



Apache Drill

INTERACTIVE, AD-HOC QUERY AT SCALE

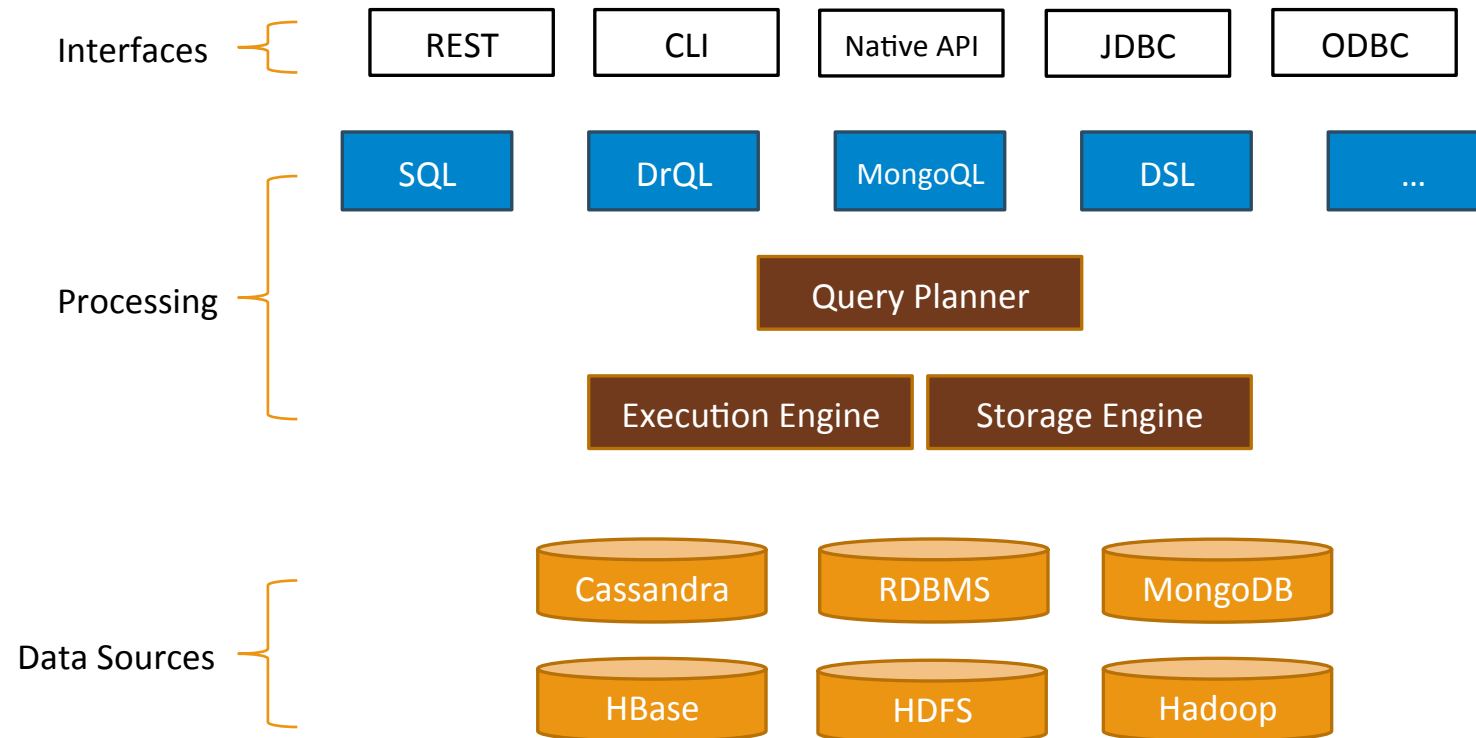
Present by Jian Fang



What is Apache Drill

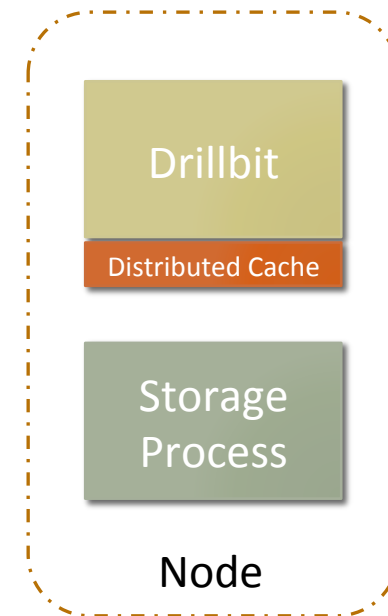
- Inspired by Google Dremel
- Interactive analysis framework of large-scale datasets
- Supports a nested data model
- Plug-able data sources
- Standard SQL 2003 support

Architecture



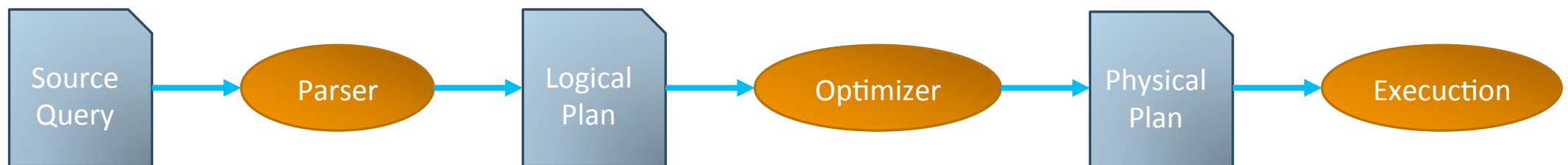
Architecture

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



Query Execution

- Source query → Parser API
- Custom operators, UDF → Logical Plan
- Serving tree, CF, topology → Physical Plan/Optimizer
- Data sources & formats → Scanner API





Status

- Under development, alpha version
- Current available features
 - Logical Plan
 - Reference interpreter
 - Basic SQL parser
 - Basic demo
- More on <http://incubator.apache.org/drill/index.html>



Thanks