October 7-10 • Ceará • Brazil

34th Brazilian Symposium on DATABASES

SBBD|2019

PROCEEDINGS

October 7-10 • Ceará • Brazil

34th Brazilian Symposium on DATABASES

PROCEEDINGS

Promotion
Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization
Departamento de Computação, Universidade Federal de Ceará– UFC

Steering Committee Chair
Bernadete Farias Lóscio (UFPE, Brazil)

Local Chair
José Maria da Silva Monteiro Filho (UFC, Brazil)

Program Committee Chairs
Full Paper: Carina F. Dorneles (UFSC, Brazil)
Short Paper: Fábio Porto (LNCC, Brazil)
Demos and Applications Chair: Robson L. F. Cordeiro (ICMC-USP, Brazil)
Thesis and Dissertation Workshop Chair: Jonice Oliveira (UFRJ, Brazil)
Tutorials Chair: Altigran Soares da Silva (UFAM, Brazil)
Short course Chair: Maria Cláudia Cavalcanti (IME, Brazil)
Workshop Chair: José Antônio Macedo (UFC, Brazil)
Thesis and Dissertation Contest Chair: Caetano Traina Jr. (USP, Brazil)
Graduation Student Workshop Chair: Ticiana Linhares (UFC, Brazil)

B839
Brazilian Symposium on Databases (SBBD 2019) (25.: 2019 october 07-10, 2019 – Fortaleza, CE)
Proceedings of 34nd Brazilan Symposium on Databases - SBBD 2019 [recurso eletrônico] / Organização: Carina Friedrich Dorneles e Fabio Andre Machado Porto – Fortaleza: SBC, 2019.

355p. v.1

Access Mode: http://sbbd.org.br/2019/

ISSN: 2016-5170

1. Computação - Congressos. 2. Base de Dados – Congressos. I. Dorneles, Carina Fiedrich. II. Sociedade Brasileira de Computação. III. Título.

CDD: 005

Elaborada pela Biblioteca da Universidade Federal do Ceará Campus Crateús (BCCR).

Message from the Local Organization Committee Chairsf

Welcome to the 34th Brazilian Symposium on Databases and to Fortaleza, Ceará! The Brazilian Symposium on Databases is the official database event of the Brazilian Computer Society (SBC) and the largest venue in Latin America for presentation and discussion of research results in the database domain. The 34th edition of the symposium (SBBD 2019) was held in Fortaleza, in the state of Ceará, from October 7th to 10th, 2019. The local organization was performed by the Federal University of Ceará (UFC) through the Computer Science Department (DC). This year, for the first time, SBBD had the Symposium on Knowledge Discovery, Mining and Learning (KDMiLe); the Brazilian Symposium on Bioinformatics (BSB) and the ACM Latin American School on Recommender Systems (LARS) as co-located events providing a rich environment for the discussion of researches of their interrelated areas.

The SBBD 2019 program offers a wide variety of activities, suited for an audience ranging from undergraduate to Ph.D. students, database professionals, practitioners and researchers. The program includes: 3 invited talks and 2 tutorials, presented by distinguished speakers from Brazil, Chile and Germany; 9 technical sessions; 4 short courses about hot topics in the area, presented by specialists in their research fields; demos and applications session; posters sessions; industrial session, thesis and dissertations workshop; the biannual thesis and dissertations contest; 2 co-located workshops; the 3rd KDDBR (Brazilian Knowledge Discovery in Databases) competition; and a panel.

The excellence of SBBD 2019 program is the result of the competence and effort of a large community, which we gratefully acknowledge. The various sections of these proceedings list in detail those that contributed to the SBBD 2019 edition. We thank the symposium chairs and our colleagues of the local organization committee who donated their precious time to made SBBD 2019 a reality. We also thank the Computer Science Department (DC) of the Federal University of Ceará (UFC) and its Post-graduation Program (MDCC), which allowed their staff and students to help on the many tasks of the event preparation. We are also grateful to the SBC board for their support and to the steering committee members for their help, advice and support. Further, we thank the program committee members and external reviewers for the high-quality reviews, and the authors who submitted their papers to SBBD 2019. Finally, we are grateful to our sponsors. Without their support we would not be able to organize this annual event that brings together our community. We hope you all enjoy SBBD 2019 in Fortaleza, Ceará!

José Maria da Silva Monteiro Filho, UFC
SBBD 2019 Local Organization Committee Chair

Table of Contents

Editorial

It is a great pleasure to introduce the Proceedings of the Brazilian Symposium on Databases (SBBD) with the full, short, vision and industry papers accepted for presentation at the 34rd edition of the symposium. SBBD 2019 was held in Fortaleza, in the state of Ceará, Brazil, from October 7th to 10th, 2019. This edition was organized by Universidade Federal do Ceará (UFC, Brazil) and Centro Universitário 7 de Setembro (UNI7, Brazil), both located in the state of Ceará. SBBD is the official database event of the Brazilian Computer Society (SBC) and the largest venue in Latin America forum for presenting and discussing research results and applications on data management. The main areas of interest include Core Database Foundations and Technology, Knowledge Modeling and Management, Information and Data Management, Knowledge Discovery in Databases and Machine Learning and Data Mining. Along with technical sessions, SBBD includes invited talks and tutorials given by distinguished speakers from the national and international research communities. SBBD regularly promotes a demos and applications session, a thesis and dissertations workshop, as well as a thesis and dissertation contest as co-located events. All papers presented in technical sessions during the event reported interesting results or proposed novel thought-provoking ideas in several subjects on the databases and related areas. For the 2019 edition, SBBD accepted six categories of submissions: JIDM articles, full papers, short papers, industrial papers, vision papers, and distinguished published papers. Submissions to the JIDM category could be made throughout the year.

The full papers track had one cycle of submissions, and the review process was performed in one round with a rebuttal phase. Authors were initially notified with the reviews and had a few days for answering the reviewers' comments during the rebuttal phase. After evaluating the rebuttal comments during the discussion period, a final decision was achieved. Out of 47 submitted papers, 14 were accepted as full papers (acceptance rate of 29,7%). The best full papers received award certificates during the event, and their authors will be invited to submit extended versions of their papers to JIDM.

Distinguished Published Papers is a category of submission introduced in SBBD 2017, aiming at attracting the best papers of the Brazilian researchers, published or accepted for publication by a first class database conference, and give the authors the opportunity to

present their work during the event. In this year, five submissions were accepted for presentation by the SBBD Steering Committee.

The topics with more full papers submissions (according to the author's indication from the Topics of Interest) were: Algorithms and Techniques for Data Mining (21 submissions), Data Analytics and Data Visualization (21, submissions), Data Cleaning, Information Filtering and Dissemination (16 submissions), Performance Evaluation and Benchmarking (12 submissions), Data Management for Machine Learning (11 submissions), Database Design and Data Semantics (11 submissions), Information Retrieval Models and Techniques (11 submissions), Information Integration and Interoperability (10 submissions), Knowledge Bases and Modeling (10 submissions).

The short, vision and industry tracks of SBBD 2019 received 38 submissions. All papers were reviewed by at least 3 reviewers and, after a discussion period, 23 were accepted. The set of accepted papers include one from each of vision and industry tracks. The papers were organized into three technical sessions for oral presentation and a poster session. The technical session gathered papers in three themes: DBMS and Big Data; Data Science and Data Model, Integration and Data Structure.

This year we organized an invited industry session counting with three keynotes and a panel from experiences in data management and analysis in large companies, such as Facebook, Petrobras and LeanXcale.

The Proceedings of SBBD are the result of the collective effort of a large community, which we gratefully acknowledge. We thank the SBBD 2019 local organization committee and its symposium chairs, who worked hard to guarantee an outstanding event. We do not have enough words to thank all committee members and external reviewers for their commitment and high-quality reviews. We are also grateful to the steering committee members for their help, advice and support. Finally, we are grateful to the authors who submitted their work to SBBD 2019.

Carina F. Dorneles (UFSC)
Program Chair – SBBD 2019 – Full Papers

Fábio Porto (LNCC)
Program Chair – SBBD 2019 – Short Papers

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

SBBD PROCEEDINGS
FULL PAPERS

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

Program Chair

Carina F. Dorneles, UFSC

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

Promotion

Brazilian Computer Society – SBC
SBD Database Steering Committee

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

SBBD Steering Commitee

Ângelo Brayner (UFC)
Bernadette Lóscio (UFPE) Steering Committee Chair
Carina Dorneles (UFSC)
Sérgio Lifschitz (PUC-Rio)
Fábio Porto (LNCC)
Carmem Hara (UFPR)

SBBD 2019 Commitee

Steering Committee Chair
Bernadette Lóscio (UFPE)

Local Chair
José Maria da Silva Monteiro Filho (UFC)

Full Paper Chair
Carina F. Dorneles (UFSC, Brazil)

Short Paper Chair
Fábio Porto (LNCC, Brazil)

Demos and Applications Chair
Robson L. F. Cordeiro (ICMC-USP, Brazil)

Thesis and Dissertation Workshop Chair
Jonice Oliveira (UFRJ, Brazil)

Tutorials Chair

Altigran Soares da Silva (UFAM, Brazil)

Short course Chair
Maria Cláudia Cavalcanti (IME, Brazil)

Workshop Chair
José Antônio Macedo (UFC, Brazil)

Thesis and Dissertation Contest Chair
Caetano Traina Jr. (USP, Brazil)

Graduation Student Workshop Chair
Ticiana Linhares (UFC, Brazil)

Local Organization Committee
SBBD Local Chair: José Maria da Silva Monteiro Filho (DC/UFC)
Leonardo Oliveira Moreira (Instituto UFC Virtual/UFC)
Marum Simão Filho (UNI7)
Angelo Roncalli de Alencar Brayner (DC/UFC)
Javam de Castro Machado (DC/UFC)

Full Papers Program Committee
Agma Traina (ICMC-USP, Brazil)
Alberto Laender (UFMG, Brazil)
Alexandre Plastino (UFF, Brazil)
Altigran Soares da Silva (UFAM, Brazil)
Ana Carolina Salgado (UFPE, Brazil)
André Santanchè (UNICAMP, Brazil)
Angelo Brayner (UFC, Brazil)
Bernadette Loscio (UFPE, Brazil)
Caetano Traina Júnior (ICMC – USP, Brazil)
Carina F. Dorneles (UFSC, Brazil)
Carmem Hara (UFPR, Brazil)
Celso Hirata (ITA, Brazil)
Clodoveu Davis (UFMG, Brazil)
Cristina Ciferri (USP, Brazil)
Damires Souza (IFPB, Brazil)
Daniel de Oliveira (UFF, Brazil)
Daniel Kaster (UEL, Brazil)
Denio Duarte (UFFS, Brazil)
Dimas C. Nascimento (UFRPE, Brazil)
Duncan Ruiz (PUCRS, Brazil)
Edleno Moura (UFAM, Brazil)
Eduardo de Almeida (UFPR, Brazil)
Eduardo Ogasawara (CEFET/RJ, Brazil)
Elaine Sousa (USP, Brazil)
Fabio Porto (LNCC, Brazil)
Fernanda Baião (PUC-Rio, Brazil)
Flavio Horita (UFABC, Brazil)

Genoveva Vargas-Solar – French Council on Scientific Research
Humberto Razente (UFU, Brazil)
Javam Machado (UFC, Brazil)
João Eduardo Ferreira (IME/USP, Brazil)
José Palazzo Moreira de Oliveira (UFRGS, Brazil)
José Antonio Macêdo (UFC, Brazil)
Julio Dos Reis (IC-UNICAMP, Brazil)
Karin Becker (UFRGS, Brazil)
Kelly Braghetto (IME/USP, Brazil)
Khalid Belhajjame (Université Paris Dauphine, France)
Lucas Augusto Carvalho (UNICAMP, Brazil)
Luciano Barbosa (UFPE, Brazil)
Marco Antonio Casanova (PUC-Rio, Brazil)
Marcos Gonçalves (UFMG, Brazil)
Maria Camila Nardini Barioni (UFU, Brazil)
Maria Claudia Cavalcanti (IME, Brazil)
Mario Nascimento (Univ. Alberta, Edmonton, Canada)
Maristela Holanda (UnB, Brazil)
Mirella Moro (UFMG, Brazil)
Mirian Halfeld-Ferrari (Université d'Orleans – LIFO, France)
Olivier Corby (INRIA, France)
Pedro Eugenio Rocha Pedreira (Facebook Inc.)
Renata Galante (UFRGS, Brazil)
Renato Fileto (UFSC, Brazil)
Ricardo Ciferri (UFSCar, Brazil)
Ricardo Torres (Unicamp, Brazil)
Ronaldo Mello (UFSC, Brazil)
Sergio Lifschitz (PUC-Rio, Brazil)
Theo Haerder (Univ. Kaiserlautern, Germany)
Tiago de Melo (UFAM, Brazil)
Valéria C. Times (UFPE, Brazil)
Vanessa Braganholo (UFF, Brazil)
Vania Bogorny (UFSC, Brazil)
Vania Vidal (UFC, Brazil)
Vaninha Vieira (UFBA, Brazil)
Wagner Meira Jr. (UFMG, Brazil)
Zoubida Kedad (UVSQ, France)

Table of Contents (Full Papers)

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

SBBD PROCEEDINGS
SHORT PAPERS

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

Program Chair

Fábio Porto, LNCC

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

SBBD Steering Commitee

Ângelo Brayner (UFC)
Bernadette Lóscio (UFPE) coordenadora da CEBD
Carina Dorneles (UFSC)
Sérgio Lifschitz (PUC-Rio)
Fábio Porto (LNCC)
Carmem Hara (UFPR)

SBBD 2019 Commitee

Steering Committee Chair
Bernadette Lóscio (UFPE)

Local Chair
José Maria da Silva Monteiro Filho (UFC, Brazil)

Full Paper Chair
Carina F. Dorneles (UFSC, Brazil)

Short Paper Chair
Fábio Porto (LNCC, Brazil)

Demos and Applications Chair
Robson L. F. Cordeiro (ICMC-USP, Brazil)

Thesis and Dissertation Workshop Chair
Jonice Oliveira (UFRJ, Brazil)

Tutorials Chair
Altigran Soares da Silva (UFAM, Brazil)

Short course Chair
Maria Cláudia Cavalcanti (IME, Brazil)

Workshop Chair
José Antônio Macedo (UFC, Brazil)

Thesis and Dissertation Contest Chair
Caetano Traina Jr. (USP, Brazil)

Graduation Student Workshop Chair
Ticiana Linhares (UFC, Brazil)

Local Organization Committee
SBBD Local Chair: José Maria da Silva Monteiro Filho (DC/UFC)
Leonardo Oliveira Moreira (Instituto UFC Virtual/UFC)
Marum Simão Filho (UNI7)
Angelo Roncalli de Alencar Brayner (DC/UFC)
Javam de Castro Machado (DC/UFC)

Short Papers Program Committee
Alessandreia Oliveira (UFJF)
Altigran Soares da Silva (UFAM)
Ana Carolina Almeida (UERJ)
Anderson Ferreira (UFOP)
Angelo Brayner (UFC)
Bernadette F. Lóscio (UFPE)
Carlos Eduardo Pires (UFCG)
Carmem Hara (UFPR)
Cristina Ciferri (USP)
Damires Souza (IFPB)
Daniel de Oliveira (UFF)
Daniel Kaster (UEL)
Daniela Barreiro Claro (UFBA)
Deise Saccol (UFSM)
Denio Duarte (UFFS)
Duncan Ruiz (PUCRS)
Eduardo Borges (FURG)
Eduardo de Almeida (UFPR)
Eduardo Ogasawara (CEFET/RJ)
Elaine Sousa (USP)
Eveline Sacramento (FUNCEME)
Fernanda Baião (UNIRIO)
Flávio R. C. Sousa (UFC)
Helena Ribeiro (UCS)
Humberto Razente (UFU)
João Eduardo Ferreira (IME/USP)

Jonas Dias (DELL EMC)
Jonice de Oliveira Sampaio Oliveira (IM/UFRJ)
José Antonio Macêdo (UFC)
José de Aguiar Moraes Filho (UNIFOR)
José Monteiro (UFC)
José Palazzo Moreira de Oliveira (UFRGS)
Karin Becker (UFRGS)
Kelly Braghetto (IME/USP)
Luciano Barbosa (UFPE)
Luiz Celso Gomes Jr (UTFPR)
Luiz Manoel Rocha Gadelha Júnior (LNCC)
Maria Camila Nardini Barioni (UFU)
Maristela Holanda (UnB)
Michele Brandão (UFMG)
Mirella Moro (UFMG)
Moisés Carvalho (UFAM)
Pedro Eugenio Rocha Pedreira (Facebook Inc.)
Raquel Stasiu (PUCPR / UTFPR)
Raqueline Penteado (UEM)
Rebeca Schroeder (UDESC)
Renata Galante (UFRGS)
Renato Fileto (UFSC)
Robson Cordeiro (USP)
Robson Fidalgo (UFPE)
Sergio Lifschitz (PUC-Rio)
Sergio Mergen (UFSM)
Thiago Henrique Silva (UTFPR)
Ticiana Coelho da Silva (UFC)
Valéria C. Times (UFPE)
Vanessa Braganholo (UFF)
Vaninha Vieira (UFBA)

ADDITIONAL REVIEWERS
Marcelo Iury S. Oliveira (UFRPE)
Eduardo Pena (UFPR)
Guilherme Queiroz Vasconcelos (USP)
Nielsen Luiz Rechia Machado (PUCRS)

Table of Contents (Short Papers)

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

TUTORIALS

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

Program Chair

Altigran Soares da Silva (UFAM, Brazil)

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

SBBD Steering Commitee

Ângelo Brayner (UFC)
Bernadette Lóscio (UFPE) coordenadora da CEBD
Carina Dorneles (UFSC)
Sérgio Lifschitz (PUC-Rio)
Fábio Porto (LNCC)
Carmem Hara (UFPR)

SBBD 2019 Commitee

Steering Committee Chair
Bernadette Lóscio (UFPE)

Local Chair
José Maria da Silva Monteiro Filho (UFC, Brazil)

Full Paper Chair
Carina F. Dorneles (UFSC, Brazil)

Short Paper Chair
Fábio Porto (LNCC, Brazil)

Demos and Applications Chair
Robson L. F. Cordeiro (ICMC-USP, Brazil)

Thesis and Dissertation Workshop Chair
Jonice Oliveira (UFRJ, Brazil)

Tutorials Chair
Altigran Soares da Silva (UFAM, Brazil)

Short course Chair
Maria Cláudia Cavalcanti (IME, Brazil)

Workshop Chair
José Antônio Macedo (UFC, Brazil)

Thesis and Dissertation Contest Chair
Caetano Traina Jr. (USP, Brazil)

Graduation Student Workshop Chair
Ticiana Linhares (UFC, Brazil)

Local Organization Committee
SBBD Local Chair: José Maria da Silva Monteiro Filho (DC/UFC)
Leonardo Oliveira Moreira (Instituto UFC Virtual/UFC)
Marum Simão Filho (UNI7)
Angelo Roncalli de Alencar Brayner (DC/UFC)
Javam de Castro Machado (DC/UFC)

Table of Contents (Tutorials)

DuckDB Autopsy: The internals of the "SQLite for Analytics"

Pedro Holanda, Mark Raasveldt

[1] Centrum Wiskunde & Informatica (CWI)
Amsterdam, The Netherlands.

{holanda,raasveld}@cwi.nl

Abstract. *The immense popularity of SQLite shows that there is a need for unobtrusive in-process data management solutions. However, there is no such system yet geared towards analytical workloads. We present DuckDB, a novel data management system designed to execute analytical SQL queries while embedded in another process. In our talk, we give an in-depth overview of the internals of DuckDB and the design choices that were made to cater to the use case of embedded analytics. DuckDB is available as Open Source software under a permissive license.*

Resumo. *A imensa popularidade do SQLite demonstra a necessidade de sistemas de gerenciamento de banco de dados (SGBD) embarcados. No entanto, ainda não existe um SGBD embarcado voltado para cargas de trabalho analíticas. Nessa palestra apresentamos o DuckDB, um novo SGBD projetado para executar consultas analíticas enquanto incorporado em outro processo. Apresentamos uma visão geral dos aspectos internos do DuckDB e das decisões de design feitas para atender as cargas analíticas em SGBDs embarcados. O DuckDB já está disponível para download e uso.*

Introduction

In this talk, we present the internal structure of our new system, *DuckDB*. DuckDB is a new purpose-built embeddable relational database management system. DuckDB is available as Open-Source software under the permissive MIT license[1]. DuckDB is no research prototype but built to be widely used, with millions of test queries run on each commit to ensure correct operation and completeness of the SQL interface. DuckDB was built specifically to support the use case of embedded analytics, and focused on fulfilling the following requirements of this use case:

- Efficient transfer of tables to and from the database is essential. Since both database and application run in the same process and thus address space, there is a unique opportunity for efficient data sharing which needs to be exploited.
- High efficiency for OLAP workloads, but without completely sacrificing OLTP performance. For example, concurrent data modification is a common use case in dashboard-scenarios where multiple threads update the data using OLTP queries, and other threads run the OLAP queries that drive visualizations simultaneously.

[1] https://github.com/cwida/duckdb

- High degree of stability, if the embedded database crashes, for example, due to an out-of-memory situation, it takes the host down with it. This can never happen. Queries need to be able to be aborted cleanly if they run out of resources, and the system needs to gracefully adapt to resource contention.
- Practical "embeddability" and portability, the database needs to run in whatever environment the host does. Dependencies on external libraries (e.g., `openssh`) for either compile- or runtime have been found to be problematic. Signal handling, calls to `exit()` and modification of singular process state (locale, working directory, etc.) are forbidden.

Outline

The talk is catered primarily towards people that have a basic understanding of core database systems, and the goal of the talk is to make users more familiar with the internals of modern columnar database systems and specifically the internals of DuckDB. The talk is divided into six sections of 30 minutes each:

1. DuckDB: Introduction and Motivation
2. Parser, Binder and Logical Planner
3. Physical Execution
4. Optimizers
5. Transactions & Storage Layer
6. Interactive Demo

DuckDB: Introduction and Motivation. In the talk, we start by giving a brief overview of DuckDB and the motivation behind embedded analytical systems.

Parser, Binder and Logical Planner. In this section we discuss the parser, binder and logical planner of the DuckDB system. We do this by taking you on a journey of the life of a query inside the database system. We show the internal structures that are created and how the query string is converted into a logical plan.

Physical Execution. After showing how the logical plan is constructed, we show how the system executes the physical plan using its vectorized execution engine [Boncz et al. 2005]. We discuss the techniques that are used for the execution of scans, indexes [Leis et al. 2013], joins, aggregates, sorting and window functions [Leis et al. 2015].

Optimizers. After the base logical plan is created, we show how the optimizers work to create a fast plan. We discuss join order optimization [Moerkotte and Neumann 2008], subquery unnesting [Neumann and Kemper 2015], filter pushdown as well as various simple scalar optimizers such as constant folding and common subexpression elimination.

Transactions & Storage Layer. We discuss the MVCC model [Neumann et al. 2015] that is used by DuckDB and how it works to maintain transactional integrity and atomicity. We also discuss the storage layer of DuckDB. We describe how the data is laid out on disk and talk about how the buffer manager loads and caches data into memory [Leis et al. 2018].

Interactive Demo. We guide people in setting up and using DuckDB in combination with Python. We have a prepared data set and demo use case in which users use DuckDB

to wrangle a set of census data concerning US voters, after which they use *sci-kit learn* to build a full machine learning pipeline that attempts to classify individual voters [Raasveldt et al. 2018].

Referências

Boncz, P. A., Zukowski, M., and Nes, N. (2005). Monetdb/x100: Hyper-pipelining query execution. In *CIDR*.

Leis, V., Haubenschild, M., Kemper, A., and Neumann, T. (2018). Leanstore: In-memory data management beyond main memory. In *ICDE*.

Leis, V., Kemper, A., and Neumann, T. (2013). The adaptive radix tree: Artful indexing for main-memory databases. In *ICDE*.

Leis, V., Kundhikanjana, K., Kemper, A., and Neumann, T. (2015). Efficient processing of window functions in analytical sql queries. *VLDB*.

Moerkotte, G. and Neumann, T. (2008). Dynamic programming strikes back. In *SIGMOD*. ACM.

Neumann, T. and Kemper, A. (2015). Unnesting arbitrary queries. *Datenbanksysteme für Business, Technologie und Web (BTW 2015)*.

Neumann, T., Mühlbauer, T., and Kemper, A. (2015). Fast serializable multi-version concurrency control for main-memory database systems. In *SIGMOD*.

Raasveldt, M., Holanda, P., Mühleisen, H., Manegold, S., et al. (2018). Deep integration of machine learning into column stores. *EDBT*.

FAT in Recommendation Systems

Denis Parra

Pontificia Universidad Católica de Chile (PUC Chile)

Abstract. *In recent years we have experienced an increasing deployment in our daily lives of applications based on Artificial Intelligence technology. Some of these applications are truly amazing, such as self driving cars, translation systems, and automatic detection of illnesses. However, we have also seen applications which are disrupting our daily lives with unintended consequences, such as recommender systems which polarize public opinion and face recognition applications which can hinder our privacy. As the researcher Michael Jordan stated: "Just as early buildings and bridges sometimes fell to the ground — in unforeseen ways and with tragic consequences — many of our early societal-scale inference-and-decision-making systems are already exposing serious conceptual flaws." In this context, a serious and systematic study of FAT (Fairness, Accountability and Transparency) in AI is a critical. The popularity of the recently created FAT conference and several related workshops in academic conferences, especially those related to topics in artificial intelligence, show evidence of how researchers and developers, as well as the society at large, is becoming aware of the effects of AI technology. In terms of recommendation systems, algorithmic transparency has been a topic of study in for at least 10 years, but only recently this area has attracted strong attention in the community. In this tutorial, Prof. Parra will present a general overview of FAT in AI, to then focus on FAT for recommendation systems with a theoretical presentation as well as practical activities.*

REALIZATION

EXECUTION

SUPPORT

SILVER SPONSOR

ACADEMIC SUPPORT