

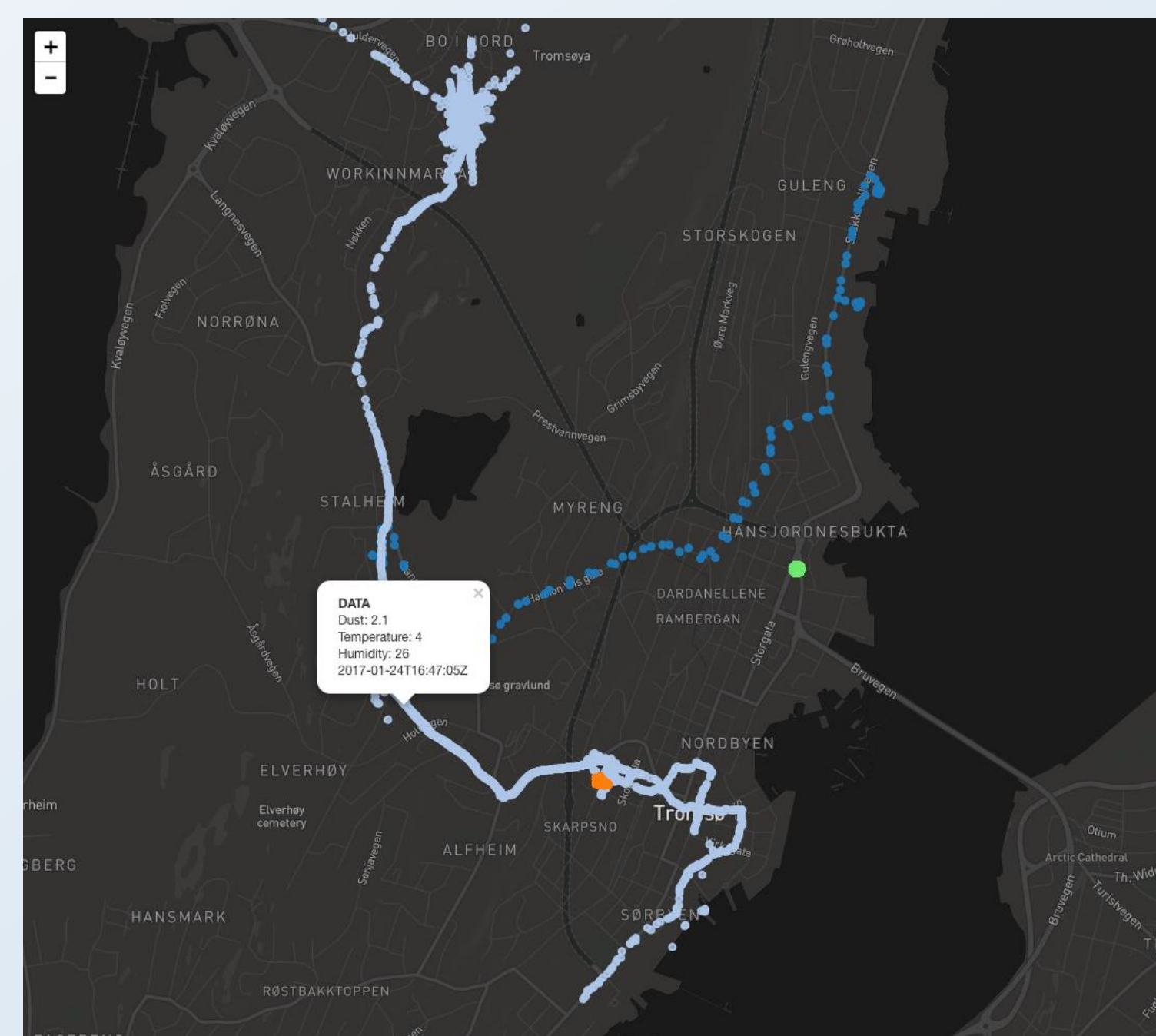
Air pollution data analysis platform for computer science education projects

Nina Angelvik
nan016@post.uit.no
Department of Computer Science, UiT – The Arctic University of Norway



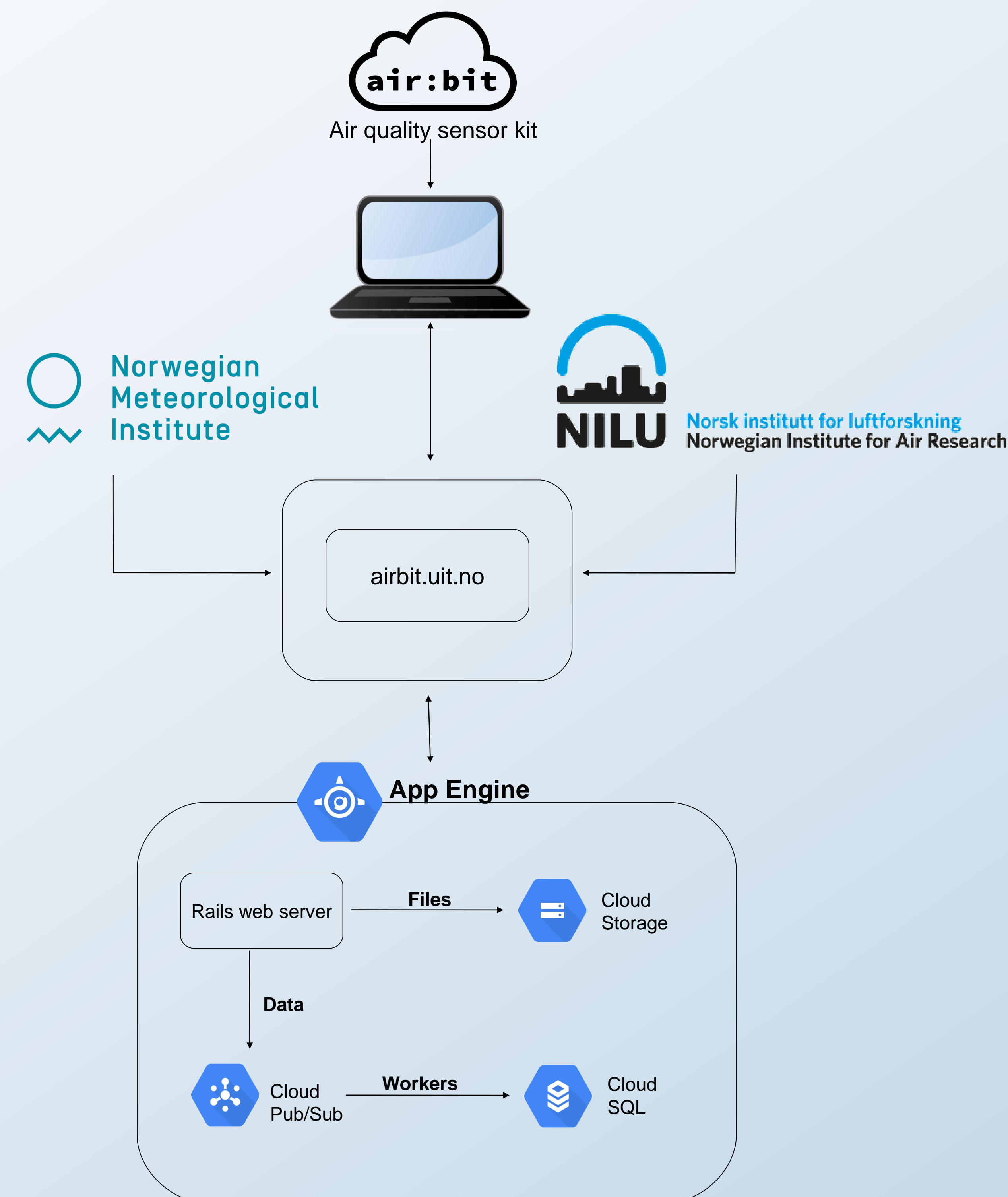
Students build and code their own air quality sensor kits

The air:bit platform is used in science education projects where upper secondary school students build and code their own air quality sensor kits, the air:bit, before investigating a research question by analyzing their collected data. The air:bit periodically registers time, location, temperature, humidity and dust particles, and stores the data as a CSV file on an on-board memory card.



Web interface for analysis of localized air quality patterns

We have built a web application where the students upload, download, query and visualize their data. It also integrates the student data with climate data from external sources, such as The Norwegian Institute for Air Research and the Norwegian Meteorological Institute.



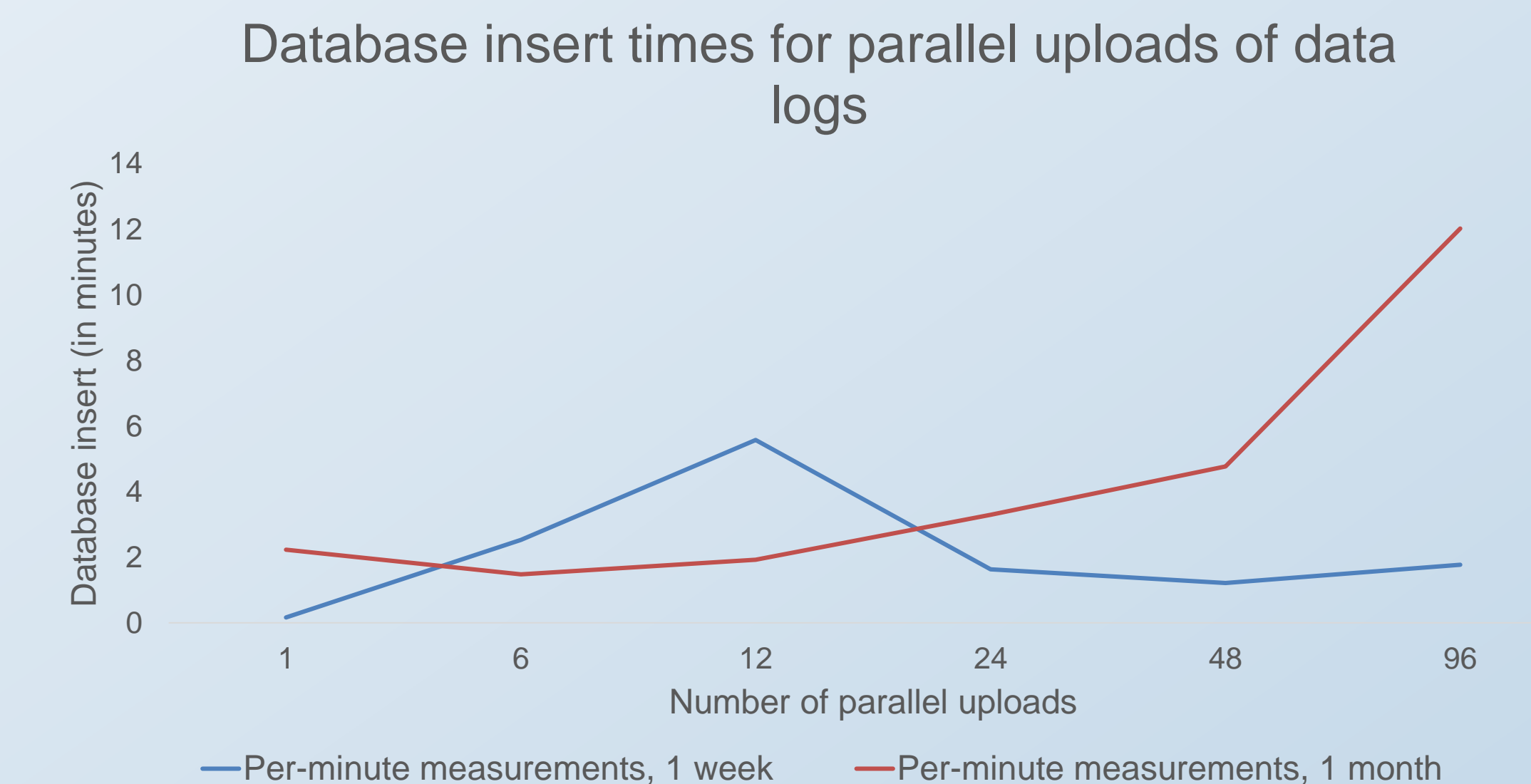
Scalable cloud based data analysis platform

The uploaded data files and web application queries are sent to our backend using HTTP REST API. The backend is deployed to the App Engine Flexible Environment, a Platform as a Service delivered by the Google Cloud Platform (GCP). It is implemented in Ruby on Rails and consists of multiple GCP services, including Google Cloud Storage, Google Cloud Pub/Sub and Google Cloud SQL. Upon receipt, the file is stored as an entity in Cloud Storage, before its content is enqueued in the Cloud Pub/Sub task queue. In an asynchronous manner, the queue is polled and the data is parsed and inserted into the Cloud SQL database. The client will receive a response message once the file has been successfully/unsuccessfully added to the queue, while query results are returned to the web application as JSON objects.

Visit our platform at:
<http://airbit.uit.no>

ACKNOWLEDGEMENTS

A big thank you to the Department of Computer Science at UiT – the Arctic University of Norway and the Centre for Science Education at UiT – The Arctic University of Norway for funding this poster presentation.



Preliminary performance results and cost estimates

The App Engine Flexible Environment automatically scales the platform up and down based on a configured CPU target. To test the scalability of the platform we have measured the data insert time for uploading a variable number of data files containing 1 week of per-minute measurements (10080 measurements, 717kB) and 1 month of per-minute measurements (44640 measurements, 3.2 MB). Initial performance results show that the average insert time for up to 96 parallel uploads is less than 6 minutes for files containing 1 week of data and 14 minutes for files containing 1 month of data. This is acceptable for our usage.

Our platform is deployed to virtual machines that are billed on a per-second basis. The main cost are the three vCPUs for the web server, a background worker and the database. The cost of the three vCPUs is ~\$115/month. The extra cost of handling bursts in uploads using additional vCPUs is low. For example the cost of using 15 vCPUs for 96 3.2 MB file uploads is \$0.25. Including the cost of memory, persistent disk and the occasional use of the extra vCPUs, the total cost is about \$135/month.



11 schools in Northern Norway are using the platform

During spring 2018, 11 school classes in Northern Norway are using the platform to study air quality patterns in their local environment.

