is not one of the initial goals, and we do not expect Drill to be a controlled export item due to the use of encryption.

Required Resources

Mailing List

- drill-private
- drill-dev
- drill-user

Subversion Directory

Git is the preferred source control system: git://git.apache.org/drill

Issue Tracking

JIRA Drill (DRILL)

Initial Committers

- Tomer Shiran <tshiran at maprtech dot com>
- Ted Dunning <tdunning at apache dot org>
- Jason Frantz <jfrantz at maprtech dot com>
- MC Srivas <mcsrivas at maprtech dot com>
- Chris Wensel <chris and concurrentinc dot com>
- Keys Botzum <kbotzum at maprtech dot com>
- Gera Shegalov <gshegalov at maprtech dot com>
- Ryan Rawson <ryan at drawntoscale dot com>

Affiliations

The initial committers are employees of MapR Technologies, Drawn to Scale and Concurrent. The nominated mentors are employees of MapR Technologies, Lucid Imagination and Nokia.

Sponsors

Champion

Ted Dunning (tdunning at apache dot org)

Nominated Mentors

- Ted Dunning <tdunning at apache dot org> – Chief Application Architect at MapR Technologies, Committer for Lucene, Mahout and ZooKeeper.
- Grant Ingersoll <grant at lucidimagination dot com> – Chief Scientist at Lucid Imagination, Committer for Lucene, Mahout and other projects.
- Isabel Drost <isabel at apache dot org> – Software Developer at Nokia Gate 5 GmbH, Committer for Lucene, Mahout and other projects.

Sponsoring Entity

Incubator