

Interroger efficacement des bases de données relationnelles avec SPARQL et Ontop

Benjamin Cogrel

KRDB Research Centre for Knowledge and Data
Université Libre de Bozen-Bolzano, Italie

 Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen-Bolzano

Optique

SemWeb.Pro 2015, Paris

Étapes

- 1 Pourquoi utiliser SPARQL



Statoil

Optique

- 2 Comment reformuler une requête SPARQL vers SQL

ontop

Analyse de données pour l'exploration pétrolière

Construction de modèles stratigraphiques

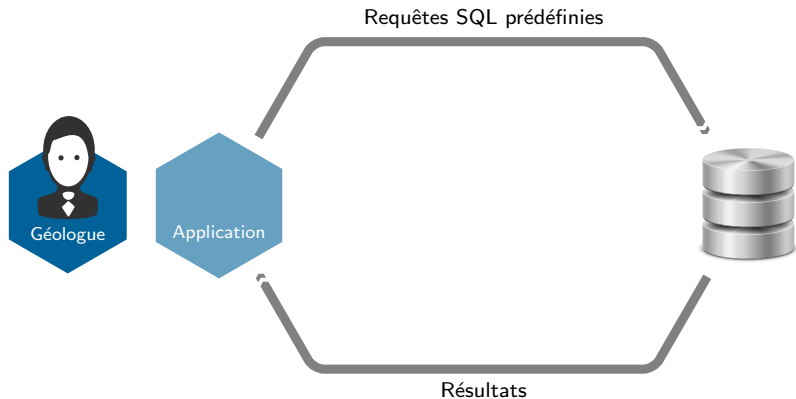
- Stratégique
- Env. 900 géologues et géophysiciens



Source principale : *Exploitation and Production Data Store (EPDS)*

- Données d'exploration et de production
- Env. 1500 tables
- Env. 1600 vues matérialisées
- Centaines de gigaoctets

Utilisation simple : requêtes prédéfinies

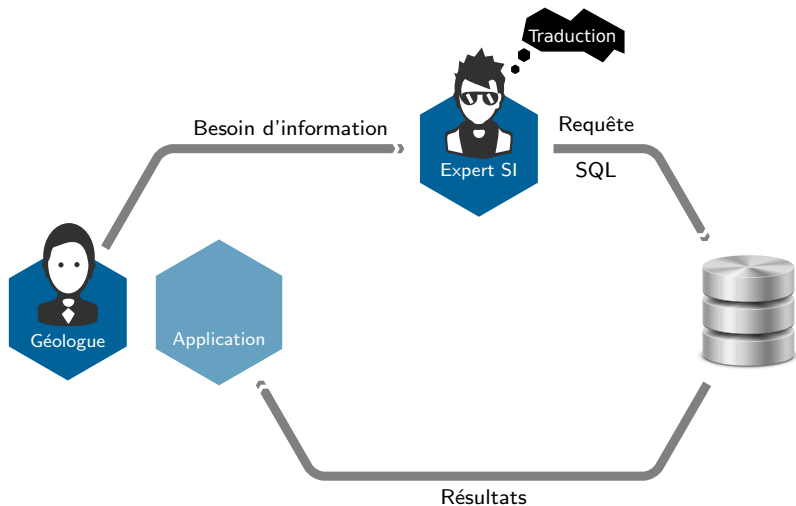


Nouvelle requête : besoins utilisateurs (simplifiés)

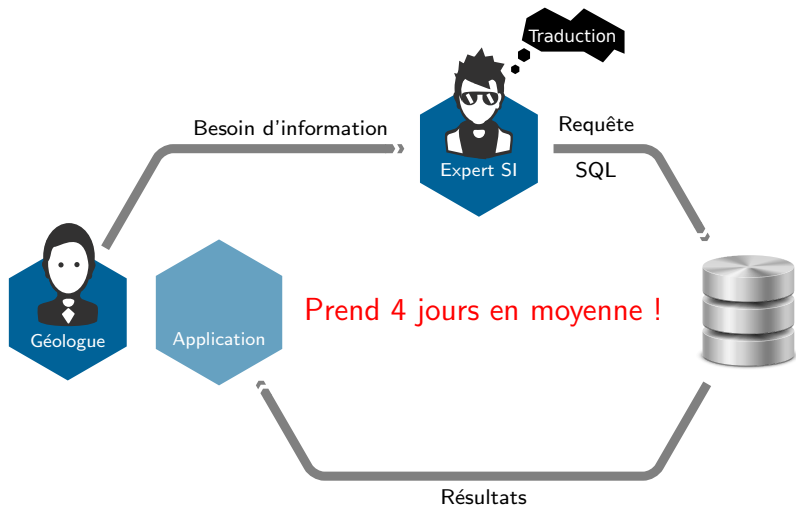
Tous les puits norvégiens de tel type
à proximité de tel lieu ayant une
perméabilité proche de telle valeur.
Attributs : date de mise en service, pro-
fondeur, etc.



Définition d'une nouvelle requête SQL



Définition d'une nouvelle requête SQL



Requête SQL typique (extrait anonymisé)

```

SELECT [...]
FROM
db_name.table1 table1,
db_name.table2 table2a,
db_name.table2 table2b,
db_name.table3 table3a,
db_name.table3 table3b,
db_name.table3 table3c,
db_name.table3 table3d,
db_name.table4 table4a,
db_name.table4 table4b,
db_name.table4 table4c,
db_name.table4 table4d,
db_name.table4 table4e,
db_name.table4 table4f,
db_name.table5 table5a,
db_name.table5 table5b,
db_name.table6 table6a,
db_name.table6 table6b,
db_name.table7 table7a,
db_name.table7 table7b,
db_name.table8 table8,
db_name.table9 table9,
db_name.table10 table10a,
db_name.table10 table10b,
db_name.table10 table10c,
db_name.table11 table11,
db_name.table12 table12,
db_name.table13 table13,
db_name.table14 table14,
db_name.table15 table15,
db_name.table16 table16
WHERE [...]

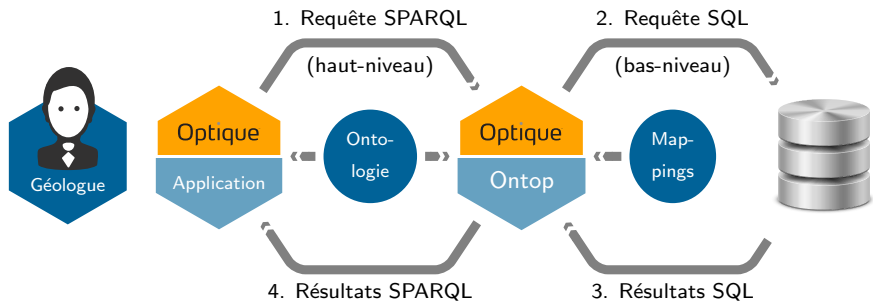
table2a.attr1='keyword' AND
table3a.attr2=table10c.attr1 AND
table3a.attr6=table6a.attr3 AND
table3a.attr9='keyword' AND
table4a.attr10 IN ('keyword') AND
table4a.attr1 IN ('keyword') AND
table5a.kinds=table4a.attr13 AND
table5b.kinds=table4c.attr74 AND
table5b.name='keyword' AND
(table6a.attr19=table10c.attr17 OR
(table6a.attr2 IS NULL AND
table10c.attr4 IS NULL)) AND
table6a.attr14=table5b.attr14 AND
table6a.attr2='keyword' AND
(table6b.attr14=table10c.attr8 OR
(table6b.attr4 IS NULL AND
table10c.attr7 IS NULL)) AND
table6b.attr19=table5a.attr55 AND
table6b.attr2='keyword' AND
table7a.attr19=table2b.attr19 AND
table7a.attr17=table15.attr19 AND
table4b.attr11='keyword' AND
table8.attr19=table7a.attr80 AND
table8.attr19=table13.attr20 AND
table8.attr4='keyword' AND
table9.attr10=table16.attr11 AND
table3b.attr19=table10c.attr18 AND
table3b.attr22=table12.attr63 AND
table3b.attr66='keyword' AND
table10a.attr54=table7a.attr8 AND
table10a.attr70=table10c.attr10 AND
table10a.attr16=table4d.attr11 AND
table4c.attr99='keyword' AND
table4c.attr1='keyword' AND

table11.attr10=table5a.attr10 AND
table11.attr40='keyword' AND
table11.attr50='keyword' AND
table2b.attr1=table1.attr8 AND
table2b.attr9 IN ('keyword') AND
table2b.attr2 LIKE 'keyword'% AND
table12.attr9 IN ('keyword') AND
table7b.attr1=table2a.attr10 AND
table3c.attr13=table10c.attr1 AND
table3c.attr10=table6b.attr20 AND
table3c.attr13='keyword' AND
table10b.attr16=table10a.attr7 AND
table10b.attr11=table7b.attr8 AND
table10b.attr13=table4b.attr89 AND
table13.attr1=table2b.attr10 AND
table13.attr20='keyword' AND
table13.attr15='keyword' AND
table3d.attr49=table12.attr18 AND
table3d.attr18=table10c.attr11 AND
table3d.attr14='keyword' AND
table4d.attr17 IN ('keyword') AND
table4d.attr19 IN ('keyword') AND
table16.attr28=table11.attr56 AND
table16.attr16=table10b.attr178 AND
table16.attr5=table14.attr56 AND
table4e.attr34 IN ('keyword') AND
table4e.attr48 IN ('keyword') AND
table4f.attr89=table5b.attr7 AND
table4f.attr45 IN ('keyword') AND
table4f.attr1='keyword' AND
table10c.attr2=table4e.attr19 AND
(table10c.attr78=table12.attr56 OR
(table10c.attr55 IS NULL AND
table12.attr17 IS NULL))

```


Automatisation avec la plateforme Optique

Ontology-Based Data Access (OBDA)



Requête SPARQL de haut-niveau

Niveau conceptuel

- Indépendant du stockage
- Terminologie du domaine formalisée (ontologie)
- Requêtes concises

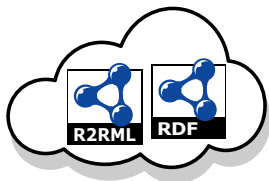
```
SELECT DISTINCT ?wellbore ?wc
                ?well ?length
WHERE {
    ?wellbore npdv:wellboreForDiscovery
                ?discovery;
                npdv:belongsToWell ?well.
    ?wc npdv:coreForWellbore ?wellbore;
        npdv:coresTotalLength ?length.

    FILTER (?length < 56796)
}
```

SPARQL et les bases de données relationnelles

Transformation d'une BD relationnelle en un graphe RDF

- Pour être requêtée avec SPARQL
- Mappings SQL-RDF (R2RML)
- Contraintes de l'ontologie (opt.)
- Virtuelle ou matérielle



Graphe RDF matérialisé

- Matérialisation du graphe RDF saturé
- Maintenance
- + Profils d'ontologie expressifs (comme OWL 2 RL)

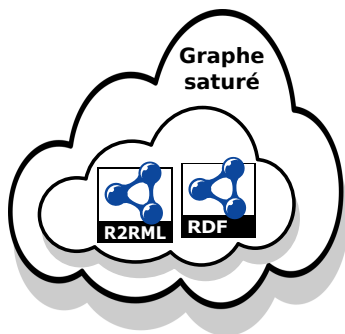
Graphe RDF virtuel

- Reformulation de requête
- + Aucune matérialisation
- Profil d'ontologie moins expressif (OWL 2 QL) *

SPARQL et les bases de données relationnelles

Transformation d'une BD relationnelle en un graphe RDF

- Pour être requêtée avec SPARQL
- Mappings SQL-RDF (R2RML)
- Contraintes de l'ontologie (opt.)
- Virtuelle ou matérielle



Graphe RDF matérialisé

- Matérialisation du graphe RDF saturé
- Maintenance
- + Profils d'ontologie expressifs (comme OWL 2 RL)

Graphe RDF virtuel

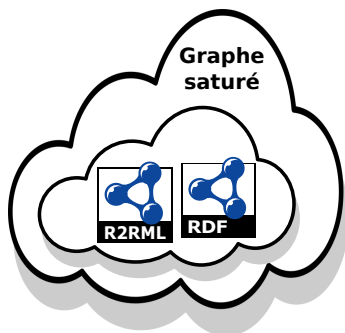
- Reformulation de requête
- + Aucune matérialisation
- Profil d'ontologie moins expressif (OWL 2 QL) *

(*) Inclut toutefois un mécanisme d'inférence absent du profil OWL 2 RL

SPARQL et les bases de données relationnelles

Transformation d'une BD relationnelle en un graphe RDF

- Pour être requêtée avec SPARQL
- Mappings SQL-RDF (R2RML)
- Contraintes de l'ontologie (opt.)
- Virtuelle ou matérielle



Graphe RDF matérialisé

- Matérialisation du graphe RDF saturé
- Maintenance
- + Profils d'ontologie expressifs (comme OWL 2 RL)

Graphe RDF virtuel

- Reformulation de requête
- + Aucune matérialisation
- Profil d'ontologie moins expressif (OWL 2 QL) *

(*) Inclut toutefois un mécanisme d'inférence absent du profil OWL 2 RL

Mapping SQL-RDF

Format natif d'Ontop

Source (SQL)

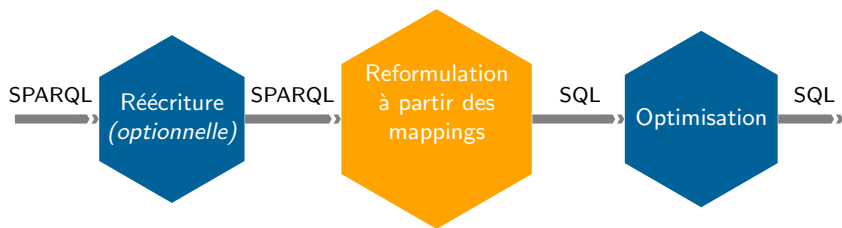
```
SELECT wellbore_s, year, month, day
FROM wellbore
WHERE wellbore.r_existence_kd_nm = 'actual'
```

Cible (RDF, Turtle-like)

```
:Wellbore-{wellbore_s} :completionDate
                        "{year}-{month}-{day}"^^xsd:date
```

Format équivalent au standard R2RML (*supporté, utilisé dans Optique*)

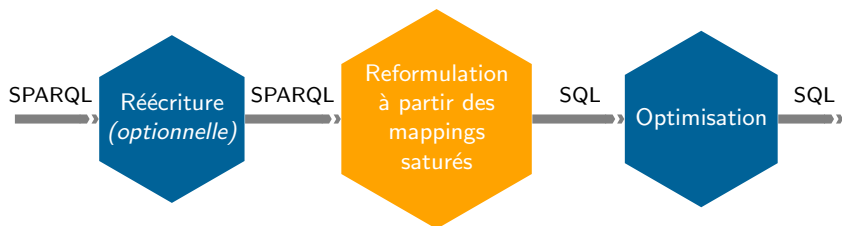
Reformulation de requête



Rôle de l'ontologie OWL 2 QL

- Marginal : réécriture de la requête SPARQL (*cas très particuliers*)
- Principal : saturation des mappings (*hors-ligne*)

Reformulation de requête



Rôle de l'ontologie OWL 2 QL

- Marginal : réécriture de la requête SPARQL (*cas très particuliers*)
- Principal : saturation des mappings (*hors-ligne*)

Optimisation de la requête SQL

Objectif : obtenir une requête SQL...

- Proche de celles écrites par des humains
- Adaptée aux planificateurs existants

Optimisations structurelles

- Jointures d'unions vers unions de jointures
- Décomposition des IRIs pour faciliter les jointures

Optimisations sémantiques

- Élimination des unions et des jointures redondantes
- À partir de contraintes fonctionnelles

Contraintes fonctionnelles

- Clés primaires, uniques ou étrangères
- Implicites dans les processus métiers (*Statoil*)
- **Vitales pour la reformulation de requête !**

Conclusion

Ontop

- Logiciel libre (Apache 2)
- Interfaces Sesame et OWLAPI
- Extension Protégé

Quelques expérimentations

- Agrégation SPARQL
- owl:sameAs via les mappings
- Support partiel de SWRL
- NoSQL

Intégration

- Plateforme Optique (FluidOps)
- Stardog 4.0 (graphes virtuels)

Liens

- Github : [ontop/ontop](https://github.com/ontop/ontop)
- ontop4obda@googlegroups.com
- Twitter : @ontop4obda
- <http://optique-project.eu>

Profil d'ontologie OWL 2 QL

Principales fonctionnalités

- Hiérarchies de classes (`rdfs:subClassOf`) et de propriétés (`rdfs:subPropertyOf`)
- Signature de propriétés (`rdfs:domain`, `rdfs:range`)
- Propriétés inverses (`owl:inverseOf`)
- Disjonction de classes (`owl:disjointWith`)
- Participation obligatoire (avancé)

Manques

- Identité d'individus (`owl:sameAs`)
- Contraintes de cardinalité (propriété fonctionnelle, etc.)
- Et bien d'autres

En résumé

- Ontologies légères
- Un peu plus que RDFS
- Traductibilité garantie vers SQL