introduction

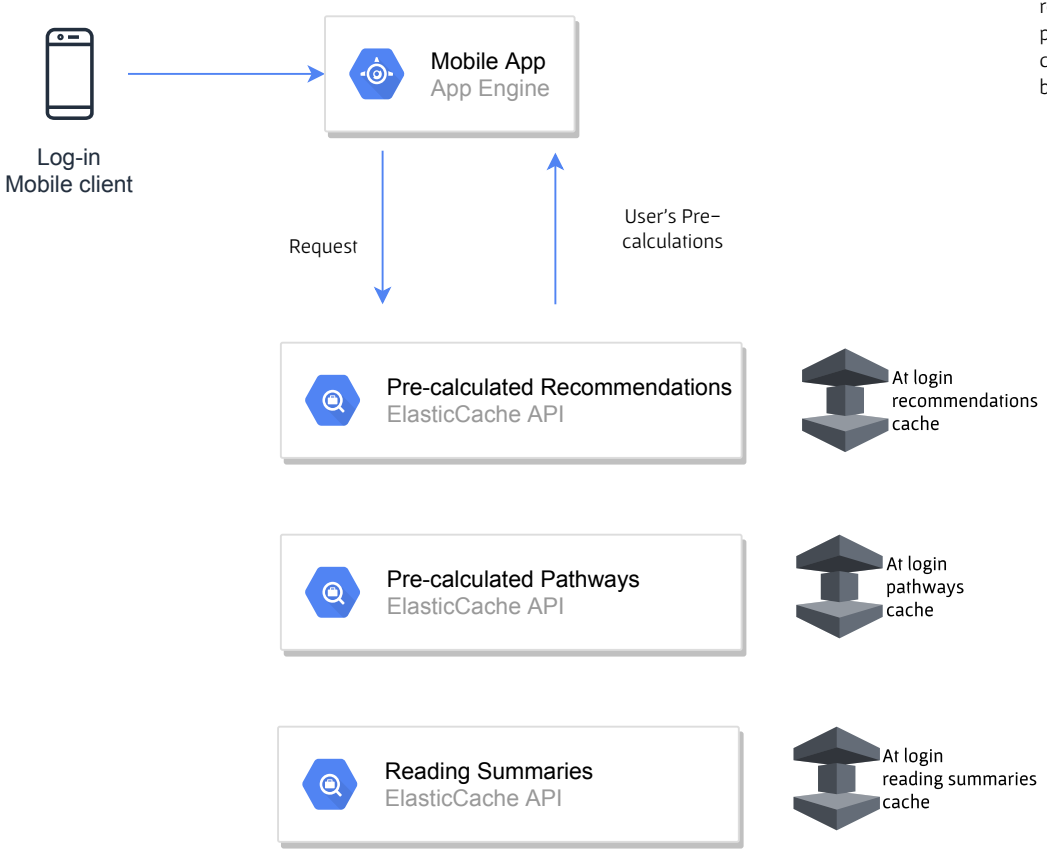
These pages outline the machine learning models that would be developed in order to proactively aid reading. It also outlines the services thereof. In order to have a good cost estimate, the underlying daily analytics and features engineering steps have also been outlined.

- Models Development: the models to be developed, in the case of clustering, further development
- Models Services: The latest version of a model is used to provide model services to users, e.g., personalised searches.
- Core Analytics & Core Features Engineering: this summarises the myriad of daily calculations required.
- Ecosystems: the pipeline ecosystem illustrates the rather dizzying array of technologies that a data pipeline might use. It illustrates the intricacies of machine learning systems. Please, do read Hidden Technical Debt in Machine Learning (https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf)

Models Services	Models Development	Core Analytics & Core Features Engineering	Ecosystems
At Login	recommendation model development	daily executions	pipeline ecosystem of tools
in-context recommendations	pathways determination & pathway assignations models		graphs & graphing engines
real-time personalised search	real-time personalised search model development		
	reading profiles clusters		

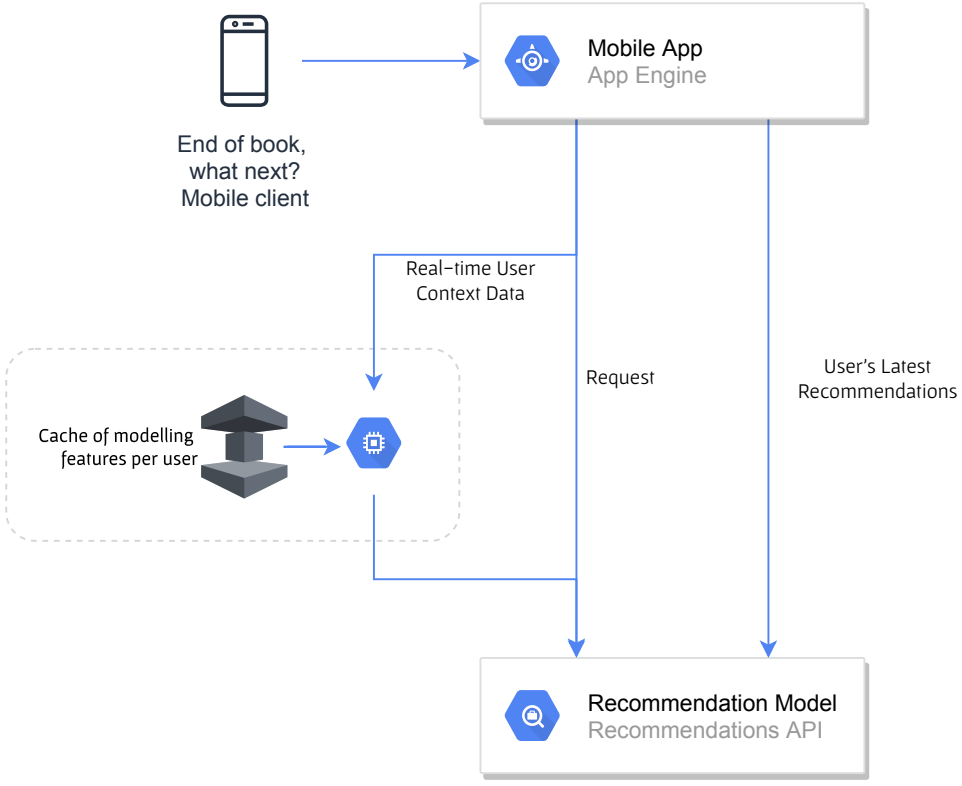
logging-in

After logging-in one of the options offered is a screen section of pre-determined recommendations. The recommendations are from the latest nightly batch predictions of recommendations per user. The same concept applies to pathways, and a user's summary of book reading thus far.

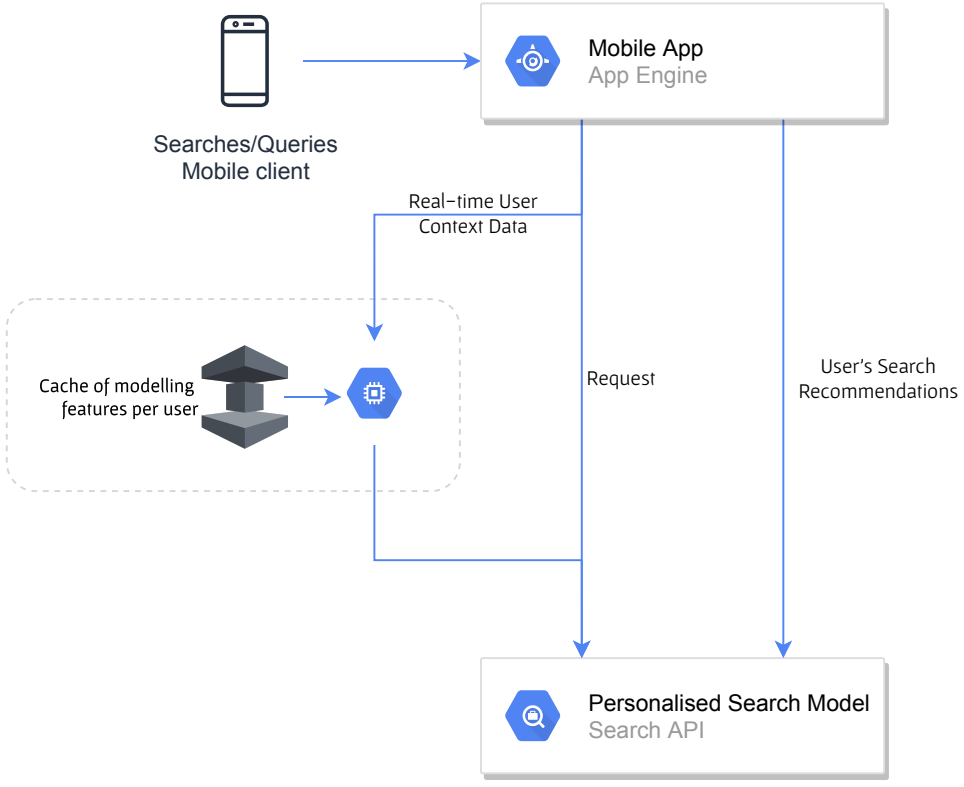


in-context recommendations

A user has just finished reading a book. Or, a user has decided to exit a book that was being read, but hasn't been finished. (If time and cost permit, the pathways option could be tried within this project.)



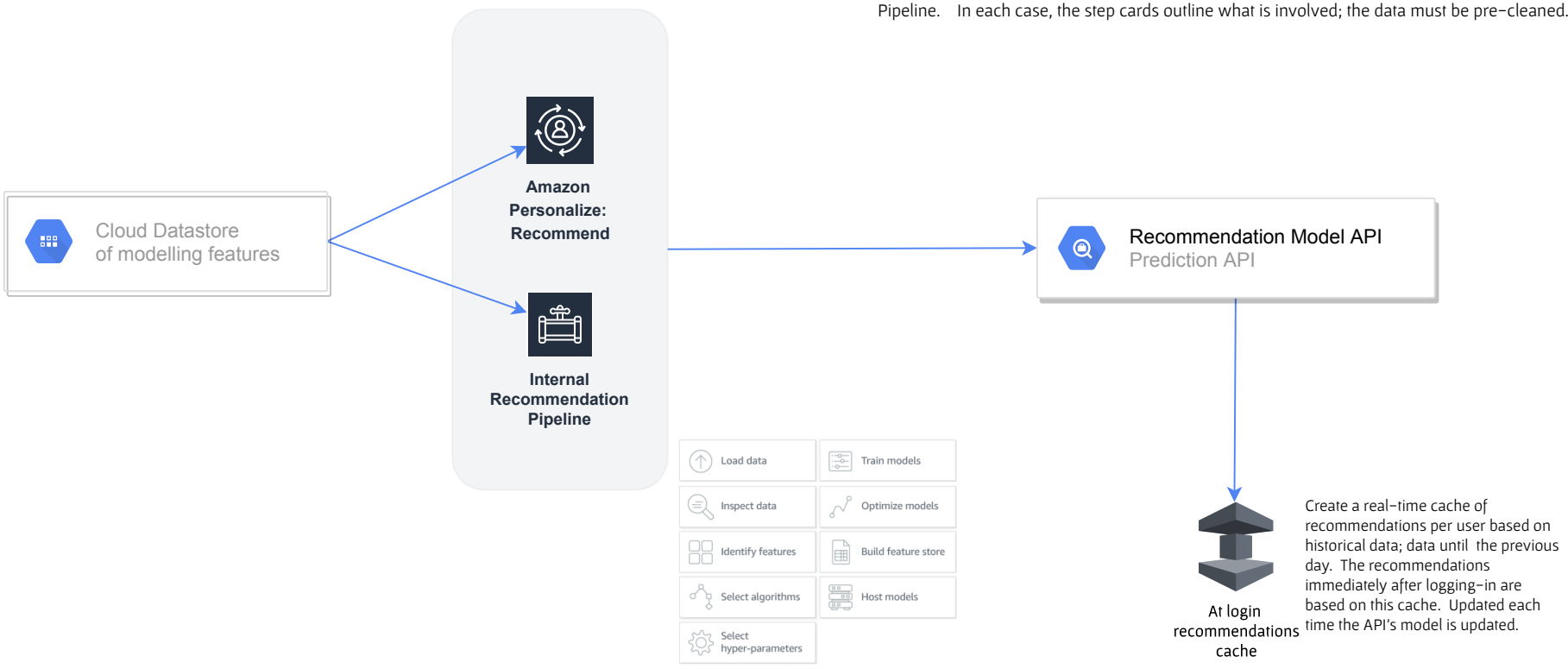
real-time personalised search



recommendation model development

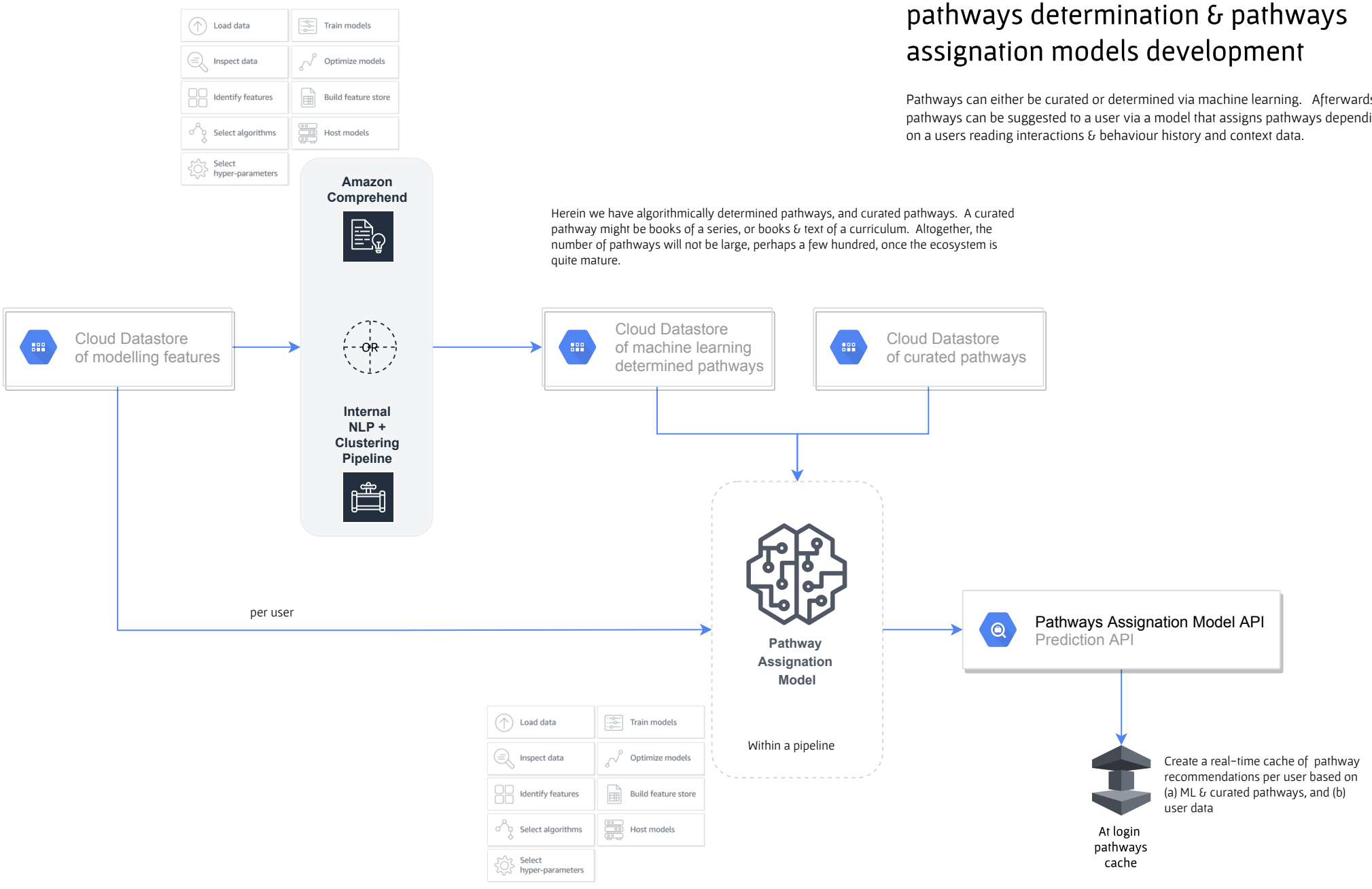
A continuously developed recommendation model. Its reommendations will be based on a user's reading interactons and context data such as geographic data.

There are two model development options: Amazon Personalize or an Internal Recommendation Pipeline. In each case, the step cards outline what is involved; the data must be pre-cleaned.



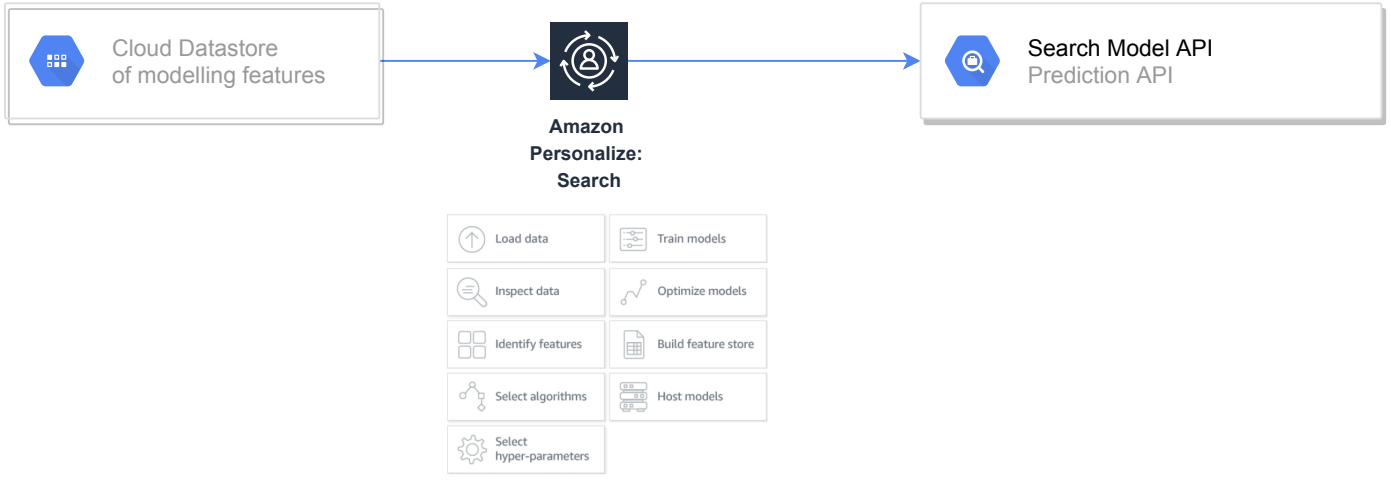
pathways determination & pathways assignation models development

Pathways can either be curated or determined via machine learning. Afterwards, pathways can be suggested to a user via a model that assigns pathways depending on a users reading interactions & behaviour history and context data.



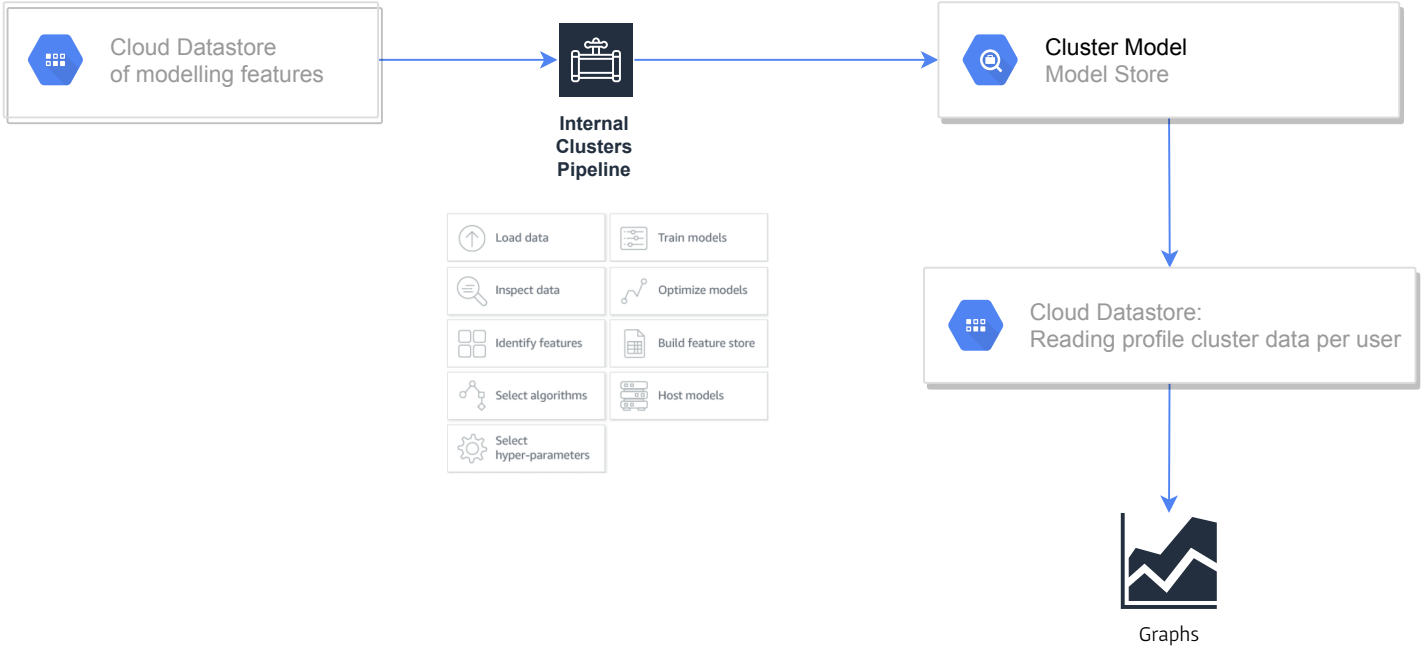
real-time personalised search model development

A continuously developed model based on **Amazon Personalize: Search**. **Amazon Personalize: Search** "can improve site search results for individual users by reranking search results using the behavioural data" of a user's "past application interactions". The step cards outline what is involved; the data must be pre-cleaned.



reading profiles clusters

Builds on the reading profiles experiment conducted for Anasoma & Kid's Android. The experiments developed reading profiles clusters based on reading behaviour – demographics data excluded.

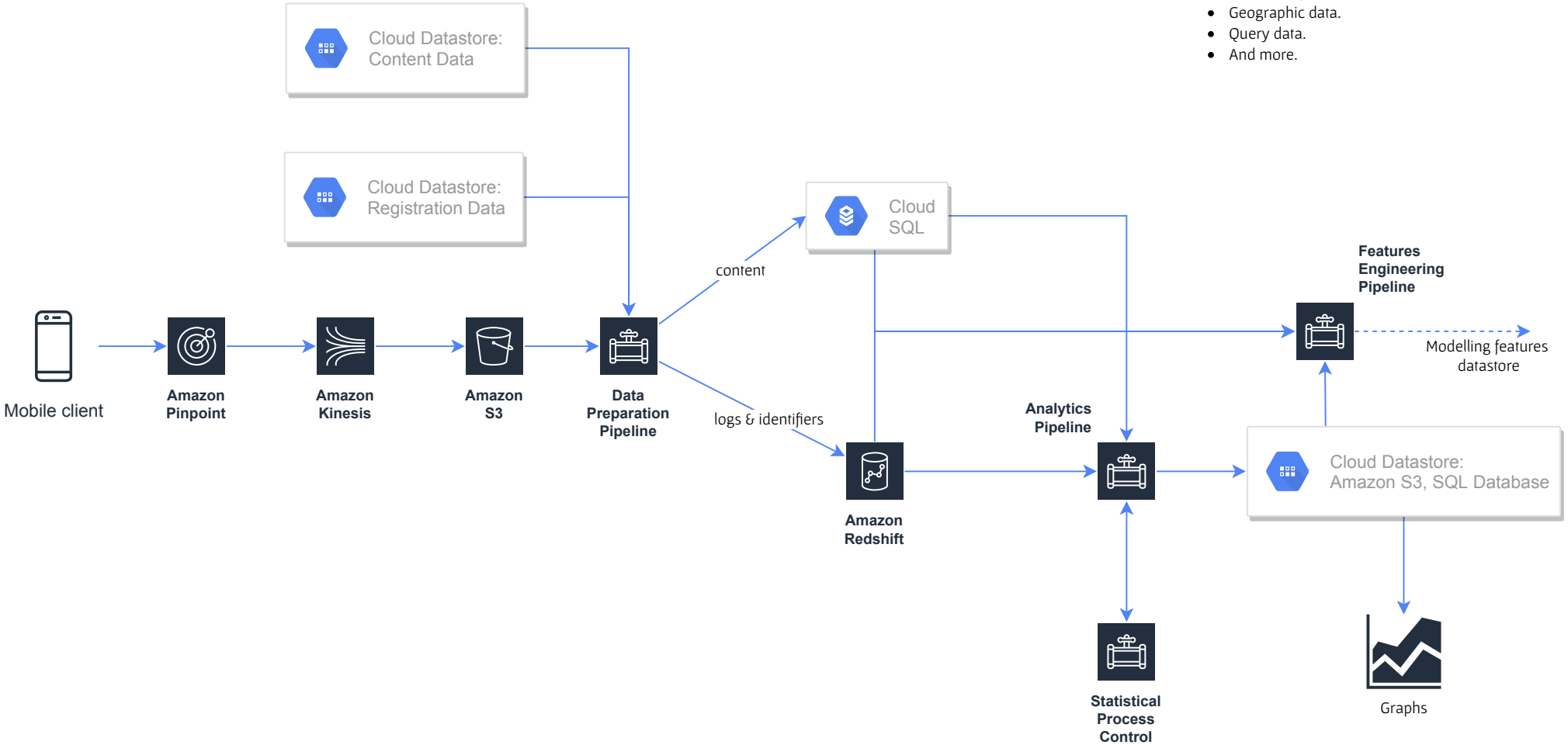


daily executions

Including (a) data preparation for modelling, analysis, and analytics, (b) analytics, (c) statistical process control, (d) features engineering, etc.

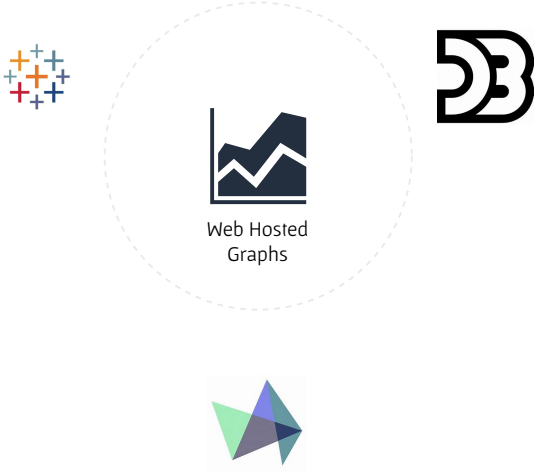
The modelling features data store is a combination of

- Derived features, e.g., reading measures and behaviour features.
- Geographic data.
- Query data.
- And more.



pipeline ecosystem

Pipelines are fed by, and use, a variety of resources.



graphs & graphing engines

For analytics and statistical process control graphs.

The best engine for a graph type will be used, but all graphs will feed into the same web pages ecosystem; S3 hosted.