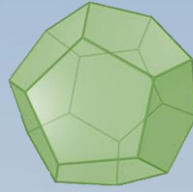




University
of Glasgow



School of Computing Science
Essence: Pervasive &
Distributed Intelligence

Thesis Presentation

Fotis Savva

Supervisors :

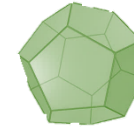
Dr. Christos Anagnostopoulos,

Prof. Peter Triantafillou

**WORLD
CHANGING
GLASGOW**



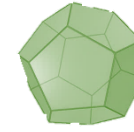
University
of Glasgow



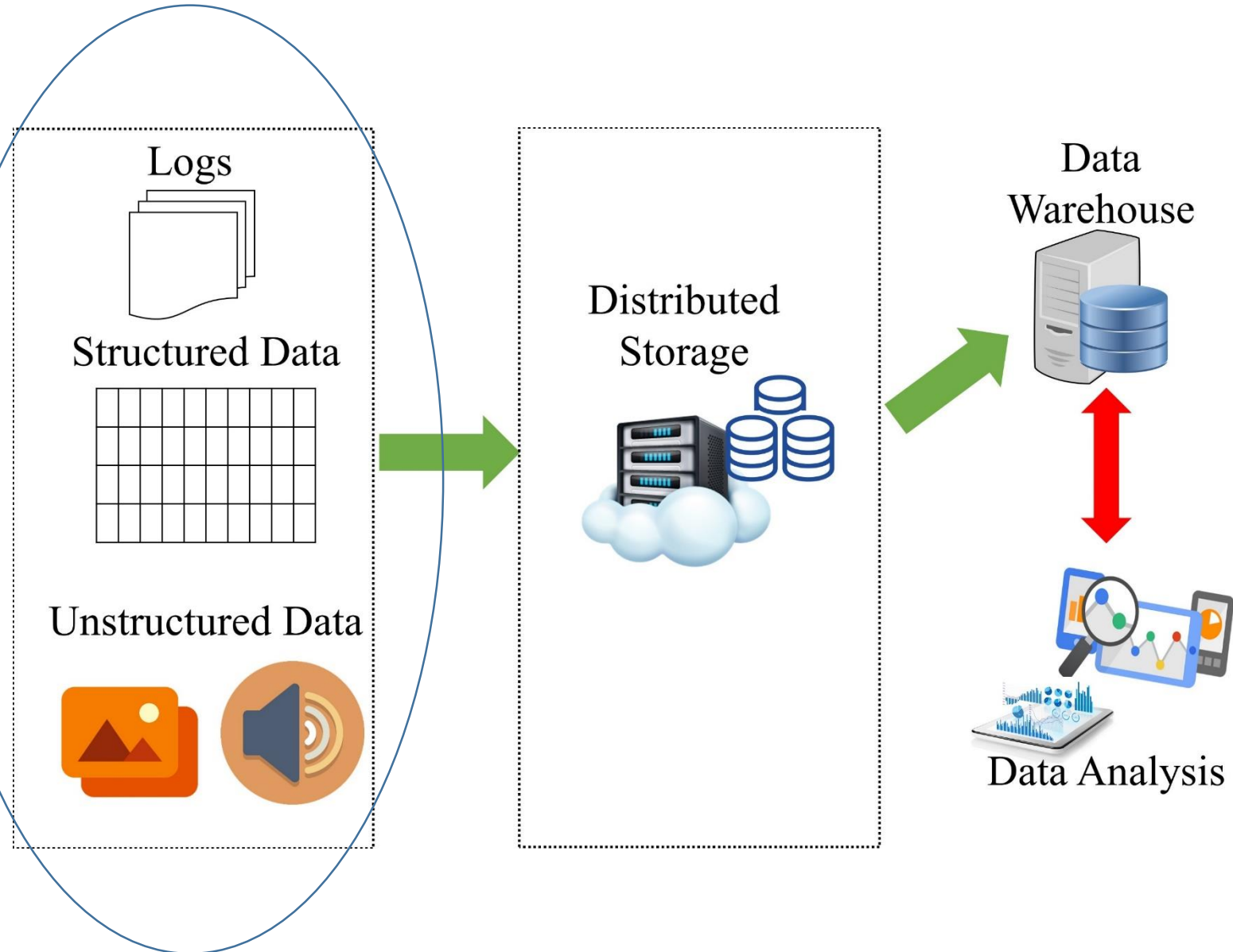
School of Computing Science
Essence: Pervasive &
Distributed Intelligence

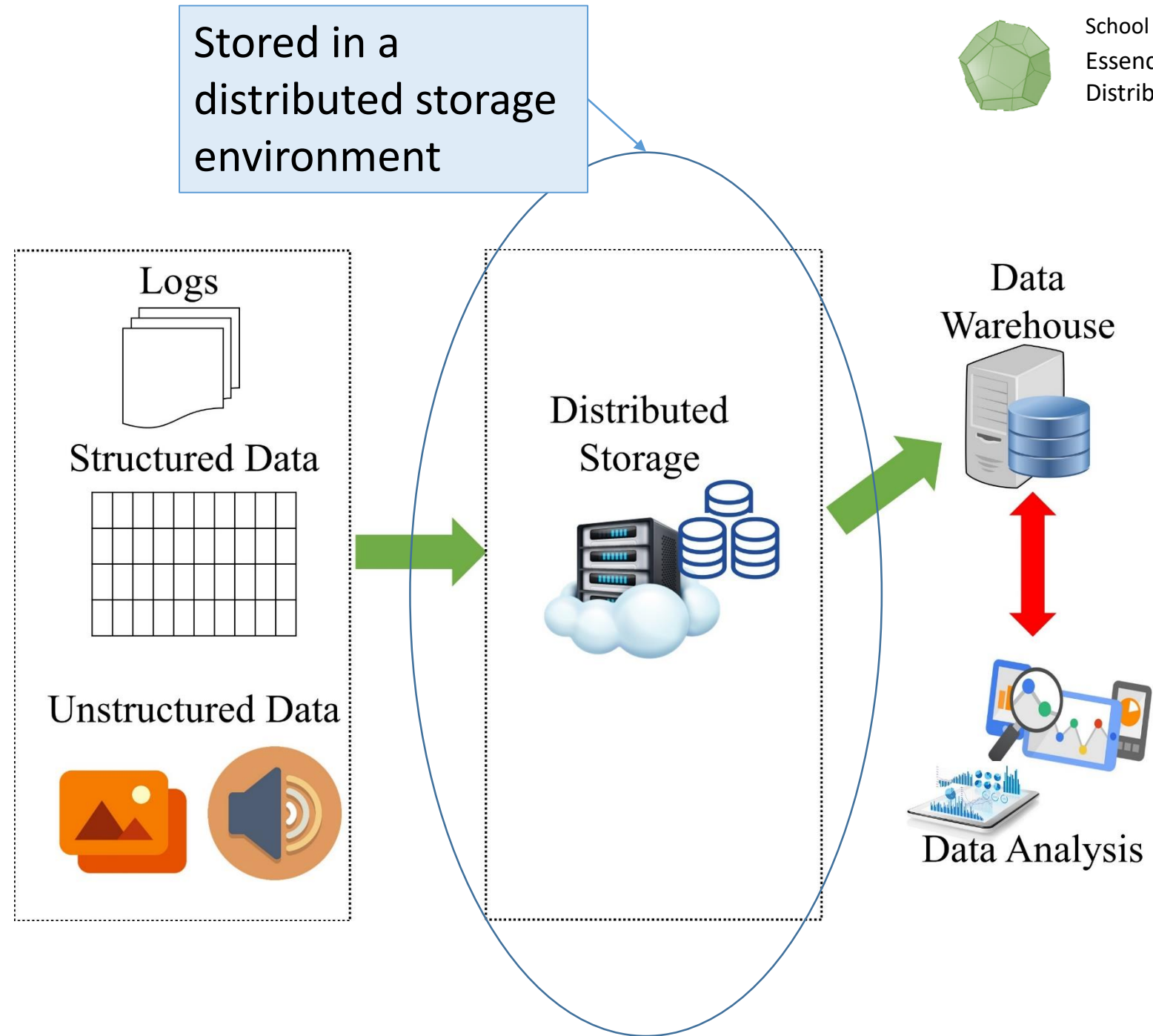
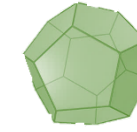
Thesis Title:

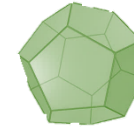
QUERY-DRIVEN LEARNING FOR
AUTOMATING EXPLORATORY
ANALYTICS IN LARGE-SCALE DATA
MANAGEMENT SYSTEMS



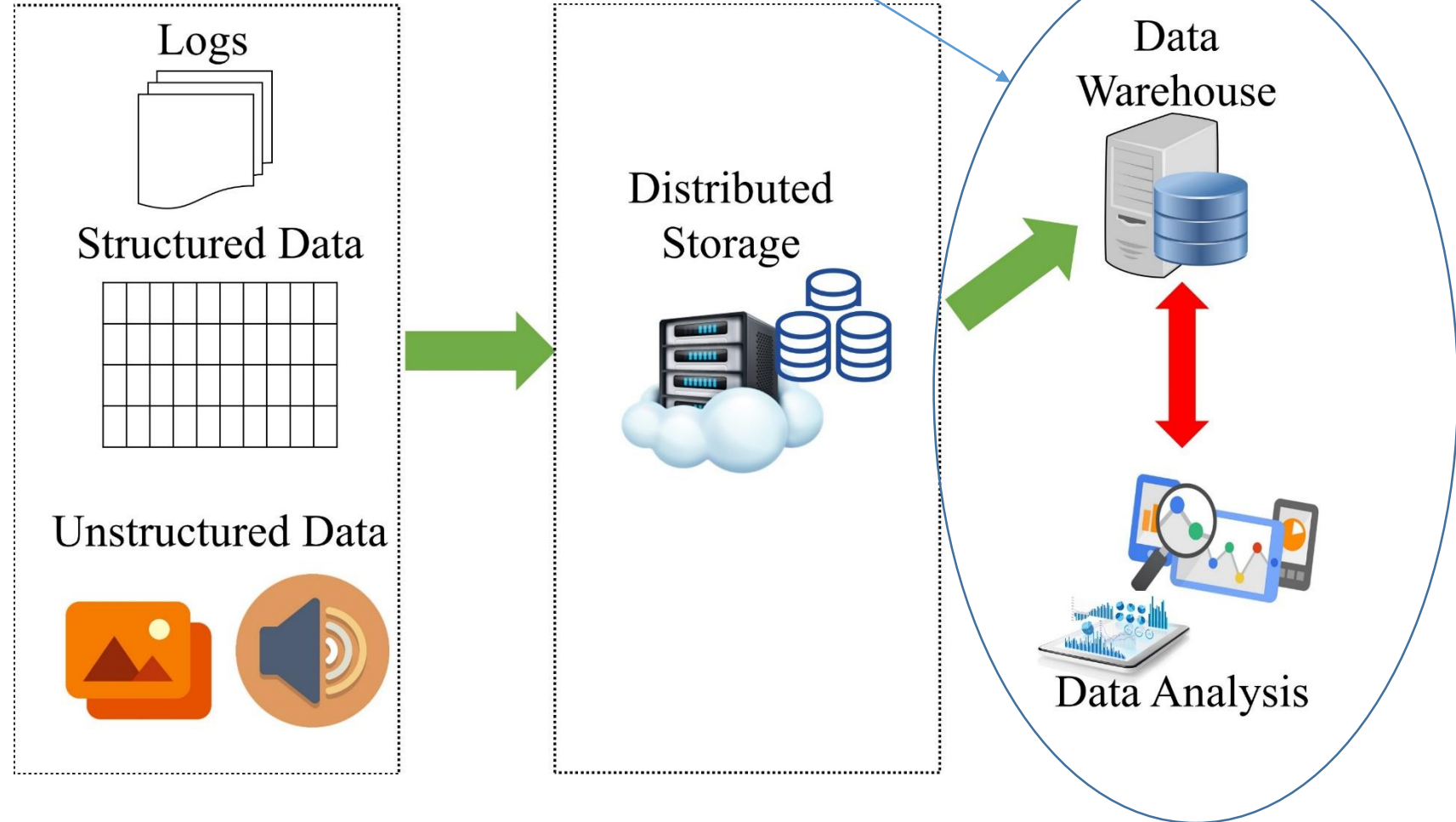
Heterogeneous
data are
initially
ingested





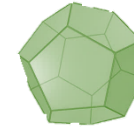


Moved to a Data
Warehouse



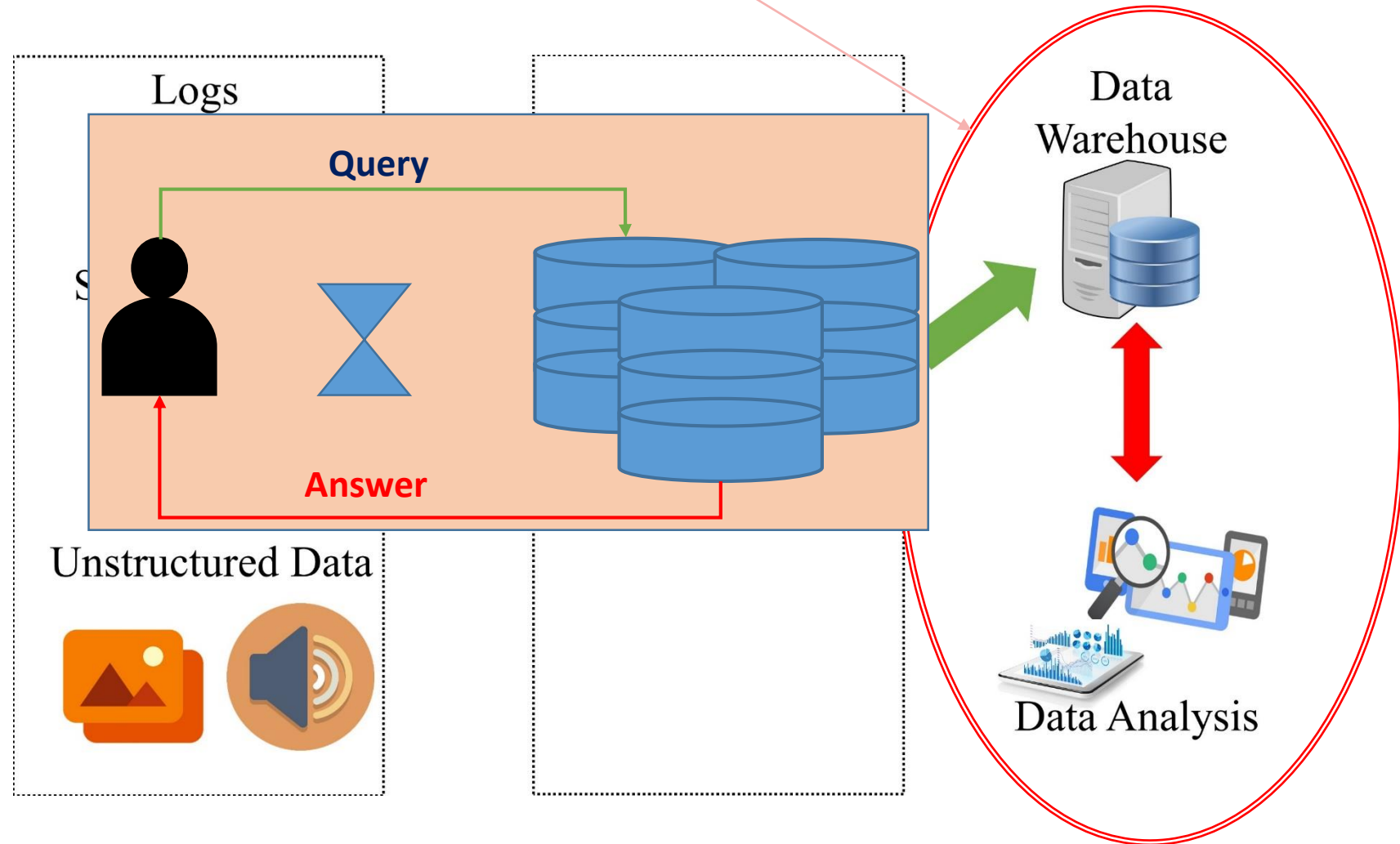


University
of Glasgow



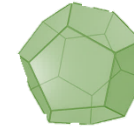
School of Computing Science
Essence: Pervasive &
Distributed Intelligence

Long and iterative
procedure



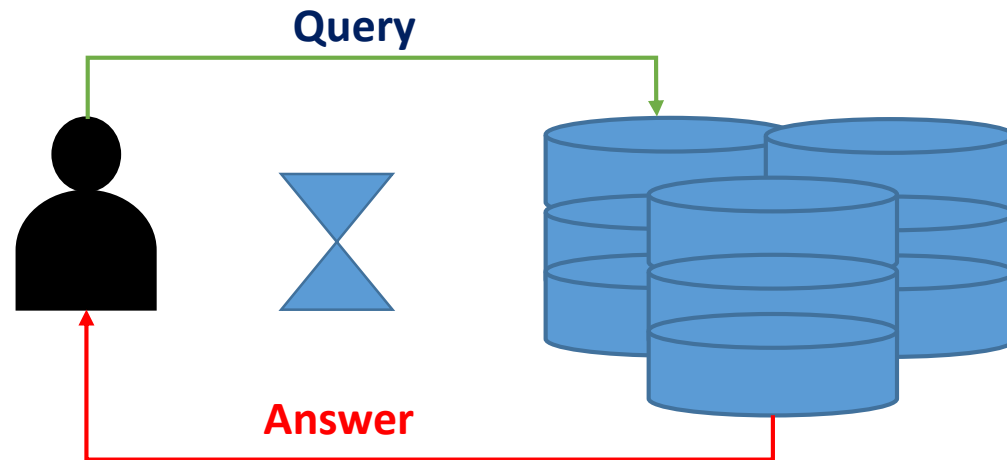
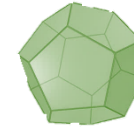


University
of Glasgow



School of Computing Science
Essence: Pervasive &
Distributed Intelligence

How can we expedite this process ?



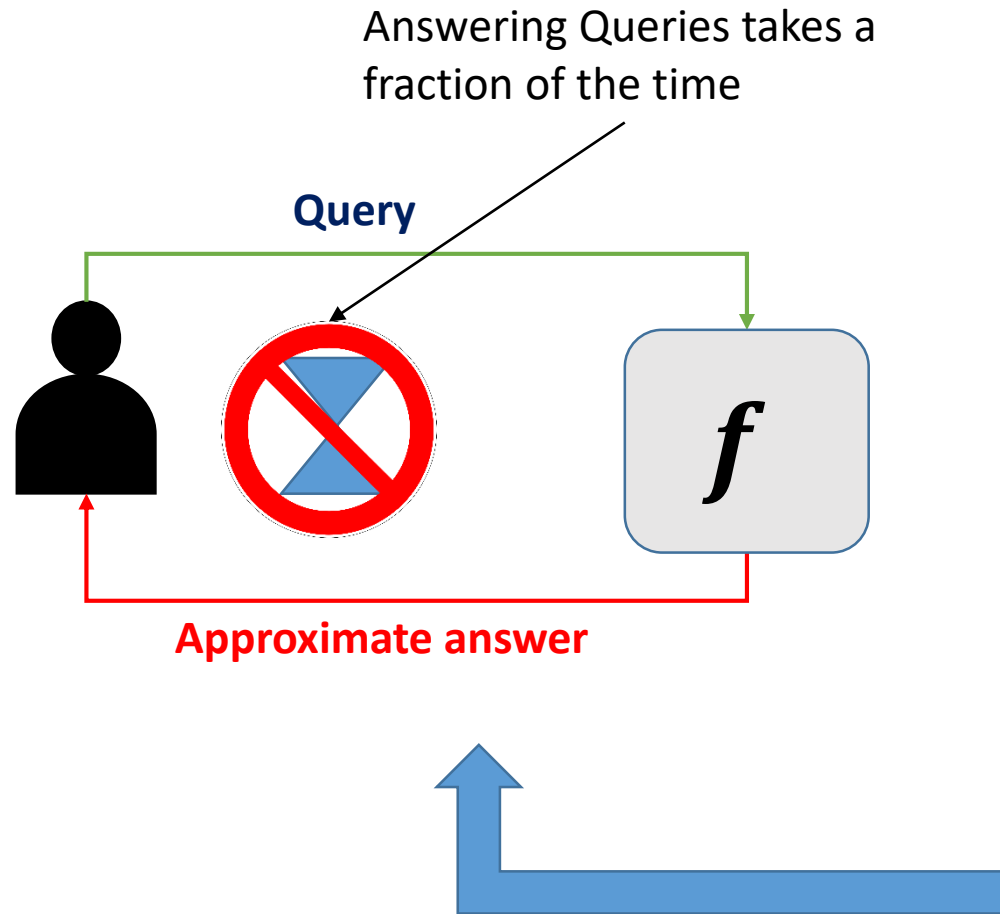
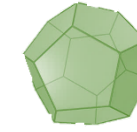
Main Idea : *Take advantage of past queries and their answers (query, answer) to build predictive models.*



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.*



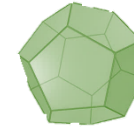
Query-Driven Learning



Key Advantage: *Decouples the data exploration process from data. Perform the same tasks with NO! data access.*

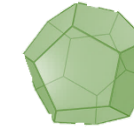


University
of Glasgow

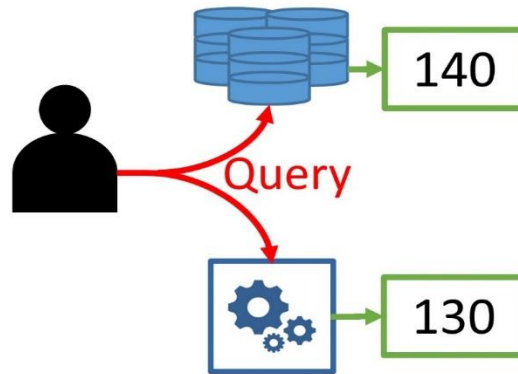


School of Computing Science
Essence: Pervasive &
Distributed Intelligence

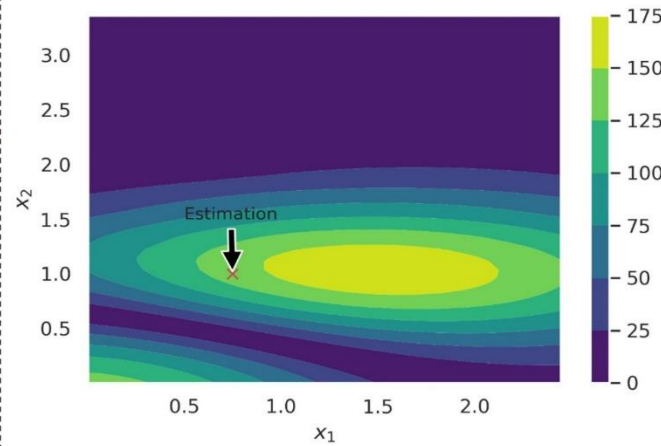
Contributions in this thesis were focused on
three pillars



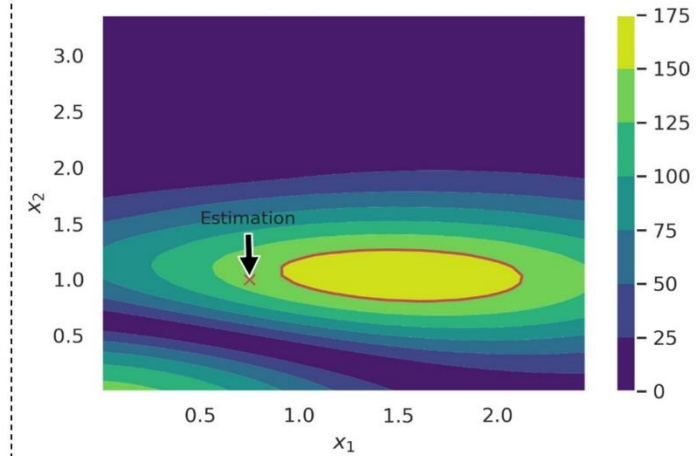
Estimation



Exploration

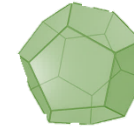


Exploitation





University
of Glasgow



School of Computing Science
Essence: Pervasive &
Distributed Intelligence

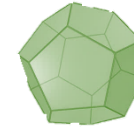
ML-AQP
for
Estimation

Explanations
for *Exploration*

Thesis
Overview

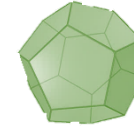
**Dynamic Data
& Workload
Adaptation**

SuRF
for
Exploitation



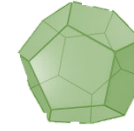
ML-AQP
for
Estimation

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 ?



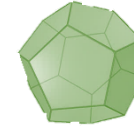
Explanations for *Exploration*

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 ?



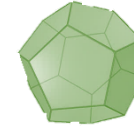
SuRF
for
Exploitation

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 ?



Dynamic Data & Workload Adaptation

1. Query-Driven Learning can be used for many tasks.
(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.

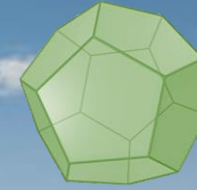


Publications

- Savva, Fotis, Christos Anagnostopoulos, and Peter Triantafillou. "ML-AQP: Query-Driven Approximate Query Processing based on Machine Learning." arXiv preprint arXiv:2003.06613 (2020)
- Savva, Fotis, Christos Anagnostopoulos, and Peter Triantafillou. "SuRF: identification of interesting data regions with surrogate models." **2020 IEEE 36th International Conference on Data Engineering (ICDE)**. IEEE, 2020.
- Savva, Fotis, Christos Anagnostopoulos, and Peter Triantafillou. "Adaptive learning of aggregate analytics under dynamic workloads." **Future Generation Computer Systems** (2020).
- Savva, Fotis, Christos Anagnostopoulos, Peter Triantafillou, and Kostas Kolomvatsos. "Large-scale data exploration using explanatory regression functions." **ACM Transactions on Knowledge Discovery from Data (2020)**.
- Savva, Fotis. "Query-Driven Learning for Next Generation Predictive Modelling & Analytics." Proceedings of the **2019 International Conference on Management of Data. (SIGMOD SRC)** 2019. Recipient of 2nd place.
- Savva, Fotis, Christos Anagnostopoulos, and Peter Triantafillou. "Aggregate query prediction under dynamic workloads." **2019 IEEE International Conference on Big Data (Big Data)**. IEEE, 2019.
- Savva, Fotis, Christos Anagnostopoulos, and Peter Triantafillou. "Explaining aggregates for exploratory analytics." **2018 IEEE International Conference on Big Data (Big Data)**. IEEE, 2018.



University
of Glasgow



School of Computing Science
Essence: Pervasive &
Distributed Intelligence

Thank You. Questions ?

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