Project Title: An Intelligent Data Integration Architecture for the automated integration of heterogenous datasets

Supervisors: Jade Lyons, Thomas Dowling

Keywords: Data Integration, data lake, Integrated data architecture, intelligent data integration

Project Summary:

The amount of data being generated and stored globally is increasing dramatically, up to 180 zettabytes by 2025. This data is generated from a variety of different sources, and to remain competitive, organisations need to be able to integrate this data effectively. It can be seen that large scale data repositories are needed to support data analytics.

In 2010 the term data lake was coined to describe a new way of storing large amounts of data in its original form, in a single location. However, there are numerous limitations to data warehouses including the fact that they aren't typically suitable for unstructured data, and they take a very long time to build because of the amount of pre-processing that is required.

Data lakes overcome some of the problems with data warehouses but, as has been seen, governance, metadata and data quality remain a problem. Thus, it is clear that improved tools for the management and integration of heterogenous data sets are required and these tools need to be better in terms of both performance and their capacity to retain the meaning of the underlying data.
Deep learning techniques can be used very effectively to resolve problems in data integration. The types of problems that may be addressed using this technology includes: Multi-modal Data Integration, Human in the Loop Data Integration and Declarative Interfaces for Data Integration.

The objective of this project is to create and evaluate an Intelligent Data Integration Architecture, based on deep earning techniques, to facilitate the automated and improved integration of heterogenous datasets.

Candidate Qualifications/Requirements:

- An Honours Degree in Computing with first or second-class (upper division) honours or an equivalent qualification.
- International Students: IELTS 7.0 or equivalent