Renan Souza

contact@renansouza.org • RenanSouza.org • renansouza1
ID: x9t36ewAAAAJ • renan-souza • Updated on April 18, 2022

Bio

Renan Francisco Santos Souza holds a Ph.D., M.Sc., and B.Sc (2009–2019) in Computer Science from the Federal University of Rio de Janeiro (UFRJ). Since 2015, he has been at IBM Research, where he is a Research Scientist in the Intelligent Cloud Technologies group. He has been working both as a software engineer and a researcher on several projects since 2010. During his B.Sc., he spent a year at Missouri State University and was an intern at Stanford University in the SLAC National Laboratory. During his Ph.D., he was a visiting researcher at Inria, France. He received the best M.Sc. thesis and an honored mention for the best Ph.D. thesis awards from SBBD, the main conference on data science in Latin America. He researches large-scale data management techniques to support the evolution of Artificial Intelligence systems in the cloud.

Research Interests

Large-scale Data Science and Data Engineering • Parallel Workflows • Data Provenance • Big Data Analytics • High Performance Computing in Clusters and Clouds • Machine Learning

Education

- **Ph.D. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil Sep 2015 – Dec 2019
  Supervised by Marta Mattoso (COPPE/UFRJ) and Patrick Valduriez (Inria).
  Title: Supporting User Steering in Large-scale Workflows with Provenance Data
- Visiting Ph.D. Student, Inria/Univ. Montpellier, France Jan 2019 – Mar 2019
  Supervised by Patrick Valduriez (Inria).
- **M.Sc. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil Jan 2013 – Jul 2015
  Supervised by Marta Mattoso (COPPE/UFRJ).
  Title: Controlling the Parallel Execution of Workflows Relying on a Distributed Database
- Computer Science exchange student, Missouri State University, U.S. Jun 2011 – Jun 2012
- **B.Sc. in Computer Science**, Federal Univ. of Rio de Janeiro, Brazil Jan 2009 – Dec 2012
  Supervised by Maria Luiza Machado Campos (DCC/UFRJ).
  Title: Linked Open Data Publication Strategies: An Application in Network Performance Data
- **Technical Degree in Information Systems**, Lemos de Castro Jan 2005 – Dec 2007

Experience

- **IBM Research** **Apr 2015 – present**
  Research Scientist, Cloud & AI Data Management **Rio de Janeiro, Brazil**
  As a Research Scientist (since 2021), he leads R&D projects in large-scale data science and data engineering to support Artificial Intelligence systems running on hybrid cloud and cluster environments with highly distributed and heterogeneous applications, data, and users. He develops software and do applied research to solve problems in different industries, such as energy, financial, and cheminformatics.
  As a Research Software Engineer (2015–2021), he participated in several R&D projects with clients in the energy field by developing techniques and systems for large-scale data integration of AI systems running on clusters and clouds. He also led the Cloud DevOps team to develop conversational AI systems.
  As a Software Engineering intern (2015), he designed and implemented big data and machine learning solutions to analyze streaming social data.

- **SLAC National Accelerator Laboratory, Stanford Univ.** **May 2013 – Dec 2014**
  Research Software Engineering intern **Menlo Park, CA**
  Developed a project to build a cloud platform that uses semantic web, big data, and data warehousing techniques to store, retrieve, visualize, and publish structured data about internet performance worldwide, enabling a rich understanding of information about the Internet quality around the world.
- **CAPGov COPPE/UFRJ** **Dec 2011 – Sep 2014**
  Software Engineer **Rio de Janeiro, Brazil**
  As a lead software engineer (2013–2014), he led the development of a system that helped the Brazilian population to have easy access to information about public services provided by the Federal Government. He also participated in the development of a system to publish linked open data of the Brazilian Federal Register ("Diário Oficial da União") on the semantic web using agile methodology, ontology data modeling, and natural language processing.
  As a Software Engineering intern (2011–2013), he participated in several R&D web systems for the Brazilian Federal Government.
- **Federal Univ. of Rio de Janeiro** **Jan 2010 – Jul 2011**
  Software Engineering intern **Rio de Janeiro, Brazil**
  Developed a system to integrate data warehouse environments with structured and unstructured data to enable more intelligent and flexible information reports.
- **Petrobras** **May 2007 – May 2008**
  IT Intern **Rio de Janeiro, Brazil**
  Helped to implement features and provided maintenance for web systems to support Petrobras employees.

Technical Knowledge

- **Languages**: Python, Java, C, C++, Shell scripting, NodeJS, Scala, Lua
- **Relational DBMS**: PostgreSQL/PostGIS, DB2, DashDB, MySQL, MySQL Cluster, MS SQL Server
- **NoSQL DBMS**: MongoDB, AllegroGraph, Jena, Blazegraph, Virtuoso, Sesame, Cloudant, CouchBase, Redis, Impala, Elasticsearch, HBase, Hive, Apache Ignite
- **Heterogeneous Data Management**: Data Integration, Multi-database Queries, Polystores, Foreign Data Wrappers
- **Big Data Frameworks**: Apache Spark: RDD, DataFrames, Streaming, MLib, GraphX, GraphFrames; Hadoop Ecosystem
- **Message Brokers**: Kafka, RabbitMQ
- **Data Science/ML Technologies**: Pandas, Jupyter Notebooks, Numpy, Matplotlib, Tensorflow, Scikit-Learn, Keras, PyTorch
- **Big Data Cluster Deployment**: YARN, Mesos, Standalone deployment
- **Business Intelligence**: MS SQL Server BI developer studio, Pentaho Solutions, Talend;
- **Semantic Web Tools/Languages**: OWL, RDF, SPARQL, Protege
- **Distributed and Concurrent Programming**: MPI, OpenMP, CUDA, Data-centric distributed and parallel programming
- **Cloud and Cluster computing**: VMs, Dockers, Kubernetes, OpenShift, HPC Clusters
- **DevOps**: Containers, Kubernetes, OpenShift, CI/CD Pipelines, GitHub, Travis, Jenkins
- **Web Development**: Python Flask/UWSGI, Java EE, Tomcat/JBoss, Spring Boot

Selected Publications

For complete list, visit: RenanSouza.org/publications

[1] **R. Souza**, V. Silva, A. A. B. Lima, D. Oliveira, P. Valduriez, M. Mattoso, "Distributed in-memory data management for workflow executions," *PeerJ Computer Science*, vol. 7, pp. 1–30, 2021. DOI: 10.7717/peerj-cs.527. [Online]. Available: https://peerj.com/articles/cs-527/.

[2] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *Concurrency*

and Computation: Practice and Experience, vol. e6544, pp. 1–21, 2021. [Online]. Available: https://doi.org/10.1002/cpe.6544.

[3] **R. Souza**, L. Azevedo, R. Thiago, E. Soares, M. Nery, M. Netto, E. V. Brazil, R. Cerqueira, P. Valduriez, M. Mattoso, "Efficient runtime capture of multiworkflow data using provenance," in *IEEE International Conference on e-Science (eScience)*, 2019, pp. 1–10. DOI: 10.1109/eScience.2019.00047. [Online]. Available: https://doi.org/10.1109/eScience.2019.00047.

[4] **R. Souza**, V. Silva, J. J. Camata, A. L. G. A. Coutinho, P. Valduriez, M. Mattoso, "Keeping track of user steering actions in dynamic workflows," *Future Generation Computer Systems*, vol. 99, pp. 624–643, 2019, ISSN: 0167-739X. DOI: 10.1016/j.future.2019.05.011. [Online]. Available: https://doi.org/10.1016/j.future.2019.05.011.

Grants and Awards

- 2nd IBM Patent Plateau (>8 patents submitted to USPTO) 2021
- SBBD Honored Mention for the Best Ph.D. Thesis Award 2021
- 1st IBM Patent Plateau (>4 patents submitted to USPTO) 2020
- SBBD Best M.Sc. Thesis Award 2017
- SBBD Honored Mention on the paper
  Spark Scalability Analysis in a Scientific Workflow 2017
- CAPES M.Sc. Grant 2013 – 2014
- Brazil Science Mobility Grant - Missouri State University 2012 – 2013
- Scientific Initiation Grant - Federal Univ. of Rio de Janeiro 2010

Teaching and Supervisions

Teaching:

- Databases Laboratory, graduate, UFRJ 2017
  Teacher assistant to Prof. Marta Mattoso
- Logics for Computer Science, undergraduate, UFRJ 2012–2013
  Teacher assistant to Prof. Mario Benevides

Supervisions of final dissertations

- Pedro Paiva Miranda, undergraduate, UFRJ, Co-supervision with Prof. Marta Mattoso 2015
  Thesis title: *A Mechanism for Fault Tolerance in Parallel Executions of Workflows supported by a Database*
- Rachel Gonçalves de Castro, undergraduate, UFRJ, Co-supervision with Prof. Marta Mattoso 2015
  Thesis title: *Publication of Workflow Provenance Data in the Semantic Web*

Talks and Events Participation

- **Brazilian Symposium on Databases (SBBD)** in Rio de Janeiro, RJ (virtual) 2021
  - User Steering Support in Large-Scale Workflows, Oral presentation , link
- **Federal Fluminense University (UFF) Computer Science Seminars** in Rio de Janeiro, RJ (virtual) 2021
  - A Knowledge-centric Approach to Support Large-scale AI Systems (in PT), Oral presentation , link
- **SIAM Conference on Computational Science and Engineering** in Forth Worth, TX (virtual) 2021
  - AI4Seismic: An AI-Driven Platform to Accelerate Geological Discoveries, Oral presentation, Highlighted by the SIAM press , link
  - Workflow Provenance in the Lifecycle of Scientific Machine Learning, Oral presentation , link
- **Open Subsurface Data Universe Development Workshop** in

- **International Conference on Management of Data (SIGMOD)** in Portland, OR. (virtual) 2020
- **Brazilian Symposium on Databases (SBBD)** in Rio (virtual) 2020
- **High-Performance Data Science workshop** in Rio (virtual) 2020
- **Seminarios de Engenharia e Ciencia Computacional** in Rio (virtual) 2020
  - Workflow Provenance in the Lifecycle of Scientific Machine Learning, Oral presentation
- **Open Subsurface Data Universe Development Workshop** in Houston, TX 2020
- **IEEE/ACM Supercomputing (SC)** in Denver, CO 2019
  Workflows in Support of Large-scale Science (WORKS)
  - Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering, Oral presentation
- **SciDISC Workshop** in Rio de Janeiro, Brazil 2019
  - Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering, Oral presentation
- **Open Subsurface Data Universe F2F Meeting** in Houston, TX 2019
- **IEEE International Conference on e-Science** in San Diego, CA 2019
  - Efficient Runtime Capture of Multiworkflow Data using Provenance, Oral presentation
- **INRIA Talks** in Montpellier, France 2019
  - Providing Online Data Analytical Support for Humans in the Loop of Computational Science and Engineering Applications, Oral presentation
- **IBM Regional Technical Exchange** in Rio de Janeiro, Brazil 2019
- **Provenance Week** in London, UK 2018
  International Provenance and Annotation Workshop (IPAW)
  - Provenance of Dynamic Adaptations in User-steered Dataflows, Oral presentation
  - Capturing Provenance for Runtime Data Analysis in Computational Science and Engineering Applications, Poster presentation

  Computational Reproducibility Workshop
  - Provenance of Dynamic Adaptations in User-steered Dataflows, Oral presentation
- **International Conference on Very Large Databases (VLDB)** in Rio de Janeiro, Brazil 2018
  Latin American Data Science Workshop
  - Tracking Hyperparameter Tuning in Deep Learning Training, Oral presentation
- **SBC Brazilian Syposium on Databases (SBBD)** in Rio de Janeiro, Brazil 2018
- **SBC Brazilian Syposium on Databases (SBBD)** in Uberlandia, Brazil 2017
  - Spark Scalability Analysis in a Scientific Workflow, Oral presentation
  - Controlling the Parallel Execution of Workflows Relying on a Distributed Database, Oral presentation
- **Federal University of Uberlandia, Brazil** in Uberlandia, Brazil 2017
  - Kubernetes, Invited talk
- **Hacker at the Smart City Cloud Hackathon OpenStack Rio** in Rio de Janeiro, Brazil 2017
- **Computer Science Week at UFRJ** in Rio de Janeiro, Brazil 2017
  - Kubernetes, Oral presentation
- **SBC Brazilian Conference on Artificial Intelligence (BRACIS)** in Recife, Brazil 2017
  - Graph Analytics with Spark, Tutorial
- **IEEE/ACM Supercomputing (SC)** in Salt Lake City, UT 2016
  Workflows in Support of Large-scale Science (WORKS)
  - Online Input Data Reduction in Scientific Workflows, Oral presentation
- **ASE BigData/SocialCom/CyberSecurity** in Stanford University, Menlo Park, CA 2014
  - Linked open data publication strategies: Application in networking performance measurement data, poster presentation

All Publications and Patents

Journal Articles

[J1] R. F. Silva, H. Casanova, K. Chard, ... **R. Souza**, et al. "Workflows community summit: Advancing the state-of-the-art of scientific workflows management systems research and development," 2021, pp. 1–24. [Online]. Available: https://arxiv.org/abs/2106.05177.

[J2] **R. Souza**, V. Silva, A. A. B. Lima, D. Oliveira, P. Valduriez, M. Mattoso, "Distributed in-memory data management for workflow executions," *PeerJ Computer Science*, vol. 7, pp. 1–30, 2021. DOI: 10.7717/peerj-cs.527. [Online]. Available: https://peerj.com/articles/cs-527/.

[J3] **R. Souza**, L. G. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Workflow provenance in the lifecycle of scientific machine learning," *Concurrency and Computation: Practice and Experience*, vol. e6544, pp. 1–21, 2021. [Online]. Available: https://doi.org/10.1002/cpe.6544.

[J4] L. G. Azevedo, **R. Souza**, R. Brandão, V. N. Lourenço, M. Costalonga, M. Machado, M. Moreno, R. Cerqueira, "Adding hyperknowledge-enabled data lineage to a machine learning workflow management system for oil and gas," *First Break*, vol. 38, no. 7, pp. 89–93, 2020. DOI: 10.3997/1365-2397.fb2020055.

[J5] **R. Souza**, V. Silva, J. J. Camata, A. L. G. A. Coutinho, P. Valduriez, M. Mattoso, "Keeping track of user steering actions in dynamic workflows," *Future Generation Computer Systems*, vol. 99, pp. 624–643, 2019, ISSN: 0167-739X. DOI: 10.1016/j.future.2019.05.011. [Online]. Available: https://doi.org/10.1016/j.future.2019.05.011.

[J6] V. Silva, L. Neves, **R. Souza**, A. L. G. A. Coutinho, D. Oliveira, M. Mattoso, "Adding domain data to code profiling tools to debug workflow parallel execution," *Future Generation Computer Systems*, pp. 624–643, 2018, ISSN: 0167-739X. DOI: 10.1016/j.future.2018.05.078.

[J7] M. G. Bayser, P. Cavalin, **R. Souza**, A. Braz, H. Candello, C. Pinhanez, J.-P. Briot, "A hybrid architecture for multi-party conversational systems," *arXiv preprint Computation and Language (cs.CL)*, pp. 1–40, 2017. [Online]. Available: https://arxiv.org/abs/1705.01214.

[J8] **R. Souza**, V. Silva, A. L. G. A. Coutinho, P. Valduriez, M. Mattoso, "Data reduction in scientific workflows using provenance monitoring and user steering," *Future Generation Computer Systems*, vol. 110, pp. 481–501, 2017, ISSN: 0167-739X. DOI: 10.1016/j.future.2017.11.028.

Conference and Workshop Papers

[C1] L. Azevedo, **R. Souza**, E. Soares, R. Thiago, A. Oliveira, M. Moreno, "Supporting polystore queries using provenance in a hyperknowledge graph," in *International Semantic Web Conference (ISWC)*, 2021, pp. 1–4.

[C2] R. L. Cunha, L. V. Real, **R. Souza**, B. Silva, M. A. Netto, "Context-aware execution migration tool for data science jupyter notebooks on hybrid clouds," in *IEEE e-Science*, 2021. DOI: 10.1109/eScience51609.2021.00013.

[C3] E. Soares, **R. Souza**, R. Thiago, M. Machado, L. Azevedo, "A recommender for choosing data systems based on application profiling and benchmarking," in *Simpósio Brasileiro de Banco de Dados (SBBD)*, 2021, pp. 265–270.

[C4] **R. Souza**, "User steering support in large-scale workflows," in *PhD Thesis Contest: Simpósio Brasileiro de Banco de Dados (SBBD)*, 2021.

[C5] L. Azevedo, **R. Souza**, E. Soares, M. Moreno, "Modern federated databases: An overview," in *International Conference on Enterprise Information Systems (ICEIS)*, 2020.

[C6] L. Azevedo, **R. Souza**, R. Thiago, E. Soares, M. Moreno, "Experiencing provlake to manage the data lineage of ai workflows," in *Meeting in Innovation in Information Systems (EISI) in Brazilian Symposium in Information Systems (SBSI)*, 2020.

[C7] R. Brandão, V. Lourenço, M. Machado, L. Azevedo, M. Cardoso, **R. Souza**, G. Lima, R. Cerqueira, M. Moreno, "A knowledge-based approach for structuring cyclic workflows," in *International Semantic Web Conference (ISWC)*, 2020.

[C8] ——, "Cycle orchestrator: A knowledge-based approach for structuring cyclic ml pipelines in the o&g industry," in *International Semantic Web Conference (ISWC)*, 2020.

[C9] **R. Souza**, J. Camata, M. Mattoso, A. Coutinho, "Runtime steering of parallel cfd simulations," in *International Conference on Parallel Computational Fluid Dynamics*, 2020.

[C10] **R. Souza**, A. Codas, J. A. Nogueira Junior, M. P. Quinones, L. Azevedo, R. Thiago, E. Soares, M. Cardoso, L. Martins, "Supporting the training of physics informed neural networks for seismic inversion using provenance," in *American Association of Petroleum Geologists Annual Convention and Exhibition (AAPG)*, 2020.

[C11] R. Thiago, **R. Souza**, L. Azevedo, E. Soares, R. Santos, W. Santos, M. De Bayser, M. Cardoso, M. Moreno, R. Cerqueira, "Managing data lineage of O&G machine learning models: The sweet spot for shale use case," in *European Association of Geoscientists and Engineers (EAGE) Digitalization Conference and Exhibition*, 2020. DOI: 10.3997/2214-4609.202032075.

[C12] **R. Souza**, L. Azevedo, V. Lourenço, E. Soares, R. Thiago, R. Brandão, D. Civitarese, E. Vital Brazil, M. Moreno, P. Valduriez, M. Mattoso, R. Cerqueira, M. A. S. Netto, "Provenance data in the machine learning lifecycle in computational science and engineering," in *Workflows in Support of Large-Scale Science (WORKS) co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, 2019, pp. 1–10. DOI: 10.1109/WORKS49585.2019.00006.

[C13] **R. Souza**, L. Azevedo, R. Thiago, E. Soares, M. Nery, M. Netto, E. V. Brazil, R. Cerqueira, P. Valduriez, M. Mattoso, "Efficient runtime capture of multiworkflow data using provenance," in *IEEE International Conference on e-Science (eScience)*, 2019, pp. 1–10. DOI: 10.1109/eScience.2019.00047. [Online]. Available: https://doi.org/10.1109/eScience.2019.00047.

[C14] **R. Souza**, E. V. Brazil, L. Azevedo, D. Ferreira, E. Soares, R. Thiago, M. Nery, V. Torres, R. Cerqueira, "Managing data traceability in the data lifecycle for deep learning applied to seismic data," in *American Association of Petroleum Geologists Annual Convention and Exhibition (AAPG)*, 2019. [Online]. Available: https://www.searchanddiscovery.com/abstracts/html/2019/ace2019/abstracts/1718.html.

[C15] M. G. Bayser, C. Pinhanez, H. Candello, M. Affonso, M. P. Vasconcelos, M. A. Guerra, P. Cavalin, **R. Souza**, "Ravel: A mas orchestration platform for human-chatbots conversations," in *International Workshop on Engineering Multi-Agent Systems (EMAS@AAMAS 2018)*, 2018.

[C16] V. Silva, **R. Souza**, J. Camata, D. Oliveira, P. Valduriez, A. L. G. A. Coutinho, M. Mattoso, "Capturing provenance for runtime data analysis in computational science and engineering applications," in *Provenance and Annotation of Data and Processes - International Provenance and Annotation Workshop (IPAW)*, ser. Lecture Notes in Computer Science (LNCS), Springer International Publishing, 2018, pp. 183–187, ISBN: 978-3-319-98379-0. DOI: 10.1007/978-3-319-98379-0_15.

[C17] **R. Souza** and M. Mattoso, "Provenance of dynamic adaptations in user-steered dataflows," in *Provenance and Annotation of Data and Processes - International Provenance and Annotation Workshop (IPAW)*, ser. Lecture Notes in Computer Science (LNCS), Springer International Publishing, 2018, pp. 16–29, ISBN: 978-3-319-98379-0. DOI: 10.1007/978-3-319-98379-0_2.

[C18] **R. Souza**, L. Neves, L. Azeredo, R. Luiz, E. Tady, P. Cavalin, M. Mattoso, "Towards a human-in-the-loop library for tracking hyperparameter tuning in deep learning development," in *Latin American Data Science (LaDaS) workshop co-located with the Very Large Database (VLDB) conference*, Rio de Janeiro, Brazil, 2018, pp. 84–87.

[C19] P. Valduriez, M. Mattoso, R. Akbarinia, H. Borges, J. Camata, A. L. G. A. Coutinho, D. Gaspar, N. Lemus, J. Liu, H. Lustosa, F. Masseglia, F. Nogueira Da Silva, V. Silva, **R. Souza**, K. Ocaña, E. Ogasawara, D. Oliveira, E. Pacitti, F. Porto, D. Shasha, "Scientific Data Analysis Using Data-Intensive Scalable Computing: the SciDISC Project," in *LADaS: Latin America Data Science Workshop*, vol. CEUR Workshop Proceedings, Rio de Janeiro, Brazil: CEUR-WS.org, 2018. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-01867804.

[C20] **R. Souza**, V. Silva, J. Camata, A. Coutinho, P. Valduriez, M. Mattoso, "Tracking of online parameter fine-tuning in scientific workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Denver, CO, 2017. [Online]. Available: `https://hal-lirmm.ccsd.cnrs.fr/lirmm-01620974`.

[C21] **R. Souza**, V. Silva, P. Miranda, A. A. B. Lima, P. Valduriez, M. Mattoso, "Spark scalability analysis in a scientific workflow," in *Simpósio Brasileiro de Banco de Dados (SBBD)*, 2017, pp. 288–293.

[C22] T. Barbosa, **R. Souza**, S. Cruz, M. Campos, R. L. Cottrell, "Applying data warehousing and big data techniques to analyze internet performance," SLAC National Accelerator Lab., Menlo Park, CA (United States), Tech. Rep., 2016.

[C23] J. Camata, J. M. Cela, D. Costa, A. L. G. A. Coutinho, D. Fernández-Galisteo, C. Jimenez, V. Kourdioumov, M. Mattoso, R. Mayo-García, T. Miras, J. A. Moríñigo, J. Navarro, P. O. A. Navaux, D. De Oliveira, M. Rodríguez-Pascual, V. Silva, **R. Souza**, P. Valduriez, "Enhancing Energy Production with Exascale HPC Methods," in *CARLA: Latin American High Performance Computing Conference*, vol. Communications in Computer and Information Science, Mexico City, Mexico: Springer, 2016, pp. 233–246. DOI: `10.1007/978-3-319-57972-6\_17`. [Online]. Available: `https://hal-lirmm.ccsd.cnrs.fr/lirmm-01654914`.

[C24] J. J. Camata, J. M. Cela, D. Costa, A. L. G. A. Coutinho, D. Fernández-Galisteo, C. Jiménez, V. Kourdioumov, M. Mattoso, R. Mayo-García, T. Miras, J. A. Moríñigo, J. Navarro, D. d. Oliveira, M. Rodríguez-Pascual, V. Silva, **R. Souza**, P. Valduriez, "Applying future exascale HPC methodologies in the energy sector," pp. 9–19, 2016. [Online]. Available: `https://upcommons.upc.edu/handle/2117/90905`.

[C25] P. Cavalin, F. Figueiredo, M. Bayser, L. Moyano, H. Candello, A. Appel, **R. Souza**, "Building a question-answering corpus using social media and news articles," in *International Conference on Computational Processing of the Portuguese Language*, 2016, pp. 353–358.

[C26] V. Silva, L. Neves, **R. Souza**, A. Coutinho, D. D. Oliveira, M. Mattoso, "Integrating domain-data steering with code-profiling tools to debug data-intensive workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Salt Lake City, USA, 2016.

[C27] **R. Souza**, V. Silva, A. Coutinho, P. Valduriez, M. Mattoso, "Online input data reduction in scientific workflows," in *Workflows in Support of Large-Scale Science (WORKS) workshop co-located with the ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, 2016, pp. 1–10. [Online]. Available: `https://hal.archives-ouvertes.fr/lirmm-01400538`.

[C28] R. Castro, **R. Souza**, V. Silva, K. Ocaña, D. Oliveira, M. Mattoso, "Uma abordagem para publicação de dados de proveniência de workflows científicos na web semântica," in *Simpósio Brasileiro de Banco de Dados (SBBD)*, 2015.

[C29] **R. Souza**, V. Silva, D. Oliveira, P. Valduriez, A. A. B. Lima, M. Mattoso, "Parallel execution of workflows driven by a distributed database management system," in *ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Salt Lake City, USA, 2015, pp. 1–3. [Online]. Available: `http://sc15.supercomputing.org/sites/all/themes/SC15images/tech_poster/tech_poster_pages/post284.html`.

[C30] **R. Souza**, L. Cottrell, B. White, M. L. Campos, M. Mattoso, "Linked open data publication strategies: Application in networking performance measurement data," in *ASE BigData/SocialCom/CyberSecurity, Stanford, CA*, 2014.

Patents

[P1] L. C. V. Real, M. N. Santos, **R. F. S. Souza**, *Continuous storage of data in a system with limited storage capacity*, US Patent App. 16/678,375, 2021.

[P2] **R. Souza**, R. Mozart, F. R. Da Silva, A. Vital, V. T. d. Silva, *Metadata-based scientific data characterization driven by a knowledge database at scale*, US Patent App. 16/527,546, 2021.

[P3] A. Braz, P. R. Cavalin, F. Figueiredo, M. G. De Bayser, **R. Souza**, *System and method for managing artificial conversational entities enhanced by social knowledge*, Granted, US Patent Application 15/265,615, 2018.

[P4] M. G. De Bayser, A. Braz, P. R. Cavalin, F. Figueiredo, **R. Souza**, *Creating coordinated multi-chatbots using natural dialogues by means of knowledge base*, Granted, US Patent Application 15/217,660, 2018.

[P5] A. P. Appel, A. Gama Leal, **R. Souza**, *Predicting user question in question and answer system*, Granted, US Patent Application 15/171,055, 2017.

Badges and Certifications

- **Machine Learning Specialist Professional** Course duration: 73h — 2022
  Exploratory Data Analysis, Regression, Classification, Deep Learning, Reinforcement Learning, Unsupervised Learning, Time Series and Survival Analysis, AI Ethics and Explainability
- **Trustworthy AI and AI Ethics** Course duration: 3.5h — 2022
- **Enterprise Design Thinking Practitioner** 2022
- **LinkedIn Skill Assessment: Python, MySQL, Linux, T-SQL, NoSQL**

Languages

- **English** - Full proficiency
  - Missouri State University, U.S. Duration: 150h — Jun 2012 – Aug 2012
    Scientific English for Graduate Students
  - Cultura Inglesa (English Culture), Rio de Janeiro, Brazil 2001 – 2009
- **Portuguese** - Native
- **Spanish** - Fluent reading, intermediate speaking and understanding, limited writing