Thesis Title:

QUERY-DRIVEN LEARNING FOR AUTOMATING EXPLORATORY ANALYTICS IN LARGE-SCALE DATA MANAGEMENT SYSTEMS
Heterogeneous data are initially ingested.
Stored in a distributed storage environment

Structure Data

Unstructured Data

Logs

Data Warehouse

Data Analysis
Moved to a Data Warehouse
Long and iterative procedure
How can we expedite this process?
Main Idea: Take advantage of past queries and their answers (query, answer) to build predictive models.

Query-Driven Learning

Key Advantage: Decouples the data exploration process from data. Perform the same tasks with NO! data access.
Main Idea: Take advantage of past queries and their answers (query, answer) to build predictive models.

Query-Driven Learning

Key Advantage: Decouples the data exploration process from data. Perform the same tasks with NO! data access.
Contributions in this thesis were focused on three pillars
Estimation

Exploration

Exploitation

School of Computing Science
Essence: Pervasive & Distributed Intelligence
Thesis Overview

ML-AQP for Estimation

Explanations for Exploration

Dynamic Data & Workload Adaptation

SuRF for Exploitation
1. How to represent SQL queries as vectors?
2. How to estimate aggregate queries?
3. How to estimate GROUP-BY queries?
4. What kind of Query-Driven Learning models to use?
5. How to perform error estimation?
Explanations for *Exploration*

1. How to explain aggregate queries?
2. How can explanations assist exploration?
3. What kind of models to use?
4. How to identify the best model for each subspace?
5. How can we use explanation functions effectively?
1. How can we automatically pinpoint regions of interest?
2. What’s a baseline algorithm for this task?
3. How can we identify multiple regions effectively?
4. How can Query-Driven Learning help expedite this task?
1. Query-Driven Learning can be used for many tasks. 
   \text{(Estimation, Exploration, Exploitation)}

2. What do we do when underlying data change?

3. What do we do when query workloads change?

4. Introduction of adaptation mechanisms.
Publications


Thank You. Questions?

http://www.dcs.gla.ac.uk/essence/