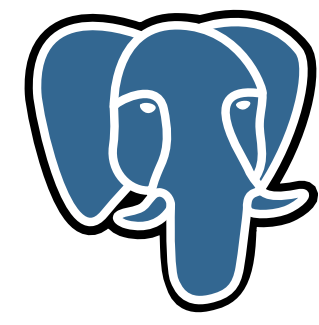


PROVSQL: PROVENANCE AND PROBABILITY MANAGEMENT IN POSTGRESQL

Pierre Senellart, Louis Jachiet, Silviu Maniu, Yann Ramusat



ProvSQL in Brief

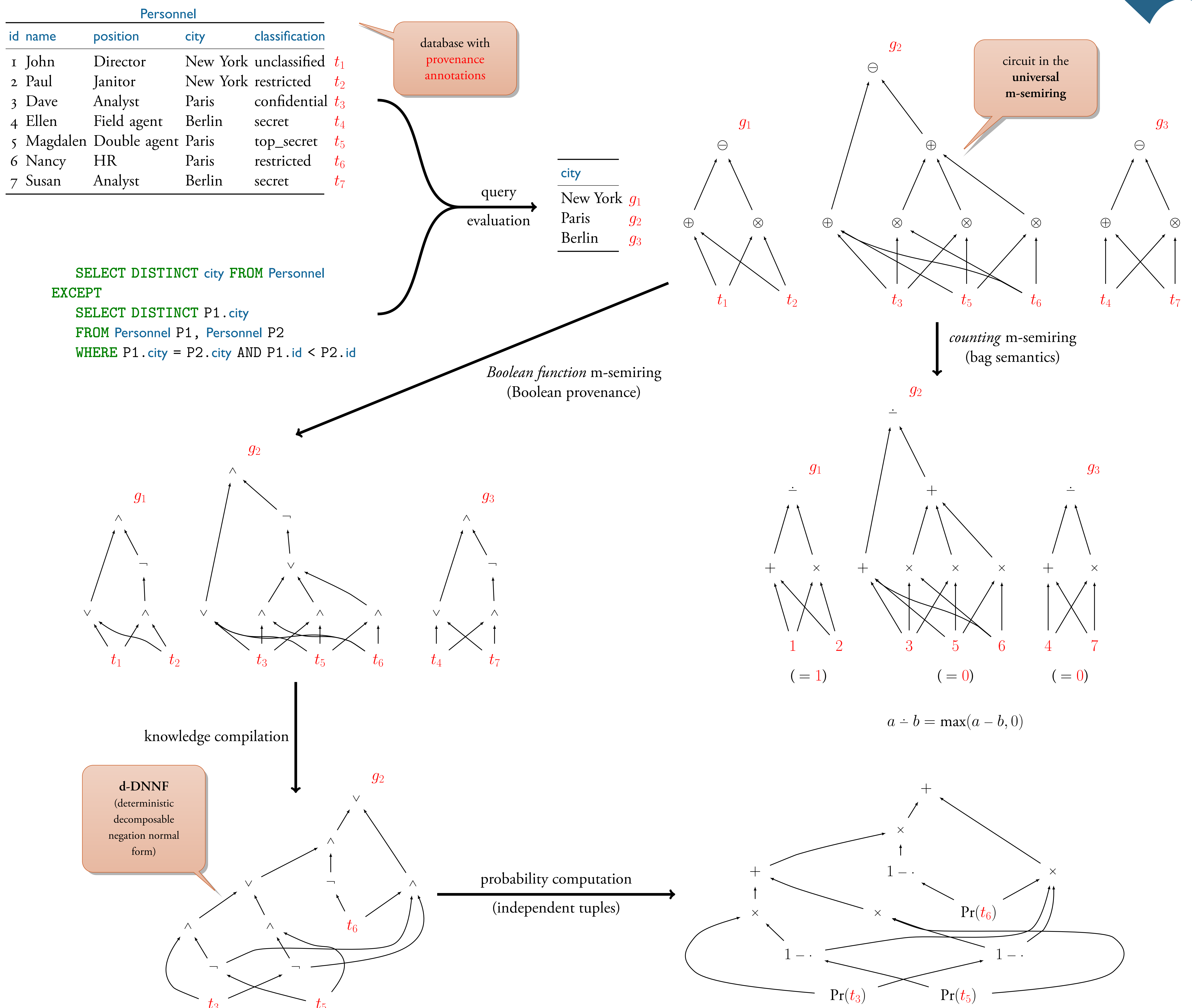


- Light-weight extension for PostgreSQL ≥ 9.5
- Transparent computation of provenance of SQL queries as a term algebra circuit
- Allows probabilistic query evaluation with various techniques:
 - Enumeration of possible worlds
 - Knowledge compilation to d-DNNF via external tools (c2d, Dsharp, d4)
 - Monte-Carlo sampling
- Open-source and available at <https://github.com/PierreSenellart/provsql>

Supported Queries and Provenance Annotations

- Large subset of (nested) SQL, without aggregates or UDFs:
 - **SELECT ... FROM ... WHERE, JOIN, DISTINCT, GROUP BY**
 - **UNION, UNION ALL, EXCEPT**
- Variety of provenance formalisms:
 - arbitrary provenance semirings
 - arbitrary provenance semirings with monus (m-semirings)
 - where-provenance

How ProvSQL Computes Various Provenance Forms and Probabilities



References

- [1] Pierre Senellart. Provenance and probabilities in relational databases: From theory to practice. *SIGMOD Record*, 46(4):5–15, 2017.
- [2] Pierre Senellart, Louis Jachiet, Silviu Maniu, and Yann Ramusat. ProvSQL: Provenance and probability management in PostgreSQL. *Proceedings of the VLDB Endowment (PVLDB)*, 11(12):2034–2037, 2018. Demonstration.