

Analizando dados abertos semiestruturados com banco convergente

Daniel Panizzo

Trilha Inovação com dados em nuvem

Analizando dados abertos semiestruturados com banco convergente

Daniel Panizzo

Trilha Inovação com dados em nuvem



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhável 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



Daniel Panizzo

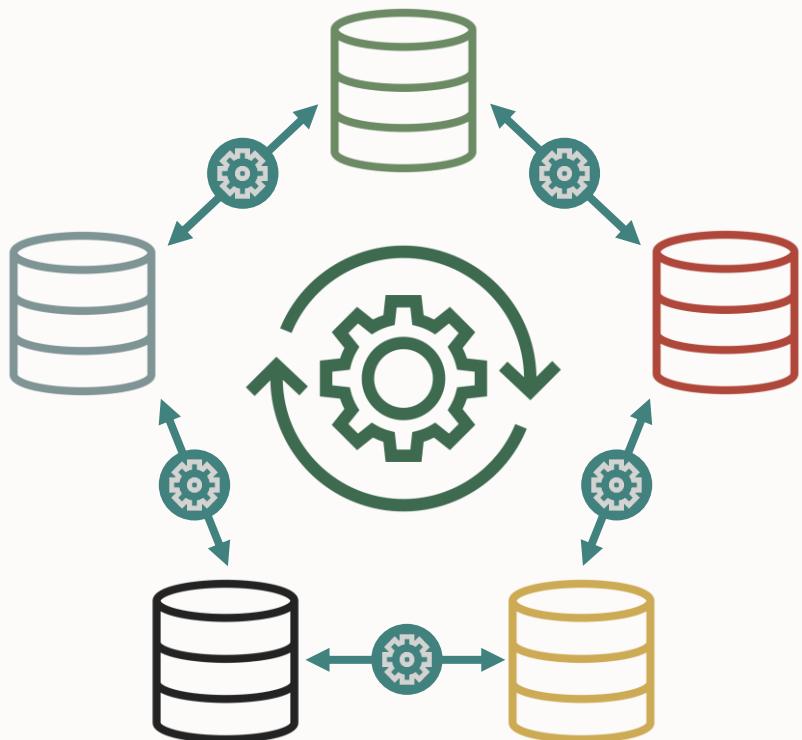
dspanizzo



Qual tipo de banco de dados devo escolher?

O dilema ao criar ou modernizar suas aplicações

Abordagem *single-purpose*

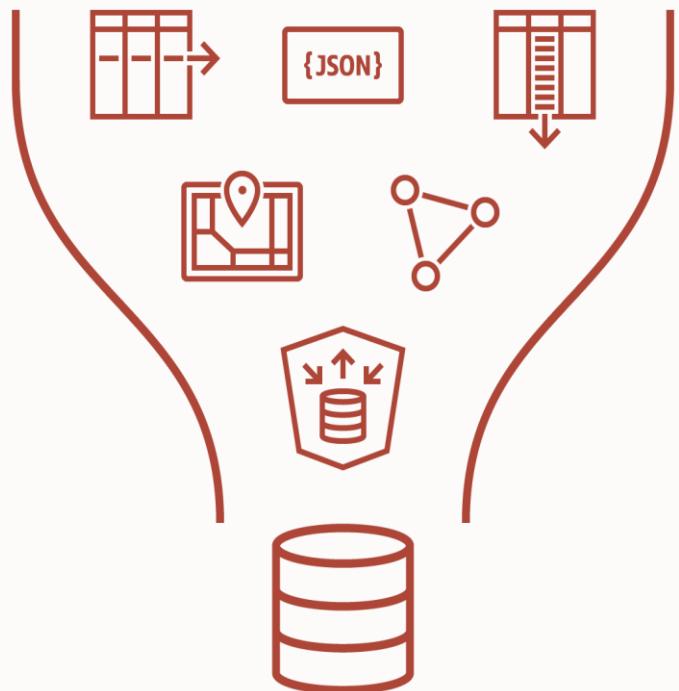


- “Preciso de um banco **documental** para a nova aplicação.”
- “Iremos monitorar os objetivos da organização com um **data warehouse**.”
- “Utilizaremos **grafos** para identificar possíveis casos de fraudes no nosso processo.”
- “Necessitamos de uma visão unificada de **segurança** para auxiliar na aderência à legislação de proteção dos dados.”
- “As métricas das atividades serão monitoradas em tempo real pelo banco de dados **in-memory**.”



Abordagem *multi-purpose*

Um banco de dados convergente



Multi-model

Suporte nativo para diversos modelos de dados e métodos de acesso.

Multitenant

Arquitetura de banco de dados em contêiner provê consolidação, isolamento e agilidade.

Multi-workload

Uma série de otimizações que trabalham isoladamente ou em conjunto para entregar a melhor performance em qualquer tipo de carga de trabalho.

Multi-model

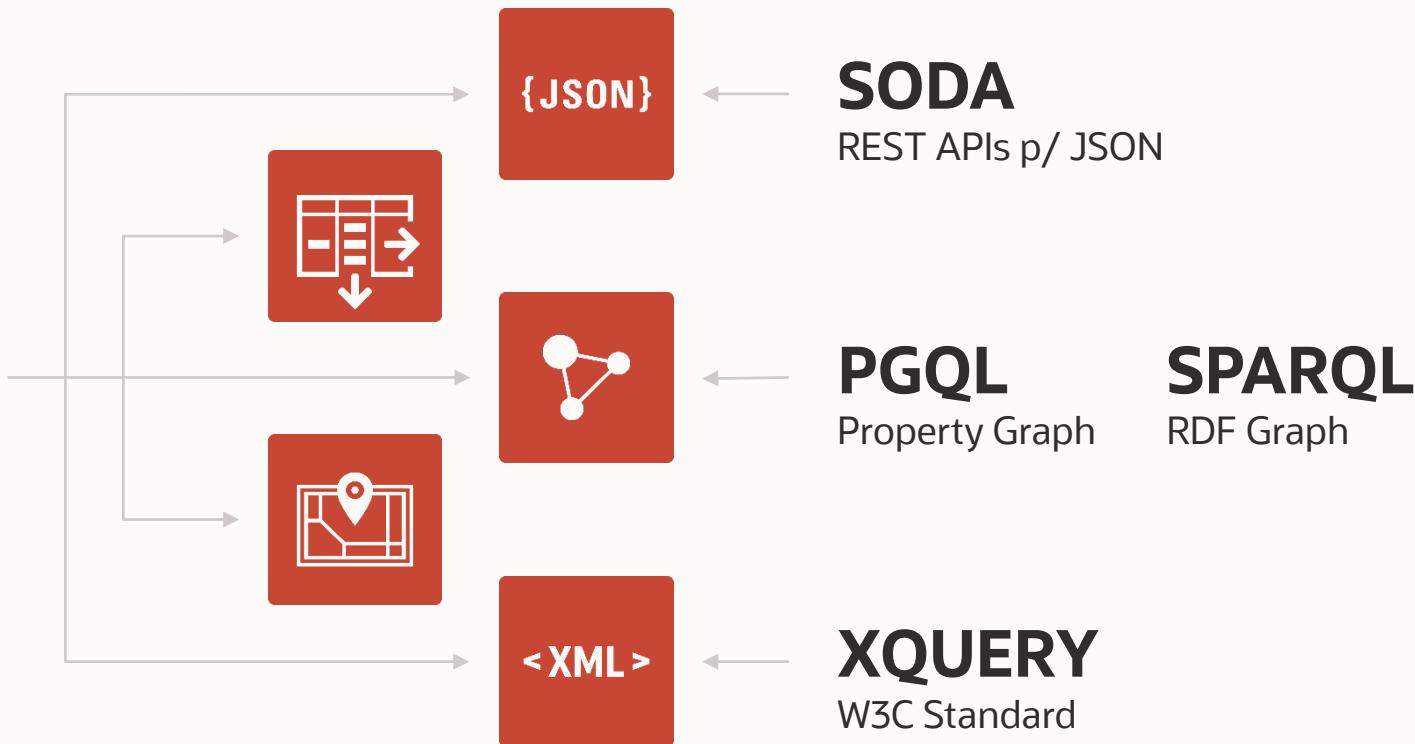
Um motor, multipas personalidades



Cross-Model Data Access

SQL & REST

Relacional, JSON,
Espacial, Grafo, XML



Model-Specific Data Access

SODA

REST APIs p/ JSON

PGQL

Property Graph

SPARQL

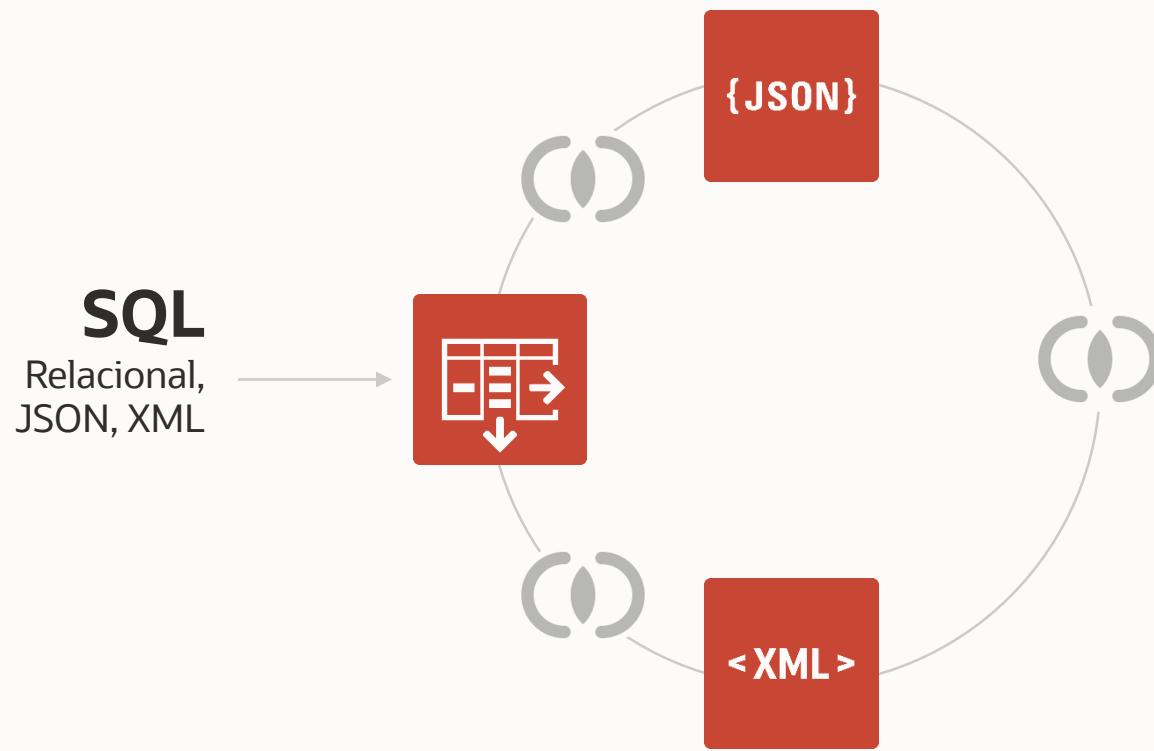
RDF Graph

XQUERY

W3C Standard

Multi-model

Escopo do workshop





Open Knowledge
Foundation

“Dados abertos são dados que podem ser livremente usados, modificados e compartilhados por qualquer um para qualquer propósito.”

Open Knowledge Foundation

Fonte: [The Open Definition](#)

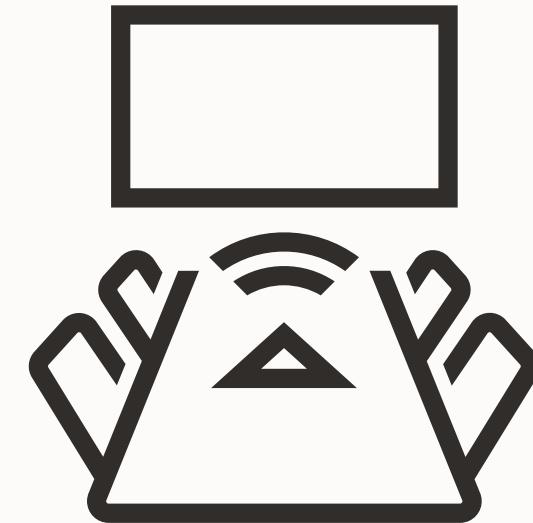
Por quê dados abertos?



Transparência



Valor social e comercial



Governança Participativa

Origem dos dados

RELACIONAL



Receita Federal

JSON



Instituto Brasileiro de
Geografia e Estatística

XML



Tribunal de Contas do
Estado do RS

Ferramentas



Autonomous
Database



SQL

Um olhar no código

As possibilidades de um banco de dados convergente

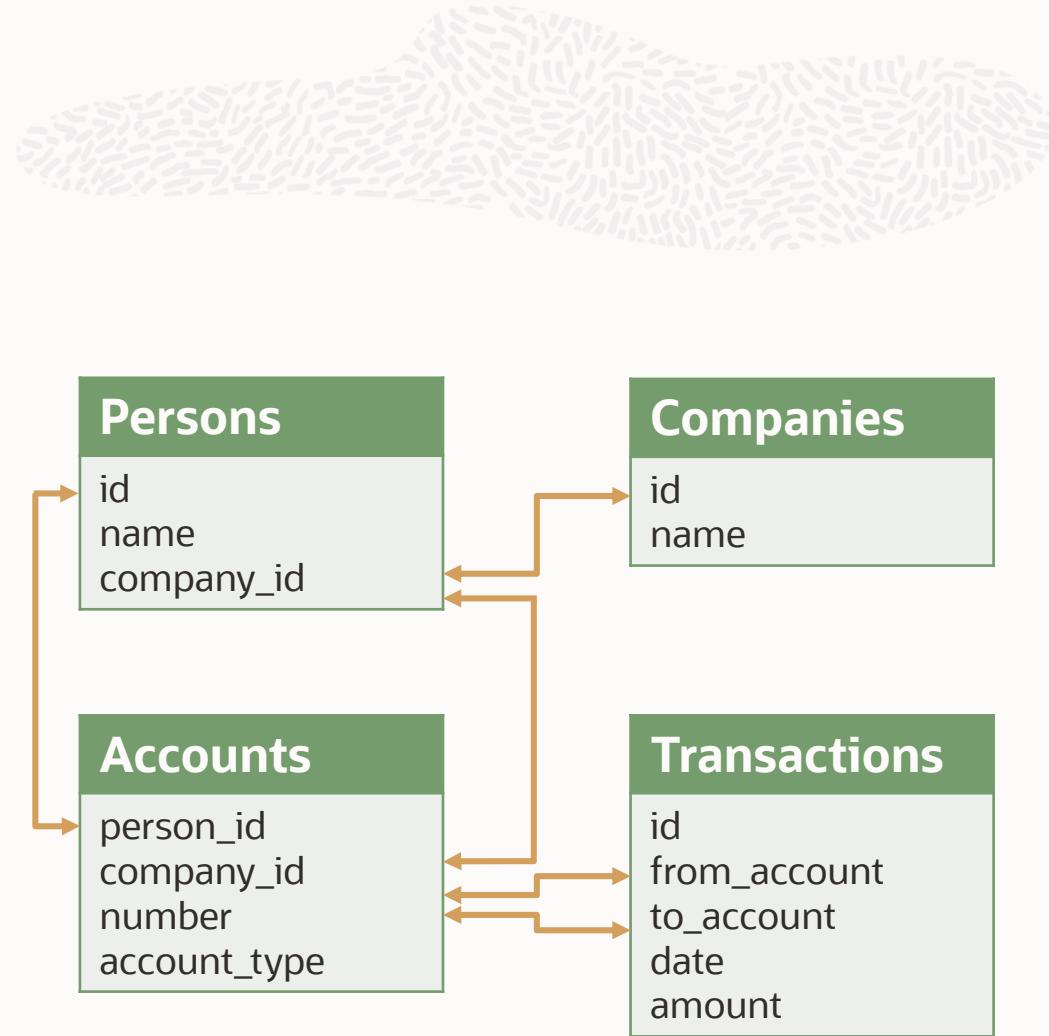
Transacional

Integridade e consistência nos dados

Modelos **normalizados** (ou que utilizam a terceira forma normal – 3FN) são os mais comuns em sistemas transacionais.

Implementam características como:

- Transações ACID
- Alta concorrência
- Baixa latência

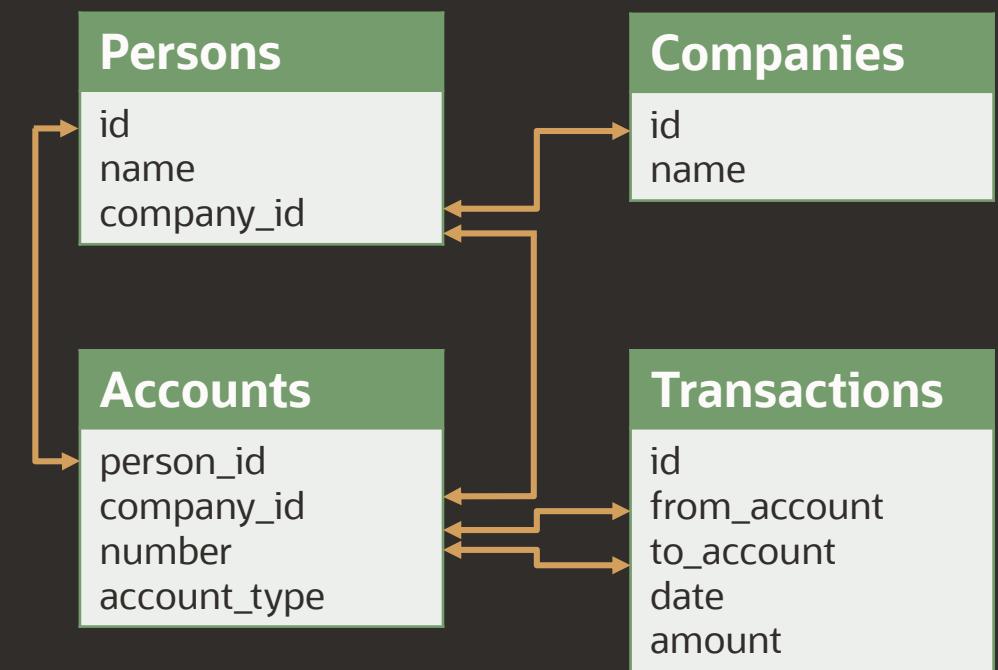


Transacional

CRUD

```
/* Create */
CREATE TABLE persons (
    id      NUMBER,
    name    VARCHAR2(50) NOT NULL,
    company_id NUMBER,
CONSTRAINT pk_persons PRIMARY KEY ( id ),
CONSTRAINT fk_company FOREIGN KEY ( company_id )
    REFERENCES companies ( id )
);

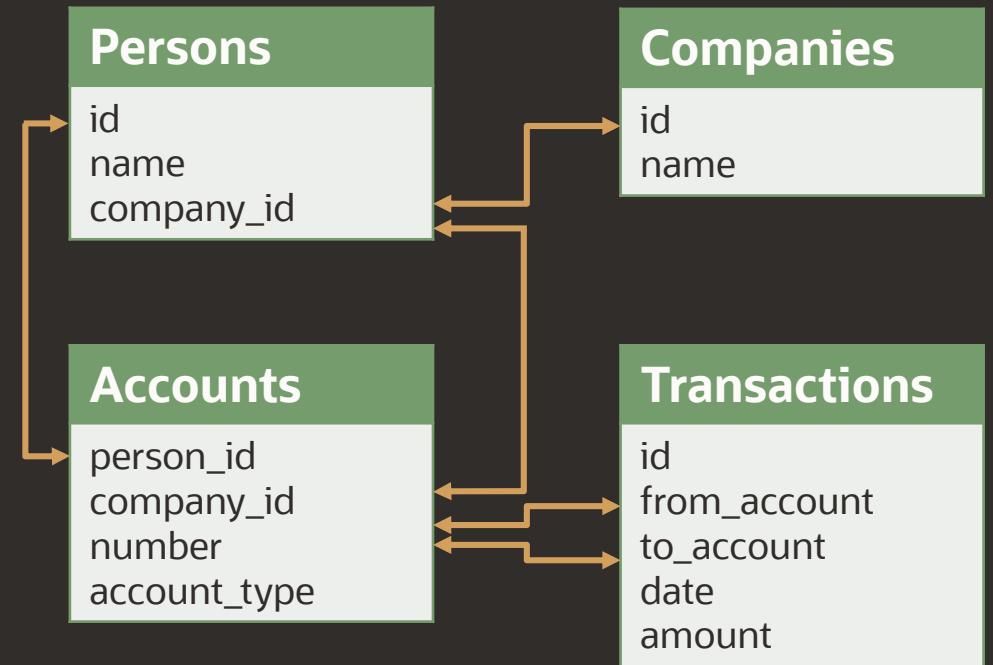
/* Insert */
INSERT INTO persons ( id , name )
VALUES ( 1 , 'Liam' );
```



Transacional

CRUD

```
/* Update */  
UPDATE persons  
    SET name = 'Liliam'  
WHERE id = 1;  
  
/* Delete */  
DELETE FROM persons WHERE id = 1;  
  
/* Drop */  
DROP TABLE persons CASCADE CONSTRAINTS;
```





Live SQL

Learn and share SQL

Tutorial
Introduction to SQL

Documento JSON

Flexibilidade e agilidade

Uma das principais categorias de banco de dados NoSQL. Ao implementar conceitos de armazenamento de dados em **chave-valor**, torna os **esquemas dinâmicos** e facilita alterações em sua estrutura.



```
{ transaction: 'aKl91qh0dFpqi',
  from: {
    id: 10039,
    name: 'Camille'
  },
  to: {
    id: 8021,
    name: 'Nikita',
    company: 'Pizza Co.'
  },
  date: '2021-11-18',
  value: 69.90
}
```

Documento

Flexibilidade e agilidade

```
/* CREATE DOCUMENT */  
CREATE TABLE invoice (  
    id    AS NUMBER,  
    doc   AS JSON  
);
```

```
{ transaction: 'aKl91qh0dFpqi',  
  from: {  
    id: 10039,  
    name: 'Camille'  
  },  
  to: {  
    id: 8021,  
    name: 'Nikita',  
    company: 'Pizza Co.'  
  },  
  date: '2030-11-17',  
  value: 69.90  
}
```

Documento

Flexibilidade e agilidade

```
/* CREATE DOCUMENT */  
CREATE TABLE invoice (  
    id    AS NUMBER,  
    doc   AS JSON  
);  
  
/* INSERT DOCUMENT */  
INSERT INTO invoice VALUES (  
    1001,  
    '  
);
```



```
{ transaction: 'aKl91qh0dFpqi',  
  from: {  
    id: 10039,  
    name: 'Camille'  
  },  
  to: {  
    id: 8021,  
    name: 'Nikita',  
    company: 'Pizza Co.'  
  },  
  date: '2030-11-17',  
  value: 69.90  
}
```

Documento

Flexibilidade e agilidade

```
/* QUERY DOCUMENT */
```

```
SELECT  
    i.doc.to.company AS COMPANY  
    ,i.doc.value      AS VALUE  
FROM invoice i;
```

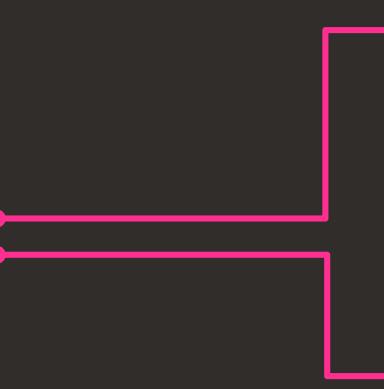
The diagram illustrates the mapping of a database query result to a JSON document. A red bracket on the left groups the columns 'COMPANY' and 'VALUE' from the query. Two red arrows point from these columns to their corresponding fields in the JSON object. The 'COMPANY' field points to the 'company' field under the 'to' key, and the 'VALUE' field points to the 'value' field.

```
{ transaction: 'aKl91qh0dFpqi',  
  from: {  
    id: 10039,  
    name: 'Camille'  
  },  
  to: {  
    id: 8021,  
    name: 'Nikita',  
    company: 'Pizza Co.'  
  },  
  date: '2030-11-17',  
  value: 69.90  
}
```

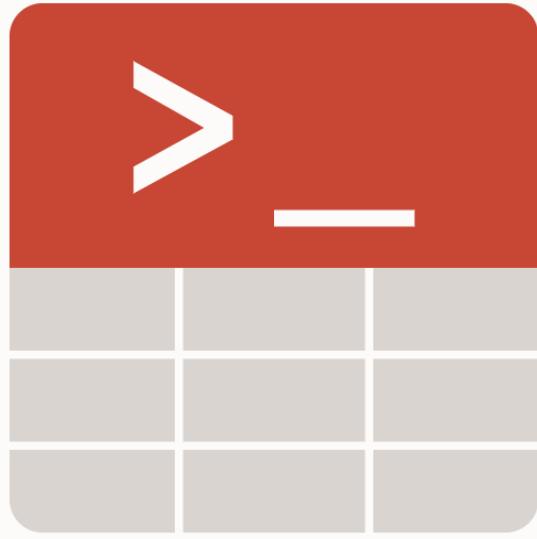
Documento

Flexibilidade e agilidade

```
/* QUERY DOCUMENT */  
SELECT  
    i.doc.to.company AS COMPANY  
    ,i.doc.value      AS VALUE  
FROM invoice i;  
  
/* UPDATE DOCUMENT */  
UPDATE invoice  
SET doc = JSON_TRANSFORM (  
    doc,  
    REMOVE '$.to.name',  
    SET '$.to.value' = 59.90  
)  
FROM invoice i  
WHERE i.id = 1001;
```



```
{ transaction: 'aKl91qh0dFpqi',  
  from: {  
    id: 10039,  
    name: 'Camille'  
  },  
  to: {  
    id: 8021,  
    name: 'Nikita',  
    company: 'Pizza Co.'  
  },  
  date: '2030-11-17',  
  value: 59.90  
}
```



Live SQL

Learn and share SQL

Tutorial

SQL/JSON Features in Database

Documento XML

Interoperabilidade e semântica

O XML é uma recomendação da W3C para gerar **linguagens de marcação** para necessidades especiais. A simplicidade e flexibilidade deste tipo de documento, que se utiliza de **tags e atributos**, permite que diferentes sistemas possam interoperar por meio de um **formato universal**.



```
<transaction id="aKl91qh0dFpqi">
  <from>
    <id>10039</id>
    <name>Camille</name>
  </from>
  <to>
    <id>8021</id>
    <name>Nikita</name>
    <company>Pizza Co.</company>
  </to>
  <date>2021-11-18</date>
  <value>69.90</value>
</transaction>
```

Documento

Interoperabilidade e semântica

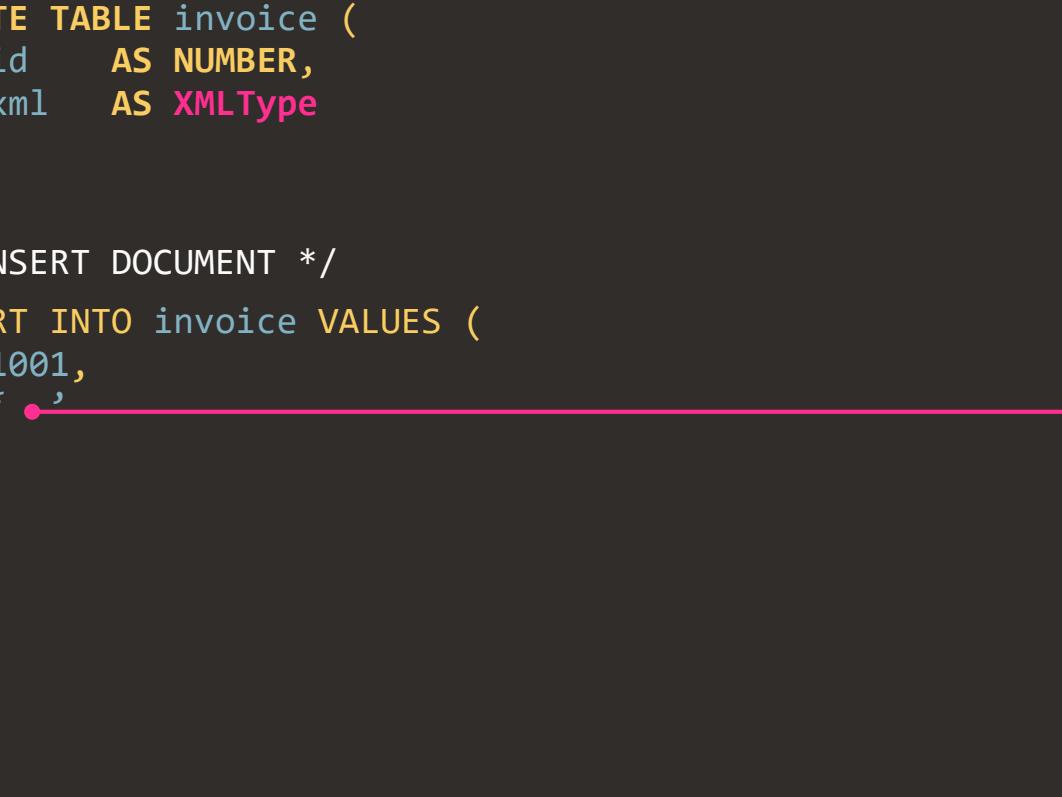
```
/* CREATE DOCUMENT */  
CREATE TABLE invoice (  
    id    AS NUMBER,  
    xml   AS XMLType  
);
```

```
<transaction id="aKl91qh0dFpqi">  
    <from>  
        <id>10039</id>  
        <name>Camille</name>  
    </from>  
    <to>  
        <id>8021</id>  
        <name>Nikita</name>  
        <company>Pizza Co.</company>  
    </to>  
    <date>2021-11-18</date>  
    <value>69.90</value>  
</transaction>
```

Documento

Interoperabilidade e semântica

```
/* CREATE DOCUMENT */  
CREATE TABLE invoice (  
    id    AS NUMBER,  
    xml   AS XMLType  
);  
  
/* INSERT DOCUMENT */  
INSERT INTO invoice VALUES (  
    1001,  
    '  
);
```



```
<transaction id="aK191qh0dFpqi">  
    <from>  
        <id>10039</id>  
        <name>Camille</name>  
    </from>  
    <to>  
        <id>8021</id>  
        <name>Nikita</name>  
        <company>Pizza Co.</company>  
    </to>  
    <date>2021-11-18</date>  
    <value>69.90</value>  
</transaction>
```

Documento

Interoperabilidade e semântica

```
/* QUERY DOCUMENT OPTION 1*/
```

```
SELECT  
    XMLQUERY('/transaction/to/company'  
        PASSING i.xml  
        RETURNING CONTENT).GETCLOBVAL() AS COMPANY
```

```
,XMLQUERY('/transaction/value'  
        PASSING i.xml  
        RETURNING CONTENT).GETCLOBVAL() AS VALUE
```

```
FROM invoice i;
```

The diagram illustrates the mapping of an XML query result to an XML document structure. On the left, a SQL query is shown that uses XMLQUERY and GETCLOBVAL functions to extract data from an XML column 'i' in the 'invoice' table. The query extracts the 'COMPANY' node from the '/transaction/to/company' path and the 'VALUE' node from the '/transaction/value' path.

On the right, the resulting XML document is displayed. It consists of a single root element: <transaction id="aK191qh0dFpqi">. This root element contains two child elements: <from> and <to>. The <from> element has an 'id' attribute of 10039 and a 'name' of Camille. The <to> element has an 'id' attribute of 8021 and a 'name' of Nikita. Both the 'from' and 'to' elements have a child element <company> named 'Pizza Co.'.

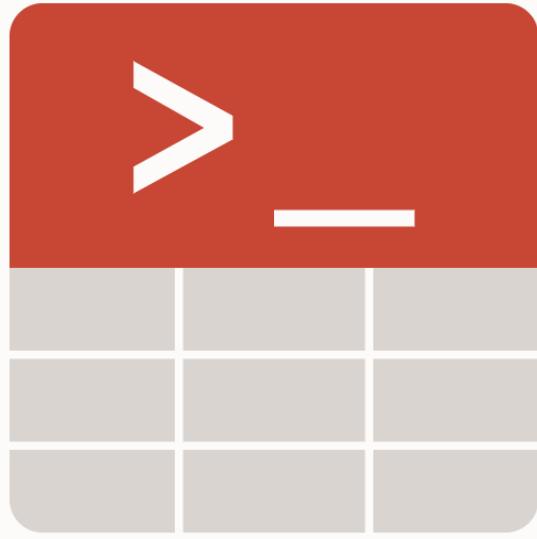
```
<transaction id="aK191qh0dFpqi">  
  <from>  
    <id>10039</id>  
    <name>Camille</name>  
  </from>  
  <to>  
    <id>8021</id>  
    <name>Nikita</name>  
    <company>Pizza Co.</company>  
  </to>  
  <date>2021-11-18</date>  
  <value>69.90</value>  
</transaction>
```

Documento

Interoperabilidade e semântica

```
/* QUERY DOCUMENT OPTION 2*/
SELECT
    inf.*
FROM invoice i
,XMLTABLE('/transaction' PASSING i.xml
          COLUMNS
            company VARCHAR2(128) PATH 'to/company'
            value    NUMBER      PATH 'value'
          ) inf
;
```

```
<transaction id="aK191qh0dFpqi">
  <from>
    <id>10039</id>
    <name>Camille</name>
  </from>
  <to>
    <id>8021</id>
    <name>Nikita</name>
    <company>Pizza Co.</company>
  </to>
  <date>2021-11-18</date>
  <value>69.90</value>
</transaction>
```



Live SQL

Learn and share SQL

Tutorial

Storing and Processing XML Documents



Vamos trilhar este
caminho juntos?



Parabéns!

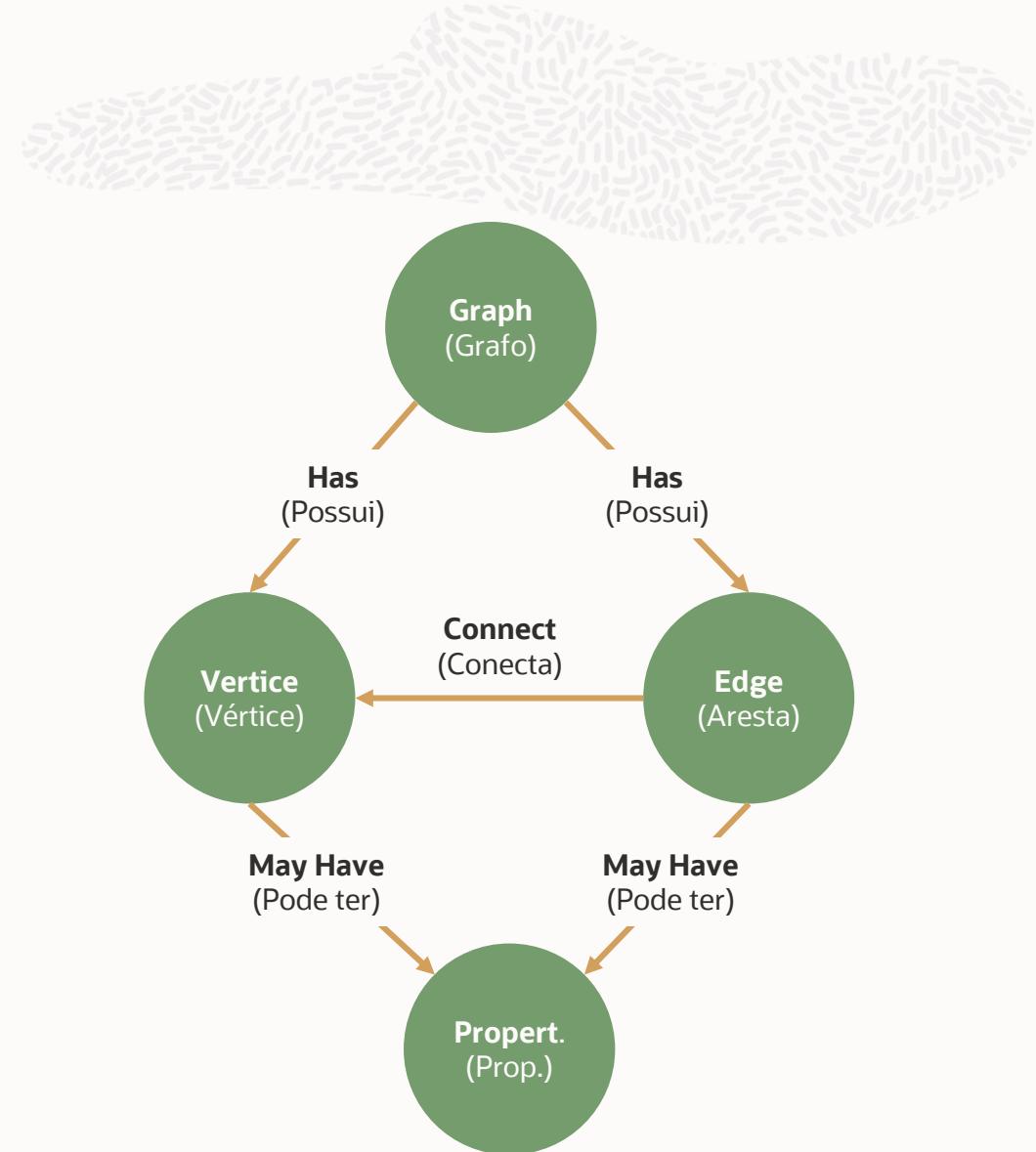
Curti! Tem mais?

Grafo

Análise de relacionamentos

Grafo é a representação de um conjunto de objetos em que seus **pares de objetos** estão, de alguma maneira, **relacionados**.

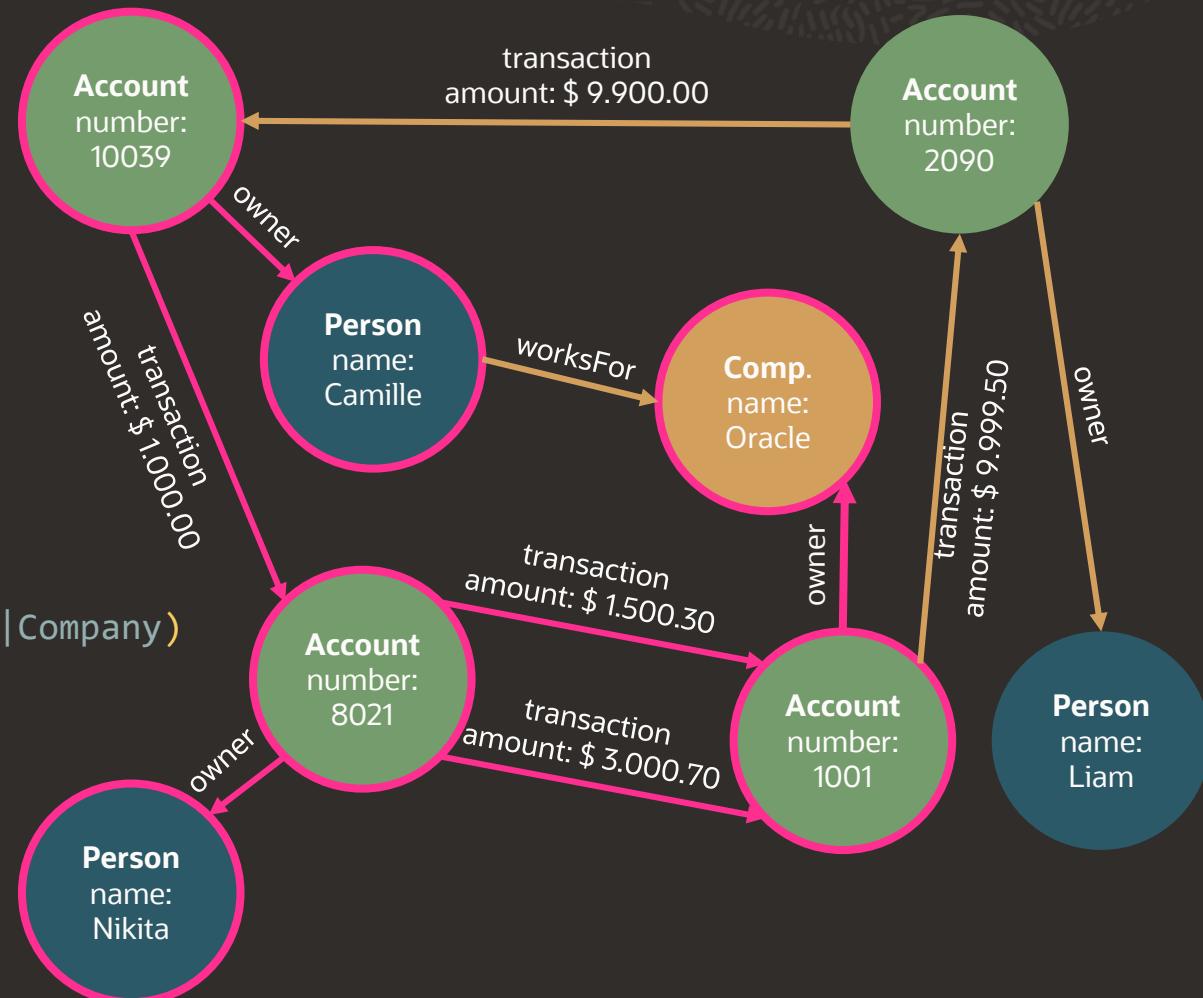
Apesar de possuir uma estrutura simples, pode representar conceitos complexos e seu maior potencial é alcançado utilizando algoritmos analíticos.



Grafos

Mapeamento de vértices e arestas

```
/* PGQL QUERY */  
SELECT  
    owner.name AS account_holder,  
    SUM(t.amount) AS total_transacted  
FROM  
    MATCH (p:Person) <-[:owner]- (account1:Account),  
    MATCH (account1) -[t:transaction]- (account2),  
    MATCH (account2:Account) -[:owner]-> (owner:Person|Company)  
WHERE p.name = 'Nikita'  
GROUP BY owner
```



Espacial

Analisando dados geoespaciais

Armazenamento, processamento e análise de **dados geoespaciais**.

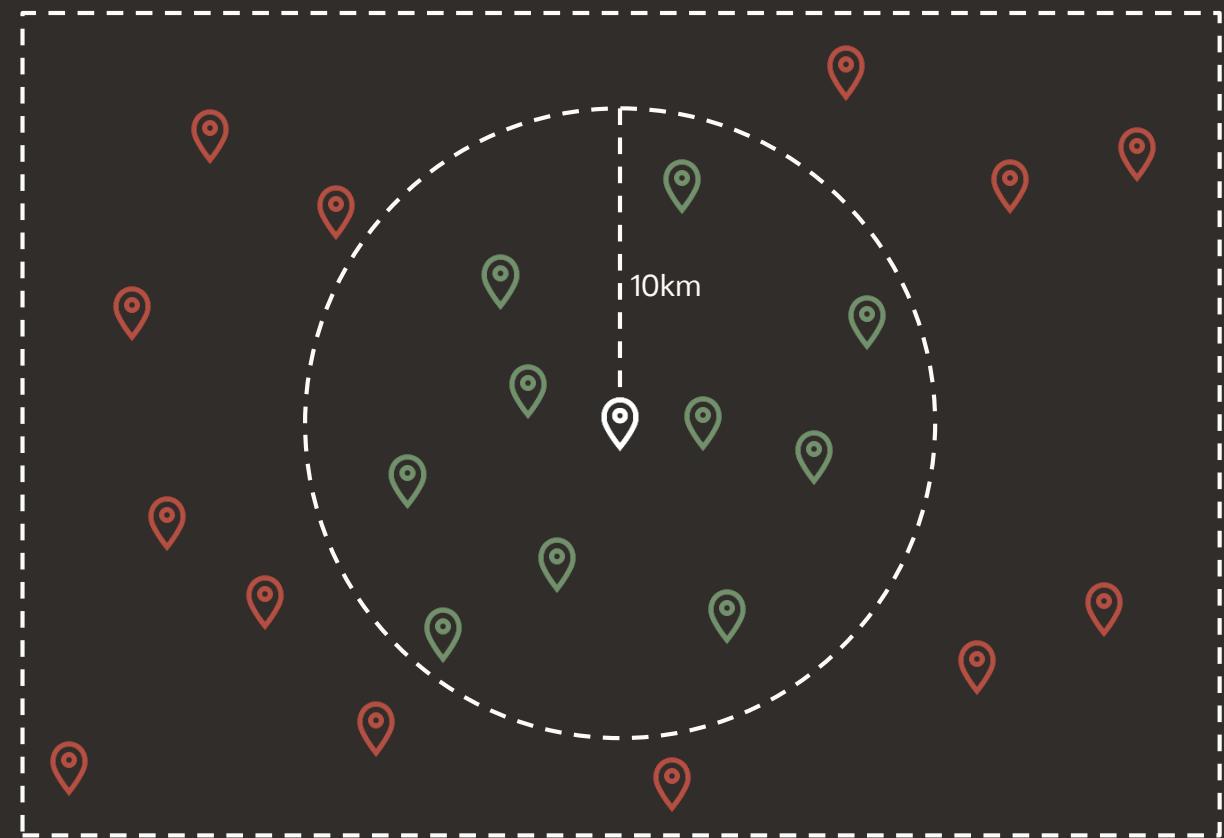
- 2D Spatial Data
- 3D Point Cloud e LiDAR
- Raster Data
- Topology Data
- Streaming Point Data



Espacial

Analisando dados geoespaciais

```
/* LOCATION-BASED QUERY */
SELECT
    c.customer_id
    ,c.cust_last_name
    ,c.gender
FROM warehouses w,
     customers c
WHERE
    w.warehouse_id = 3
AND SDO_WITHIN_DISTANCE(
    c.cust_geo_location
    ,w.wh_geo_location
    ,'DISTANCE=10 UNIT=KM') = 'TRUE'
;
```

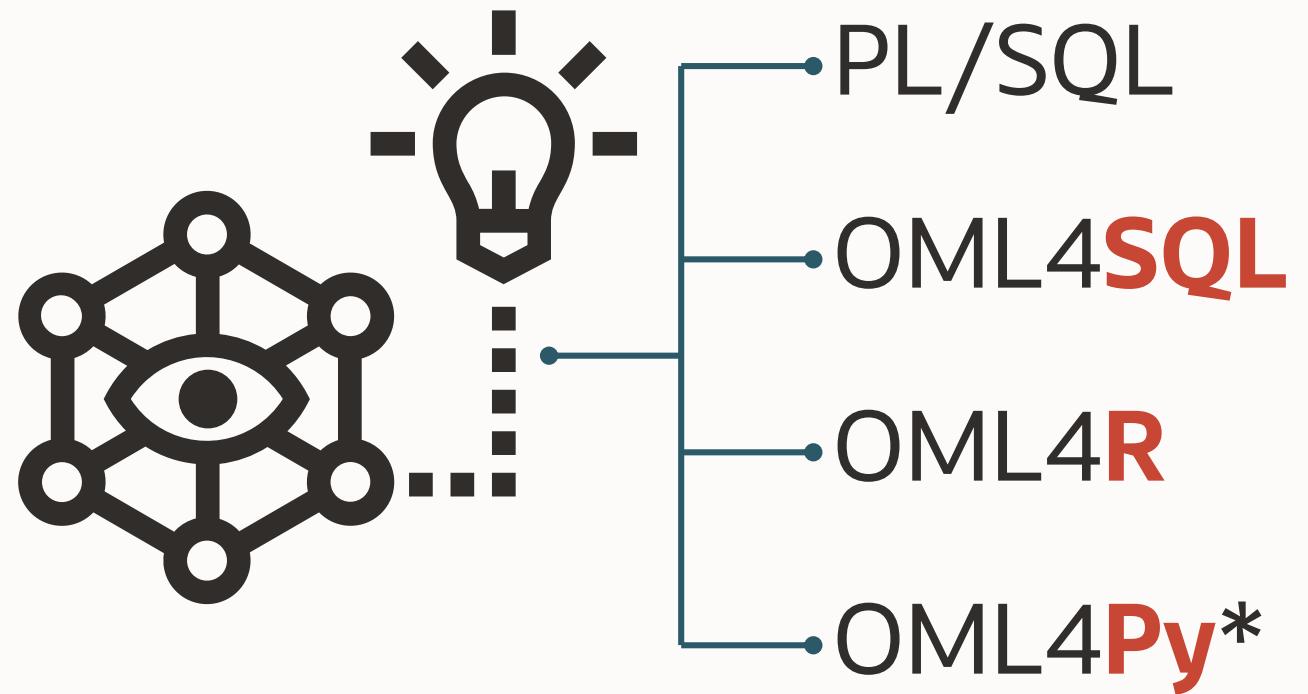


Machine Learning

Treinamento de modelos e predição de valores



Com a funcionalidade de **Machine Learning**, a Oracle move os algoritmos para os dados, processando-os onde estiverem localizados - **minimizando ou eliminando a movimentação de dados**, obtendo escalabilidade, preservando a segurança e acelerando a implementação até o modelo.



Machine Learning

Treinando um modelo



```
/* CREATE THE MODEL */  
  
BEGIN  
  DBMS_DATA_MINING.CREATE_MODEL(  
    model_name      => 'SVMC_SH_Clas_sample',  
    mining_function => dbms_data_mining.classification,  
    data_table_name  => 'mining_data_build_v',  
    case_id_column_name => 'cust_id',  
    target_column_name  => 'affinity_card',  
    settings_table_name => 'svmc_sh_sample_settings');  
END;  
/  
/* USE THE MODEL */
```

```
SELECT cust_gender,  
       COUNT(*) AS cnt,  
       ROUND(AVG(age)) AS avg_age  
  FROM mining_data_apply_v  
 WHERE PREDICTION(SVMC_SH_Clas_sample  
                   USING *) = 1  
 GROUP BY cust_gender  
 ORDER BY cust_gender;
```

Blockchain Table

Prevenção de manipulações nos dados

Tabelas blockchain utilizam **criptografia para encadear** seus registros. Os registros tornam-se imutáveis, auxiliando na identificação e **prevenção de manipulações não autorizadas.** Permite que a cadeia possa ser verificada e assinada por seus participantes.

ID	User	Value	Hash
1	Tom	500	ADSJS
2	Carol	176	%SHS
3	Steve	500	SH@1
4	John	176	DHD3
5	Mike	332	*EGG
6	Sarah	632	AH11
7	Eve	25	LIO\$
8	Prisha	850	SHS4

BLOCKCHAIN TABLE



Blockchain Table

Prevenção de manipulações nos dados

```
/* CREATE A BLOCKCHAIN TABLE */
CREATE BLOCKCHAIN TABLE ledger_emp (
    employee_id NUMBER
    , employee_name VARCHAR2(128)
    , salary NUMBER
)
NO DROP UNTIL 31 DAYS IDLE
NO DELETE LOCKED
HASHING USING "SHA2_512"
VERSION "v1";
```

ID	User	Value	Hash	Link
1	Tom	500	ADSJS	
2	Carol	176	%SHS	
3	Steve	500	SH@1	
4	John	176	DHD3	
5	Mike	332	*EGG	
6	Sarah	632	AH11	
7	Eve	25	LIO\$	
8	Prisha	850	SHS4	

BLOCKCHAIN TABLE



github.com/dspanizzo



linkedin.com/in/dspanizzo

Obrigado!



Vamos juntos nesta trilha!



linkedin.com/groups/8984009



youtube.com/c/Inovaçãocomdadosemnuvem



anchor.fm/inova-dados-nuvem



github.com/taborda-cbip/inovacao-com-dados-em-nuvem