

Query-Driven Learning for Next Generation Analytics

Fotis Savva

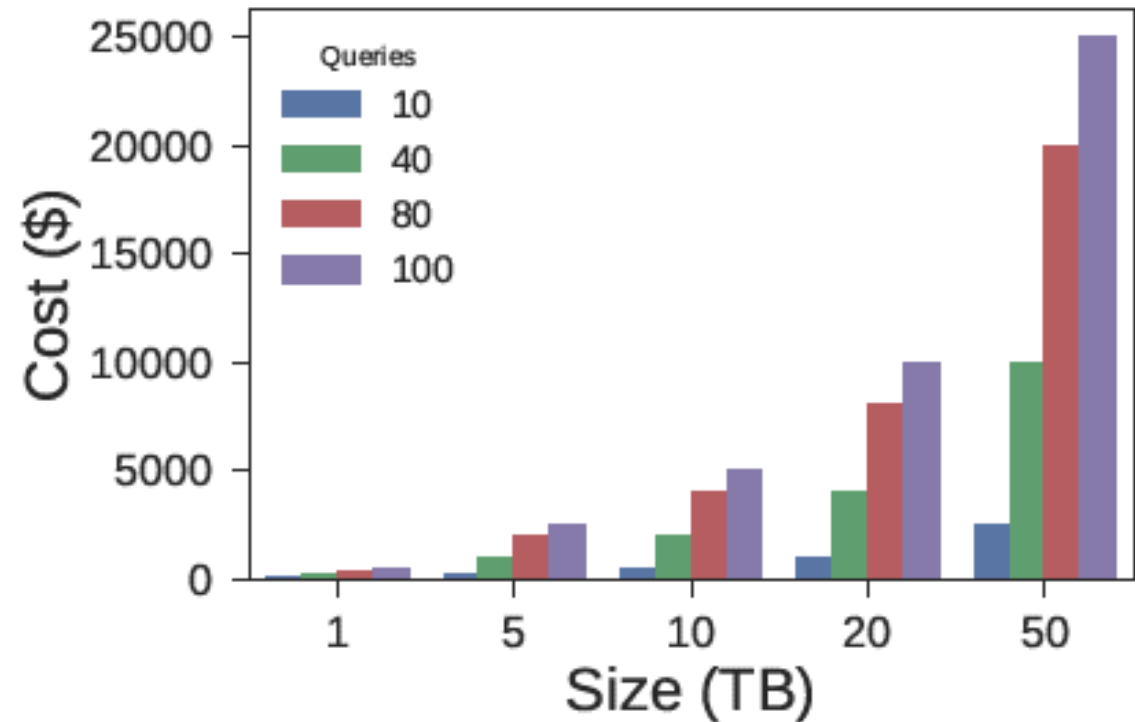
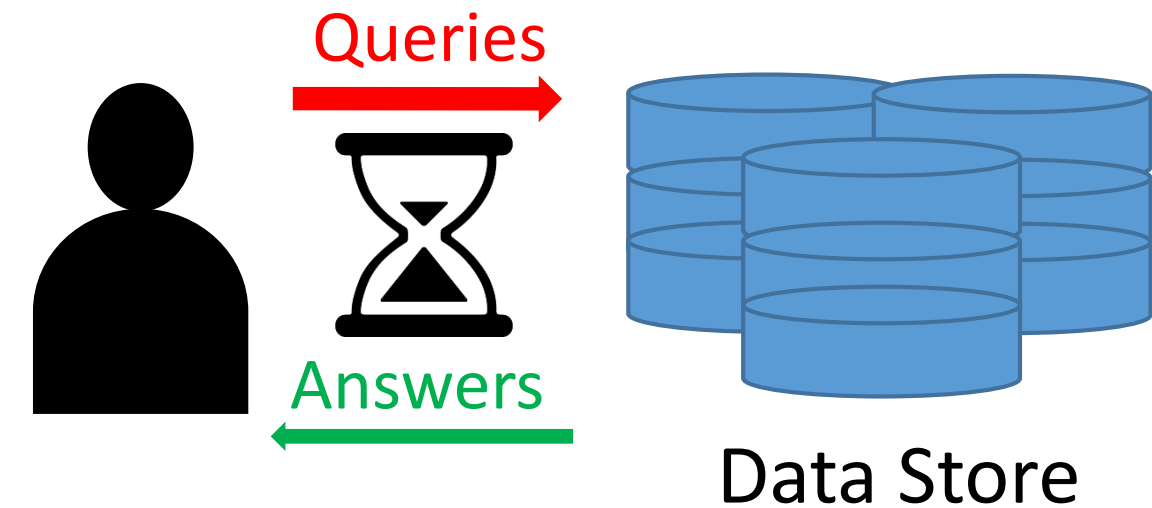
University of Glasgow

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

02/07/2019

Exponential Data = Exponential Cost

Monetary Cost



Computational Cost

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

Finding solutions which address both costs is crucial

Approximate Query Processing (AQP)

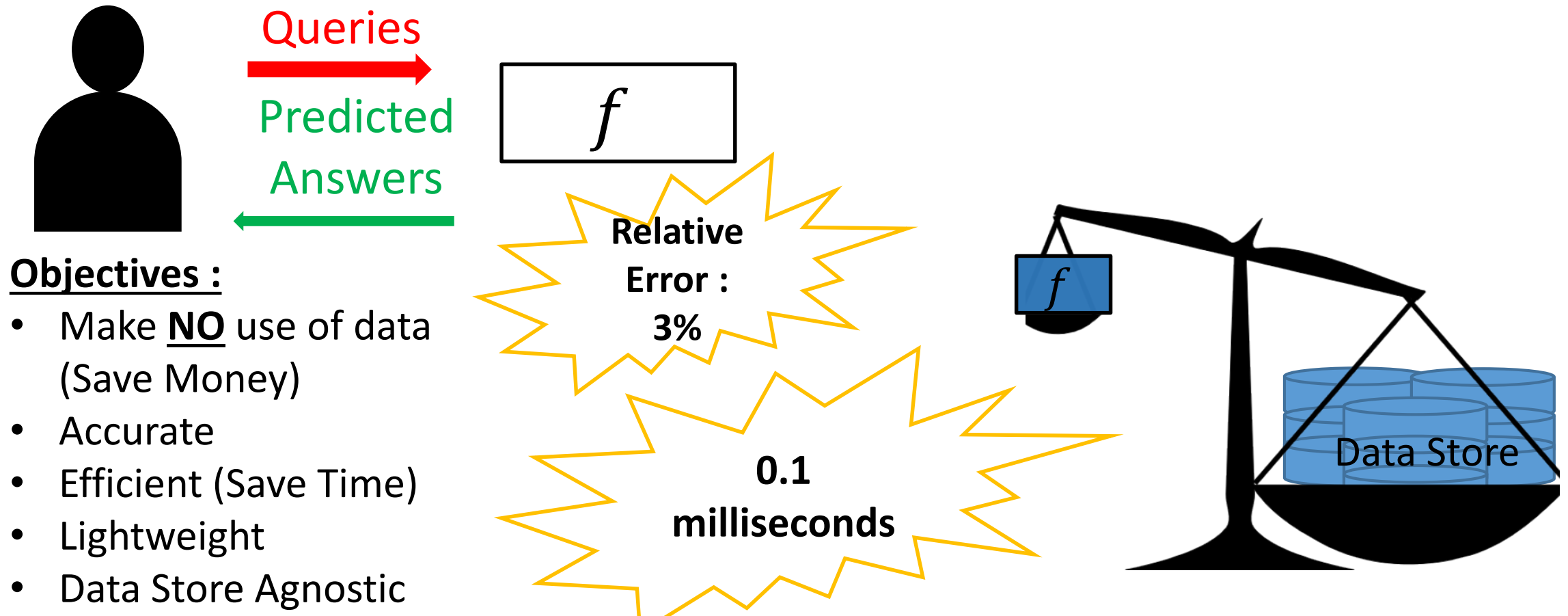
- Provides **approximate** answers at a fraction of the time needed for the computation of the actual answer



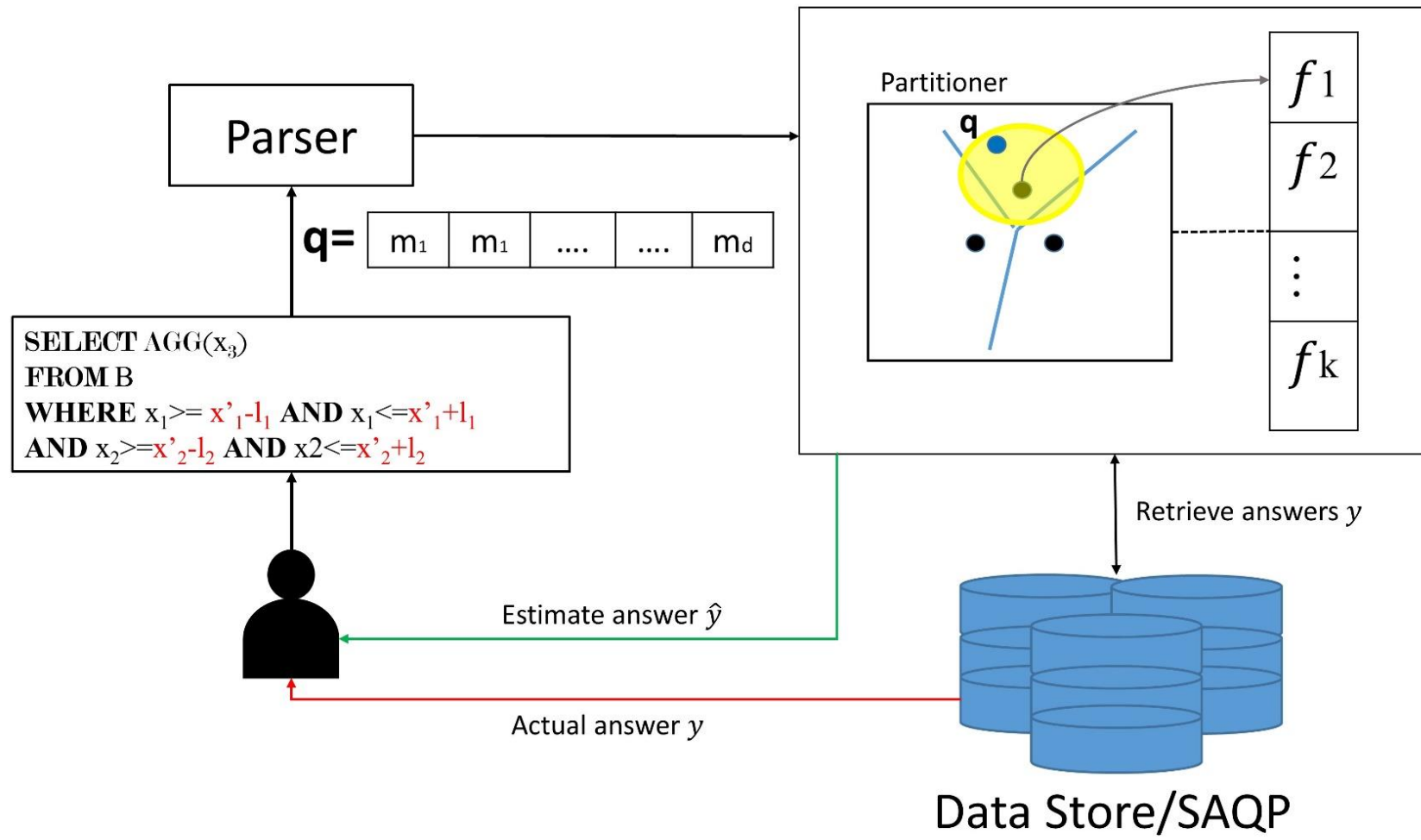
- Make use of samples; still require storage
- Trade-off accuracy – sampling ratio
- Can break interactivity constraint
- Make use of same infrastructure

Query-Driven Learning (QDL)

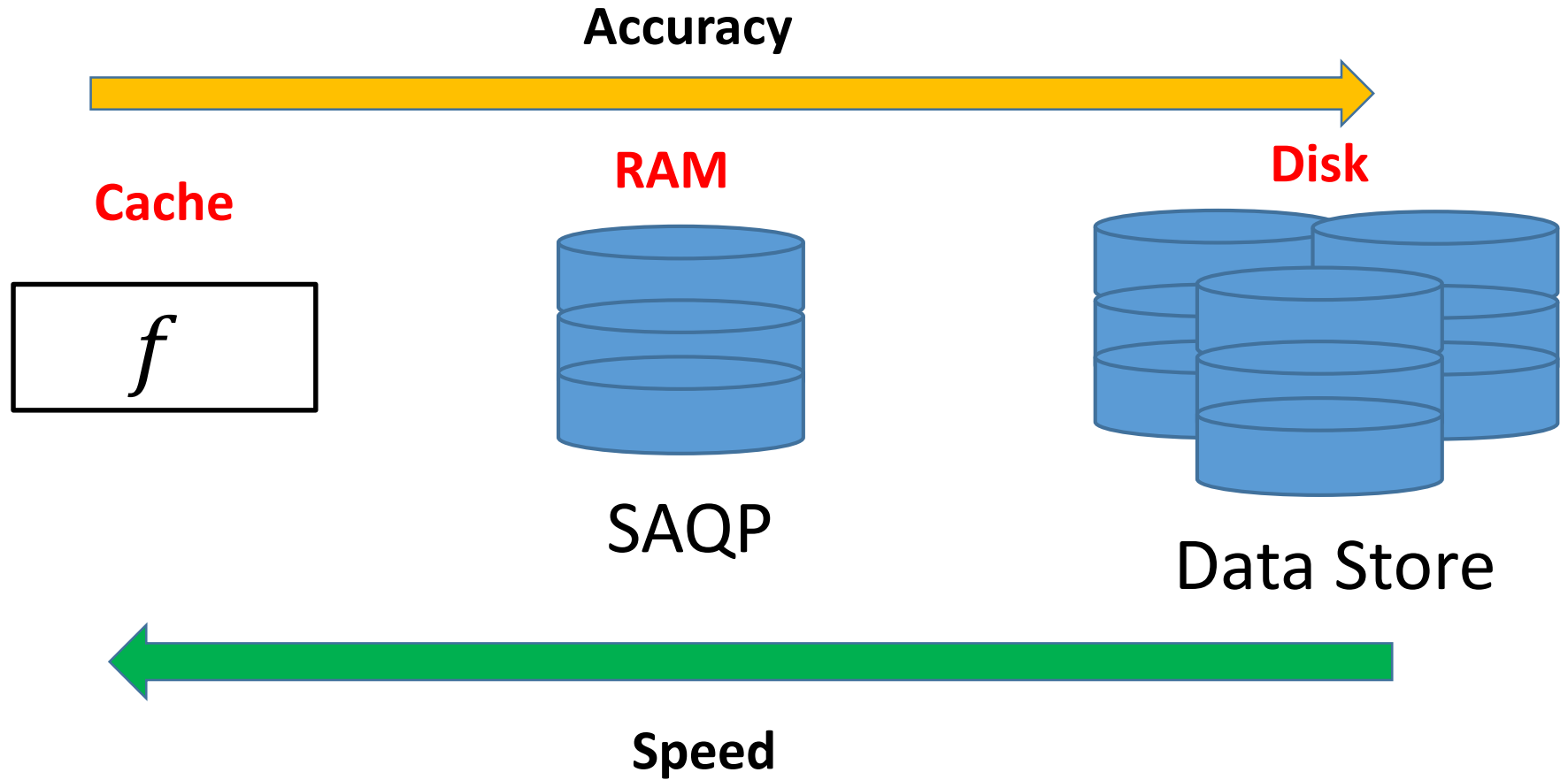
- Use past queries and train Machine Learning models to predict answers



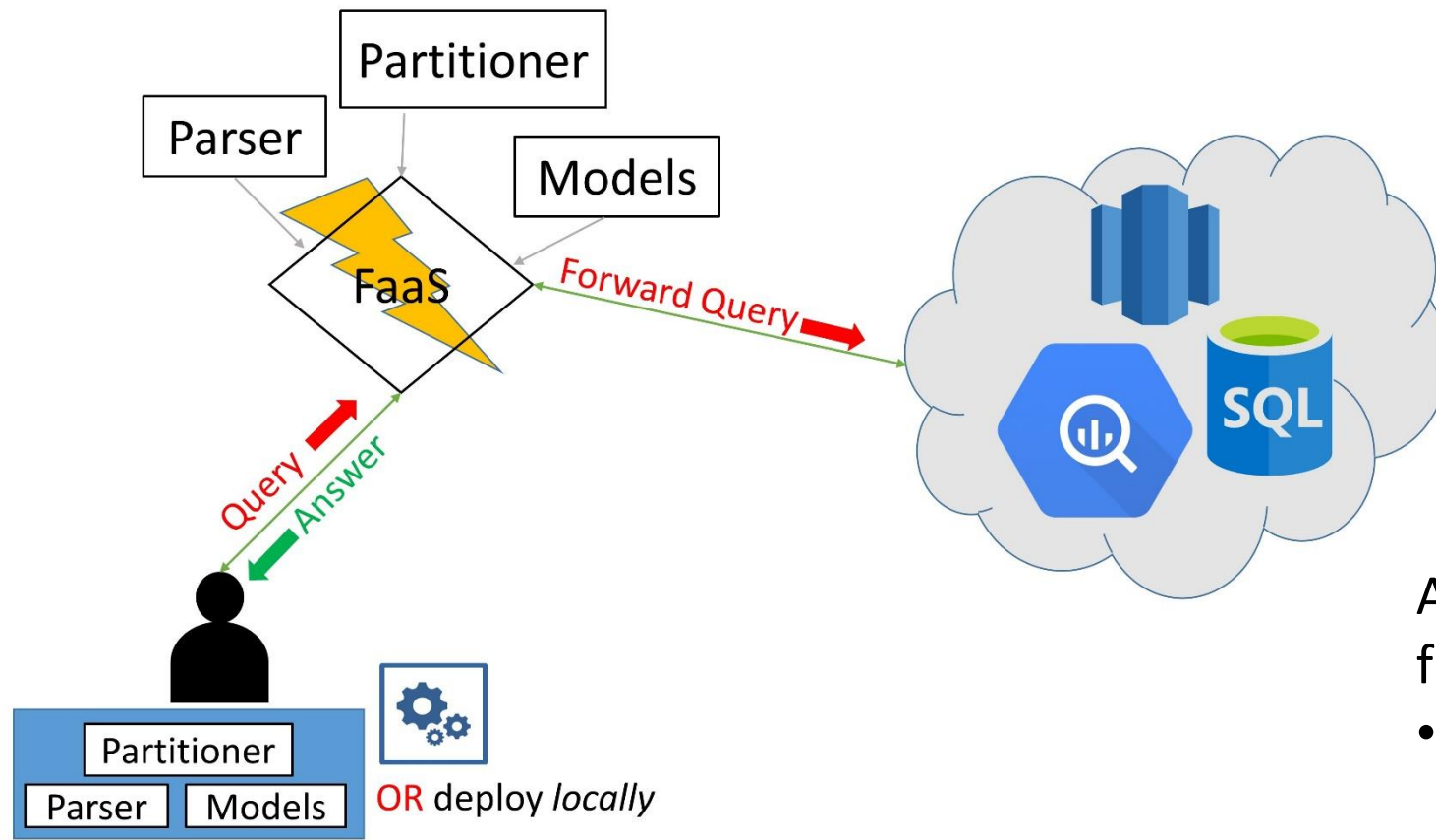
System Overview



Analytics Stack



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