

## Deep specification and verification of SQL compilation chain

Junior Conference on Data Science and Engineering

September 2017

#### Data systems and languages, proofs

Until now there have been little usages of formal methods towards SQL and DBMS, yet, the compilation chain of SQL queries is complex.



### Certified parsing, from SQL to SQLCoq

SELECT s.sname FROM sailors s WHERE EXISTS (SELECT \* FROM reserves r WHERE  $\rightarrow$  r.bid = 103 AND s.sid = r.sid);

- SQL parser in menhir/OCamllex
- certified using the CompCert parser generator
- automatic translation into SQLCoq (in OCaml)

# Involving DBMS, from SQL to unified execution plan

- PostgreSQL and Oracle query execution plan parser
- both translated to our own unified plan format

#### From unified plan to Extended Algebra

- Extended Algebra (ExtAlg) is similar to Relationnal Algebra but more powerful
- translation from an unified plan to an ExtAlg expression
- automatic annotation of ExtAlg expression with algorithms
- ability to execute our Coq version of those algorithms



- translation from SQLCoq to ExtAlg
- two ExtAlg expressions, one from the DBMS, one from the SQL query/SQLCoq
- proof that two ExtAlg expression have the same semantic
- strong guarantees about the fact that the DBMS kept the semantic of the SQL query
- algorithms used by the DBMS have not been proved yet
- Coq version of some algorithms

