

Apache Spark

Andrés Gómez Ferrer
andresgomezfrr@gmail.com

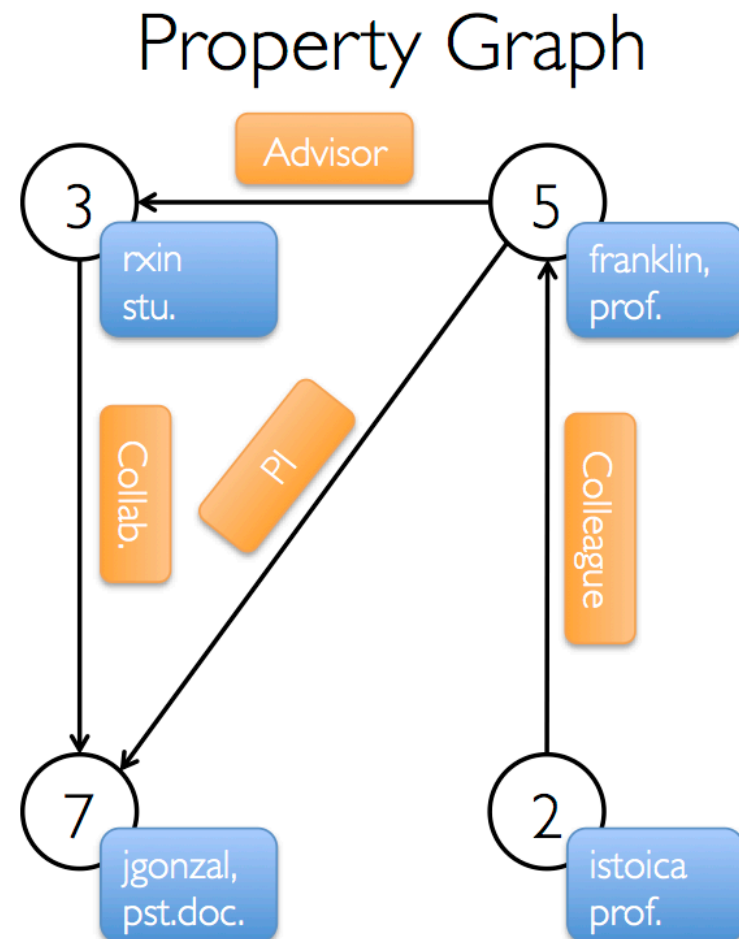


Spark GraphX



■ Spark GraphX

- GraphX es una librería en Spark para trabajar con gráficos de manera distribuida.
- Un grafo se compone de vertex y edge.



Vertex Table

Id	Property (V)
3	(rxin, student)
7	(jgonzal, postdoc)
5	(franklin, professor)
2	(istoica, professor)

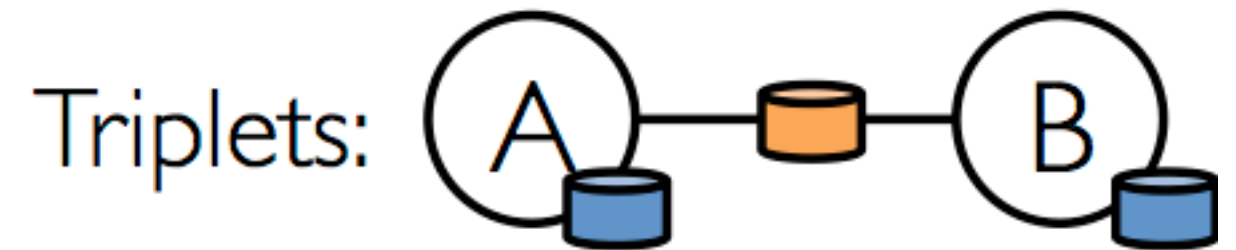
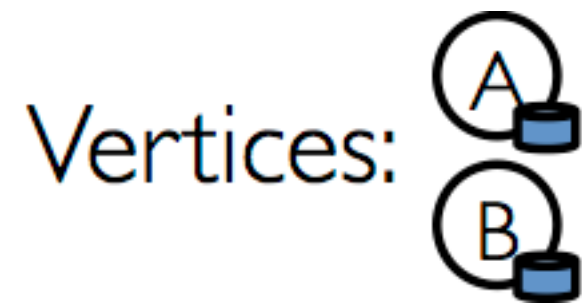
Edge Table

SrcId	DstId	Property (E)
3	7	Collaborator
5	3	Advisor
2	5	Colleague
5	7	PI



■ Spark GraphX

- Spark Graph nos permite trabajar con:



- **Vertices:** Vertices existentes:
- **Edges:** Relaciones existentes:
- **Triplets:** Relación entre dos vértices y su edge.



■ Spark GraphX: Algorithms

- Spark GraphX nos proporciona una serie de algoritmos para trabajar:
 - **PageRank**: Mide la importancia de cada vertex en el grafo. Considerante el numero de relaciones que tiene cada vértice, directas o indirectas.
 - **Connected Componentes**: Obtiene los componentes conectados.
 - **Triangle Counting**: Obtiene los componentes que están conectados formando un triángulo.



Exercise 1

School GraphX

