Spark GraphFrames

Anurag Nagar, Ph.D.

Big Data Class
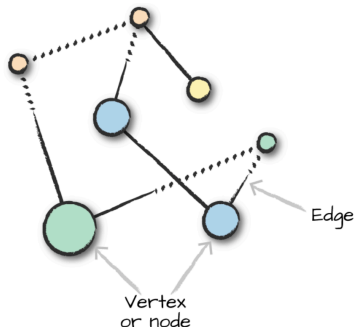
Outline

Background

GraphFrames

- graph processing library for GraphX
- graph library based on DataFrames
- uniform
  API for data and graph processing
- easier way
  to crate graphs from tabular data

Edge
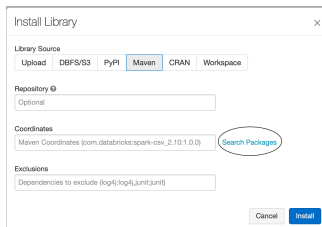
Vertex
or node

Installation on Databricks

Steps

1. Start Databricks cluster

2. Go to clusters links and hover your mouse over the cluster

3. Click "Libraries" link
and then on the next page, click "Install New" button.

[⚙ Install New]

Installation on Databricks

4. On the popup window, go to the
Maven tab and then click on Search package

5. Search for "GraphFrame"
Spark package and choose the release
that matches your cluster configuration.

6. Finally click on Install and your cluster
should contain the required jar dependency.

Installation on a SBT project

Install the appropriate version according to your Spark and Scala versions
from
https://mvnrepository.com/artifact/graphframes/graphframes
or
https://spark-packages.org/package/graphframes/graphframes

Example 1

Outline

Example 1

Example

Let's create a DataFrame of nodes

```scala
import org.graphframes.GraphFrame
val v = sqlContext.createDataFrame(List(
  ("a", "Alice", 34),
  ("b", "Bob", 36),
  ("c", "Charlie", 30),
  ("d", "David", 29),
  ("e", "Esther", 32),
  ("f", "Fanny", 36),
  ("g", "Gabby", 60)
)).toDF("id", "name", "age")
```

Example 1

Example

Let's create a DataFrame of edges (relationships)

```
val e = sqlContext.createDataFrame(List(
  ("a", "b", "friend"),
  ("b", "c", "follow"),
  ("c", "b", "follow"),
  ("f", "c", "follow"),
  ("e", "f", "follow"),
  ("e", "d", "friend"),
  ("d", "a", "friend"),
  ("a", "e", "friend")
)).toDF("src", "dst", "relationship")
```

Example 1

Example

Finally, let's create the graph and run some queries.

```scala
val g = GraphFrame(v, e)

// indegrees of each node
display(g.inDegrees)

// youngest person
val youngest = g.vertices.groupBy().min("age")
display(youngest)
```

Example 1

Example

Go through the complete example at:
https://graphframes.github.io/graphframes/docs/_site/user-guide.html

Example 2

Outline

Example 2

Example 2

We will work through further examples from Chapter 30 of Spark: The Definitive Guide
https://github.com/databricks/Spark-The-Definitive-Guide