

Query-Driven Learning for Next Generation Predictive Modelling & Analytics

Fotis Savva

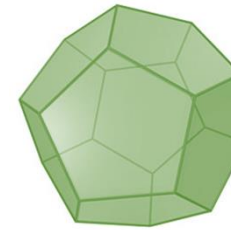
University of Glasgow, School of Computing Science

Essence: Pervasive & Distributed Intelligence

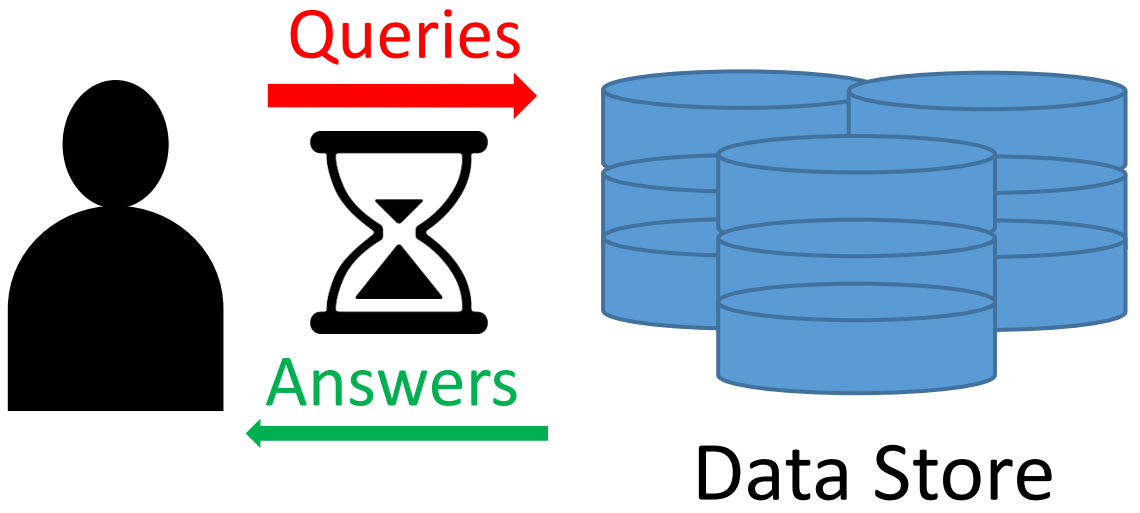
f.savva.1@research.gla.ac.uk

02/07/2019

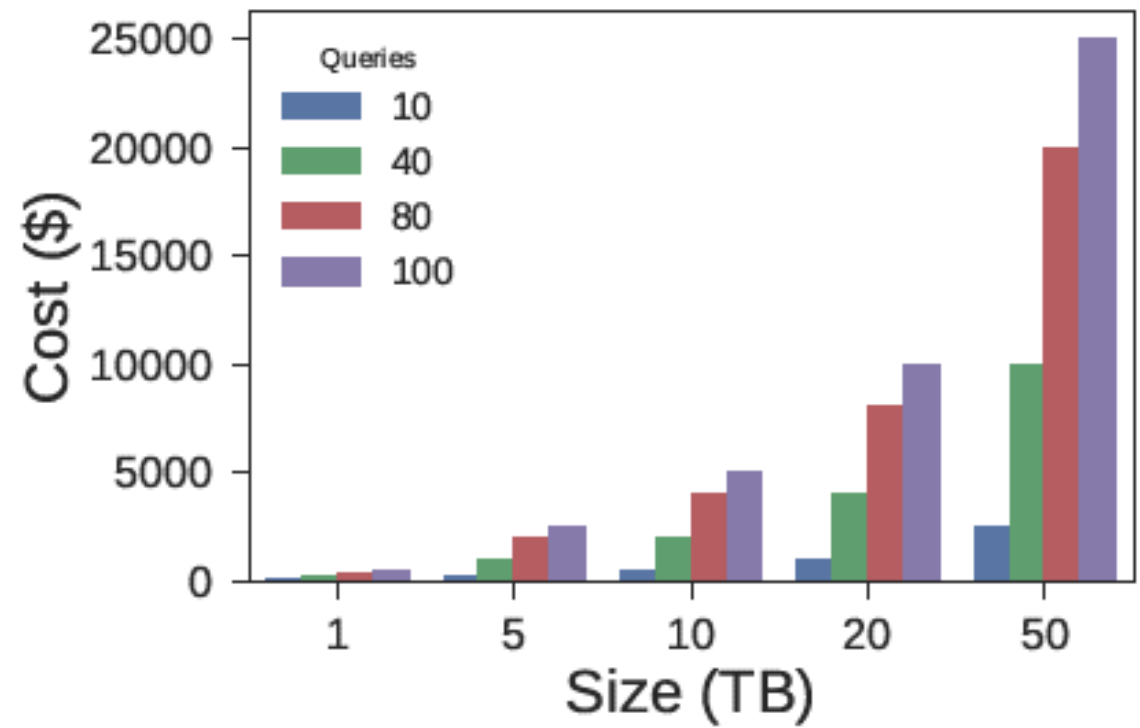
<http://www.dcs.gla.ac.uk/essence/>



Exponential Data = Exponential Cost

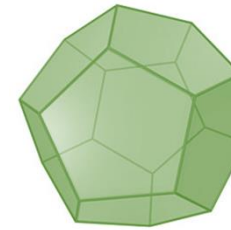


Monetary Cost



Computational Cost

Increased waiting times affect productivity. Interactivity constraint 500ms – 2s [1]



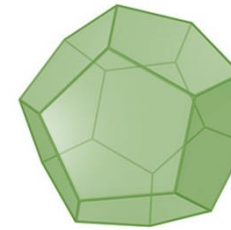
Approximate Query Processing (AQP)

- Provides **approximate** answers at a fraction of the time



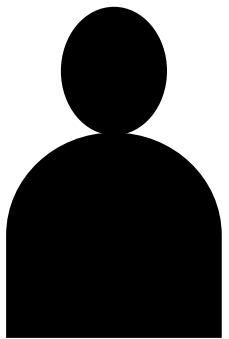
Observations:

- Make use of samples; still require storage
- Trade-off accuracy – sampling ratio
- Can break interactivity constraint
- Make use of same infrastructure



Query-Driven Learning (QDL)

- Use past queries and train Machine Learning models to predict answers



Queries



Predicted
Answers



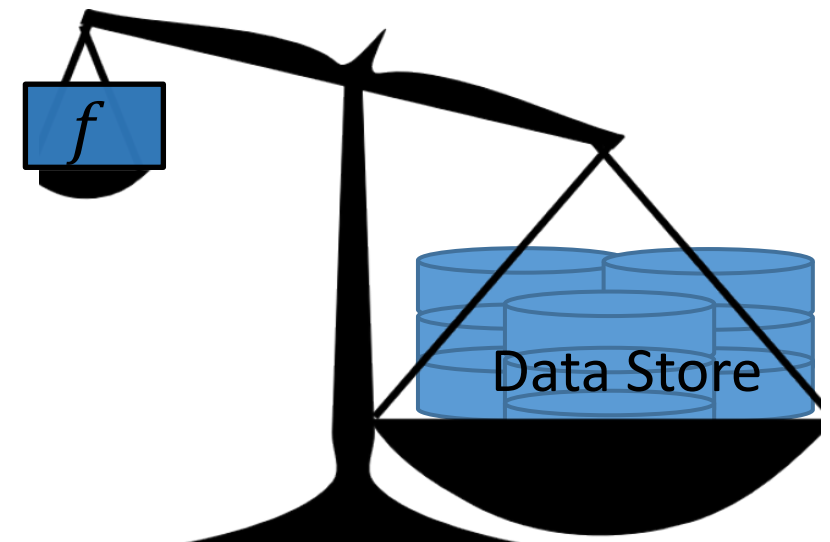
f

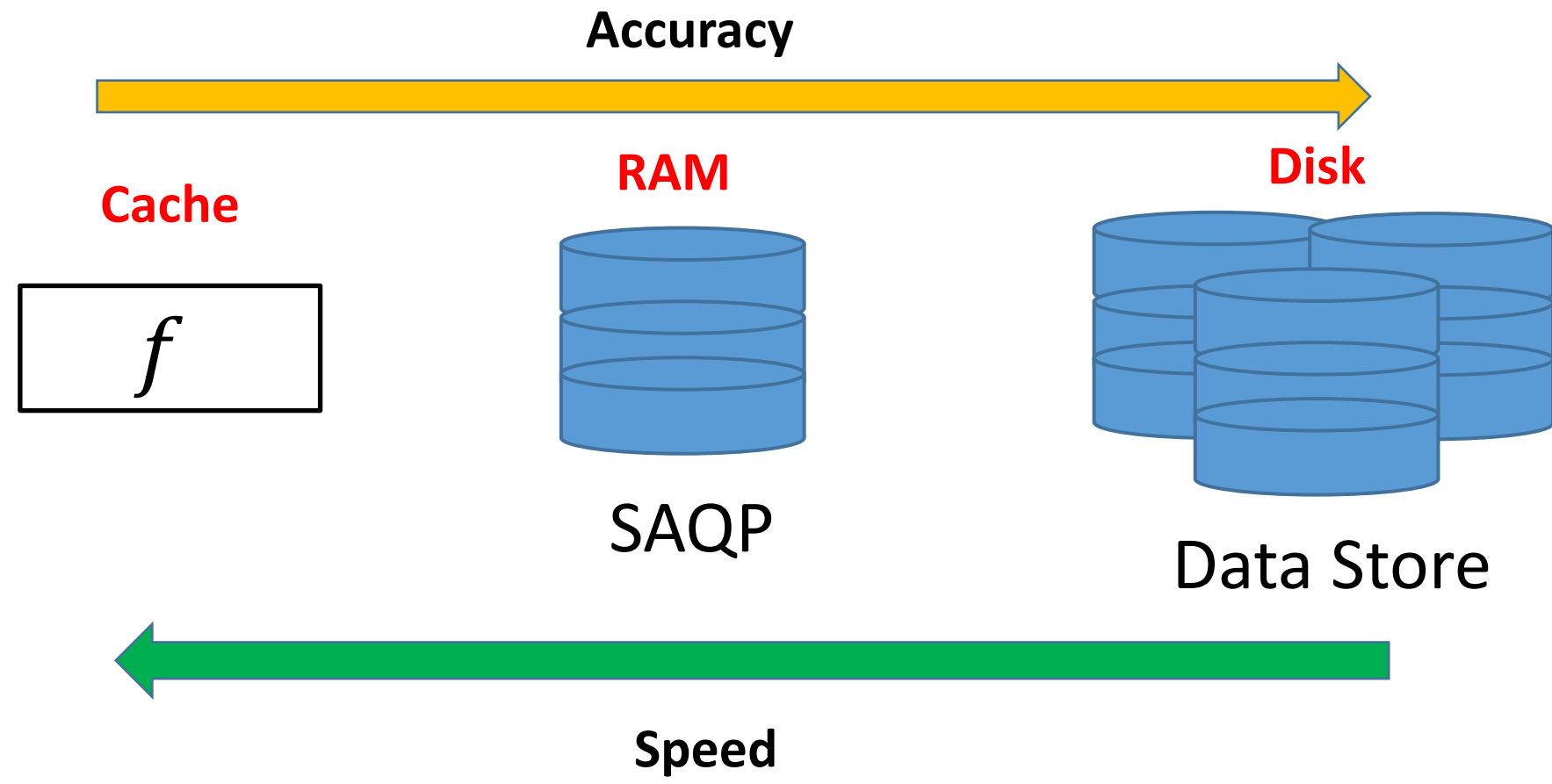
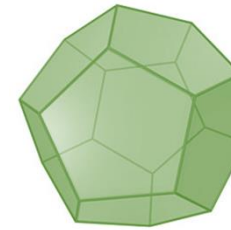
Relative
Error :
3%

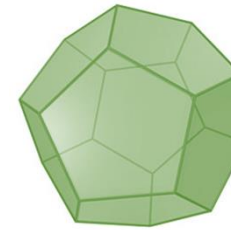
0.0001
seconds

Objectives :

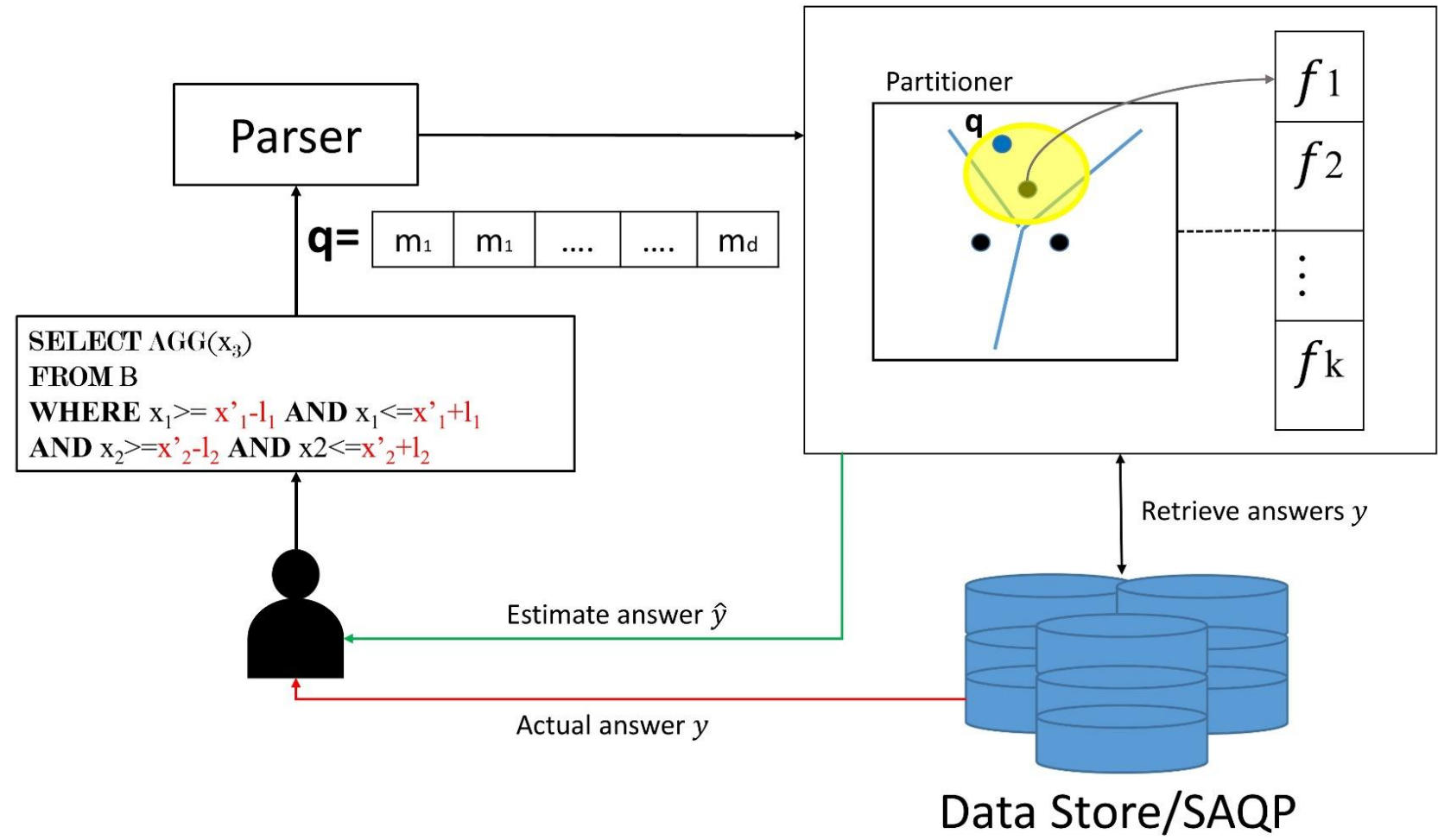
- Make **NO** use of data (Save Money)
- Accurate
- Efficient (Save Time)
- Lightweight
- Data Store Agnostic

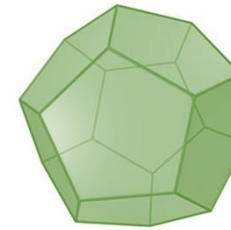




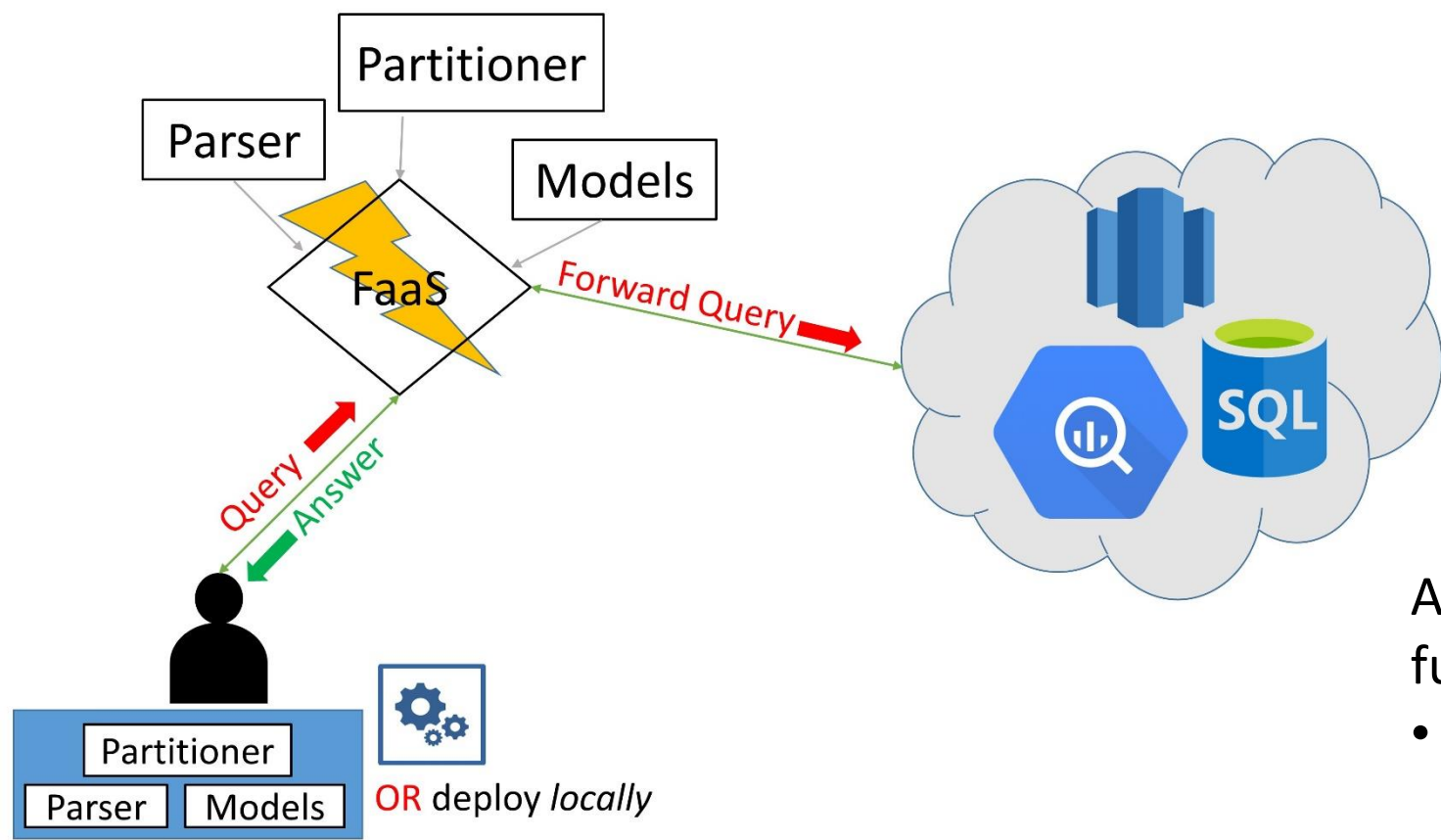


System Overview





Possible Deployment



BigQuery Pricing :

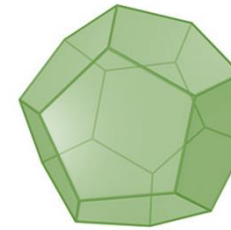
- On-demand : \$5 per TB

Amazon Redshift :

- On-demand : \$0.25 per hour (cheapest)
 - Spectrum : \$5 per Terabyte
- Both also offer **flat rate** options.

Analytics using **QDL** and Cloud functions (Function as a Service - **FaaS**)

- Google Cloud Functions : \$0.4 per **1 million** function calls (=queries)
- AWS Lambda : \$0.2 per **1 million**



Thank you for your attention

Essence: Pervasive & Distributed Intelligence

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