

Status update 16/09/2016

From Spark CSVs to FiloDB



FiloDB + Cassandra

- Replaces (CSV)DataFrames
- Solves issue with appending to DataFrames
- Only primitive types (dates to longs)

<https://github.com/tuplejump/FiloDB>

Issues and optimizations

- Cassandra JVM increasing garbage collection times (Young and Old)
- Cache edge weights (shared edges)
- Partitioning/Segmentation in FiloDB/Cassandra
- Rootpost DataFrame → X
- DataFrame join order
- Cache inserts until select
- Remove data from selects (simplified validator)

FAIR scheduling

- Spark scheduling
- Making code thread-safe
- Synchronized voids vs fine grained locks
- 'Consistent' world view discussion