

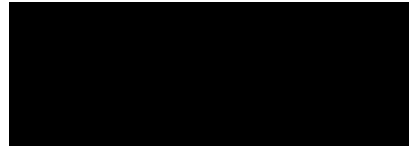
IEEE BigData 2014 Coimbra Satellite Session ADVANCE PROGRAM

2014 International BigData Coimbra Satellite Session

<http://www.ieeebigdata.org/2014/satellite/coimbra>

2014 3rd International Congress on Big Data (BigData 2014)

<http://www.ieeebigdata.org>



Organization Committee

■ Program Committee Chairs:

Pedro Furtado
Department of Informatics Engineering
University of Coimbra, Portugal
Email: pnf@dei.uc.pt

Ladjel Bellatreche
National Engineering School for Mechanics and Aerotechnics (ENSMA), France
Email: bellatreche@ensma.fr

■ Publicity Chairs:

Jorge Bernardino
Instituto Superior de Engenharia de Coimbra, Portugal

Nuno Laranjeiro
University of Coimbra, Portugal

■ Program Committee Members:

Jose Orlando Pereira, Universidade do Minho, Portugal
Sven Groppe, Universitat zu Lubeck, Germany
Tilmann Rabl, University of Toronto, Canada
Marco Vieira, University of Coimbra, Portugal
Joao Gama, University of Porto, Portugal
Carson K. Leung, University of Manitoba, Canada
Robert Wrembel, Poznan University of Technology, Poland
Luis Veiga, Instituto Superior Tecnico, Portugal
Filipe Araujo, University of Coimbra, Portugal
Alfredo Cuzzocrea, ICAR-CNR and University of Calabria, Italy
Orlando Belo, universidade do Minho, Portugal

May 16, 2014 (Friday)		
Keynote Speech Session 1		
Time	Keynote Speech	Keynote Speaker
09:00 - 10:00	Using Machine Learning and Big-Data to Fight Payment Fraud	Paulo Marques/Pedro Bizarro - FeedZai
10:00 - 10:15	Coffee Break	
Research Track Session 1		
Time	Topic	
10:15 - 10:45	Real-Time Integration of Building Energy Data	
10:45 - 11:15	Geometrical and Topological Modelling: A Fast Computation of Spatial 3D TLS Data Selections	
11:15 - 11:45	Using Multimedia Ontologies for Automatic Image Annotation and Classification	
11:45 - 11:55	Poster Time (first show)	
	<p>Decision Support Encryption Performance on MySQL database</p> <p>Graph Databases and Neo4j</p> <p>Comparing Neo4J with mySQL</p> <p>Benchmarking Scalability of Cloud Engines for Decision Support Systems</p> <p>Benchmarking Scalability of Secure Decision Support Systems</p> <p>Decision Support System Scalability on the Cloud</p> <p>Cassandra: What it Does and What it Doesn't</p>	
11:55 - 13:00	Lunch Break (Canteen, costs you ~5 to 10 euros)	

Keynote Speech Session 2		
Time	Keynote Speech	Keynote Speaker
13:00 - 14:00	“Small and Big Data” – Opposite Vectors on Dimensional Modelling	Orlando Belo, U. Minho
14:00 – 14:10	Break	
Research Track Session 2		
Time	Topic	
14:10 - 14:50	On Automatically scaling the ETL process for freshness-preserving in high-rate data warehousing	
14:50 – 15:10	SQL vs NoSQL: Performance and Scalability	
15:10 – 15:40	A DW model for the bigData era	
15:40– 15:50	Coffee Break	
WiP and Poster Papers		
Time	Topic	
15:50 – 16:05	Testing MySQL Cluster with Star Schema Benchmark	
16:05 - 16:20	Benchmarking Scalability for Keyword Search in BigData	
16:20 – 16:35	Clustering over SSB	
16:35 – 16:50	Bigdata Scalable Security	
16:50 – 17:00	A low cost private cloud infrastructure using OpenStack	
Poster Time (second show)/ Break		
17:00 – 17:15	Decision Support Encryption Performance on MySQL database	
	Graph Databases and Neo4j	
	Comparing Neo4J with mySQL	

	<p>Benchmarking Scalability of Cloud Engines for Decision Support Systems</p> <p>Benchmarking Scalability of Secure Decision Support Systems</p> <p>Decision Support System Scalability on the Cloud</p> <p>Cassandra: What it Does and What it Doesn't</p>	
<p>Keynote Speech Session 3</p>		
Time	Keynote Speech	Keynote Speaker
17:15 - 18:15	Data Warehousing and OLAP over Big Data	Alfredo Cuzzocrea, ICAR-CNR & University of Calabria, Rende, Cosenza, Italy.
18:15	<p>End of Event</p>	

IEEE BigData 2014

Coimbra Satellite Session

Keynote/Invited Talks

Paulo Marques
FeedZai

Using Machine Learning and Big-Data to Fight Payment Fraud

Payment Fraud and data breaches are on a rise. Every year millions of credit-card numbers are compromised. In the recent Target breach between 70 and 110 million people say their information stolen and purchases being made in their name. In this talk we will discuss how at Feedzai we leverage Machine Learning and Big-Data Techniques to make commerce safe. In particular, we'll discuss fraud trends, challenges with dealing with high-volume high-velocity skewed data, and how machine learning and big-data provide a strong foundation for fighting fraudsters.

Orlando Belo

“Small and Big Data” – Opposite Vectors on Dimensional Modelling

Department of Informatics
ALGORITMI R&D Centre
School of Engineering
University of Minho

The implementation of a data warehousing system depends largely from the way it is designed following business requirements and decision-makers' needs, accompanied by a huge host of hellish conditions and operational aspects. However, if we consider the usefulness and the utility of the system there is a very important aspect that we should take strongly into consideration, not only because it represents the materialization of business and analytical exploration requirements, but also the performance of the system in meeting users' ad hoc operating procedures. We are talking, of course, to the way we do dimensional modelling. It provides an effective way to develop successful multidimensional data schemes, having the ability to host all analysis dimensions as well as the facts that sustain sophisticated temporal analysis over them. Stars, snowflakes and constellations are common terms used to represent all types of settings that dimensional schemas can assume as configurations. By default, dimensional schemes are designed to accommodate large volumes of data. Usually they are arranged to support and satisfy any query (almost) immediately, even when it is very complex and demanding such as a star-join. With the advent of "big data" - intensive data processing -, today the volume of information managed by a data warehouse seems to be positioned in other kind of dimension: "smaller". However, the influence of application scenarios, labelled as big data, is not very clearly, and we do not know what is its real impact over data warehouses components. Thus, in this talk we are interested to identify such influence in a more concrete way, taking a typical big

data scenario, and discussing some of the aspects more relevant in the development of a data warehouse system, giving particular emphasis, obviously, to the design and implementation of dimensional data models.

Alfredo Cuzzocrea,
ICAR-CNR & University of Calabria, Rende, Cosenza, Italy.

Data Warehousing and OLAP over Big Data

Data Warehousing and OLAP over Big Data is becoming one of the emergent challenges for next-generation research, with special emphasis on data-intensive Cloud infrastructures. As a consequence, several studies are focusing the attention on this relevant issue, and various open problems arise. This evidence has inspired our study, which provide a comprehensive overview on actual open research problems in the context of Data Warehousing and OLAP over Big Data, along with a deep critical discussion on future research directions to be taken under this so-challenging road.

Pedro Martins
University of Coimbra

Automatically scaling the ETL process for freshness-preserving in high-rate data warehousing

We investigate how to deal with the problem of providing scalability and data freshness automatically, and how to deal with high-rate data efficiently in very large data warehouses. In general, data freshness is not guaranteed in those contexts, since data loading, transformation and integration are heavy tasks that are done only periodically, instead of row by row.

Many current data warehouse deployments are designed to work as single server, although for many applications problems related with data volume processing times, data rates, and requirements for fresh and fast responses, increasingly make this approach less useful. The solution is to use/build parallel architectures and mechanisms to speed-up data integration and handle fresh data efficiently.

Desirably users developing data warehouses should need to concentrate solely on the conceptual and logic design (e.g. business driven requirements, logical warehouse schemes, workloads analysis and ETL process), while physical details including mechanisms for scalability, freshness and integration of high-rate data should be left to automated tools.

We propose a universal data warehouse parallelization solution, that is, an approach that enables the automatic scalability and freshness of any data warehouse and ETL.

BigDW: A DW model for the bigData era
João Costa (jcosta@isec.pt), Pedro Furtado
(pnf@dei.uc.pt)

Abstract:

bigData introduces data and processing issues to DW processing, due the variety of the data and the velocity that it is continuously being produced and have to be integrated and became available for OLAP analysis. The way that typical DWs store and process data is unable to efficiently handle such scalable data volumes, with freshness and timeliness guarantees. In this paper we discuss these issues and present an approach that massively partitions data (both horizontally and vertically), and distributes the inbound query processing load among processing nodes, by orchestrating them in a mix of distributed and parallel tasks to cooperating nodes for processing common tasks and thus providing timely results.

Veronika Abramova
University of Coimbra
“SQL vs NoSQL: Performance and Scalability”

With an increase of interest in non-relational database technology, was developed a variety of NoSQL databases. Currently there are over 150 diverse NoSQL databases that are divided accordingly to used mechanisms and purpose. However, it is important to understand the aim of different databases as well as its main characteristics is what determines entire system performance. Also with the increase of data storage and extraction, databases should be able to efficiently manage high demand. Therefore we decided to evaluate two aspects of databases that we consider important nowadays: performance and scalability. For the first part of our evaluation we tested one of the most popular NoSQL databases: Cassandra. We evaluated how database performance is affected by load and whether adding more cluster nodes result in lower execution time of requests. Second part of our experiment was focused on decision support systems as one of the main purposes of databases. We decided to verify if non-relational databases are suited for decision support systems and capable of executing complex requests. For that we used standard Star Schema Benchmark and executed queries over Hadoop framework, using Hive, and MySQL Cluster.

Research Papers

Real-Time Integration of Building Energy Data(#6854), *Diogo Anjos (INESC-ID Lisboa) Paulo Carreira (INESC-ID Lisboa) Alexandre Francisco (INESC-ID Lisboa)*.

Geometrical and Topological Modelling: A Fast Computation of Spatial 3D TLS Data Selections(#6826), *Jose Rodrigues (University of Algarve PT) Mauro Figueiredo (University of Algarve) Ivo Silvestre (University of Algarve) Cristina Veiga-Pires (University of Algarve)*.

Using Multimedia Ontologies for Automatic Image Annotation and Classification, *Antonio M. Rinaldi (DIETI-UNINA IT)*.

Scalability and Speedup evaluation of Cassandra, *Veronika Abramova (DEI/CISUC), Jorge Bernardino (IPC/ISEC), Pedro Furtado (DEI/CISUC)*.

WiP (Student) Papers

Testing MySQL Cluster with Star Schema Benchmark

Rafael Almeida, IPC-ISEC, Coimbra, Portugal.
Jorge Bernardino, IPC-ISEC, Coimbra, Portugal.
Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Benchmarking scalability for Keyword Search in BigData

Mariana Lourenço, DEI, University of Coimbra, Coimbra, Portugal.
Adriana Ferrugento, DEI, University of Coimbra, Coimbra, Portugal.
Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Clustering over SSB

Filipe Assunção, DEI, University of Coimbra, Coimbra, Portugal.
Manuel Levi, DEI, University of Coimbra, Coimbra, Portugal.
Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

A low cost private cloud infrastructure using Openstack

Tiago Rosado, IPC-ISEC, Coimbra, Portugal,
Jorge Bernardino, IPC-ISEC, Coimbra, Portugal.

NoSQL Databases – Voldemort vs Riak

Ricardo Neves, IPC-ISEC, Coimbra, Portugal,
Jorge Bernardino, IPC-ISEC, Coimbra, Portugal,
Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Bigdata Scalable Security

Joao Correia, DEI, University of Coimbra, Coimbra, Portugal.
Jose Grilo, DEI, University of Coimbra, Coimbra, Portugal.
Joao Sa, University of Coimbra, Coimbra, Portugal.
Tiago Mateus, University of Coimbra, Coimbra, Portugal.
Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Posters (Student)

Decision Support Encryption Performance on MySQL database,

João Matos, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Jorge Bernardino, IPC-ISEC, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Graph Databases and Neo4j,

José Guia, IPC-ISEC, Coimbra, Portugal.

Jorge Bernardino, IPC-ISEC, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Comparing Neo4J with mySQL

Mariana Moutinho, DEI, University of Coimbra, Coimbra, Portugal.

Telmo Neves, DEI, University of Coimbra, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Benchmarking Scalability of Cloud Engines for Decision Support Systems

Daniel Frutuoso, DEI, University of Coimbra, Coimbra, Portugal.

João Macedo, DEI, University of Coimbra, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Benchmarking Scalability of Decision Support Systems on the Cloud

Tiago Antunes, DEI, University of Coimbra, Coimbra, Portugal.

Filipe António, DEI, University of Coimbra, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Benchmarking Decision Support Scalability for Secure BigData

Pais Marta, DEI, University of Coimbra, Coimbra, Portugal.

José Ramos, DEI, University of Coimbra, Coimbra, Portugal.

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Cassandra: What it Does and What it Doesn't

Melyssa Barata, IPC-ISEC, Coimbra, Portugal,

Jorge Bernardino, IPC-ISEC, Coimbra, Portugal,

Pedro Furtado, DEI/CISUC, University of Coimbra, Coimbra, Portugal.

Event Location Information

The event will take place in room B1 of “Departamento de Engenharia Informática da Universidade de Coimbra”. Participants go directly to the conference room.



http://www.uc.pt/en/fctuc/dei/informacao_visitantes/#
<http://www.uc.pt/fctuc/dei/contactos>

Contacts

Adriana Ferrugento aferr@student.dei.uc.pt
Antonio M. Rinaldi amrinald@unina.it
Cuzzocrea cuzzocrea@si.dimes.unical.it
Daniel Frutuoso daniel@student.dei.uc.pt
Diogo Anjos diogo.silva.anjos@tecnico.ulisboa.pt
Filipe António fantonio@student.dei.uc.pt
Filipe Assunção fga@student.dei.uc.pt
Joao Correia jrcorr@student.dei.uc.pt
João Matos jmatos@student.dei.uc.pt
João Pedro Costa jpcosta@dei.uc.pt
Jorge Bernardino jorge@isec.pt
Jose Grilo jgrilo@student.dei.uc.pt
José Guia
José Ramos jrramos@student.dei.uc.pt
Jose Rodrigues jirodrig@ualg.pt
Manuel Levi eiro@student.dei.uc.pt
Mariana Lourenço mrlouren@student.dei.uc.pt
Marta Pais martaluisapais@gmail.com
Paulo Carreira paulo.carreira@ist.utl.pt
Pedro Furtado pnf@dei.uc.pt
Pedro Martins pmom@dei.uc.pt
Rafael Almeida rafcorreia@hotmail.com
Ricardo Neves ricfrneves@hotmail.com
Tiago Rosado a20105059@alunos.isec.pt
Veronika Abramova veronika@student.dei.uc.pt