Good news: ***Effective Syntax*** [2]

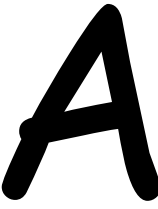
[1] W. Fan, F. Geerts, Y. Cao, T. Deng, P. Lu: Querying Big Data by Accessing Small Data, PODS 2015

- Input: A Query Q and an *access schema* $\mathcal{A}$
- Question: Is Q boundedly evaluable with

10

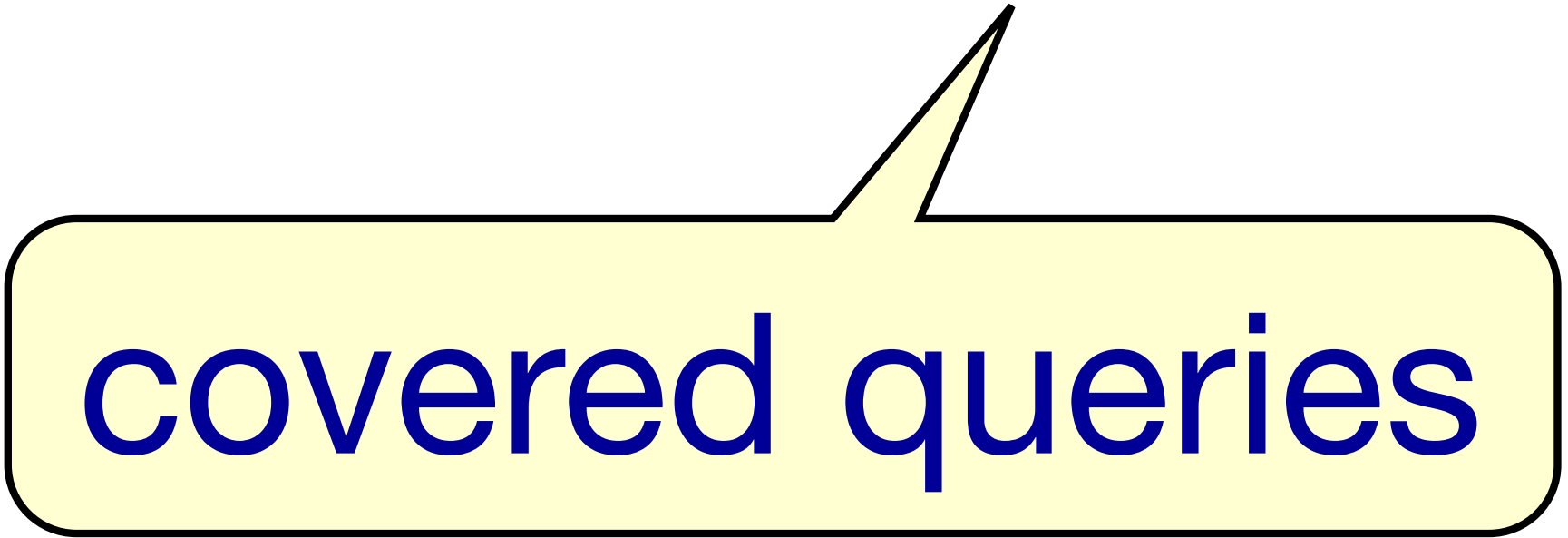
Bounded Evaluability

a set of access constraints

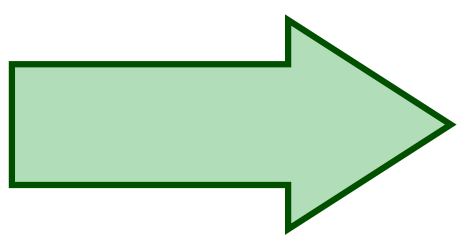


No matter how desirable, the problem is hard [1]

- Undecidable for first-order logic (relational algebra) queries
- EXPSPACE-hard for conjunctive queries (SPC queries)


covered queries

- a **proof** of Q is covered



bounded plan for $Q^{[3]}$

Reducing undecidable to PTIME without loosing expressive power up to equivalence

[2] Y. Cao, W. Fan: An Effective Syntax for Bounded Relational Queries, SIGMOD 2016

- FO Q is bounded *iff* $Q \equiv Q'$ and Q' is covered
- PTIME to check whether Q is covered

[3] Y. Cao, W. Fan, W. Yu: Bounded Conjunctive Queries, VLDB 2014

Bounded Evaluability

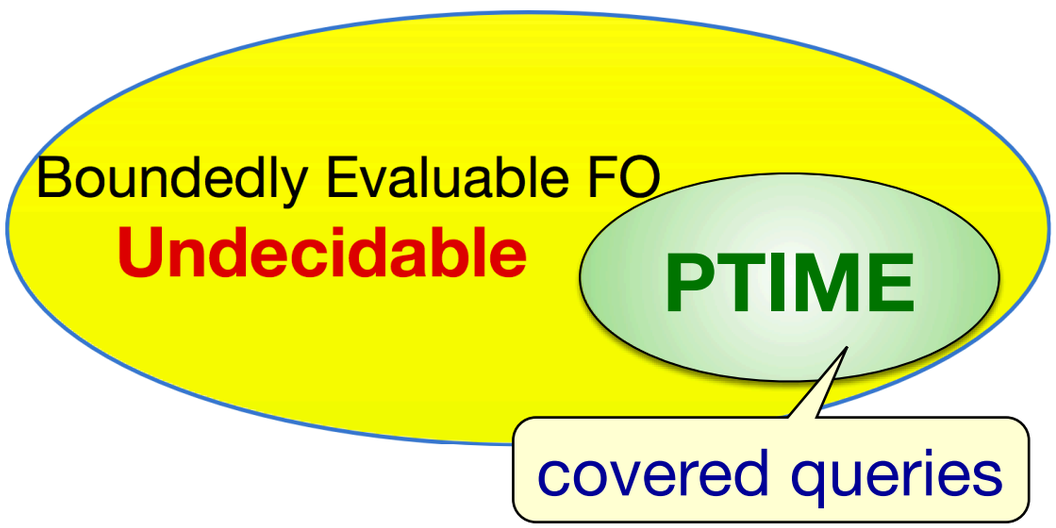
a set of access constraints

- Input: A Query Q and an *access schema* $\mathcal{A}$
- Question: Is Q boundedly evaluable with $\mathcal{A}$

No matter how desirable, the problem is hard [1]

- Undecidable for first-order logic (relational algebra) queries
- EXPSPACE-hard for conjunctive queries (SPC queries)

Good news: ***Effective Syntax*** [2]



- FO Q is bounded *iff* $Q \equiv Q'$ and Q' is covered
- PTIME to check whether Q is covered

Reducing undecidable to PTIME without loosing expressive power up to equivalence

- a **proof** of Q is covered ⇨ **bounded plan** for Q [3]

[1] W. Fan, F. Geerts, Y. Cao, T. Deng, P. Lu: Querying Big Data by Accessing Small Data, PODS 2015
[2] Y. Cao, W. Fan: An Effective Syntax for Bounded Relational Queries, SIGMOD 2016
[3] Y. Cao, W. Fan, W. Yu: Bounded Conjunctive Queries, VLDB 2014

Effectiveness of Bounded Evaluation

- Huawei query workload

Query Type	Query Number	Original Complexity	Accelerate Ratio
Bounded query without join	2	Low	Medium
Simple aggregate query without join	2	Low	Medium
Join on key attributes	4	Medium	Low
Join on non-key attributes	3	High	High
Unbounded Query	2	-	-

- Performance comparison with DBMS

	Bounded non-join	Aggregate non-join	Join on Key	Join on non-key
Query count:	2	2	4	2
PostgreSQL:	1.58 sec	7.53 sec	629.96 sec	$>$ 3 hours
BEAS:	0.01 sec	0.02 sec	15.14 sec	0.04 sec
Accelerate ratio:	158	376.5	41.61	$> 10^5$