

# A systematic overview of data federation systems — A simplified version

Zhenzhen Gu<sup>a</sup>, Francesco Corcoglioniti<sup>a</sup>, Davide Lanti<sup>a</sup>, Alessandro Mosca<sup>a</sup>, Guohui Xiao<sup>b,c,d,\*</sup>,  
Jing Xiong<sup>a</sup> and Diego Calvanese<sup>a,d,e</sup>

<sup>a</sup> *KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy*

*E-mail: <name>.<surname>@unibz.it*

<sup>b</sup> *Department of Information Science and Media Studies, University of Bergen, Norway*

*E-mail: guohui.xiao@uib.no*

<sup>c</sup> *Department of Informatics, University of Oslo, Norway*

*E-mail: guohuix@ifi.uio.no*

<sup>d</sup> *Ontopic S.r.l, Italy*

*E-mail: <name>.<surname>@ontopic.ai*

<sup>e</sup> *Department of Computing Science, Umeå University, Sweden*

*E-mail: diego.calvanese@umu.se*

This document is a simplified version of the paper published in the Semantic Web Journal and provided a systematic overview of 50 data federation systems. The motivation of this document is to provide a way of making users, such as data federation consumers and developers, to check or get knowledge of the characteristics of the considered data federation systems quickly rather than reading a paper with 60 pages.

---

\*Corresponding author. E-mail: guohui.xiao@uib.no.

Table 1  
Summary of the selected data federation systems. Academic systems in *italics*

System	Provider	Description
AllegroGraph [1]	Franz Inc.	Distributed graph & document DB supporting OWL, SPARQL, SHACL and federation
Amazon Athena [2]	Amazon.com, Inc.	Inter. cloud query service for Amazon S3 data, based on Presto [3]
Amazon Neptune [4]	Amazon.com, Inc.	Fully-managed cloud graph DB (property graph, RDF), part of Amazon AWS
AnzoGraph DB [5]	Cambridge Semantics	Massively-parallel distributed graph DB (property-graph, RDF) for large-scale analytics
Apache Drill [6, 7]	Apache Software Foundation	Distributed schema-free engine for interactive SQL queries on heterogeneous & nested data, inspired by Dremel [8]
Apache Jena [9]	Apache Software Foundation	SPARQL query engine of Jena framework and TDB triple store, supporting federation
Apache Spark [10, 11]	Apache Software Foundation	Multi-lang. (incl. SQL) distributed engine for large-scale data processing & analytics
<i>BigDAWG</i> [12, 13]	Intel Science & Technology Center for Big data	Polystore with heterogeneous storage engines for time series (SciDB), text (Accumulo) and relational data (PostgreSQL)
Blazegraph [14]	Systap	Triple store supporting SPARQL 1.1 federation and powering Wikidata (via a fork)
<i>CloudMdsQL</i> [15, 16]	Inria & LIRMM	Polystore integrating heterogeneous storage engines (incl. RDBMS, NoSQL, HDFS)
<i>Comunica</i> [17]	Univ. Ghent	Modular JS federated query engine for heterogeneous web sources, incl. SPARQL endpoints
<i>CostFed</i> [18]	Univ. Leipzig	Index-assisted, cost-based data federation system for SPARQL endpoints
<i>DARQ</i> [19]	Univ. HU Berlin	Earliest data federation system for SPARQL endpoints, cost-based
Data Virtuality [20]	Data Virtuality GmbH	Heterogeneous data integration solution combining data virtualization and ETL
Denodo [21]	Denodo Technologies Inc.	Data virtualization solution for heterogeneous sources, also available as cloud service
Dremio [22]	Dremio Corporation	Data “lakehouse” (lake + warehouse) solution supporting heterogeneous data sources
<i>FEDRA</i> [23]	Univ. Nantes (LINA lab.)	Data federation system for SPARQL endpoints exploiting data replication
FedX (RDF4J) [24, 25]	fluid Operations AG	On-demand (no statistics, query-time) data federation system for SPARQL endpoints
GraphDB [26]	Ontotext	Triple store featuring OWL reasoning, SPARQL federated queries & RDBMS access
<i>HiBISCuS</i> [27]	Univ. Leipzig	Source selection for SPARQL data federation (DARQ, FedX & SPLENDID extension)
IBM Cloud Pak for Data [28]	IBM	Data federation system with data discovery, governance, security and privacy solutions, also available as cloud service (formerly IBM Cloud Private for Data)
IBM Db2 Big SQL [29]	IBM	Massively-parallel Hadoop SQL engine for heterogeneous sources (formerly IBM SQL)
IBM InfoSphere Federation Server [30]	IBM	SQL-based data federation system for heterogeneous sources (formerly WebSphere Federation Server)
JBoss Data Virtualization [31]	Red Hat, Inc.	Data federation system based on Teiid and providing read/write access to heterogeneous sources, data security, and multiple user interfaces / APIs
Metaphactory [32, 33]	metaphacts GmbH	KG platform on top of SPARQL endpoints with two federation engines (Ephedra, FedX)
<i>Myria</i> [34]	Univ. Washington	Cloud service for big data management/analytics with parallel & federated query engine
Neo4j (Fabric) [35]	Neo4j, Inc.	Federation solution of Neo4J graph DB (Cypher [36] queries on property graph model)
<i>Obi-Wan</i> [37, 38]	Inria & Polytechnic Institute of Paris	Ontology-Based Data Access (OBDA) [39] system on top of Tatooine [40] mediator for heterogeneous sources
<i>Odyssey</i> [41]	Univ. Aalborg & Univ. Nantes	Statistics & cost-based optimizer for SPARQL data federation (FedX extension)
<i>Ontario</i> [42]	L3S Research Center	Heuristics-based system using RDF Molecule Templates (introduced by its predecessor MULDER [43]) to describe/map source content as star-shaped RDF instance descriptions
<i>Onto-KIT</i> [44]	Univ. Toulouse	Data federation system focusing on Earth Observation data with hypergraph-based data model and query processing techniques
Oracle Big Data SQL [45]	Oracle Corporation	Data federation system for Oracle DB that accesses Hadoop storage & processing
Oracle DB (Spatial & Graph) [46, 47]	Oracle Corporation	Oracle DB component for semantic technologies with data federation capabilities (RDF views) over relational, graph, and RDF (SPARQL) sources
<i>PolyWeb</i> [48, 49]	Univ. NUI Galway	SPARQL-based data federation system for different sources on the Web (RDF & CSV data, RDBMS), focusing on source selection, query optimization & execution
Presto [3, 50]	Presto Foundation	SQL-based distributed query engine suitable to interactive (big) data analytics
Queron Data Virtualization [51]	YouNeedIT Sp. z o.o. Sp. k.	Data federation system for a variety of heterogeneous sources, based on Apache Spark and targeting big data analytics with the support of main BI tools
RDFLib [52]	RDFLib team	A pure Python package for working with RDF, supporting SPARQL 1.1 federation
<i>SAFE</i> [53]	Insight SFI Research Centre for Data Analytics	Data federation system for SPARQL endpoints exposing RDF data cubes with sensitive data, featuring access policy-aware data summaries, source selection & query execution
<i>SAGE</i> [54]	Univ. Nantes	SPARQL engine with “web preemption” (i.e., query suspend/resume) & federation capabilities
SAP HANA [55]	SAP SE	In-memory DB targeting with data federation capabilities, also available as cloud service
SAS Federation Server [56]	SAS Institute	Data federation system featuring data caches, masking, encryption & quality functions
<i>SemaGrow</i> [57]	IIT NCSR ‘Demokritos’	Data federation system for SPARQL endpoints with statistics-based query optimization
<i>SPLENDID</i> [58]	Univ. Koblenz-Landau	Data federation system for SPARQL endpoints that provide VOID [59] data statistics
SQL Server (PolyBase) [60]	Microsoft Corporation	SQL Server component for data federation supporting Hadoop and Azure storage
<i>Squerall</i> [61]	Univ. Bonn	Data federation system for heterogeneous sources built on Spark, Presto, and RML mappings
Starburst [62]	Starburst Data, Inc.	Commercial distribution of Trino, extra security features, available on-premise/on-cloud
Stardog [63]	Stardog Union	KG platform including data federation of heterogeneous sources & query-time inference
Teiid [64]	Red Hat, Inc.	SQL-based engine for data federation of heterogeneous sources
TIBCO Data Virtualization [65]	TIBCO Software Inc.	Data federation system for heterogeneous sources, with data caching & security, massively parallel processing & GUI tools (formerly Composite, then Cisco Data Virtualization)
Trino [66]	Trino Software Foundation	SQL-based query distributed engine for interactive big data analytics, forked from Presto
Virtuoso [67–69]	OpenLink Software	Multi-model DB (object-relational, RDF, XML) with data federation facilities

Table 2

Evaluation of query language and data source sub-dimensions. Academic systems in *italics*. “–” denotes feature/information not found in the systems’ official documentation, websites, or academic publications, to the best of our efforts

System	Query language			Data source					
	SPARQL	SQL	Other	Relational	Graph-based	Aggregate-oriented	Structured Files	Web Service Paradigms	Other
AllegroGraph	✓	–	Prolog	–	✓	–	–	–	–
Amazon Athena	–	✓	–	✓	✓	✓	✓	–	✓
Amazon Neptune	✓	–	–	–	✓	–	–	–	–
AnzoGraph DB	✓	–	Cypher	✓	✓	–	✓	✓	–
Apache Drill	–	✓	–	✓	–	✓	✓	✓	✓
Apache Jena	✓	–	–	–	✓	–	–	–	–
Apache Spark	–	✓	–	✓	–	–	✓	–	–
<i>BigDAWG</i>	–	–	BigDAWG Query	✓	–	✓	–	–	✓
Blazegraph	✓	–	–	–	✓	–	–	–	–
<i>CloudMdsQL</i>	–	–	CloudMdsQL	✓	✓	✓	–	–	–
<i>Comunica</i>	✓	–	GraphQL	–	✓	–	✓	–	–
<i>CostFed</i>	✓	–	–	–	✓	–	–	–	–
<i>DARQ</i>	✓	–	–	–	✓	–	–	–	–
Data Virtuality	–	✓	–	✓	✓	✓	✓	✓	✓
Denodo	–	✓	GraphQL	✓	–	✓	✓	✓	✓
Dremio	–	✓	–	✓	–	✓	✓	–	–
<i>FEDRA</i>	✓	–	–	–	✓	–	–	–	–
FedX (RDF4J)	✓	–	–	–	✓	–	–	–	–
GraphDB	✓	✓	Cypher	✓	✓	–	–	–	–
<i>HiBISCuS</i>	✓	–	–	–	✓	–	–	–	–
IBM Cloud Pak for Data	–	✓	–	✓	–	✓	✓	✓	✓
IBM Db2 Big SQL	–	✓	–	✓	–	✓	✓	–	✓
IBM InfoSphere Federation Server	–	✓	–	✓	–	–	✓	✓	✓
JBoss Data Virtualization	–	✓	–	✓	–	✓	✓	✓	✓
Metaphactory	✓	–	–	✓	✓	✓	–	✓	–
<i>Myria</i>	–	✓	MyriaL	–	✓	✓	✓	–	✓
Neo4j (Fabric)	–	–	Cypher	–	✓	–	–	–	–
<i>Obi-Wan</i>	✓	–	–	✓	✓	✓	–	–	–
<i>Odyssey</i>	✓	–	–	–	✓	–	–	–	–
<i>Ontario</i>	✓	–	–	✓	✓	✓	✓	–	–
<i>Onto-KIT</i>	✓	–	–	–	–	–	✓	–	–
Oracle Big Data SQL	–	✓	–	✓	–	✓	✓	–	✓
Oracle DB (Spatial & Graph)	✓	–	–	✓	✓	–	–	–	–
<i>PolyWeb</i>	✓	–	–	✓	✓	–	✓	–	–
Presto	–	✓	–	✓	–	✓	–	–	✓
Querona Data Virtualization	–	✓	–	✓	–	✓	✓	–	✓
RDFLib	✓	–	–	–	✓	–	–	–	–
<i>SAFE</i>	✓	–	–	–	✓	–	–	–	–
<i>SAGE</i>	✓	–	–	–	✓	–	–	–	–
SAP HANA	–	✓	–	✓	–	–	–	–	✓
SAS Federation Server	–	–	FedSQL	✓	–	–	–	–	✓
<i>SemaGrow</i>	✓	–	–	–	✓	–	–	–	–
<i>SPLendid</i>	✓	–	–	–	✓	–	–	–	–
SQL Server (PolyBase)	–	✓	–	✓	–	✓	✓	–	–
<i>Squerall</i>	✓	–	–	✓	–	✓	✓	–	–
Starburst	–	✓	–	✓	–	✓	✓	–	✓
Stardog	✓	–	–	✓	✓	✓	✓	–	✓
Teiid	–	✓	–	✓	–	✓	✓	✓	✓
TIBCO Data Virtualization	–	✓	–	✓	–	✓	✓	✓	✓
Trino	–	✓	–	✓	–	✓	–	–	✓
Virtuoso	✓	✓	–	✓	✓	–	–	–	–
<b>Number</b>	<b>27</b>	<b>22</b>	<b>10</b>	<b>32</b>	<b>30</b>	<b>24</b>	<b>24</b>	<b>10</b>	<b>20</b>

Table 3

Evaluation of the data security dimension. Academic systems in *italics*. “–” denotes feature/information not found in the systems’ official documentation, websites, or academic publications, to the best of our efforts. Subscript <sub>ng</sub> denotes the use of named graph-based solutions to hide (mask) sensitive information in selected graphs to certain users, and possibly (for AnzoGraph DB) expose sanitized named graph views

System	Data security				
	Authentication	Authorization	Auditing	Encryption	Data masking
AllegroGraph	✓	✓	–	–	–
Amazon Athena	✓	✓	✓	✓	–
Amazon Neptune	✓	✓	✓	✓	–
AnzoGraph DB	✓	✓	–	–	✓ <sub>ng</sub>
Apache Drill	✓	✓	–	✓	–
Apache Jena	✓	–	–	–	–
Apache Spark	✓	✓	–	✓	–
<i>BigDAWG</i>	–	–	–	–	–
Blazegraph	–	–	–	–	–
<i>CloudMdsQL</i>	–	–	–	–	–
<i>Comunica</i>	–	–	–	–	–
<i>CostFed</i>	–	–	–	–	–
<i>DARQ</i>	–	–	–	–	–
Data Virtuality	✓	✓	–	–	–
Denodo	✓	✓	✓	✓	✓
Dremio	✓	✓	–	✓	✓
<i>FEDRA</i>	–	–	–	–	–
FedX (RDF4J)	✓	–	–	–	–
GraphDB	✓	✓	✓	✓	–
<i>HiBISCuS</i>	–	–	–	–	–
IBM Cloud Pak for Data	✓	✓	✓	✓	✓
IBM Db2 Big SQL	✓	✓	✓	✓	–
IBM InfoSphere Federation Server	✓	✓	–	✓	–
JBoss Data Virtualization	✓	✓	✓	✓	–
Metaphactory	✓	✓	–	–	–
<i>Myria</i>	–	–	–	–	–
Neo4j (Fabric)	✓	✓	–	–	–
<i>Obi-Wan</i>	–	–	–	–	–
<i>Odyssey</i>	–	–	–	–	–
<i>Ontario</i>	–	–	–	–	–
<i>Onto-KIT</i>	–	–	–	–	–
Oracle Big Data SQL	✓	✓	–	–	–
Oracle DB (Spatial & Graph)	✓	✓	–	✓	✓
<i>PolyWeb</i>	–	–	–	–	–
Presto	✓	✓	✓	–	–
Queron Data Virtualization	✓	✓	–	✓	✓
RDFLib	–	–	–	–	–
<i>SAFE</i>	✓	✓	–	–	–
<i>SAGE</i>	–	–	–	–	–
SAP HANA	✓	–	–	–	–
SAS Federation Server	✓	✓	–	✓	✓
<i>SemaGrow</i>	–	–	–	–	–
<i>SPLendid</i>	–	–	–	–	–
SQL Server (PolyBase)	✓	✓	✓	✓	–
<i>Squerall</i>	–	–	–	–	–
Starburst	✓	✓	✓	✓	–
Stardog	✓	✓	–	–	✓ <sub>ng</sub>
Teiid	✓	✓	–	✓	–
TIBCO Data Virtualization	✓	✓	–	✓	–
Trino	✓	✓	–	✓	–
Virtuoso	✓	✓	–	✓	–
Number	32	29	10	20	8

Table 4

Evaluation of the *interface* dimension. Academic systems in *italics*. “–” denotes feature/information not found in the systems’ official documentation, websites, or academic publications, to the best of our efforts

System	Graphical interface	Command line interface	Application programming interface				
			JDBC Driver	ODBC Driver	Web API	ADO.NET	SPARQL HTTP API
AllegroGraph	✓	✓	–	–	✓	–	✓
Amazon Athena	✓	✓	✓	✓	–	–	–
Amazon Neptune	✓	✓	✓	–	✓	–	–
AnzoGraph DB	✓	✓	–	–	✓	–	✓
Apache Drill	✓	✓	✓	✓	✓	–	–
Apache Jena	–	✓	✓	–	–	–	✓
Apache Spark	✓	✓	✓	✓	–	–	–
<i>BigDAWG</i>	–	✓	–	–	✓	–	–
Blazegraph	✓	✓	–	–	✓	–	–
<i>CloudMdsQL</i>	–	✓	✓	–	–	–	–
<i>Comunica</i>	✓	✓	–	–	–	–	✓
<i>CostFed</i>	✓	–	–	–	–	–	–
<i>DARQ</i>	–	✓	–	–	–	–	–
Data Virtuality	✓	✓	✓	✓	✓	–	–
Denodo	✓	✓	✓	✓	✓	✓	–
Dremio	✓	✓	✓	✓	✓	–	–
<i>FEDRA</i>	–	–	–	–	–	–	–
FedX (RDF4J)	✓	✓	–	–	✓	–	✓
GraphDB	✓	✓	✓	–	✓	–	✓
<i>HiBISCuS</i>	–	–	–	–	–	–	–
IBM Cloud Pak for Data	✓	✓	–	–	✓	–	–
IBM Db2 Big SQL	✓	✓	✓	✓	–	–	–
IBM InfoSphere Federation Server	✓	✓	✓	–	✓	–	–
JBoss Data Virtualization	✓	✓	✓	✓	✓	–	–
Metaphactory	✓	✓	–	–	✓	–	✓
<i>Myria</i>	✓	✓	–	–	✓	–	–
Neo4j (Fabric)	✓	✓	✓	–	✓	–	–
<i>Obi-Wan</i>	–	✓	–	–	–	–	✓
<i>Odyssey</i>	–	✓	–	–	–	–	–
<i>Ontario</i>	–	✓	–	–	–	–	–
<i>Onto-KIT</i>	✓	–	–	–	–	–	–
Oracle Big Data SQL	✓	✓	–	–	–	–	–
Oracle DB (Spatial & Graph)	✓	✓	–	–	✓	–	✓
<i>PolyWeb</i>	–	–	–	–	–	–	–
Presto	✓	✓	✓	✓	✓	–	–
Querona Data Virtualization	✓	✓	✓	✓	–	✓	–
RDFLib	–	✓	–	–	–	–	–
<i>SAFE</i>	–	–	–	–	–	–	–
<i>SAGE</i>	✓	✓	–	–	–	–	–
SAP HANA	✓	✓	✓	✓	✓	✓	–
SAS Federation Server	✓	✓	✓	✓	✓	–	–
<i>SemaGrow</i>	✓	✓	–	–	–	–	–
<i>SPLendid</i>	–	✓	–	–	–	–	–
SQL Server (PolyBase)	✓	✓	✓	✓	–	✓	–
<i>Squerall</i>	✓	✓	–	–	–	–	–
Starburst	✓	✓	✓	✓	✓	–	–
Stardog	✓	✓	–	–	✓	–	✓
Teiid	✓	✓	✓	✓	✓	✓	–
TIBCO Data Virtualization	✓	✓	✓	✓	✓	✓	–
Trino	✓	✓	✓	✓	✓	–	–
Virtuoso	✓	✓	✓	✓	✓	✓	✓
<b>Number</b>	<b>38</b>	<b>45</b>	<b>24</b>	<b>18</b>	<b>27</b>	<b>7</b>	<b>11</b>

Table 5

Evaluation of *development* dimension. Academic systems in *italics*. “F.” and “L.” denote “First” and “Latest” respectively. Subscript letters further qualify available deployment options: *n* = native; *c* = containerized; *a* = Amazon AWS; *m* = Microsoft Azure; *g* = Google Cloud Platform. “–” denotes feature/information not found in the systems’ official documentation, websites, or academic publications, to the best of our efforts

System	Main development language			Deployment			Comm. support	Open source	Release			
	C/C++	Java	Others	On-prem	IaaS/PaaS	SaaS			F. Year	F. Version	L. Year	L. Version
AllegroGraph	–	✓	Lisp	✓ <sub>nc</sub>	✓ <sub>am</sub>	–	✓	–	2004	6.4.0	2021	7.2.0
Amazon Athena	–	✓	–	–	–	✓ <sub>a</sub>	✓	–	2017	–	2021	–
Amazon Neptune	–	✓	–	–	–	✓ <sub>a</sub>	✓	–	2018	1.0.1.0	2021	1.0.5.1
AnzoGraph DB	✓	–	–	✓ <sub>nc</sub>	✓ <sub>a</sub>	–	✓	–	–	2.0	2021	2.3
Apache Drill	–	✓	–	✓ <sub>nc</sub>	–	–	–	✓	2012	M1	2021	1.19
Apache Jena	–	✓	–	✓ <sub>nc</sub>	–	–	–	✓	2012	2.7.0	2021	4.2.0
Apache Spark	–	–	Scala	✓ <sub>nc</sub>	–	–	–	✓	2014	1.0	2021	3.2.1
BigDAWG	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2015	–	2017	0.0.5
Blazegraph	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2019	2.1.5	2020	2.1.6rc
CloudMdsQL	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2017	–	2017	–
Comunica	–	–	JavaScript	✓ <sub>nc</sub>	–	–	–	✓	2018	1.0.0	2021	1.22.3
CostFed	–	✓	–	✓ <sub>nc</sub>	–	–	–	✓	2016	–	2018	–
DARQ	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2006	–	2008	–
Data Virtuality	–	–	–	✓ <sub>nc</sub>	–	–	✓	–	–	–	2021	2.4
Denodo	–	–	–	✓ <sub>nc</sub>	✓ <sub>amg</sub>	–	✓	–	2002	1.0	2020	8.0
Dremio	–	✓	–	✓ <sub>nc</sub>	✓ <sub>am</sub>	–	✓	✓	2017	1.1	2021	19.0
FEDRA	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2015	–	2015	–
FedX (RDF4J)	–	✓	–	✓ <sub>n</sub>	–	–	✓	✓	2011	–	2021	3.7.4
GraphDB	–	✓	–	✓ <sub>nc</sub>	–	–	✓	–	2015	6.2	2021	9.10
HiBISCuS	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2014	1	2014	1
IBM Cloud Pak for Data	–	–	–	✓ <sub>c</sub>	✓ <sub>amg</sub>	✓	✓	–	2018	2.1.0	2021	4.0
IBM Db2 Big SQL	–	✓	–	✓ <sub>n</sub>	–	✓	✓	–	2017	–	2020	7.1.0
IBM InfoSphere Federation Server	–	–	–	✓ <sub>n</sub>	–	–	✓	–	–	–	2019	10.5.0
JBoss Data Virtualization	–	✓	–	✓ <sub>nc</sub>	–	–	✓	✓	2014	6.0.0	2018	6.4.0
Metaphactory	–	–	–	✓	✓ <sub>a</sub>	–	–	–	2015	–	2021	4.3.0
Myria	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2014	1	2017	1
Neo4j (Fabric)	–	✓	–	✓ <sub>nc</sub>	✓ <sub>amg</sub>	✓	✓	✓	2020	4.0.11	2021	4.3.7
Obi-Wan	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2020	–	2020	–
Odyssey	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2016	–	2019	–
Ontario	–	–	Python	✓ <sub>n</sub>	–	–	–	✓	2018	–	2021	–
Onto-KIT	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2020	–	2020	–
Oracle Big Data SQL	–	–	–	✓ <sub>n</sub>	–	–	–	–	–	3.0.1	2021	4.1.1
Oracle DB (Spatial & Graph)	–	–	–	✓ <sub>nc</sub>	–	–	✓	–	2016	–	2021	21c
PolyWeb	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2017	–	2017	–
Presto	–	✓	–	✓ <sub>nc</sub>	–	–	✓	✓	2013	0.54	2021	0.265.1
Querona Data Virtualization	–	–	–	✓ <sub>n</sub>	–	–	✓	–	2015	–	2020	–
RDFLib	–	–	Python	✓ <sub>n</sub>	–	–	–	✓	2002	1.1.1	2021	6.1.1
SAFE	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2017	–	2017	–
SAGE	–	–	Python	✓ <sub>nc</sub>	–	–	–	✓	2019	1.1	2021	2.3
SAP HANA	✓	–	–	✓ <sub>nc</sub>	✓ <sub>ag</sub>	–	✓	–	2018	1.0.SPS12	2020	2.0.SPS05
SAS Federation Server	✓	–	–	✓ <sub>n</sub>	–	–	✓	–	2013	3.2	2021	4.4
SemaGrow	–	✓	–	✓ <sub>nc</sub>	–	–	–	✓	2014	1.0	2021	2.2.1
SPLendid	–	✓	–	✓ <sub>n</sub>	–	–	–	✓	2011	–	2011	–
SQL Server (PolyBase)	✓	–	–	✓ <sub>nc</sub>	–	–	✓	–	2016	2016	2019	2019
Squerall	–	–	Python	✓ <sub>n</sub>	–	–	–	✓	2018	0.1	2019	0.2
Starburst	–	✓	–	✓ <sub>nc</sub>	✓ <sub>amg</sub>	–	✓	–	2019	0.188-e	2021	364-e LTS
Stardog	–	✓	–	✓ <sub>nc</sub>	✓ <sub>a</sub>	✓	✓	–	2011	0.7.3	2021	7.7.3
Teiid	–	✓	–	✓ <sub>nc</sub>	–	–	✓	✓	2009	6.0.0	2020	16.0.0
TIBCO Data Virtualization	–	–	–	✓ <sub>nc</sub>	✓ <sub>am</sub>	–	✓	–	2007	7.0.5	2021	8.4.0
Trino	–	✓	–	✓ <sub>nc</sub>	–	–	✓	✓	–	0.54	2021	364
Virtuoso	✓	–	–	✓ <sub>nc</sub>	✓ <sub>am</sub>	–	✓	–	–	–	2020	8.3
Number	5	31	7	49	12	6	26	30	–	–	–	–

The following Table 6 lists the specific sources supported by each investigated data federation system, obtained from available systems' documentation and publications. Sources are classified on a *local, per-system* basis, along the source types defined in Section 6 of the original paper, with additional source information — such as the specific kind(s) of relational, graph-based or aggregate-oriented system — reported next to the source name via subscript letters (see table caption for legend). We remark the following:

- Some sources correspond to data access interfaces that can be configured to connect additional systems beyond the ones explicitly listed in the table. In particular, companies such as CData<sup>1</sup> and Progress<sup>2</sup> commercialize *connectors* for the relational SQL-based JDBC, ODBC, ADO.NET and OLE DB interfaces that can be used to access a myriad of heterogeneous data sources, possibly different from the ones listed in Table 6 (e.g., GraphQL sources via specific connectors<sup>3</sup>) and possibly using a different data model that is transparently adapted to the relational one by the connector (e.g., via flattening of nested data). In Table 6, besides the supported data access interfaces, we explicitly list only the sources that are directly and natively supported by a system without relying on such third party connectors / adapters.
- Structured files are distinguished from other source types with the same data model (e.g., relational sources for CSV files, aggregate-oriented — specifically, document-based — for JSON files) by virtue of direct access to raw file contents by the data federation system. In some cases, however, access to stored structured files may require metadata services external to the filesystem (e.g., Hive Metadata Store) for locating and interpreting file contents, or may leverage processing services (e.g., from Hadoop) co-located with the nodes storing the file in a distributed filesystem (e.g., HDFS), for instance to *push down* data access operations and computations (e.g., filtering, sorting) close to where raw file data reside, this way reducing communication costs.
- Some of the data federation systems investigated in this survey are also listed as supported sources (marked with \* subscript) of other systems in Table 6, reflecting the fact that the virtual data sources obtained through data federation can be used themselves in downstream federations. As a limit case (e.g., AllegroGraph), a system may list only itself as a supported data source, which occurs when the system offers both storage and data federation capabilities, and the latter are restricted to instances of the same system.
- Test sources (e.g., emulating /dev/null) and system-specific connectors used to access configuration, performance or log data of the system itself are omitted in Table 6, for simplicity.

Table 6

Supported data sources. Academic systems in *italics*. Additional source information in subscript position: \* = investigated system; a = specialized web API; r = RDF triple store; g = property graph store; k = key-value store; w = wide-column store; d = document store; s = search engine; h = hardware + software appliance; m = MDX (MultiDimensional eXpressions) support. SPARQLp denotes the SPARQL protocol

System	Relational	Graph-based	Aggregate-oriented	Structured Files	Web Service Paradigms	Other
AllegroGraph		Allegro-Graph <sub>r*</sub>				
Amazon Athena	Amazon Redshift, MySQL, PostgreSQL, Vertica	Amazon Neptune <sub>rg*</sub>	Amazon DocumentDB <sub>d</sub> , Amazon DynamoDB <sub>d</sub> , Amazon OpenSearch <sub>s</sub> , HBase <sub>w</sub> , Redis <sub>k</sub>	Common Log Format, CSV, JSON, ORC, Parquet		Amazon AWS System Manager Inventory <sub>a</sub> , Amazon CloudWatch <sub>a</sub> , Amazon Timestream
Amazon Neptune		SPARQLp				
AnzoGraph DB	Derby, Google BigQuery, Hive, HSQLDB, IBM DB2, Impala, JDBC, MariaDB, MS SQL Server <sub>*</sub> , MySQL, PostgreSQL, SAP ASE	SPARQLp		CSV, JSON, Parquet, SAS7BDAT, SAS XPT, XML	HTTP / REST	

<sup>1</sup><https://www.cdata.com/drivers/>

<sup>2</sup><https://www.progress.com/connectors>

<sup>3</sup><https://www.cdata.com/drivers/graphql/>

System	Relational	Graph-based	Aggregate-oriented	Structured Files	Web Service Paradigms	Other
Apache Drill	Derby, Druid, Hive, H2, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL		Cassandra <sub>w</sub> , Elasticsearch <sub>g</sub> , HBase <sub>w</sub> , MapR-DB <sub>w</sub> , MongoDB <sub>d</sub> , Splunk <sub>s</sub>	Avro, Common Log Format, CSV, Excel, JSON, Parquet, SequenceFile, XML	HTTP / REST	Kafka, OpenTSDB
Apache Jena		Jena API, SPARQLp				
Apache Spark	Hive, JDBC			any file (content field + metadata), Avro, CSV, JSON, ORC, Parquet		
BigDAWG	PostgreSQL		Accumulo <sub>w</sub>			SciDB
Blazegraph		SPARQLp				
CloudMdsQL	Derby	Sparksee <sub>g</sub>	MongoDB <sub>d</sub>			
Comunica		SPARQLp, TPF		RDF		
CostFed		SPARQLp				
DARQ		SPARQLp				
Data Virtuality	Amazon Redshift, ClickHouse, Data Virtuality*, Derby, Exasol, Google BigQuery, Greenplum, Hive, HSQLDB, H2, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , Ingres, JDBC, MDX <sub>m</sub> , MetaMatrix*, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL, SAP ASE, SingleStore, Snowflake, Teradata	Neo4j <sub>g</sub> *	MongoDB <sub>d</sub> , Redis <sub>k</sub>	CSV, Excel, JSON, XML	HTTP / REST	DHL Track & Trace <sub>a</sub> , Google Ads <sub>a</sub> , Google Analytics <sub>a</sub> , InterSystems Caché, Kdb+, LDAP, ModeShape, Salesforce <sub>a</sub>
Denodo	Amazon Athena*, Amazon Redshift, Databricks, Denodo*, Derby, Google BigQuery, Greenplum, Hive, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , Impala, JDBC, MS Analysis Service <sub>m</sub> , MS Azure SQL Database, MS SQL Server*, MS Azure Synapse Analytics, Mondrian <sub>m</sub> , MySQL, Oracle DB*, Oracle Essbase <sub>m</sub> , Oracle TimesTen, PostgreSQL, Presto*, SAP ASE, SAP Business Warehouse <sub>m</sub> , SAP HANA*, Snowflake, Teradata, Trino*, Vertica, Yellowbrick <sub>h</sub>		Amazon OpenSearch <sub>s</sub> , Cassandra <sub>w</sub> , Elasticsearch <sub>g</sub> , MongoDB <sub>d</sub>	CSV, Excel, JSON, XML	SOAP / WSDL	ITPilot (website wrapper generator), LDAP, Salesforce <sub>a</sub> , SAP Business <sub>a</sub>
Dremio	Amazon Redshift, Hive, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL, Teradata		Amazon OpenSearch <sub>s</sub> , Elasticsearch <sub>g</sub> , HBase <sub>w</sub> , MongoDB <sub>d</sub>	CSV, Excel, JSON, Parquet		
FEDRA		SPARQLp				
FedX (RDF4J)		RDF4J API, SPARQLp				
GraphDB	IBM DB2, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL	GraphDB <sub>r</sub> , SPARQLp				
HiBISCuS		SPARQLp				
IBM Cloud Pak for Data	Amazon Redshift, Derby, Google BigQuery, Greenplum, Hive, IBM DB2, IBM Db2 Big SQL*, IBM Db2 Warehouse, IBM DVM, IBM Informix, IBM Netezza <sub>h</sub> , Impala, MariaDB, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL, SAP ASE, SAP HANA*, Snowflake, Teradata		MongoDB <sub>d</sub>	CSV, Excel	OData	IBM Db2 Event Store, Salesforce <sub>a</sub> , SAP Gateway OData <sub>a</sub>
IBM Db2 Big SQL	Amazon Athena*, Amazon Redshift, Derby, Google BigQuery, Greenplum, Hive, IBM DB2, IBM Db2 Big SQL*, IBM Db2 Warehouse, IBM DVM, IBM Informix, IBM Integrated Analytics System <sub>h</sub> , IBM Netezza <sub>h</sub> , IBM PureData <sub>h</sub> , Impala, MariaDB, MS Azure SQL Database, MS SQL Server*, MySQL, Oracle DB*, PostgreSQL, SAP ASE, SAP HANA*, Teradata		Amazon OpenSearch <sub>s</sub> , CouchDB <sub>d</sub> , MongoDB <sub>d</sub>	Parquet		IBM MQ, Salesforce <sub>a</sub>
IBM InfoSphere Federation Server	IBM DB2, IBM Informix, MS SQL Server*, Oracle DB*, SAP ASE, Datacom/DB, Teradata, IBM Netezza <sub>h</sub>			Excel, XML	SOAP / WSDL	BioRS, IBM MQ, IDMS, IMS



System	Relational	Graph-based	Aggregate-oriented	Structured Files	Web Service Paradigms	Other
JBoss Data Virtualization	Actian Vector, Amazon Redshift, Exasol, Greenplum, Hive, Hive, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , Impala, Ingres, JBoss Data Virtualization <sub>*</sub> , MariaDB, MetaMatrix <sub>*</sub> , MS Access, MS SQL Server <sub>*</sub> , Mondrian <sub>m</sub> , MySQL, Oracle DB <sub>*</sub> , PostgreSQL, Presto <sub>*</sub> , SAP ASE, SAP HANA <sub>*</sub> , SAP IQ, Teradata, Vertica		Accumulo <sub>w</sub> , Amazon OpenSearch <sub>s</sub> , Cassandra <sub>w</sub> , Couchbase <sub>d</sub> , HBase <sub>w</sub> , MongoDB <sub>d</sub> , Red Hat Data Grid <sub>k</sub> , Solr <sub>s</sub>	CSV, Excel, XML	HTTP / REST, OData, SOAP / WSDL	Google Sheets <sub>a</sub> , LDAP, ModeShape, OSIssoft PI, Red Hat Directory Server, Salesforce <sub>a</sub> , SAP Gateway OData <sub>a</sub>
Metaphactory	JDBC	Amazon Neptune <sub>rg*</sub> , GraphDB <sub>r</sub> , SPARQLp, Stardo <sub>gr*</sub> , Virtuoso <sub>r*</sub>	Elasticsearch <sub>s</sub>		HTTP / REST	
Myria		SPARQLp	Amazon OpenSearch <sub>s</sub>	CSV		SciDB
Neo4j (Fabric)		Neo4j <sub>g*</sub>				
Obi-Wan	PostgreSQL	Jena TDB <sub>r</sub>	MongoDB <sub>d</sub> , Redis <sub>k</sub>			
Odyssey		SPARQLp				
Ontario	MySQL	Neo4j <sub>g*</sub> , SPARQLp	MongoDB <sub>d</sub>	CSV, XML		
Onto-KIT				CSV, ENVI, JSON		
Oracle Big Data SQL	Hive		HBase <sub>w</sub> , Oracle NoSQL <sub>k</sub>	Avro, CSV, JSON, ORC, Parquet, XML		Kafka
Oracle DB (Spatial & Graph)	Oracle DB <sub>*</sub>	SPARQLp				
PolyWeb	MySQL	SPARQLp		CSV		
Presto	Amazon Redshift, Druid, Google BigQuery, Hive, Iceberg, Kudu, MS SQL Server <sub>*</sub> , MySQL, Oracle DB <sub>*</sub> , Pinot, PostgreSQL		Accumulo <sub>w</sub> , Cassandra <sub>w</sub> , Elasticsearch <sub>s</sub> , MongoDB <sub>d</sub> , Redis <sub>k</sub>			Kafka, Prometheus
Querona Data Virtualization	Actian Matrix, Actian Vector, ADO.NET, Alibaba AnalyticDB for MySQL, Alibaba Data Lake Analytics, Amazon Athena <sub>*</sub> , Amazon Aurora, Amazon Redshift, ClickHouse, Databricks, dBASE, Denodo <sub>*</sub> , Drill <sub>*</sub> , Exasol, Google BigQuery, IBM DB2, JDBC, MariaDB, MS Access, MS SQL Server <sub>*</sub> , MS Azure Synapse Analytics, MySQL, ODBC, OLE DB, Oracle DB <sub>*</sub> , PostgreSQL, SAP HANA <sub>*</sub> , SAS Scalable Performance Data Server, Spark <sub>*</sub> , Teradata, Teradata Aster, Vertica		Amazon OpenSearch <sub>s</sub> , DataStax <sub>w</sub>	CSV, Excel, MSG/EML (email), PDF (metadata)		Kafka
RDFLib		SPARQLp				
SAFE		SPARQLp				
SAGE		SPARQLp				
SAP HANA	Amazon Athena <sub>*</sub> , Google BigQuery, IBM DB2, IBM Netezza <sub>h</sub> , MS SQL Server <sub>*</sub> , Oracle DB <sub>*</sub> , SAP ASE, SAP HANA <sub>*</sub> , SAP IQ, SAP MaxDB, Teradata					SAP HANA Streaming Analytics
SAS Federation Server	dBASE, Greenplum, Hive, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , Impala, MS Access, MS SQL Server <sub>*</sub> , MySQL, Oracle DB <sub>*</sub> , Paradox, PostgreSQL, Progress OpenEdge RDBMS, SAP ASE, SAP HANA <sub>*</sub> , SAS Federation Server <sub>*</sub> , SAS Scalable Performance Data Server, Teradata					Btrieve, Salesforce <sub>a</sub> , SAP RFC <sub>a</sub>
SemaGrow		SPARQLp				
SPLendid		SPARQLp				
SQL Server (PolyBase)	MS SQL Server <sub>*</sub> , ODBC, Oracle DB <sub>*</sub> , Teradata		MongoDB <sub>d</sub>	CSV, JSON, ORC, Parquet, RCFile		
Squerall	MySQL		Cassandra <sub>w</sub> , Couchbase <sub>d</sub> , Elasticsearch <sub>s</sub> , MongoDB <sub>d</sub>	CSV, Parquet		

System	Relational	Graph-based	Aggregate-oriented	Structured Files	Web Service Paradigms	Other
Starburst	Amazon Redshift, ClickHouse, Druid, Google BigQuery, Greenplum, Hive, IBM DB2, IBM Netezza <sub>h</sub> , Iceberg, JDBC, Kudu, MS SQL Server <sub>s</sub> , MS Azure Synapse Analytics, MySQL, Oracle DB <sub>s</sub> , Pinot, PostgreSQL, SAP HANA <sub>s</sub> , SingleStore, Snowflake, Starburst <sub>s</sub> , Teradata, Vertica		Accumulo <sub>w</sub> , Amazon DynamoDB <sub>d</sub> , Cassandra <sub>w</sub> , Elasticsearch <sub>s</sub> , HBase <sub>w</sub> , MongoDB <sub>d</sub> , Redis <sub>k</sub> , Splunk <sub>s</sub>	Avro, CSV, JSON, ORC, Parquet, RCFile, SequenceFile		Amazon Kinesis, Google Sheets <sub>a</sub> , Kafka, Prometheus, Salesforce <sub>a</sub>
Stardog	Amazon Athena <sub>s</sub> , Amazon Aurora, Amazon Redshift, Derby, Exasol, Google BigQuery, Hive, H2, IBM DB2, Impala, MariaDB, MS SQL Server <sub>s</sub> , MySQL, Oracle DB <sub>s</sub> , PostgreSQL, SAP ASE, SAP HANA <sub>s</sub> , Snowflake, Teradata	SPARQLp, Stardog <sub>r</sub>	Amazon OpenSearch <sub>s</sub> , Cassandra <sub>w</sub> , DataStax <sub>w</sub> , Elasticsearch <sub>s</sub> , MS Azure Cosmos DB <sub>d</sub> , MongoDB <sub>d</sub> , Splunk <sub>s</sub>	CSV, JSON		Google Sheets <sub>a</sub> , Jira <sub>a</sub> , LDAP, Salesforce <sub>a</sub>
Teiid	Actian Vector, Amazon Athena <sub>s</sub> , Amazon Redshift, Derby, Exasol, Greenplum, Hive, HSQLDB, H2, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , Impala, Ingres, JDBC, MariaDB, MDX <sub>m</sub> , MetaMatrix <sub>s</sub> , MS Access, MS SQL Server <sub>s</sub> , Mondrian <sub>m</sub> , MySQL, Oracle DB <sub>s</sub> , PostgreSQL, Presto <sub>s</sub> , SAP ASE, SAP HANA <sub>s</sub> , SAP IQ, Teiid <sub>s</sub> , Teradata, Vertica		Accumulo <sub>w</sub> , Amazon OpenSearch <sub>s</sub> , Amazon SimpleDB <sub>kw</sub> , Cassandra <sub>w</sub> , Couchbase <sub>d</sub> , HBase <sub>w</sub> , Infinispan <sub>k</sub> , MongoDB <sub>d</sub> , Solr <sub>s</sub>	CSV, Excel, JSON, XML	HTTP / REST, OData, OpenAPI, SOAP / WSDL	Google Sheets <sub>a</sub> , InterSystems Caché, JPA/JPQL sources, LDAP, MS Active Directory, ModeShape, OSIssoft PI, Red Hat Directory Server, Salesforce <sub>a</sub> , SAP Gateway OData <sub>a</sub>
TIBCO Data Virtualization	Amazon Redshift, Drill <sub>s</sub> , Google BigQuery, Greenplum, Hive, HP Neoview <sub>h</sub> , HSQLDB, IBM DB2, IBM Informix, IBM Netezza <sub>h</sub> , MS Access, MS SQL Server <sub>s</sub> , MySQL, Oracle DB <sub>s</sub> , PostgreSQL, SAP ASE, SAP Business Warehouse <sub>m</sub> , SAP Business Warehouse <sub>m</sub> , SAP HANA <sub>s</sub> , Snowflake, Teradata, Tibco ComputeDB, TibcoDataVirtualization <sub>s</sub> , Vertica		Amazon DynamoDB <sub>d</sub> , Amazon OpenSearch <sub>s</sub> , Amazon SimpleDB <sub>kw</sub> , Cassandra <sub>w</sub> , Couchbase <sub>d</sub> , Elasticsearch <sub>s</sub> , HBase <sub>w</sub> , MarkLogic <sub>d</sub> , MS Azure Cosmos DB <sub>d</sub> , MongoDB <sub>d</sub> , Splunk <sub>s</sub>	CSV, Excel, JSON, XML	HTTP / REST, OData, SOAP / WSDL	Eloqua <sub>a</sub> , Facebook <sub>a</sub> , Google Ads <sub>a</sub> , Google Analytics <sub>a</sub> , Google Calendar <sub>a</sub> , Google Contacts <sub>a</sub> , Google Sheets <sub>a</sub> , HubSpot <sub>a</sub> , IMAp, Marketo <sub>a</sub> , MS Sharepoint <sub>a</sub> , MS Sharepoint Excel Services <sub>a</sub> , NetSuite <sub>a</sub> , RSS, Salesforce <sub>a</sub> , SAP RFC <sub>a</sub> , Twitter <sub>a</sub>
Trino	Amazon Redshift, ClickHouse, Druid, Google BigQuery, Hive, Iceberg, Kudu, MS SQL Server <sub>s</sub> , MySQL, Oracle DB <sub>s</sub> , Pinot, PostgreSQL, SingleStore		Accumulo <sub>w</sub> , Cassandra <sub>w</sub> , Elasticsearch <sub>s</sub> , HBase <sub>w</sub> , MongoDB <sub>d</sub> , Redis <sub>k</sub>			Amazon Kinesis, Google Sheets <sub>a</sub> , Kafka, Prometheus
Virtuoso	Firebird, IBM DB2, IBM Informix, Ingres, MS SQL Server <sub>s</sub> , MySQL, Oracle DB <sub>s</sub> , PostgreSQL, Progress OpenEdge RDBMS, SAP ASE	SPARQLp				

## References

- [1] AllegroGraph, Accessed 18 November 2021. <https://allegrograph.com/>.
- [2] Amazon Athena, Accessed 18 November 2021. <https://docs.aws.amazon.com/athena/latest/ug/work-with-data-stores.html>.
- [3] Presto, Accessed 18 November 2021. <https://prestodb.io/>.
- [4] Amazon Neptune, Accessed 18 November 2021. <https://aws.amazon.com/neptune/>.
- [5] Anzograph, Accessed 17 November 2021. <https://www.cambridgesemantics.com/anzograph/>.
- [6] Apache Drill, Accessed 18 November 2021. <https://drill.apache.org/>.
- [7] M. Hausenblas and J. Nadeau, Apache Drill: Interactive Ad-Hoc Analysis at Scale, *Big Data* **1**(2) (2013), 100–104. doi:10.1089/big.2013.0011.
- [8] S. Melnik, A. Gubarev, J.J. Long, G. Romer, S. Shivakumar, M. Tolton and T. Vassilakis, Dremel: interactive analysis of web-scale datasets, *Communications of the ACM* **54**(6) (2011), 114–123. doi:10.1145/1953122.1953148.
- [9] Jena, Accessed 18 November 2021. <https://jena.apache.org/documentation/query/>.
- [10] Spark SQL, Accessed 18 November 2021. <https://spark.apache.org/sql/>.
- [11] M. Armbrust, R.S. Xin, C. Lian, Y. Huai, D. Liu, J.K. Bradley, X. Meng, T. Kaftan, M.J. Franklin, A. Ghodsi and M. Zaharia, Spark SQL: Relational Data Processing in Spark, in: *Proc. of ACM SIGMOD Int. Conf. on Management of Data (SIGMOD)*, ACM, 2015, pp. 1383–1394. doi:10.1145/2723372.2742797.
- [12] J. Duggan, A.J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson and S.B. Zdonik, The BigDAWG Polystore System, *SIGMOD Record* **44**(2) (2015), 11–16. doi:10.1145/2814710.2814713.
- [13] V. Gadepally, K. O'Brien, A. Dziedzic, A.J. Elmore, J. Kepner, S. Madden, T. Mattson, J. Rogers, Z. She and M. Stonebraker, BigDAWG version 0.1, in: *Proc. of IEEE High Performance Extreme Computing Conf. (HPEC)*, IEEE, 2017, pp. 1–7. doi:10.1109/HPEC.2017.8091077.
- [14] Blazegraph, Accessed 6 December 2021. <https://blazegraph.com/>.

- [15] B. Kolev, C. Bondiombouy, P. Valduriez, R. Jiménez-Peris, R. Pau and J. Pereira, The CloudMdsQL Multistore System, in: *Proc. of ACM SIGMOD Int. Conf. on Management of Data (SIGMOD)*, ACM, 2016, pp. 2113–2116. doi:10.1145/2882903.2899400.
- [16] B. Kolev, P. Valduriez, C. Bondiombouy, R. Jiménez-Peris, R. Pau and J. Pereira, CloudMdsQL: querying heterogeneous cloud data stores with a common language, *Distributed Parallel Databases* **34**(4) (2016), 463–503. doi:10.1007/s10619-015-7185-y.
- [17] R. Taelman, J.V. Herwegen, M.V. Sande and R. Verborgh, Comunica: A Modular SPARQL Query Engine for the Web, in: *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part II*, Lecture Notes in Computer Science, Vol. 11137, Springer, 2018, pp. 239–255. doi:10.1007/978-3-030-00668-6\_15.
- [18] M. Saleem, A. Potocki, T. Soru, O. Hartig and A.-C. Ngonga Ngomo, CostFed: Cost-Based Query Optimization for SPARQL Endpoint Federation, in: *Proc. of Int. Conf. on Semantic Systems (SEMANTICS)*, Procedia Computer Science, Vol. 137, Elsevier, 2018, pp. 163–174. doi:10.1016/j.procs.2018.09.016.
- [19] B. Quilitz and U. Leser, Querying Distributed RDF Data Sources with SPARQL, in: *Proc. of European Semantic Web Conf. (ESWC)*, LNCS, Vol. 5021, Springer, 2008, pp. 524–538. doi:10.1007/978-3-540-68234-9\_39.
- [20] Data Virtuality, Accessed 17 November 2021. <https://datavirtuality.com/>.
- [21] Denodo, Accessed 17 November 2021. <https://www.denodo.com/en>.
- [22] Dremio, Accessed 17 November 2021. <https://www.dremio.com/>.
- [23] G. Montoya, H. Skaf-Molli, P. Molli and M.-E. Vidal, Federated SPARQL Queries Processing with Replicated Fragments, in: *Proc. of Int. Semantic Web Conf. (ISWC)*, LNCS, Vol. 9366, Springer, 2015, pp. 36–51. doi:10.1007/978-3-319-25007-6\_3.
- [24] A. Schwarte, P. Haase, K. Hose, R. Schenkel and M. Schmidt, FedX: A Federation Layer for Distributed Query Processing on Linked Open Data, in: *Proc. of Extended Semantic Web Conference (ESWC)*, LNCS, Vol. 6644, Springer, 2011, pp. 481–486. doi:10.1007/978-3-642-21064-8\_39.
- [25] A. Schwarte, P. Haase, K. Hose, R. Schenkel and M. Schmidt, FedX: Optimization Techniques for Federated Query Processing on Linked Data, in: *Proc. of Int. Semantic Web Conf. (ISWC)*, LNCS, Vol. 7031, Springer, 2011, pp. 601–616. doi:10.1007/978-3-642-25073-6\_38.
- [26] GraphDB, Accessed 17 November 2021. <https://graphdb.ontotext.com/>.
- [27] M. Saleem and A.-C. Ngonga Ngomo, HiBISCuS: Hypergraph-Based Source Selection for SPARQL Endpoint Federation, in: *Proc. of European Semantic Web Conf. (ESWC)*, LNCS, Vol. 8465, Springer, 2014, pp. 176–191. doi:10.1007/978-3-319-07443-6\_13.
- [28] IBM Cloud Pak for Data, Accessed 17 November 2021. <https://www.ibm.com/products/cloud-pak-for-data>.
- [29] IBM Db2 Big SQL, Accessed 18 November 2021. <https://www.ibm.com/products/db2-big-sql>.
- [30] IBM InfoSphere Federation Server, Accessed 18 November 2021. <https://www.ibm.com/docs/en/iis/11.7?topic=components-infosphere-federation-server>.
- [31] JBoss Data Virtualization, Accessed 17 November 2021. <https://developers.redhat.com/products/datavirt/overview>.
- [32] Metaphactory, Accessed 18 November 2021. <https://metaphacts.com/product>.
- [33] P. Haase, D.M. Herzig, A. Kozlov, A. Nikolov and J. Trame, metaphactory: A platform for knowledge graph management, *Semantic Web* **10**(6) (2019), 1109–1125. doi:10.3233/SW-190360.
- [34] J. Wang, T. Baker, M. Balazinska, D. Halperin, B. Haynes, B. Howe, D. Hutchison, S. Jain, R. Maas, P. Mehta, D. Moritz, B. Myers, J. Ortiz, D. Suci, A. Whitaker and S. Xu, The Myria Big Data Management and Analytics System and Cloud Services, in: *Proc. of Biennial Conf. on Innovative Data Systems Research (CIDR)*, www.cidrdb.org, 2017.
- [35] Neo4j, Accessed 17 November 2021. <https://neo4j.com/>.
- [36] N. Francis, A. Green, P. Guagliardo, L. Libkin, T. Lindaaker, V. Marsault, S. Plantikow, M. Rydberg, P. Selmer and A. Taylor, Cypher: An Evolving Query Language for Property Graphs, in: *Proc. of ACM SIGMOD Int. Conf. on Management of Data (SIGMOD)*, ACM, 2018, pp. 1433–1445. doi:10.1145/3183713.3190657.
- [37] M. Buron, F. Goasdoué, I. Manolescu and M.-L. Mugnier, Ontology-Based RDF Integration of Heterogeneous Data, in: *Proc. of 23rd Int. Conf. on Extending Database Technology (EDBT)*, OpenProceedings.org, 2020, pp. 299–310. doi:10.5441/002/edbt.2020.27.
- [38] M. Buron, F. Goasdoué, I. Manolescu and M.-L. Mugnier, Obi-Wan: Ontology-Based RDF Integration of Heterogeneous Data, *Proc. of VLDB Endowment* **13**(12) (2020), 2933–2936. doi:10.14778/3415478.3415512.
- [39] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati and M. Zakharyashev, Ontology-Based Data Access: A Survey, in: *Proc. of 27th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, ijcai.org, 2018, pp. 5511–5519. doi:10.24963/ijcai.2018/777.
- [40] R. Bonaque, T.D. Cao, B. Cautis, F. Goasdoué, J. Letelier, I. Manolescu, O. Mendoza, S. Ribeiro, X. Tannier and M. Thomazo, Mixed-instance querying: a lightweight integration architecture for data journalism, *Proc. of VLDB Endowment* **9**(13) (2016), 1513–1516. doi:10.14778/3007263.3007297.
- [41] G. Montoya, H. Skaf-Molli and K. Hose, The Odyssey Approach for Optimizing Federated SPARQL Queries, in: *Proc. of Int. Semantic Web Conf. (ISWC)*, LNCS, Vol. 10587, Springer, 2017, pp. 471–489. doi:10.1007/978-3-319-68288-4\_28.
- [42] K.M. Endris, P.D. Rohde, M.-E. Vidal and S. Auer, Ontario: Federated Query Processing Against a Semantic Data Lake, in: *Proc. of Int. Conf. on Database and Expert Systems Applications (DEXA)*, LNCS, Vol. 11706, Springer, 2019, pp. 379–395. doi:10.1007/978-3-030-27615-7\_29.
- [43] K.M. Endris, M. Galkin, I. Lytra, M.N. Mami, M.-E. Vidal and S. Auer, Querying Interlinked Data by Bridging RDF Molecule Templates, *Trans. Large Scale Data Knowledge Centered Systems* **39** (2018), 1–42. doi:10.1007/978-3-662-58415-6\_1.
- [44] M. Masmoudi, S.B.A.B. Lamine, H.B. Zghal, B. Archimède and M.-H. Karray, Knowledge hypergraph-based approach for data integration and querying: Application to Earth Observation, *Future Generation Computer Systems* **115** (2021), 720–740. doi:10.1016/j.future.2020.09.029.
- [45] Oracle Big Data SQL, Accessed 18 November 2021. <https://www.oracle.com/database/technologies/datawarehouse-bigdata/bigdata-sql.html>.

- [46] Oracle Spatial and Graph, Accessed 16 November 2021. <https://www.oracle.com/database/technologies/spatialandgraph.html>.
- [47] L. Jayapalan, Oracle Spatial and Graph RDF Knowledge Developer's Guide, Technical Report, Oracle, 2021. <https://docs.oracle.com/en/database/oracle/oracle-database/19/rdfm/spatial-and-graph-rdf-knowledge-graph-developers-guide.pdf>.
- [48] Y. Khan, A. Zimmermann, A. Jha, V. Gadepally, M. d'Aquin and R. Sahay, One Size Does Not Fit All: Querying Web Polystores, *IEEE Access* **7** (2019), 9598–9617. doi:10.1109/ACCESS.2018.2888601.
- [49] Y. Khan, A. Zimmermann, A. Jha, D. Rebholz-Schuhmann and R. Sahay, Querying web polystores, in: *Proc. of IEEE Int. Conf. on Big Data (IEEE BigData)*, IEEE Computer Society, 2017, pp. 3190–3195. doi:10.1109/BigData.2017.8258299.
- [50] R. Sethi, M. Traverso, D. Sundstrom, D. Phillips, W. Xie, Y. Sun, N. Yegitbasi, H. Jin, E. Hwang, N. Shingte and C. Berner, Presto: SQL on Everything, in: *Proc. of 35th Int. Conf. on Data Engineering (ICDE)*, IEEE, 2019, pp. 1802–1813. doi:10.1109/ICDE.2019.00196.
- [51] Querona Data Virtualization, Accessed 17 November 2021. <https://www.querona.io/>.
- [52] RDFLib, Accessed 26 July 2022. <https://rdflib.readthedocs.io/en/stable/>.
- [53] Y. Khan, M. Saleem, M. Mehdi, A. Hogan, Q. Mehmood, D. Rebholz-Schuhmann and R. Sahay, SAFE: SPARQL Federation over RDF Data Cubes with Access Control, *J. Biomedical Semantics* **8**(1) (2017), 5:1–5:22. doi:10.1186/s13326-017-0112-6.
- [54] T. Minier, H. Skaf-Molli and P. Molli, SaGe: Web Preemption for Public SPARQL Query Services, in: *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, L. Liu, R.W. White, A. Mantrach, F. Silvestri, J.J. McAuley, R. Baeza-Yates and L. Zia, eds, ACM, 2019, pp. 1268–1278. doi:10.1145/3308558.3313652.
- [55] SAP HANA (Smart Data Access), Accessed 18 November 2021. [https://help.sap.com/docs/SAP\\_HANA\\_PLATFORM/6b94445c94ae495c83a19646e7c3fd56/a07c7ff25997460bbcb73099fb59007d.html?locale=en-US&version=2.0.05](https://help.sap.com/docs/SAP_HANA_PLATFORM/6b94445c94ae495c83a19646e7c3fd56/a07c7ff25997460bbcb73099fb59007d.html?locale=en-US&version=2.0.05).
- [56] SAS Federation Server, Accessed 18 November 2021. <https://support.sas.com/en/software/federation-server-support.html>.
- [57] A. Charalambidis, A. Troumpoukis and S. Konstantopoulos, SemaGrow: optimizing federated SPARQL queries, in: *Proc. of 11th Int. Conf. on Semantic Systems (SEMANTICS)*, ACM, 2015, pp. 121–128. doi:10.1145/2814864.2814886.
- [58] O. Görlitz and S. Staab, SPLENDID: SPARQL Endpoint Federation Exploiting VOID Descriptions, in: *Proc. of 2nd Int. Workshop on Consuming Linked Data (COLD)*, CEUR Workshop Proceedings, Vol. 782, CEUR-WS.org, 2011.
- [59] K. Alexander, R. Cyganiak, M. Hausenblas and J. Zhao, Describing Linked Datasets, in: *Proc. of Int. Workshop on Linked Data on the Web (LDOW)*, CEUR Workshop Proceedings, Vol. 538, CEUR-WS.org, 2009.
- [60] SQL Server (PolyBase), Accessed 18 November 2021. <https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver15>.
- [61] M.N. Mami, D. Graux, S. Scerri, H. Jabeen, S. Auer and J. Lehmann, Squerall: Virtual Ontology-Based Access to Heterogeneous and Large Data Sources, in: *Proc. of Int. Semantic Web Conf. (ISWC)*, LNCS, Vol. 11779, Springer, 2019, pp. 229–245. doi:10.1007/978-3-030-30796-7\_15.
- [62] Starburst, Accessed 18 December 2021. <https://www.starburst.io/>.
- [63] Stardog, Accessed 17 November 2021. <https://www.stardog.com/>.
- [64] Teiid, Accessed 16 November 2021. <https://teiid.io/>.
- [65] TIBCO Data Virtualization, Accessed 17 November 2021. <https://www.tibco.com/products/data-virtualization>.
- [66] Trino, Accessed 18 November 2021. <https://trino.io/>.
- [67] Virtuoso, Accessed 17 November 2021. <https://virtuoso.openlinksw.com/>.
- [68] O. Erling, Virtuoso, a Hybrid RDBMS/Graph Column Store, *IEEE Data Engineering Bull.* **35**(1) (2012), 3–8.
- [69] O. Erling and I. Mikhailov, RDF Support in the Virtuoso DBMS, in: *Proc. of Conf. on Social Semantic Web (CSSW)*, LNI, Vol. P-113, GI, 2007, pp. 59–68.