

Lesson 3

Graph Databases' Resource Description Framework (RDF) and SparQL

Triple-Store

- Represents graph database in Resource Description Framework (RDF)
- [A triple instance identifier-property name-property value uses a Universal Resource Indicator (URI) for instance identifier.]

Instance Identifier-property Name-property Value Triple

- Represents the nodes and edges
- An instance identifier in triple— like an entity in a SQL database
- Property name— like a key
- Property value— like a value in a field.

RDF

- Is a graph database in the format triple-store
- Is a simple, yet very effective language for representing information using triples
- Is a W3C (World Wide Consortium) standard for storing a graph database

Features in RDF

1. An RDF data file similar to three columns of triples: subject-predicate-objects
 - Also similar to triplets of document-key-values in the MongoDB.
 -

...Features in RDF

2. An RDF does not depend on a schema and is thus flexible

A standard RDF schema provides definitions of classes and relationships between the properties and classes;

...Features in RDF

3. RDF provides for inclusion of new entities and relationships, just as a tabular database provides for inclusion of additional rows, or a columnar-family database for additional columns.

...Features in RDF

4. RDF provides for inclusion of additional properties to the relationship, association and attribute in a triple

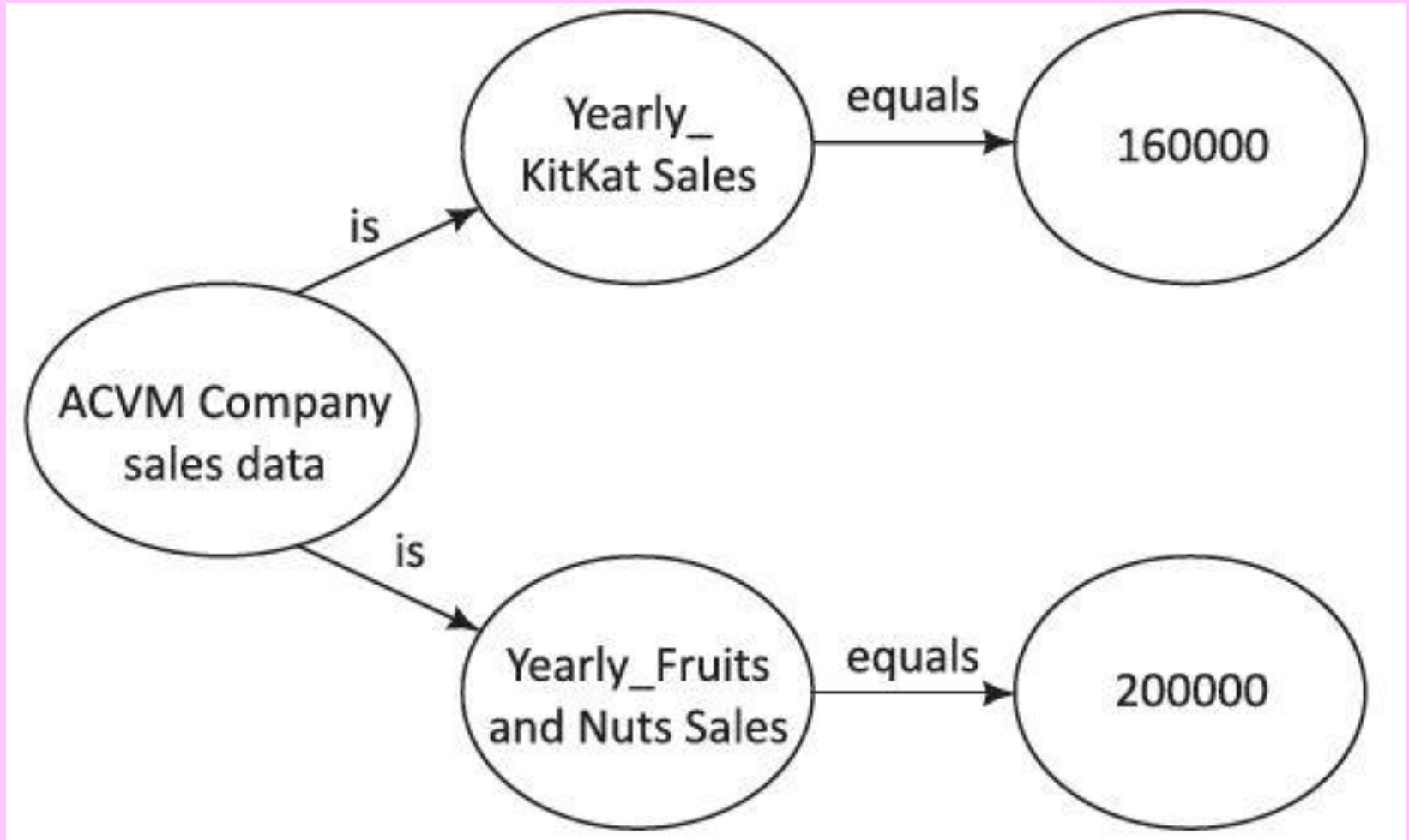
...Features in RDF

5. Simple concatenation combines multiple datasets, used as a whole.
6. Easy sharding as splitting the triples into multiple lines does not change the collective meaning

RDF

- Refer Example 8.3 for how to write
- graph database of Automatic Chocolate Vending Machine (ACVM) Company daily and yearly sales

Figure 8.2: Graph for yearly total sales of an ACVM company



Spark Query Language (SPARQL)

- A query language for RDF graph database
- Provides for querying data using graph traversal along a path
- Traversal may be single step, path expression, or full recursion

SparQL Features

- Allows taking the data without definition for separate schema
- Considers a schema as part of the data itself
- (Schema information may be provided separately, which enable joining of datasets without any problem)

SparQL Features

- Provides query operators needed during graph analytics— JOIN, SORT, AGGREGATE operators
- Provides syntax for specific graph path traversals
- Includes queries for conjunctions, disjunctions, triple patterns and optional patterns in triples

Example 8.4 SparQL

- Refer Example 8.4 for how to write a SPARQL query for output of RDF triples and the results of query processor

Summary

We learnt:

- Graph Database as RDF, a triple-store storing triples using Universal Resource Indicators (URIs)
- RDF Features
- Example of yearly total sales graph-database of an ACVM company in RDF

Summary

We learnt:

- SparQL, a query language for RDF graph database

End of Lesson 3 on Graph Databases' Resource Description Framework (RDF) and SparQL