

DataFusion Donation - IP Clearance

Background

12 people (other than myself) contributed to the original DataFusion POC which was based on a fork of Apache Arrow 0.10

On 11/26/18 I deleted all Rust source code from the master branch and started implementing DataFusion from scratch, leveraging some of the work that had happened in the Arrow 0.11 and 0.12 releases (such as the new `array_ops` code for primitive compute on arrays including min/max/sum and basic math ops). Some work (particularly the query planner) did use existing code from the original POC.

The SQL tokenizer and parser code in the original POC was split out into a separate crate (<https://github.com/andygrove/sqlparser-rs>) and is not being donated.

Audit of Contributions

Committer: Malu

I have attempted to contact malu about signing an ICLA but have not heard back yet, so I removed the contribution, which implemented the query planner and execution of SQL LIMIT clauses. The DataFusion rewrite had included the query planner part but not the execution yet. Around 50 lines of code were removed.

There was also a bug fix to the CREATE EXTERNAL TABLE code which was part of the donation so I removed all code related to that feature.

Commit ID	Change	Status
c5d9108f3fa4461804a117e465b56d29fdec4dcf	Handle sql query LIMIT src/exec.rs src/parser.rs src/re1.rs src/sqltore1.rs	Removed from donation
62e90b77bcac1123833800bb5e57689b9b464472	Fixed SQL parser issue re CREATE EXTERNAL TABLE (not part of donation - sql parser was split out to a separate crate)	Removed from donation

	src/parser.rs	
55702048d17a2b515ca71f906bd94a0bf945fedf	Fixed a SQL parser test src/parser.rs	not part of donation - sql parser was split out to a separate crate
840c25dd261a276816121efe04e730f6ac3342b6	Fixed a SQL parser test src/parser.rs	not part of donation - sql parser was split out to a separate crate

Committer: Srenatus

Commit ID	Change	Status
6c898776b463b12b4b5c016f8d216d41f0f45934	MIN and MAX with nulls src/functions/max.rs src/functions/min.rs	Not part of donation - Array min max got reimplemented directly in arrow and DF was reimplemented to delegate to Arrow
29cb4b25b926a6a9d5d5cab8a9a9f7e1f3b95de0	Bug fix in console src/bin/console/linereader.rs	Console is not part of donation
ad6ddb5408945ce4c62148594d482eb9e89263dc	Fix typo in min/max src/functions/min.rs	Not part of donation - Array min max got reimplemented directly in arrow

Committer: KenSuenobo

Commit ID	Change	Status
ed7b098c074d49bba0d760fed4205a696b4b2761	Cci quickfix .circleci/config.yml	Not part of donation (switched to Travis CI)
2ae47327cbd220dbeec250ab16e96e3e5a2b645d	CircleCI 2.0 Build Rules	Not part of donation (switched to Travis CI)

	<code>.circleci/config.yaml</code> <code>.circleci/config.yml</code> <code>scripts/circle/*.sh</code> <code>scripts/docker/build.sh</code>	
010e9a1d6ab82599f9695fcb64e27fb58239e915	Docker packaging improvements <code>.travis-docker.yml</code>	Not part of donation (only Rust source is donated, no Docker images)

Committer: Gnieto

Commit ID	Change	Status
b657d480de22f013af18c20dbb002846eb454e32	Add support for wildcard in projection <code>src/sqlplanner.rs</code>	Not part of donation - this feature is not yet implemented in the DF rewrite
8f12eefdf8c50e95bb02002e78c6488e06bda33b	Fix bug in SQL console <code>src/bin/console/main.rs</code>	SQL Console is not part of donation

Committer: Hntd187

Commit ID	Change	Status
0166d4ad57570e465b2d992984bbd5851a5d7bad	First pass of basic operators finished <code>src/bin/console/main.rs</code> <code>src/datasource/csv.rs</code> <code>src/exec.rs</code> <code>src/lib.rs</code> <code>src/sqlparser.rs</code> <code>src/sqltokenizer.rs</code> <code>src/types.rs</code> <code>test/data/expected/*</code> <code>tests/all_types.rs</code>	Not part of donation - Arrow now has array_ops implemented as part of compute module and DataFusion rewrite uses that
b685bf613d3305b727733701babd5a66263d3efa	Sum Aggregate function <code>Cargo.toml</code>	Not part of donation DataFusion aggregate logic

	src/bin/console/linereader.rs src/bin/console/main.rs src/exec.rs src/functions/mod.rs src/functions/sum.rs src/lib.rs src/relations/aggregate.rs src/sqlplanner.rs src/types.rs test/data/expected/* tests/all_types.rs	was rewritten from scratch and now delegates to Arrow for sum operations.
--	--	---

Committer: Sunchao

Has signed ICLA already.

Commit ID	Change	Status
79c48fff2429fa52b5720255bfc6be51feb70f0a	Add travis-cargo to CI and enable code coverage	Part of donation - ICLA in place.
b8857a33b3493634ef027c88d3210a2a619f0f81	Add code coverage for CI	Part of donation - ICLA in place.

Committer: Sidred

Commit ID	Change	Status
	use lazy static to create hashmap for sql keywords at compile time Cargo.toml src/lib.rs src/sqltokenizer.rs	Sql parser - not part of donation
	Smoketest script scripts/smoketest.sh	Not part of donation

Committer: Maccam912

Commit ID	Change	Status
555f02bfd65a15834de2b8c70f20f2b248becb6c	got criterion benchmarks working again benches/dataframe_bench.rs benches/sql_bench.rs	Benchmarks are not part of the donation
81ecab3d41c1ac81994ae50cb36eacbe2ce3b3a2	use criterion to benchmark performance Cargo.toml benches/dataframe_bench.rs benches/sql_bench.rs src/exec.rs	Benchmarks are not part of the donation

Committer: Pmaciolek

Commit ID	Change	Status
93c533061b0620fbbf7d33785ed732b0b596ce3a	Fix Utf8 to Float64 cast datatype src/exec.rs	Not part of donation. Cast was reimplemented.

Committer: Atouchet

Commit ID	Change	Status
0837d1ea3302b709a3619f82991abd46633b4879	Capitalization and spelling fixes (to README) README.md	Not relevant

Committer: Mgxmx

Commit ID	Change	Status
01e530c7f3dc3645e01fc35b2a ce300457f4159d	Implement a new console Cargo.toml src/bin/console/linereader.rs src/bin/console/main.rs	SQL Console is not part of the donation

Committer: bgard6977

Commit ID	Change	Status
6e3c552f9c061657db299c1cfe 3f64f67cdd26cb	Clean up gitignore .gitignore	Not relevant