



Minicurso:

Interoperabilidade entre Bancos de Dados Relacionais e Bancos de Dados NoSQL

Geomar A. Schreiner
Ronaldo S. Mello

Departamento de Informática e Estatística (INE)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Santa Catarina (UFSC)



Sumário

1. Introdução
2. Bancos de Dados (BDs) nas Nuvens e o Movimento NoSQL
3. Gerência de Dados Relacionais na Nuvem
4. Abordagens para Interoperabilidade entre BDs Relacionais e BDs NoSQL
 - i. Abordagens do tipo *Layer*
 - ii. Abordagens do tipo *Storage Engine*
5. Considerações Finais

Sumário

1. Introdução

2. Bancos de Dados (BDs) nas Nuvens e o Movimento NoSQL

3. Gerência de Dados Relacionais na Nuvem

4. Abordagens para Interoperabilidade entre BDs Relacionais e BDs NoSQL

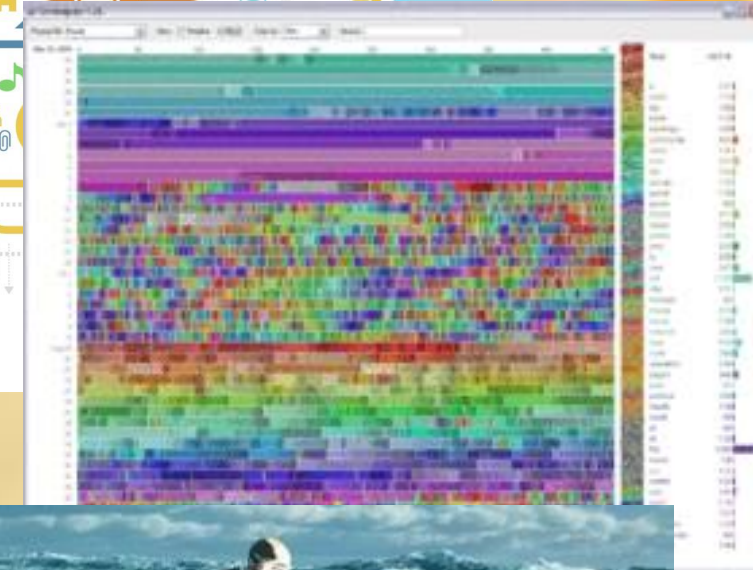
i. Abordagens do tipo *Layer*

ii. Abordagens do tipo *Storage Engine*

5. Considerações Finais

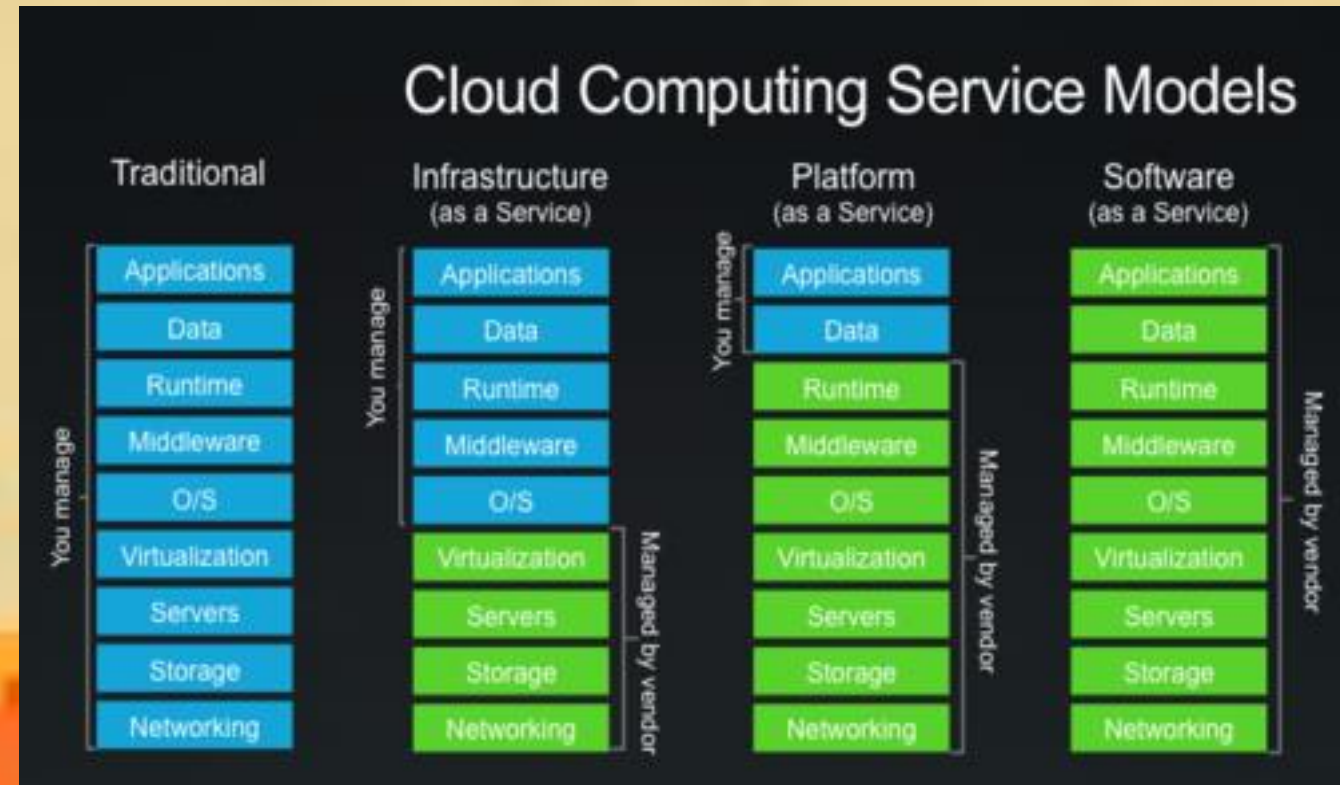
Introdução

- **Big Data**
 - Dados cujas características geram desafios de gerenciamento
 - Volume: Tb, Pb, Zb, (!) ...
 - Velocidade: dados gerados e acessados com (alta!) frequência
 - Variedade: diferentes formatos (E, SE, NE)
 - Exemplos de aplicações
 - *Social networks, data streaming, BI, Web-based systems (crawling, extração, integração, ..., de dados)*



Introdução

- **Computação nas nuvens**
 - Oferta de serviços remotos de computação em diferentes níveis
 - *IaaS, PaaS, SaaS (DBaaS)*
 - Serviços gratuitos ou pagos (*pay-as-you go*)
 - Vantagens
 - Custo de implantação e manutenção de sistemas
 - Elasticidade
 - Paradigma adequado ao gerenciamento de *Big Data*
 - BDs nas Nuvens



Introdução

- **BDs relacionais (tradicionais)**
 - Não são adequados ao gerenciamento de *Big Data*

SGBDs Relacionais (SGBDRs)	Gerenciamento de Big Data
Grande volume de dados	(Muito!) Gigantesco volume de dados
Esquema fixo	Ausência de esquema ou esquema simples/flexível
Escalabilidade limitada	Alta escalabilidade
Consistência forte (ACID)	Consistência eventual (foco em disponibilidade)
Métodos de acesso com suporte ao processamento de consultas complexas (otimização e processamento de junções)	Métodos de acesso simples (foco em escalabilidade e disponibilidade)

 **SGBDs NoSQL**

Introdução

- **Questão**

- *“Muitas aplicações mantêm seus dados em BDRs, seus dados estão “se tornando” Big Data e elas gostariam de contar com o suporte de um BD NoSQL... O que fazer?”*

Introdução

- **Questão**

- *“Muitas aplicações mantêm seus dados em BDRs, seus dados estão “se tornando” Big Data e elas gostariam de contar com o suporte de um BD NoSQL... O que fazer?”*

- **Possíveis soluções**

Solução	Desvantagem(ns)
Migrar para uma solução nativa <i>“relacional-in-the-cloud”</i> ou extensões de SGBDRs comerciais para a nuvem	<ul style="list-style-type: none">• Custo de aquisição• Rigidez do modelo relacional para representar dados NE ou SE

Introdução

- **Questão**

- *“Muitas aplicações mantêm seus dados em BDRs, seus dados estão “se tornando” Big Data e elas gostariam de contar com o suporte de um BD NoSQL... O que fazer?”*

- **Possíveis soluções**

Solução	Desvantagem(ns)
Migrar para uma solução nativa <i>“relacional-in-the-cloud”</i> ou extensões de SGBDRs comerciais para a nuvem	<ul style="list-style-type: none">• Custo de aquisição• Rigidez do modelo relacional para representar dados NE ou SE
Substituir o BDR por um BD NoSQL	<ul style="list-style-type: none">• Curva de aprendizagem de uma nova tecnologia• Custo de modificação das interfaces de acesso da aplicação (SQL → NoSQL)

Introdução

- **Questão**

- *“Muitas aplicações mantêm seus dados em BDRs, seus dados estão “se tornando” Big Data e elas gostariam de contar com o suporte de um BD NoSQL... O que fazer?”*

- **Possíveis soluções**

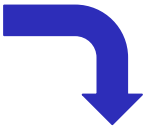
Solução	Desvantagem(ns)
Migrar para uma solução nativa <i>“relacional-in-the-cloud”</i> ou extensões de SGBDRs comerciais para a nuvem	<ul style="list-style-type: none">• Custo de aquisição• Rigidez do modelo relacional para representar dados NE ou SE
Substituir o BDR por um BD NoSQL	<ul style="list-style-type: none">• Curva de aprendizagem de uma nova tecnologia• Custo de modificação das interfaces de acesso da aplicação (SQL → NoSQL)
Prover uma camada de interoperabilidade entre os métodos de acesso da aplicação (SQL) e os métodos de acesso do BD NoSQL	Custo de desenvolvimento desta camada

Introdução

- **Questão**

- *“Muitas aplicações mantêm seus dados em BDRs, seus dados estão “se tornando” Big Data e elas gostariam de contar com o suporte de um BD NoSQL... O que fazer?”*

- Possíveis soluções

Solução	Desvantagem(ns)
Migrar para uma solução nativa <i>“relacional-in-the-cloud”</i> ou extensões de SGBDRs comerciais para a nuvem	<ul style="list-style-type: none">• Custo de aquisição• Rigidez do modelo relacional para representar dados NE ou SE
Substituir o BDR por um BD NoSQL Foco deste minicurso 	<ul style="list-style-type: none">• Curva de aprendizagem de uma nova tecnologia• Custo de modificação das interfaces de acesso da aplicação (SQL → NoSQL)
Prover uma camada de interoperabilidade entre os métodos de acesso da aplicação (SQL) e os métodos de acesso do BD NoSQL	Custo de desenvolvimento desta camada

Introdução

- **Objetivo deste minicurso**

- Apresentação e comparação de abordagens que suportam a interoperabilidade de esquemas relacionais e operações SQL para esquemas e métodos de acesso de BDs NoSQL

- **Justificativas**

- Evita manutenção da aplicação
- Evita conhecimento dos métodos de acesso NoSQL pelos desenvolvedores/mantenedores da aplicação
- Permite o gerenciamento eficiente do armazenamento de dados com representações heterogêneas na nuvem
 - Exemplo: registros cujo conjunto de atributos é variável

Sumário

1. Introdução
2. Bancos de Dados (BDs) nas Nuvens e o Movimento NoSQL
3. Gerência de Dados Relacionais na Nuvem
4. Abordagens para Interoperabilidade entre BDs Relacionais e BDs NoSQL
 - i. Abordagens do tipo *Layer*
 - ii. Abordagens do tipo *Storage Engine*
5. Considerações Finais

Bancos de Dados nas Nuvens

- SGBDs oferecidos como serviço na nuvem (DBaaS)
- Vantagens
 - Redução de custos com aquisição/manutenção de SGBD
 - Delegação de tarefas de administração de dados
 - Exemplos: *tuning*, *backup*, ..., do BD
 - Processamento de *Big Data*
 - Arquiteturas de alto desempenho baseadas em
 - *Data Centers* (nodos provedores de dados)
 - Tecnologias de processamento paralelo e distribuído (Ex.: *Map-Reduce*)



BDs nas Nuvens - Características

- **Escalabilidade**
 - *Data centers, Map-Reduce, ...*
- **Disponibilidade**
 - Replicação de dados, consistência relaxada, ...
- **APIs simples para acesso**
 - Ausência/baixo *overhead* com *parsing*/execução de instruções de linguagens de BD
- **Suporte à alternância de *workload***
 - Elasticidade na demanda de operações sobre dados e alocação de recursos



BDs nas Nuvens - Taxonomia

Nativo

Não-Nativo

Relacional

SQL Azure Amazon RDS Relational Cloud	Oracle 12 c DB2 on the Cloud PostgreSQL Plus
Amazon S3 (BDs chave-valor) Voldemort (BDs colunares) HBase Cassandra	Couch DB (BDs documento) Mongo DB (BDs documento) Neo4j (BDs grafo) Orient DB

Não-Relacional

BDs nas Nuvens - Taxonomia

Nativo

Não-Nativo

Relacional

SQL Azure	Oracle 12 c
Amazon RDS	DB2 on the Cloud
Relational Cloud	PostgreSQL Plus
Amazon S3	Couch DB (BDs documento)
	(BDs grafo) Neo4j Orient DB

Não-

SGBDRs projetados para operar na nuvem, ou seja, oferecem funcionalidades de gerência de dados adequadas a um SGBD na nuvem

BDs nas Nuvens - Taxonomia

	Nativo	Não-Nativo
Relacional	SQL Azure Amazon RDS Relational Cloud	Oracle 12 c DB2 on the Cloud PostgreSQL Plus
Não-Relacional	Amazon S3 (BDs chave-valor) Voldemort (BDs colunares) HBase Cassandra	

SGBDRs não concebidos para a nuvem, mas que podem ser executados na nuvem através da utilização de serviços de gerenciamento adicionais

BDs nas Nuvens - Taxonomia

Nativo

Não-Nativo

Relacional

SGBDs não-relacionais concebidos para a nuvem, ou seja, baseados em modelos de dados propostos para a nuvem

		Oracle 12 c
		2 on the Cloud
	Relational Cloud	PostgreSQL Plus
	Amazon S3 Voldemort (BDs chave-valor)	Couch DB Mongo DB (BDs documento)
	(BDs colunares) HBase Cassandra	(BDs grafo) Neo4j Orient DB

Não-Relacional

BDs nas Nuvens - Taxonomia

Nativo

Não-Nativo

Relacional

SQL A
 Amazon
 Relational Cloud

Postgre

SGBDs não-relacionais não concebidos originalmente para a nuvem, mas são utilizados em ambientes na nuvem

Não-Relacional

Amazon S3 (BDs chave-valor)
 Voldemort
 (BDs colunares) HBase
 Cassandra

Couch DB (BDs documento)
 Mongo DB
 (BDs grafo) Neo4j
 Orient DB

BDs nas Nuvens - Taxonomia

Nativo

Não-Nativo

Relacional

SQL Azure	Oracle 12 c
Amazon RDS	DB2 on the Cloud
Relational Cloud	PostgreSQL Plus

Não-Relacional

Amazon S3 (BDs chave-valor) Voldemort	Couch DB (BDs documento) Mongo DB
(BDs colunares) HBase Cassandra	(BDs grafo) Neo4j Orient DB

BDs NoSQL

NoSQL (*NOT* only SQL)



- Movimento pelo desenvolvimento de SGBDs não-relacionais para o gerenciamento de dados na nuvem
- Principais características
 - Métodos de acesso simples
 - APIs baseadas em acesso por chave
 - Não suporta junções
 - Maioria não suporta relacionamentos entre dados e integridade referencial
 - Modelos de dados heterogêneos
 - Esquemas flexíveis
 - Falta de padronização



NoSQL (*NOt only SQL*)

- Obs.: não confunda NoSQL com **NewSQL!** 😊




- **NewSQL**

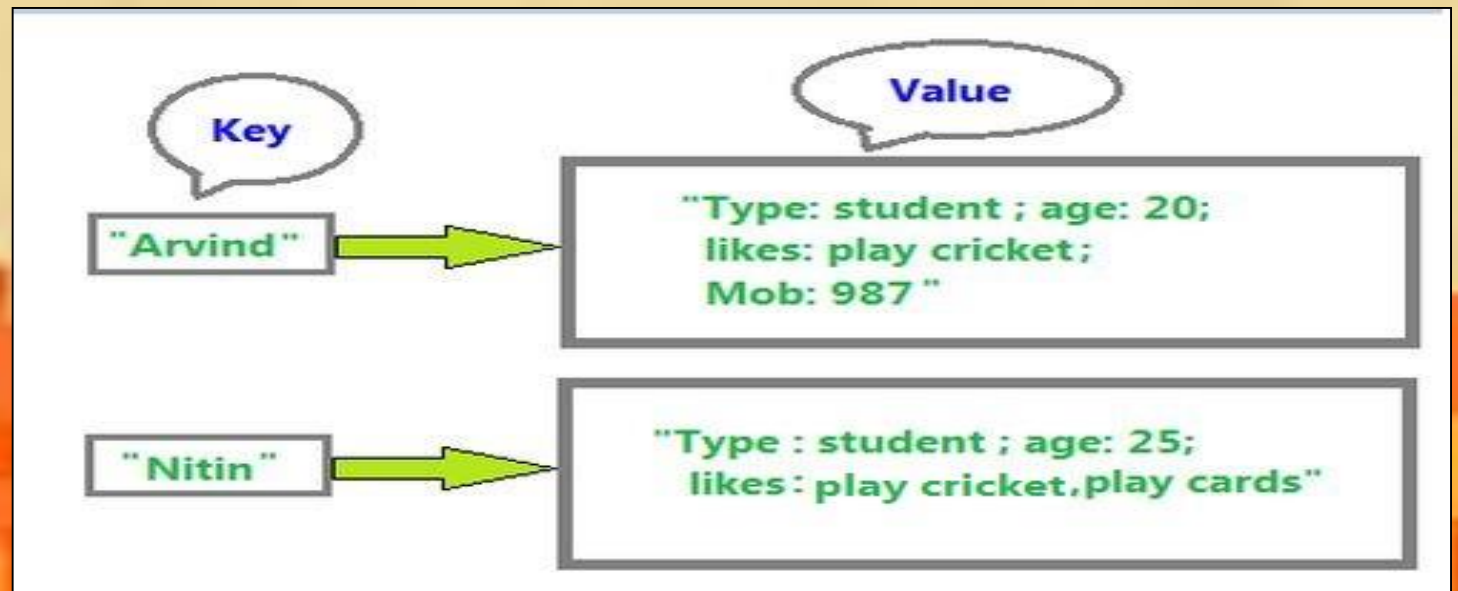
- Movimento pelo desenvolvimento de SGBDRs
 - Escaláveis para o tratamento de Big Data
 - Suporte a transações e propriedades ACID
 - Não são necessariamente serviços na nuvem
 - Exemplos: *VoltDB, MySQL cluster, NuoDB*

Modelos de Dados para NoSQL

- Chave-Valor
- Colunares
- Documento
- Grafo

Modelos de Dados para NoSQL

- Chave-Valor 
 - Colunares
 - Documento
 - Grafo
- Modelo simples similar a uma estrutura de indexação
 - Chave identifica um conteúdo mono ou multivalorado
 - API simples
 - *get(key)*, *put(key, value)*, *delete(key)*
 - Não suporta
 - Definição de esquemas
 - Relacionamentos entre dados
 - Linguagem de consulta



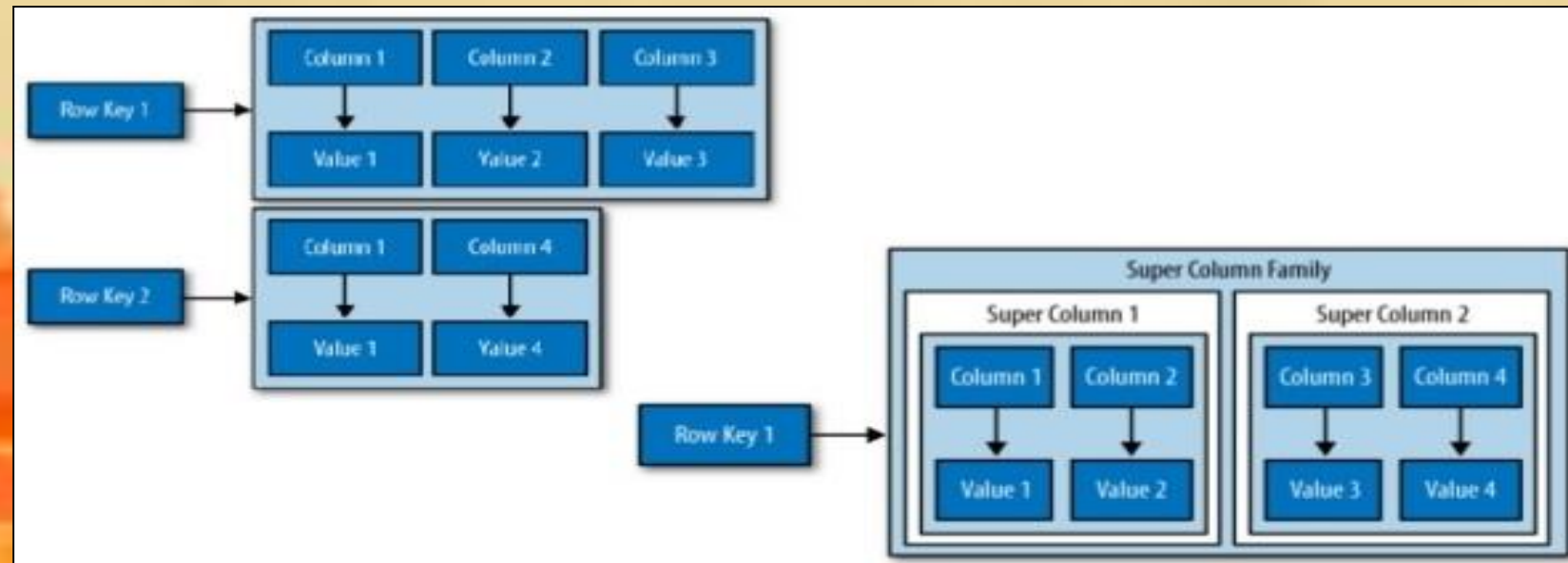
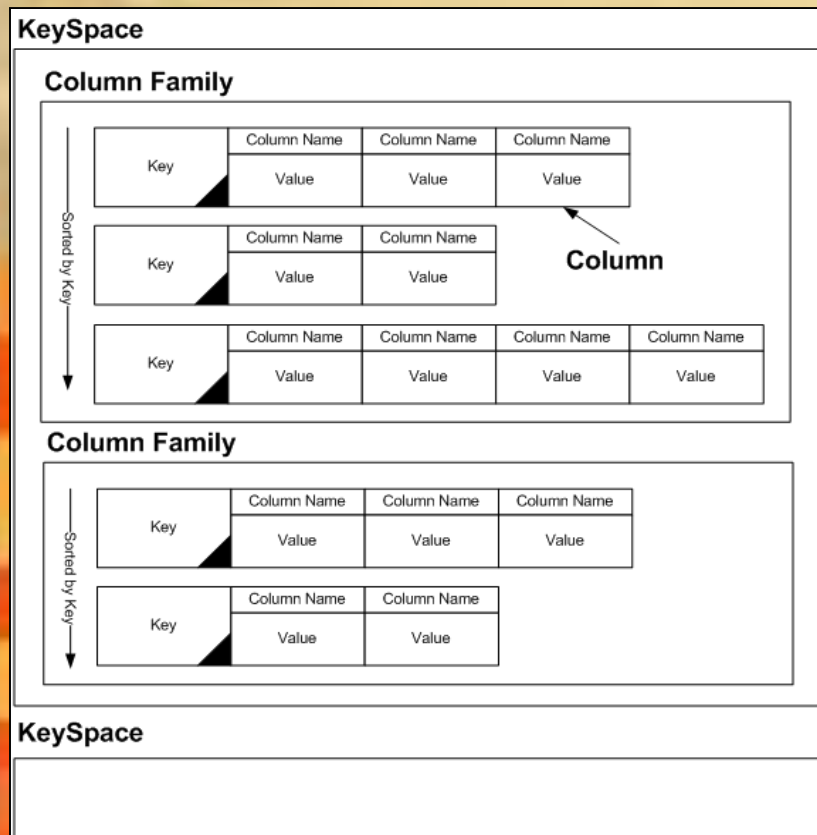
Modelos de Dados para NoSQL

- Chave-Valor

- Colunares



- Modelo mais complexo que o chave-valor
 - Conceitos: *keyspace* (\equiv BD), *column family* (\equiv tabela) e um *conjunto de colunas* (\equiv registro)
 - Uma coluna possui um *nome* e um *valor*
 - Um conjunto de colunas é acessado por uma *chave*
 - Itens de dados (“registros”) podem ter colunas diferentes
 - Suporte a colunas multivaloradas e super-colunas
- APIs proprietárias e/ou linguagens de consulta simples
- Não suporta relacionamentos entre dados



Modelos de Dados para NoSQL

- Chave-Valor

- Colunares

- Documento

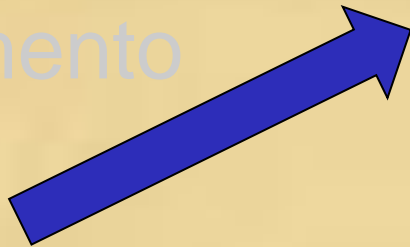


- Modelo adequado à representação de objetos complexos
 - Um objeto (“documento”) possui uma **chave** e um **conjunto de atributos**
 - Atributos podem ter domínios **atômicos** ou **complexos** (listas, tuplas, conjuntos)
- APIs proprietárias e/ou linguagens de consulta simples
- Não suporta relacionamentos entre dados
- Falta de padronização
 - Exemplos: MongoDB (JSON), Amazon SimpleDB (Domínio→Item →Atributo →{Valor})

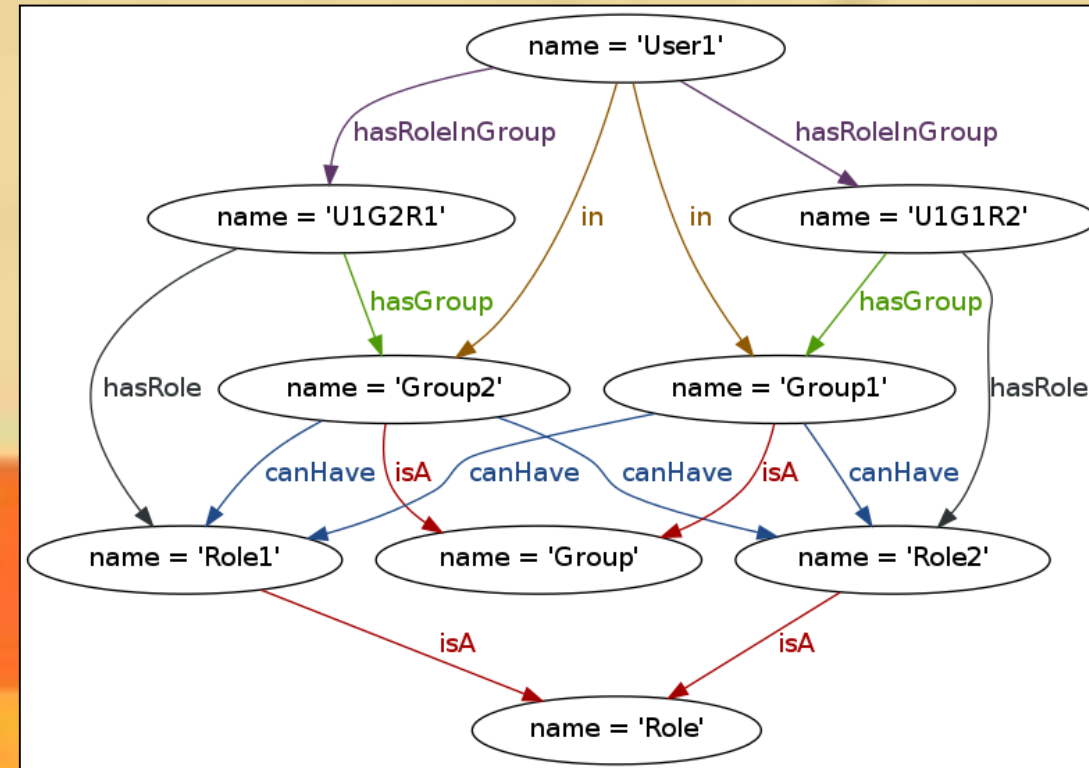
```
{ "_id":"discussion_tables",  
  "_rev":"D1C946B7",  
  "Sunrise":true,  
  "Sunset":false,  
  "FullHours":[1,2,3,4,5,6,7,8,9,10],  
  "Activities": [  
    {"Name":"Football", "Duration":2, "DurationUnit":"Hours"},  
    {"Name":"Breakfast", "Duration":40, "DurationUnit":"Minutes",  
     "Attendees":["Jan", "Damien", "Laura", "Gwendolyn", "Roseanna"]} ] }  
JSON
```

Modelos de Dados para NoSQL

- Chave-Valor
- Colunares
- Documento
- Grafo



- Modelo composto por nodos, arestas e atributos
 - **Nodo**: um item de dado (“registro”) composto por atributos
 - **Aresta**: relacionamento entre nodos, composto por um rótulo e atributos opcionais
 - **Atributo**: composto por nome e valor (domínio atômico ou multivalorado)
- APIs proprietárias e/ou linguagens de consulta simples



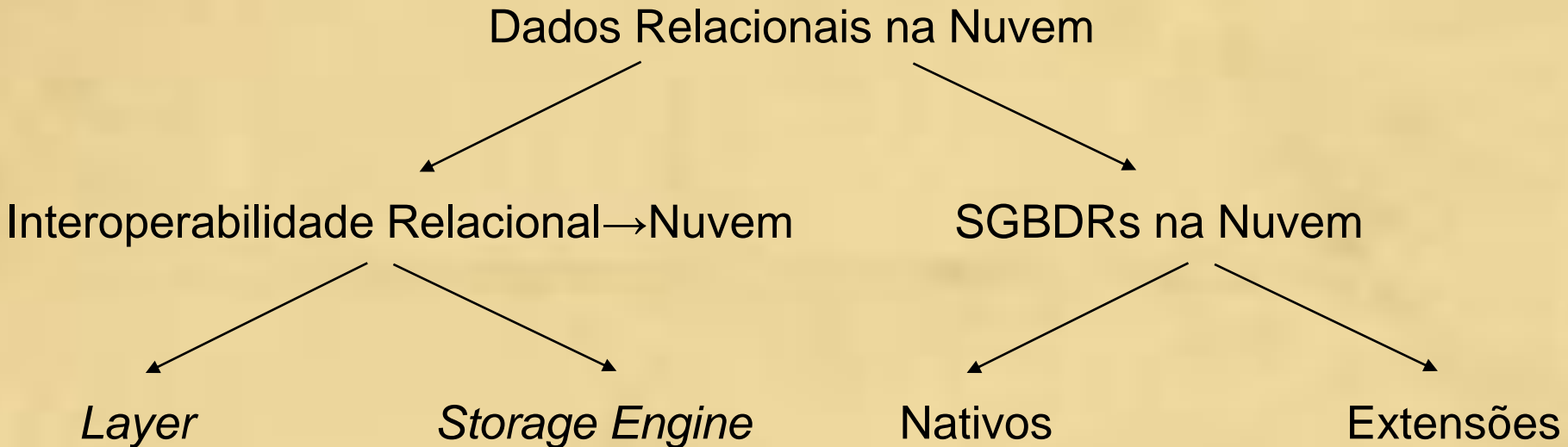
Sumário

1. Introdução
2. Bancos de Dados (BDs) nas Nuvens e o Movimento NoSQL
3. **Gerência de Dados Relacionais na Nuvem**
4. Abordagens para Interoperabilidade entre BDs Relacionais e BDs NoSQL
 - i. Abordagens do tipo *Layer*
 - ii. Abordagens do tipo *Storage Engine*
5. Considerações Finais

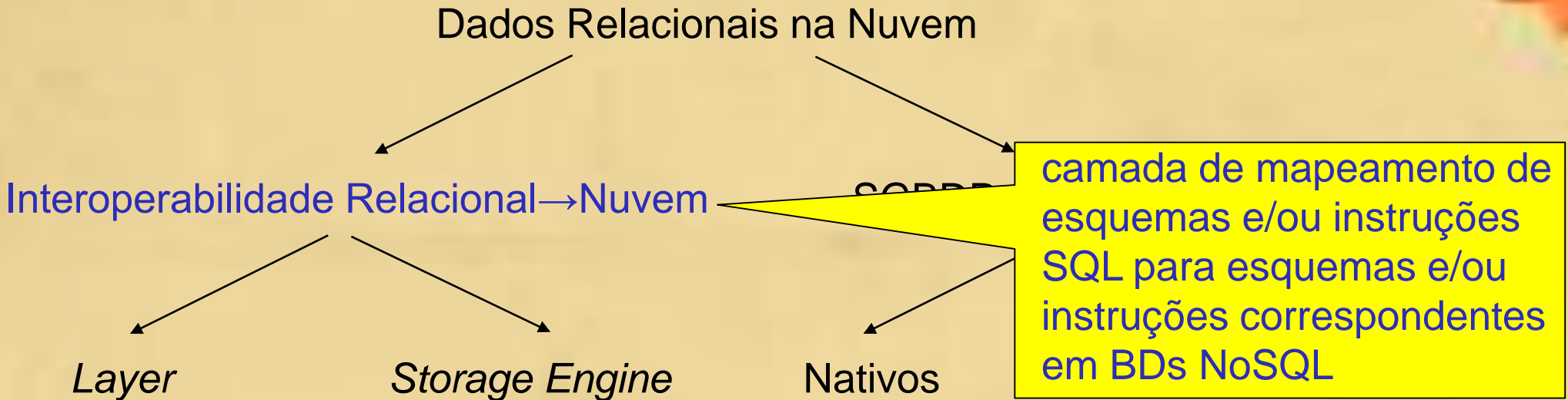
Gerência de Dados Relacionais na Nuvem

- Abordagens para gerenciamento de *Big Data relacional* na nuvem
 - Ênfase em escalabilidade, disponibilidade e elasticidade
- Foco
 - Movimento *relacional* → *cloud* (e não o contrário!)
- Abordagens são organizadas em uma *taxonomia*
 - Conforme a arquitetura do solução de gerência de dados

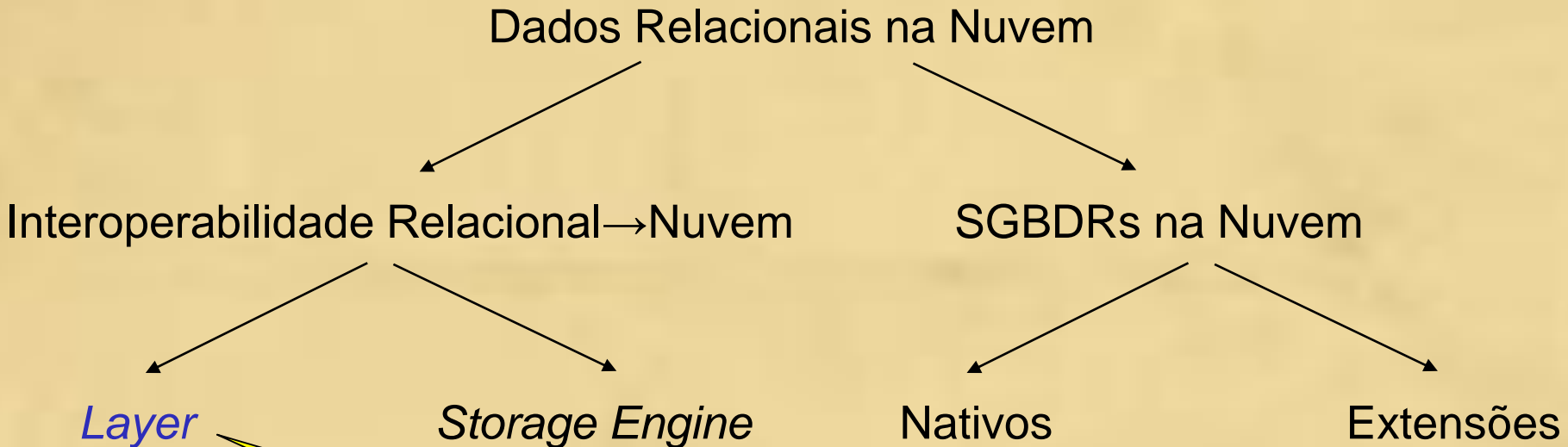
Gerência Relacional na Nuvem - Taxonomia



Gerência Relacional na Nuvem - Taxonomia

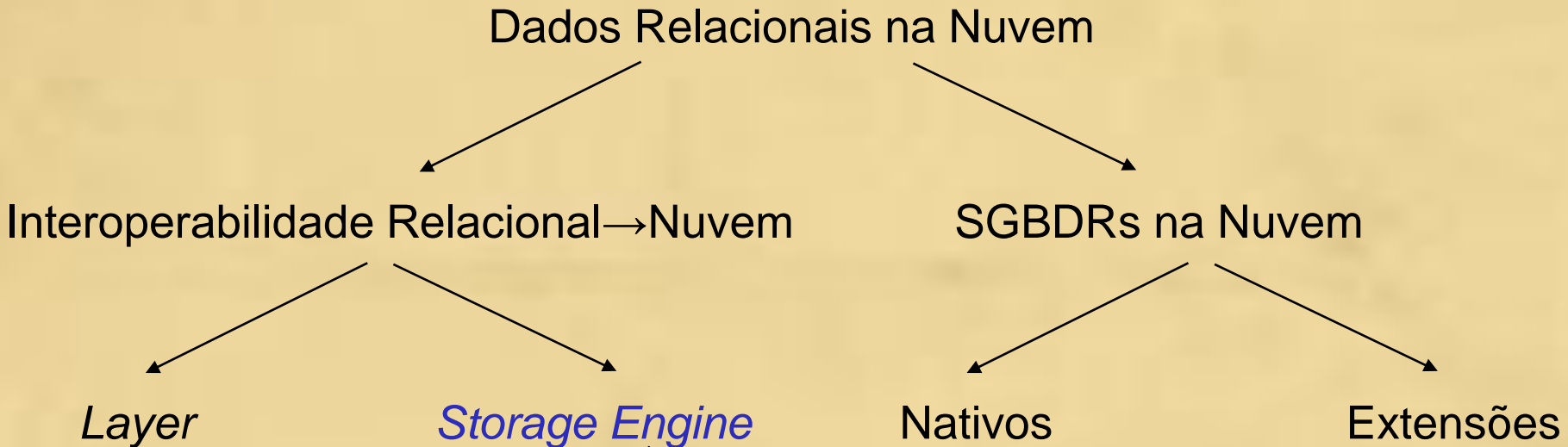


Gerência Relacional na Nuvem - Taxonomia



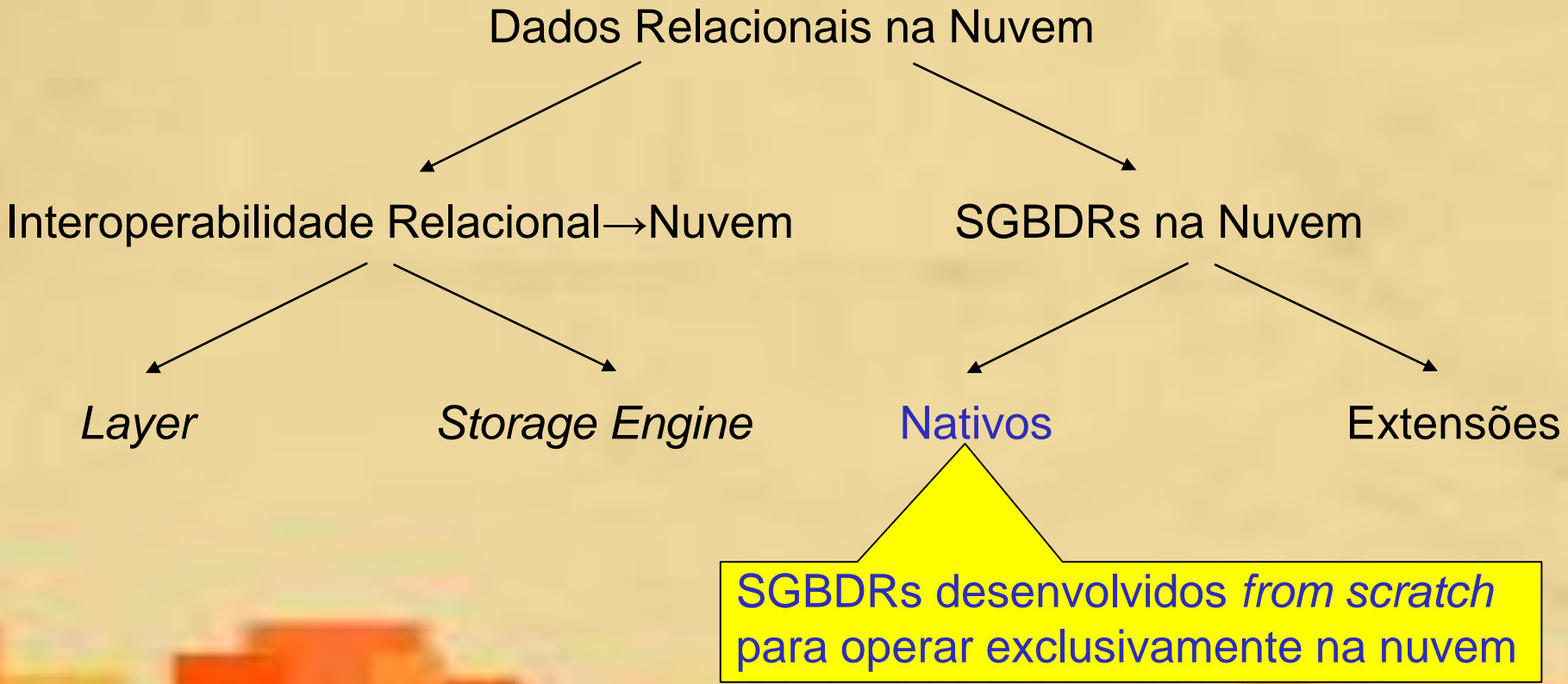
Camada de mapeamento é um módulo externo que pode ser acoplado entre um BDR e um BD NoSQL

Gerência Relacional na Nuvem - Taxonomia

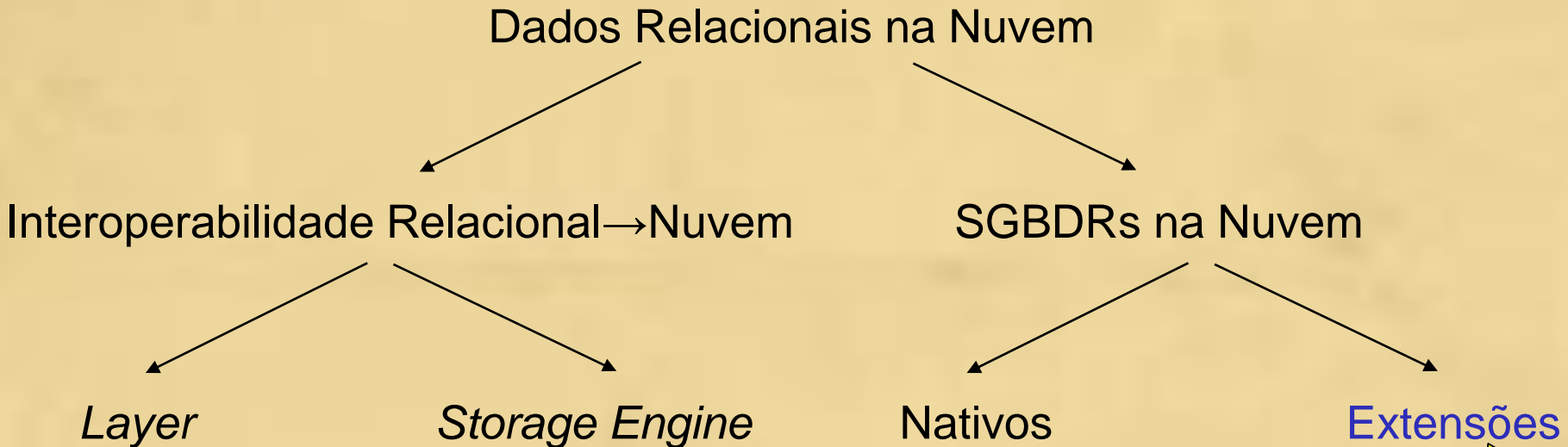


Camada de mapeamento é um módulo interno no *kernel* do SGBDR que realiza o mapeamento e a persistência em um SGBD NoSQL

Gerência Relacional na Nuvem - Taxonomia

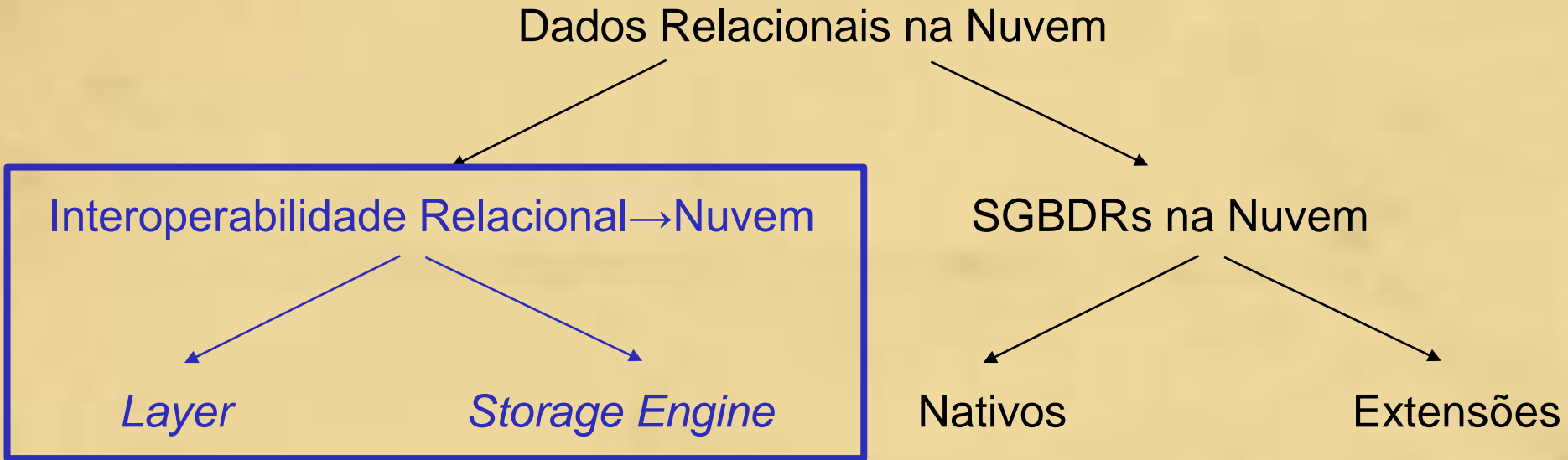


Gerência Relacional na Nuvem - Taxonomia



SGBDRs comerciais com *packages* para o gerenciamento de dados na nuvem

Gerência Relacional na Nuvem - Taxonomia



Foco deste minicurso



Minicurso:

Interoperabilidade entre Bancos de Dados Relacionais e Bancos de Dados NoSQL

Geomar A. Schreiner - schreiner.geomar@posgrad.ufsc.br

Ronaldo S. Mello – r.mello@ufsc.br

Departamento de Informática e Estatística (INE)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Universidade Federal de Santa Catarina (UFSC)

