

An Indian-Australian research partnership

Project Title:	Constraint Solving for Checking Correctness of SQL Queries	
Project Number	IMURA0803	
Monash Main Supervisor (Name, Email Id, Phone)	Peter Stuckey peter.stuckey@monash.edu	Full name, Email
Monash Co-supervisor(s) (Name, Email Id, Phone)		
Monash Head of Dept/Centre (Name,Email)	Alan Dorin	Full name, email
Monash Department:	Faculty of Information Technology	
Monash ADGR (Name,Email)	Sue Mckemmish	Full name, email
IITB Main Supervisor (Name, Email Id, Phone)	S. Sudarshan sudarsha@cse.iitb.ac.in	Full name, Email
IITB Co-supervisor(s) (Name, Email Id, Phone)		Full name, Email
IITB Head of Dept (Name, Email, Phone)	Uday Khedker head@cse.iitb.ac.in	Full name, email
IITB Department:	Computer Science and Engineering	

Research Clusters:

Research Themes:

Highlight which of the Academy's CLUSTERS this project will address? <i>(Please nominate JUST one. For more information, see www.iitbmonash.org)</i>		Highlight which of the Academy's Theme(s) this project will address? <i>(Feel free to nominate more than one. For more information, see www.iitbmonash.org)</i>	
1	Material Science/Engineering (including Nano, Metallurgy)	1	Advanced computational engineering, simulation and manufacture
2	Energy, Green Chem, Chemistry, Catalysis, Reaction Eng	2	Infrastructure Engineering
3	Math, CFD, Modelling, Manufacturing	3	Clean Energy
4	CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control	4	Water
5	Earth Sciences and Civil Engineering (Geo, Water, Climate)	5	Nanotechnology
6	Bio, Stem Cells, Bio Chem, Pharma, Food	6	Biotechnology and Stem Cell Research
7	Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng	7	Humanities and social sciences
8	HSS, Design, Management	8	Design

The research problem

The bulk of the worlds information is stored in databases, and the most common approach to accessing this information is using the SQL query language. Complex SQL queries are widely used today, but it is rather difficult to check if a complex query has been written correctly. Formal verification based on comparing a specification with an implementation is not applicable, since SQL queries are essentially a specification without any implementation. Queries are usually checked by running them on sample datasets and checking that the correct result is returned; there is no guarantee that all possible errors are detected.

The XData system, built at IIT Bombay, generates test datasets for a query by creating sets of constraints designed to catch specific types of errors, and then generates test data using a constraint solver. While the system already handles a large class of SQL queries and error types, and has been tested extensively on student SQL assignment queries, more complex queries are still hard to handle. IITB is working on handling such queries but there are many technical challenges.. Monash university has world-leading experts on discrete optimization and constraint solving. We believe a joint effort can help expand the class of queries that the XData system can handle, helping generate test data for complex industrial SQL queries. Conversely, we trust that the need to handle complex SQL queries can drive research on constraint solving techniques, including string solving.

Project aims

A primary aim of the system is to develop techniques to generate test data for testing complex SQL queries. Such queries are widely used in the commercial world, and there is always a lurking fear that the queries are not quite correct, and that errors in the queries can lead to wrong business decisions. We believe wide sections of the data analysis industry can benefit from the research results and tools that the project aims to generate.

A secondary goal of the project is to generate data for testing student SQL queries, which will benefit students who are learning SQL, and instructors who need to grade SQL queries.

Expected outcomes

The project will produce new approaches to checking complex SQL queries, and new approaches to constraint solving generated by the new challenges that arise in this complex domain. Given the expertise in IITB in database query optimization and test data generation, and the expertise at Monash in discrete optimization, in particular in string constraint solving, we have no doubt that new and exciting results will be obtained.

How will the project address the Goals of the above Themes?

The advanced computational engineering theme seeks to reason about large complex systems. In order to reason about such systems we need a range of engineering-based models and technologies. This is usually accomplished by handling extensive amounts of business data about a system or process. Crucial to handling large amounts of data effectively is the use of databases. But correctness of database operations are hard to verify. In this project we will develop new approaches and new solving technologies in order to verify the correctness of database queries that are essential to such systems.

Capabilities and Degrees Required

Bachelors/Masters degree in Computer Science or in related areas but with demonstrable knowledge of computer science fundamentals including logic, algorithms, databases, etc.

Potential Collaborators

In addition to Peter Stuckey at Monash and S. Sudarshan at IITB:

At IITB, there is a formal methods group, some members of which, particularly Supratik Chakraborty who works on SMT solvers, could be collaborators.

At Monash there is a large discrete optimization group, including for example Graeme Gange who is an expert on string constraint solving.

Select up to **(4)** keywords from the Academy's approved keyword list (**available at <http://www.iitbmonash.org/becoming-a-research-supervisor/>**) relating to this project to make it easier for the students to apply.

Data Science, optimisation, algorithms 6