Scalability in RDF Stream Processing Systems

Alejandro Llaves and Oscar Corcho

Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
{allaves,ocorcho}@fi.upm.es

There is no universal definition of scalability in the context of software systems [6]. In the field of RDF Stream Processing (RSP), we can consider that a system is scalable if it is able to process queries over a growing amount of input data efficiently, i.e. without losing quality of service. As a community interested in RSP, we have to (1) identify the causes and effects that affect the performance of a RSP system and (2) understand their relationship to keep the system performance as desired [4].

Our research focuses on scalable approaches to execute queries over RDF data streams. We investigate real-time processing technologies, e.g. Storm\(^1\) or Spark\(^2\), to parallelize the execution in distributed environments. We are also interested in self-adaptive strategies that allow systems to react to changes in the input data (e.g. frequency, number of streams, or heterogeneity), variable query complexity, and failure of processing nodes.

Dr. Oscar Corcho is an associate professor and leads the Data Integration group at the Ontology Engineering Group, Universidad Politécnica de Madrid. He has broader experience in the field of RSP, ranging from novel methods to access streams, definition of sensor observation ontologies, and RDF compression techniques [10,1,3,2,5].

Dr. Alejandro Llaves works as a postdoctoral researcher at Corcho’s group. His background is on event processing in the Semantic Sensor Web [11], with publications related to semantic interoperability among information communities [7,9,8].

We are active participants of the W3C RSP community group\(^3\) and believe that our contribution to this workshop is relevant to the ESWC audience because of the cross-domain nature of RSP, addressing fields like Distributed Systems, Social Networks, and Internet of Things.

Acknowledgements

This research has been funded by Ministerio de Economía y Competitividad (Spain) under the project ”4V: Volumen, Velocidad, Variedad y Validez en la Gestión Innovadora de Datos” (TIN2013-46238-C4-2-R).

---

\(^1\) https://storm.apache.org/
\(^2\) https://spark.apache.org/streaming/
\(^3\) https://www.w3.org/community/rsp/
References