Relational Databases Ingestion into a NoSQL Data Warehouse

Authors : Rym JEMMALI
Fatma ABDELHEDI
Gilles ZURFLUH

Presented By: Rym JEMMALI
Toulouse Institute of Computer Science Research (IRIT),
CBI²- Trimane, Paris, France

Email: rym.jemmali@trimane.fr

3-7/10/2021

Presenter's Resume

Rym JEMMALI

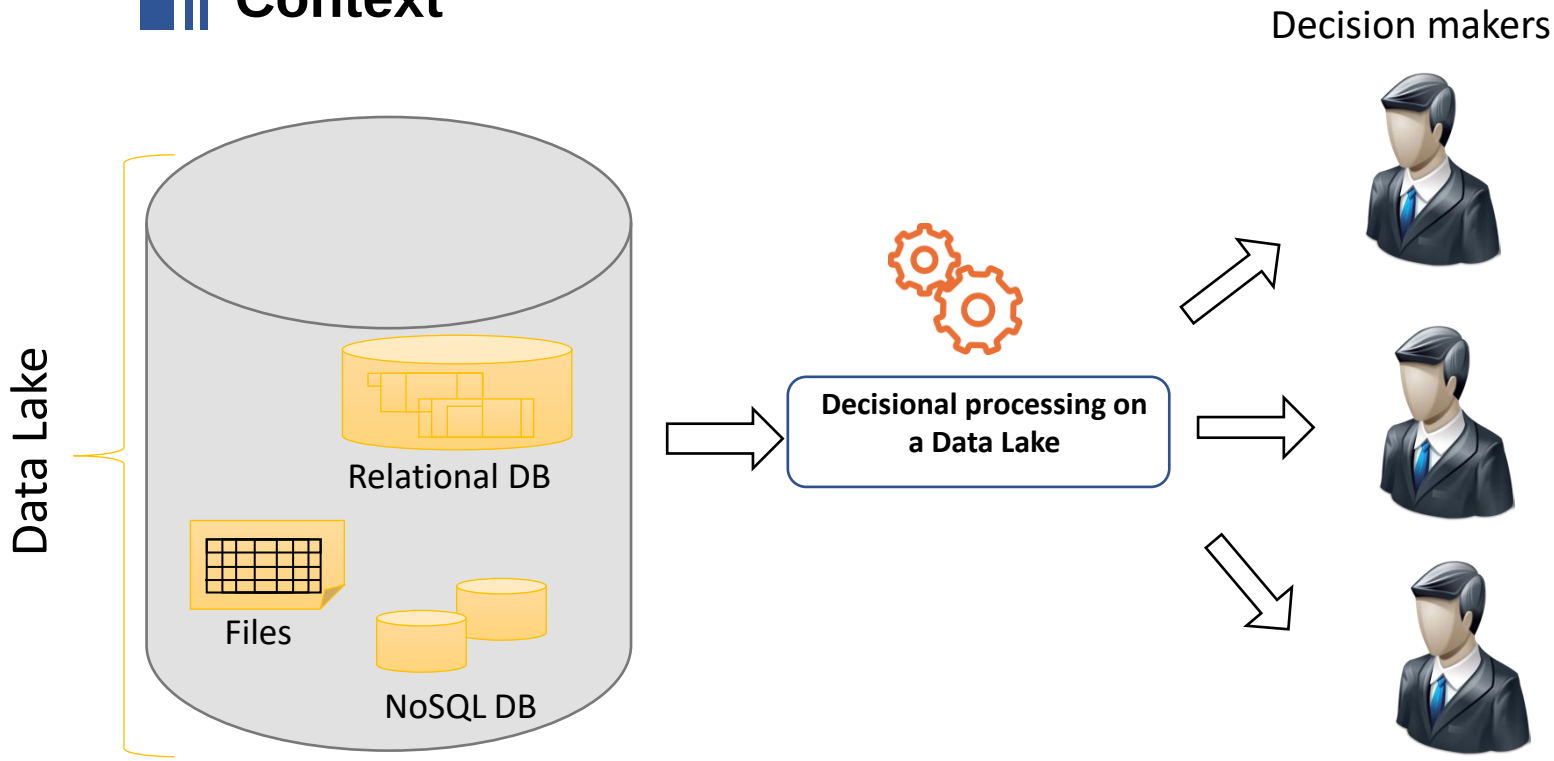
- Engineer diploma in computer science (2019)
- 2nd year PhD student : PhD in computer science, Big Data and Business Intelligence

CBI²- Trimane, Paris, France
&
Toulouse Institute of Computer Science Research (IRIT)

Summary

1. Context & issue
2. Case study
3. Related work
4. Contribution & implementation
5. Conclusion & perspectives

Context


Decision makers
Data Lake
Relational DB
Files
NoSQL DB
Decisional processing on a Data Lake

Issue

- The diversity of data types and formats
- The volumes stored which can reach several terabytes
- The raw nature of the data in a Data Lake

Decisional processing on a Data Lake

Case study

Medical application

Hospitalizations
Analysis

IRM
Radiology
Interpretation

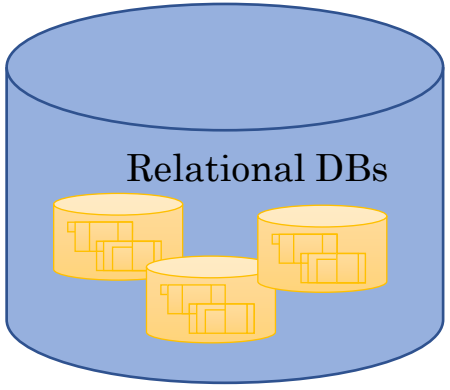
ENS

Diagnosis
Prescriptions

Health insurance
companies

Case study

Medical application



ENS

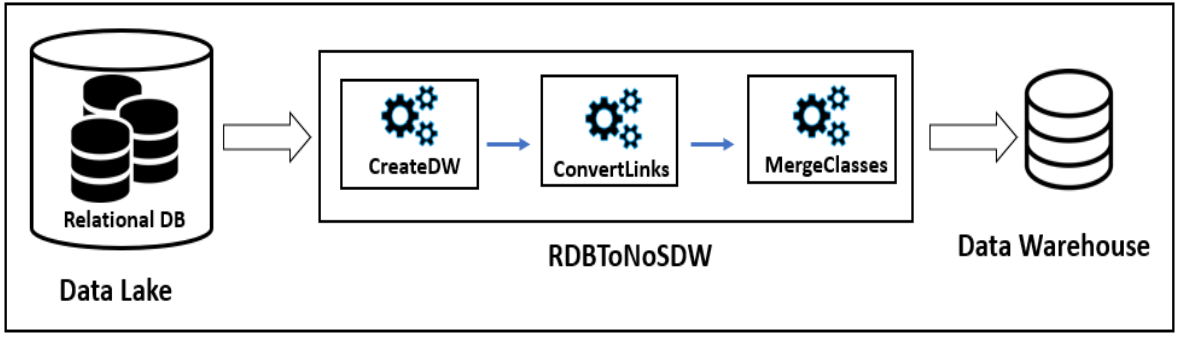
Create a Data Warehouse from a Data Lake

Private health insurance companies

Related work

	Data ingestion from a Data Lake					No data ingestion (without restructuring data)
	Data management			Links processing		Polystores
	Proposing metamodels	Data transferring	Restructuring Data	Document nesting	References	
Diamantini et al. (2018)			X (linking data and create a graph)			
Candel et al. (2021)	X					
Hanine et al. (2015)			X	X		
Stanescu et al. (2016)			X	X		
Chikerur et al. (2015)			X	No	No	
Liyanaarachchi et al. (2016)		X		X	X	
Duggan et al. (2015)						X
Alotaibi et al. (2020)						X

Data ingestion process



Architecture of the data ingestion process from a Data Lake: RDBToNoSDW

MDA (Model Driven Architecture)

- **OMG** (Object Management Group)
- **Objectives :** Functional specifications **VS** technical specifications
- **Principle**
  - ✓ Use of models
  - ✓ Automation of transformations between models
- **Concepts**
  - ✓ Model
  - ✓ Meta-model

Implementation tools

EMF (Eclipse Modeling Framework)

- ✓ Definition of a metamodel
- ✓ Creation of models
- ✓ Automation of transformations

Technical tools

- ✓ Ecore
- ✓ XMI
- ✓ QVT

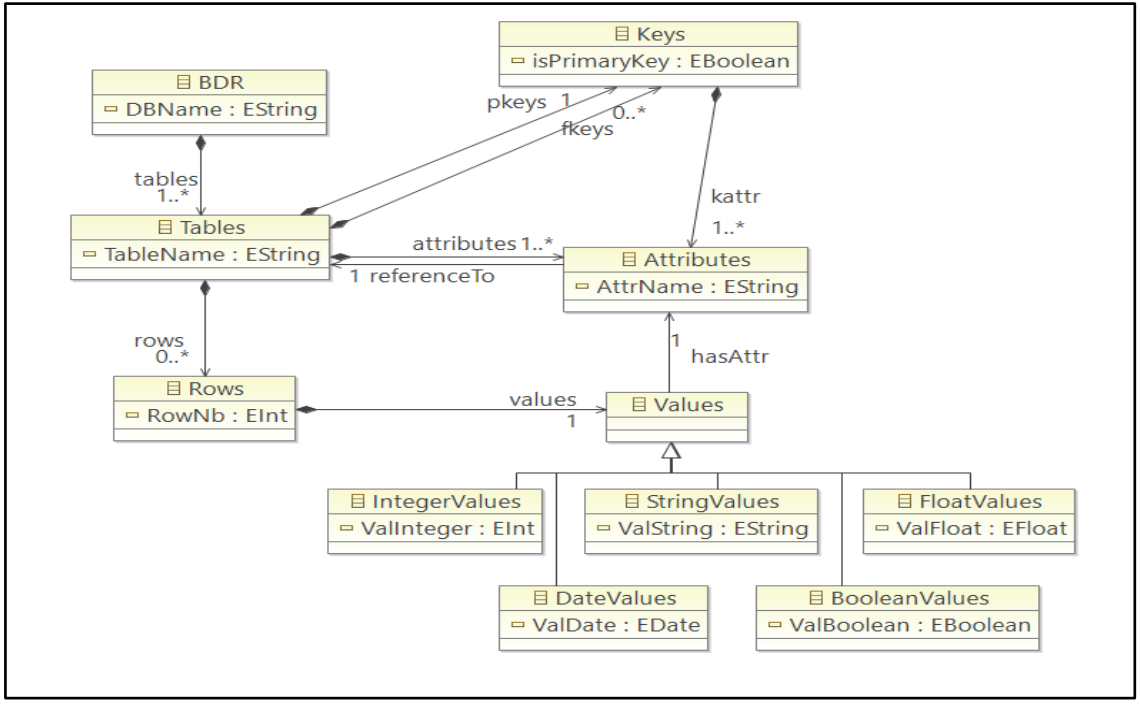
Implementation tools

Eclipse IDE (Java Integrated Development Environment)

- ✓ Java coding
- ✓ Algorithmic

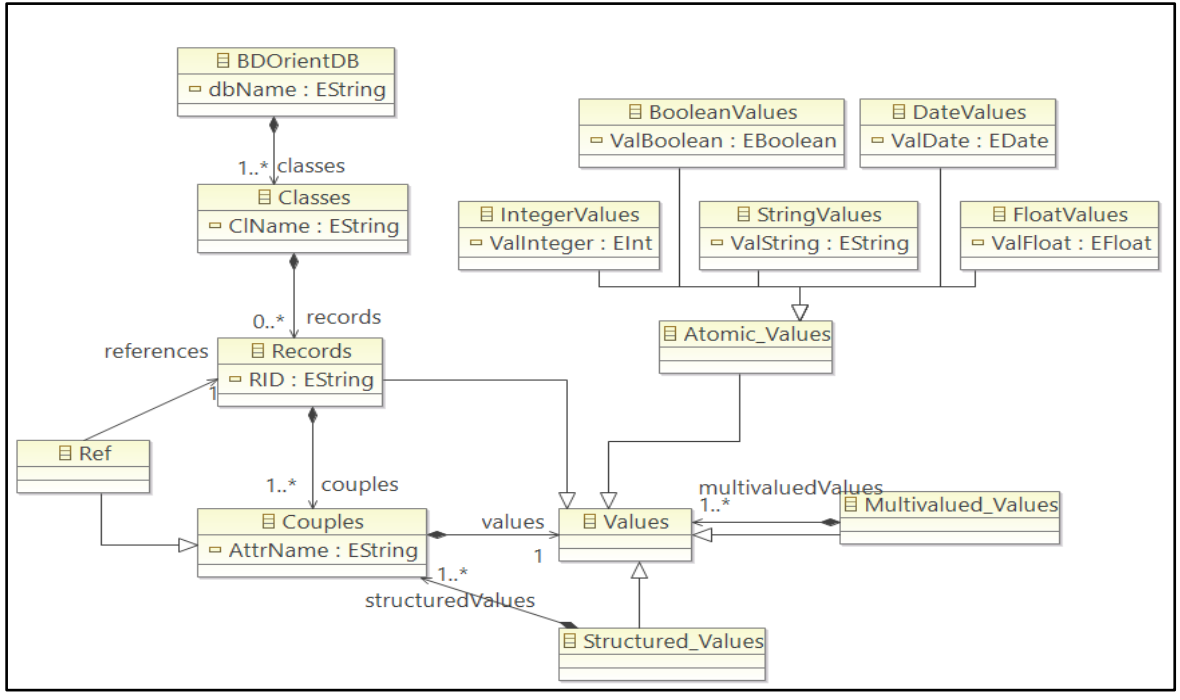
CreateDW Module

Source metamodel:



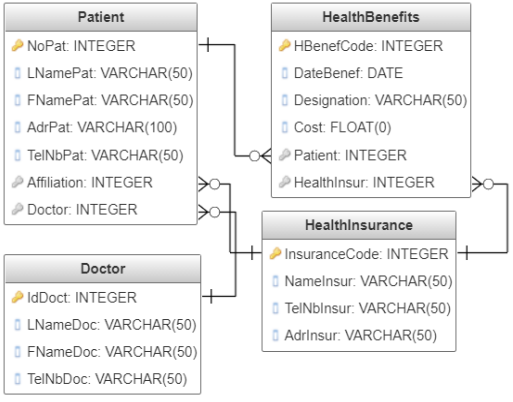
CreateDW Module

Target metamodel:

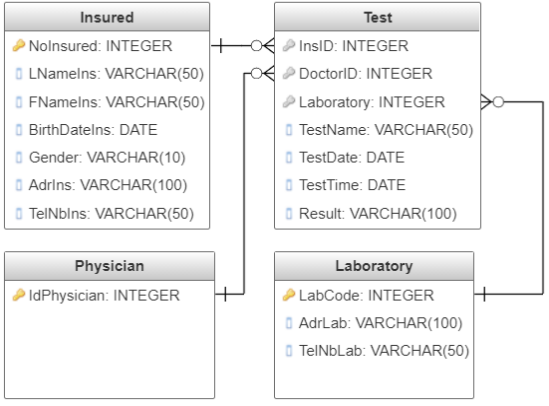


ConvertLinks Module

Experimentation : Medical application



The « Service Provision » database



The « Analysis » database

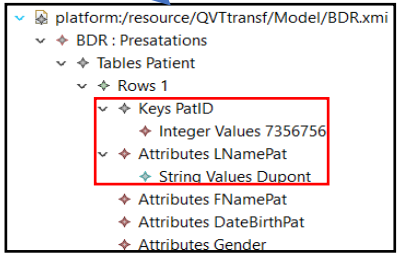
Extracts from the relational schemas of the two Data Lake databases

CreateDW Module

Experimentation : Medical application

Data Lake

Instantiate RDB with XMI



XMI instantiation file of a Relational DB

Apply QVT rules

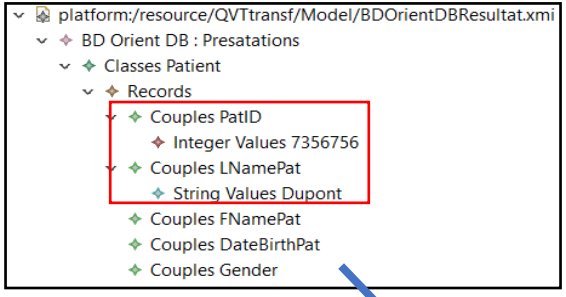
```
//Transform a relational database into an OrientDB database
mapping BDR::RDBtoODB(): BDOrientDB{
dbName:=self.DBName;
classes:= self.tables.map toTable();
}

//Transform a relational table into an OrientDB class
mapping RDB::Tables:: toTable():ODB::Classes{
ClName := self.TableName;
records:= self.rows.map toDoc();
}

//Transform a row into an OrientDB Record
mapping RDB::Rows:: toDoc():ODB::Records{
couples:=self.hasAttributes.map toCouple();
```

Extract of QVT rules

Result XMI file of NoSQL DB



Create NoSQL DW

Data Warehouse

CreateDW Module

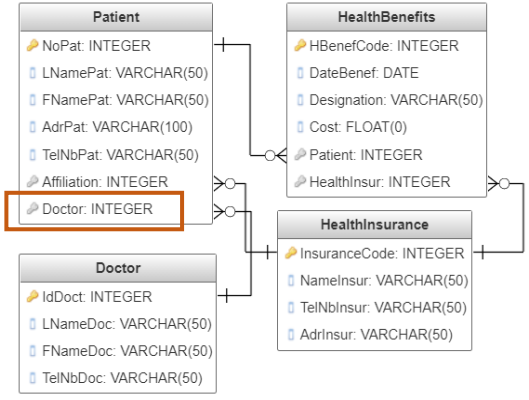
Experimentation : Medical application

Name	Color	SuperClasses	Alias	Abstract	Clusters	Default Cluster	Cluster Selection	Records	Actions
Analysis_Insured					[34, 35, 36, 37]	34	round-robin	5	RENAME QUERY ALL NEW RECORD DROP
Analysis_Physician					[42, 43, 44, 45]	42	round-robin	9	RENAME QUERY ALL NEW RECORD DROP
Doctor_DW					[66, 67, 68, 69]	66	round-robin	78	RENAME QUERY ALL NEW RECORD DROP
ServiceProvision_Doctor					[22, 23, 24, 25]	22	round-robin	9	RENAME QUERY ALL NEW RECORD DROP
Analysis_Insured					[26, 27, 28, 29]	26	round-robin	3	RENAME QUERY ALL NEW RECORD DROP

Extract from the list of the Data Warehouse classes stored in OrientDB

ConvertLinks Module

Experimentation : Medical application



The « Service Provision » database

Record from the «SeviceProvision_Patients» class

```
{
  "@type": "d",
  "@rid": "#26:0",
  "@version": 1,
  "@class": "ServiceProvision_Patients",
  "Email": "ramon.saadi@gmail.com",
  "FNamePDoc": "Ramon",
  "LNameDoc": "Saadi"
  "NoPat": "45657709",
  "Doctor": "5685983"
}
```

```
{
  "@type": "d",
  "@rid": "#42:0",
  "@version": 1,
  "@class": "ServiceProvision_Doctor",
  "IdDoc": "5685983",
  "FNameDoc": "Olivier",
  "LNameDoc": "Durand",
  "TelNbDoc": "06977899"
}
```

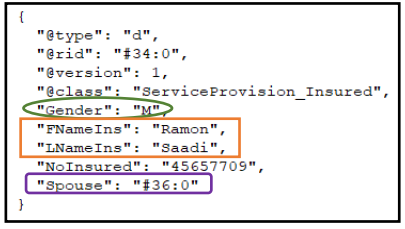
Record from the «SeviceProvision_Doctor» class

```
{
  "@type": "d",
  "@rid": "#26:0",
  "@version": 1,
  "@class": "ServiceProvision_Patients",
  "Email": "ramon.saadi@gmail.com",
  "FNamePat": "Ramon",
  "LNamePat": "Saadi"
  "NoPat": "45657709",
  "Doctor": "#42:0"
}
```

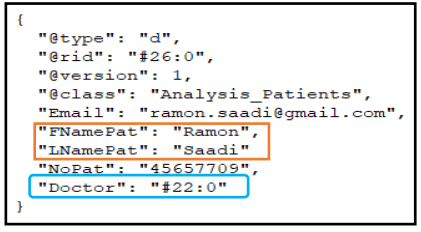
Record from the «SeviceProvision_Patients» class after converting links

MergeClasses Module

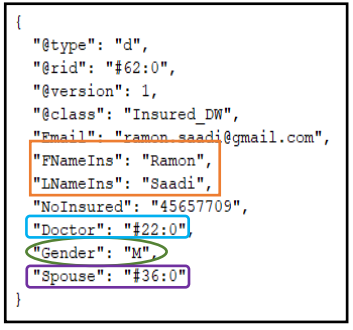
Experimentation : Medical application



Record from the « ServiceProvision_Insured » class



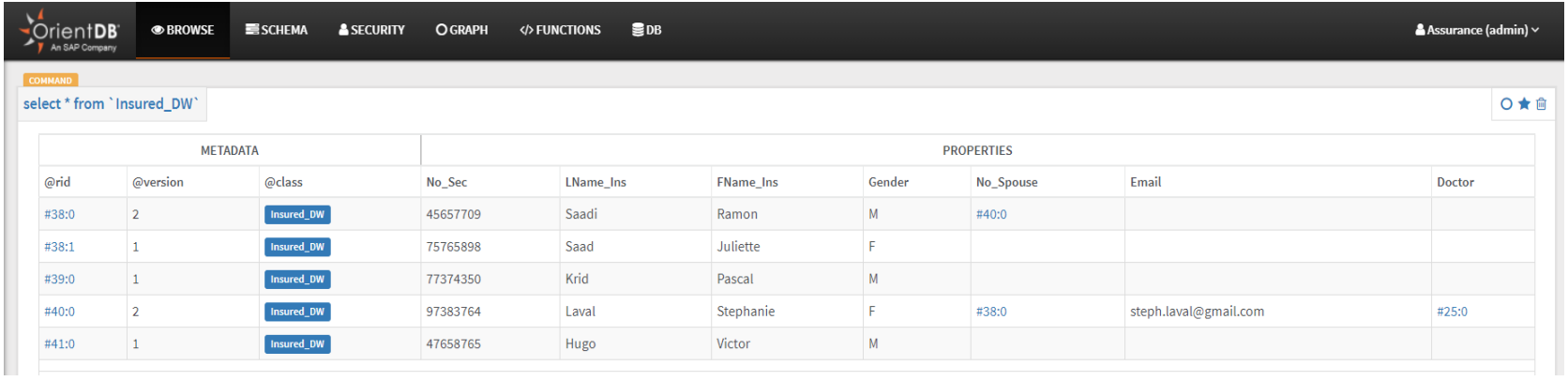
Record from the « Analysis_Patients » class



Record from the new created « Insured_DW » class

MergeClasses Module

Experimentation : Medical application



METADATA			PROPERTIES						
@rid	@version	@class	No_Sec	LName_Ins	FName_Ins	Gender	No_Spouse	Email	Doctor
#38:0	2	Insured_DW	45657709	Saadi	Ramon	M	#40:0		
#38:1	1	Insured_DW	75765898	Saad	Juliette	F			
#39:0	1	Insured_DW	77374350	Krid	Pascal	M			
#40:0	2	Insured_DW	97383764	Laval	Stephanie	F	#38:0	steph.laval@gmail.com	#25:0
#41:0	1	Insured_DW	47658765	Hugo	Victor	M			

Extract from the new created "Insured_DW" class

Conclusion

- Creation of a NoSQL Data Warehouse from a Data Lake
- Source: Relational databases
- Target: NoSQL Data Warehouse

Perspectives

- Extending the Data Lake to other types of sources
- Data processing and ingestion from these sources

Thank you for your attention