Data analytics at MessageBird

Aleksandar Aleksandrov & Félix Mattrat

NOVEMBER 2018

About us

Data engineers & Team leads

Aleksandar Aleksandrov

Félix Mattrat

ABOUT

Introducing MessageBird

MessageBird is a cloud communications platform that empowers consumers to communicate with your business in the same way they communicate with their friends - seamlessly, on their own timeline and with the context of previous conversations.

For additional information visit: **www.messagebird.com**

225+ Agreements

We have 225+ direct-to-carrier agreements with operators worldwide.

15,000+ Customers

Customers in over 60+ countries, across a great variety of industries.

180+ Employees

More than 180 employees speaking over 20 languages based in the Americas, Europe & Asia.

Our offices

Our bird's nests around the world provide customers
in every time zone with 24/7 support.

AMSTERDAM

SINGAPORE

HAMBURG

LONDON

SAN FRANCISCO

SHANGHAI

SYDNEY

What’s on the menu?

01. Data at MessageBird
02. The past - Age Of Darkness
03. Enlightenment - Analytics today

Needs

Mostly about statistics and reporting

Internal needs

- State of the system
- Routing SMS
- Training algorithms
- ML Models

External needs

- Customer dashboard
- Reporting API

The landscape

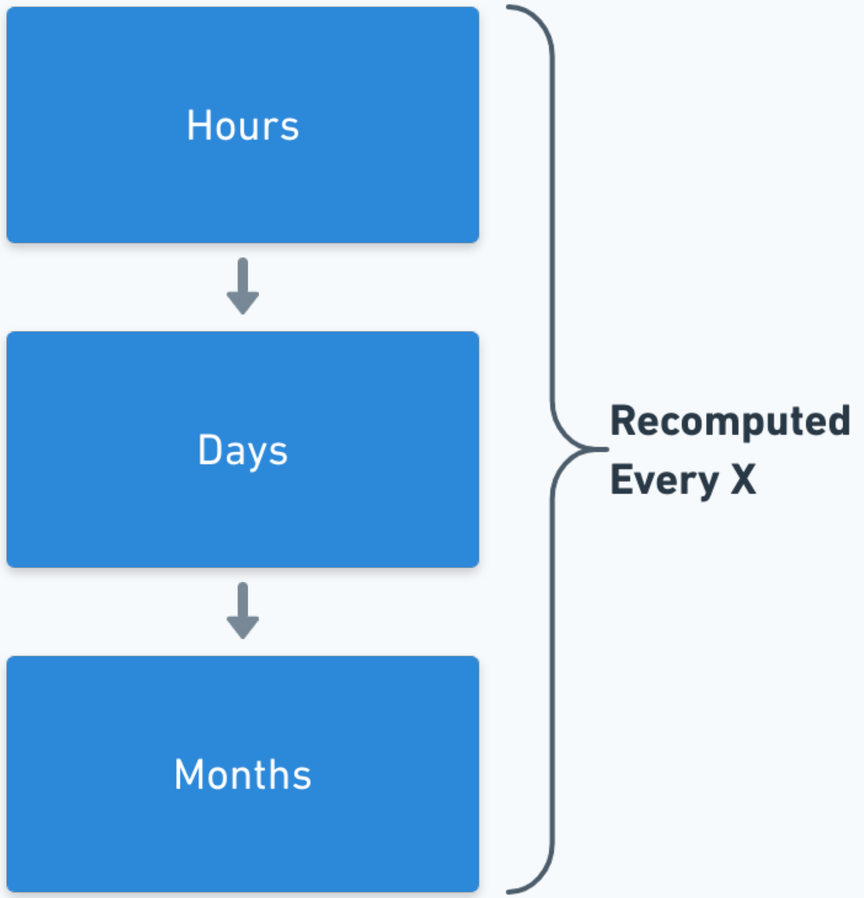
- Multiple carriers is messy - no uniformity of the data
- SMS messages go through many state changes up to months into the past
- Pricing (both carrier and customer) changes retro-actively

Age of Darkness

AGE OF DARKNESS

Hello CRON my old friend

- MySQL based
- Aggregates re-computed every X period of time
- Served us well for +5 years



AGE OF DARKNESS

Scaling problems

- The system had difficulty scaling and was often lagging
- Loss of granularity with pre-aggregation
- Performed poorly while doing analytical queries
- Inaccuracies

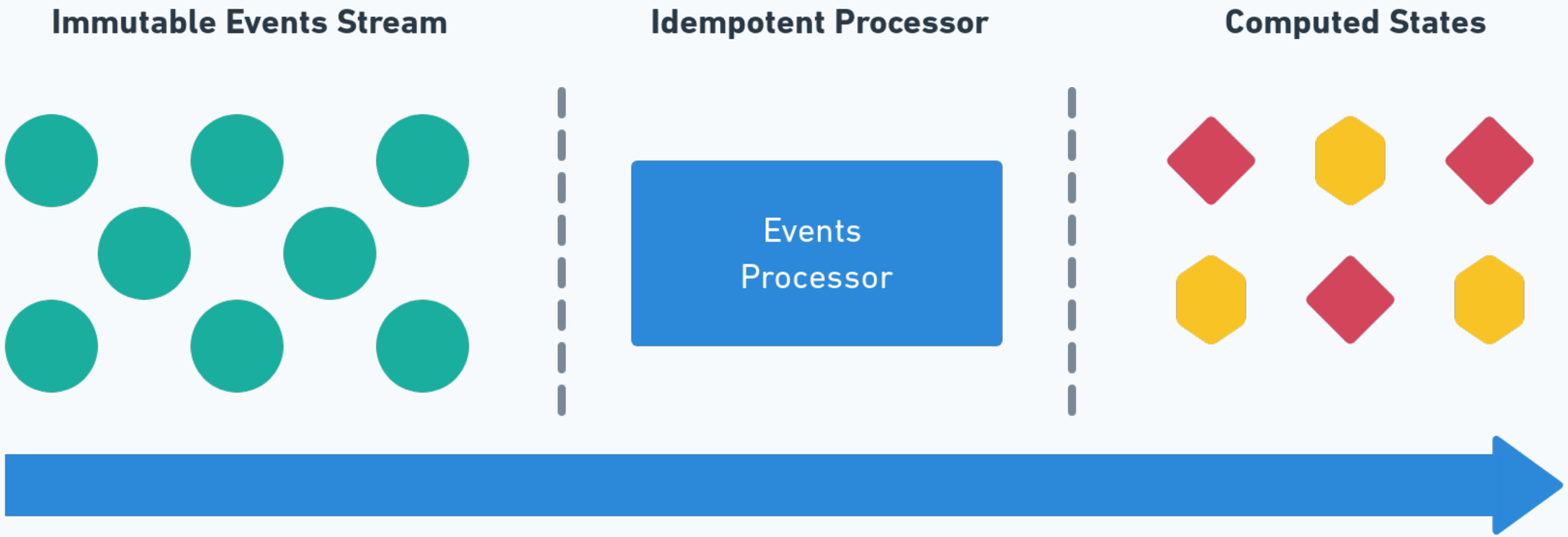
ENLIGHTENMENT - QUEST FOR AN ALTERNATIVE

Re-thinking data collection

- Able to keep up with continuously changing SMS message states
- In real time*
- Scalable to handle MessageBird’s global growth
- More flexible to accommodate wider use of data

Introducing event sourcing

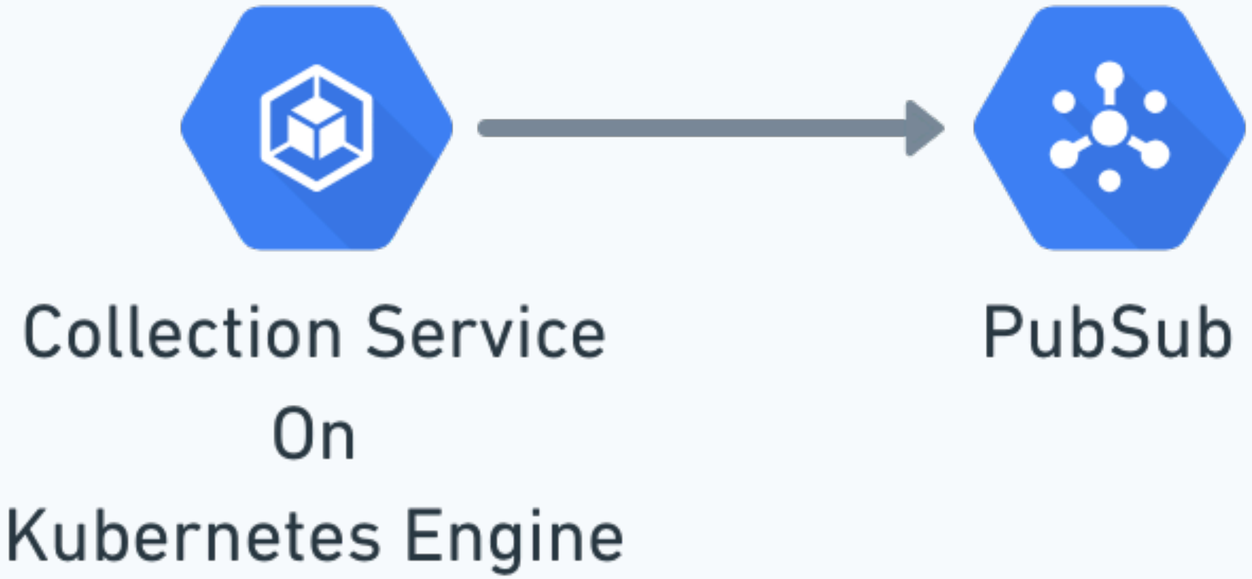
- Event sourcing, fairly common technique
- An immutable stream of events from which all states can be derivate


Immutable Events Stream
Idempotent Processor
Computed States
Events
Processor

Introducing event sourcing

- Problem: now we have increased our data by an order of magnitude.

How to collect all those events?



Pub/Sub

- Publisher/Subscriber queuing system
- Fully managed
- Globally available, consume your data anywhere in the world
- Not ordered and at least once guarantee

How to analyse all those events?

First design, dumping all in BigQuery

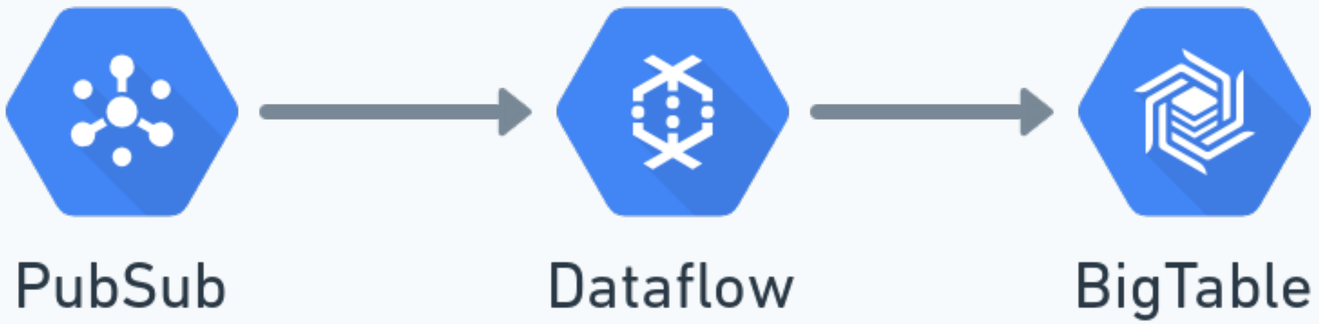
BigQuery

- Analytical database
- Fully managed
- Can ingest and query over very large dataset
- Nice standard SQL interface

Dumping all in BigQuery

- Not made for sub-second querying
- “Update” is expensive
- Can become costly fast

Second design, aggregate all things



Dataflow & BigTable

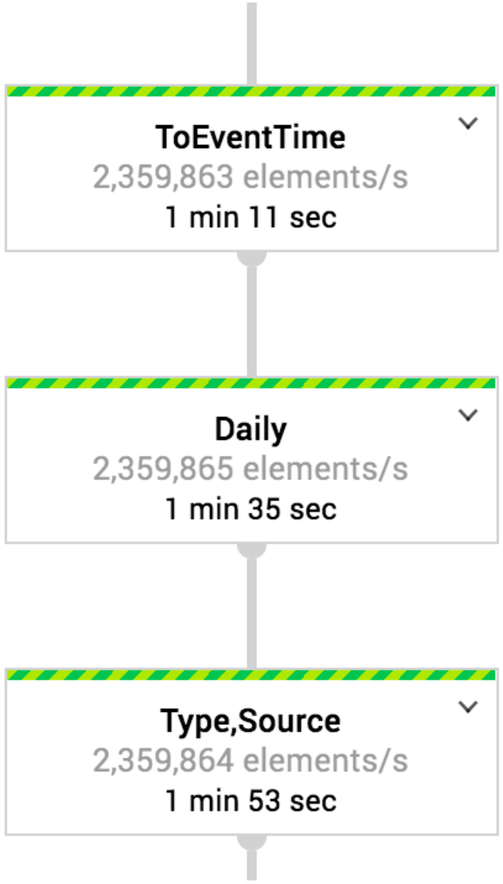
Dataflow

- “MapReduce in streaming”
- A common way of expressing stream and batch processing
- Integration with Pub/Sub

BigTable

- Massively scalable key-value store
- Read large consecutive key range
- Read / Write random key-value

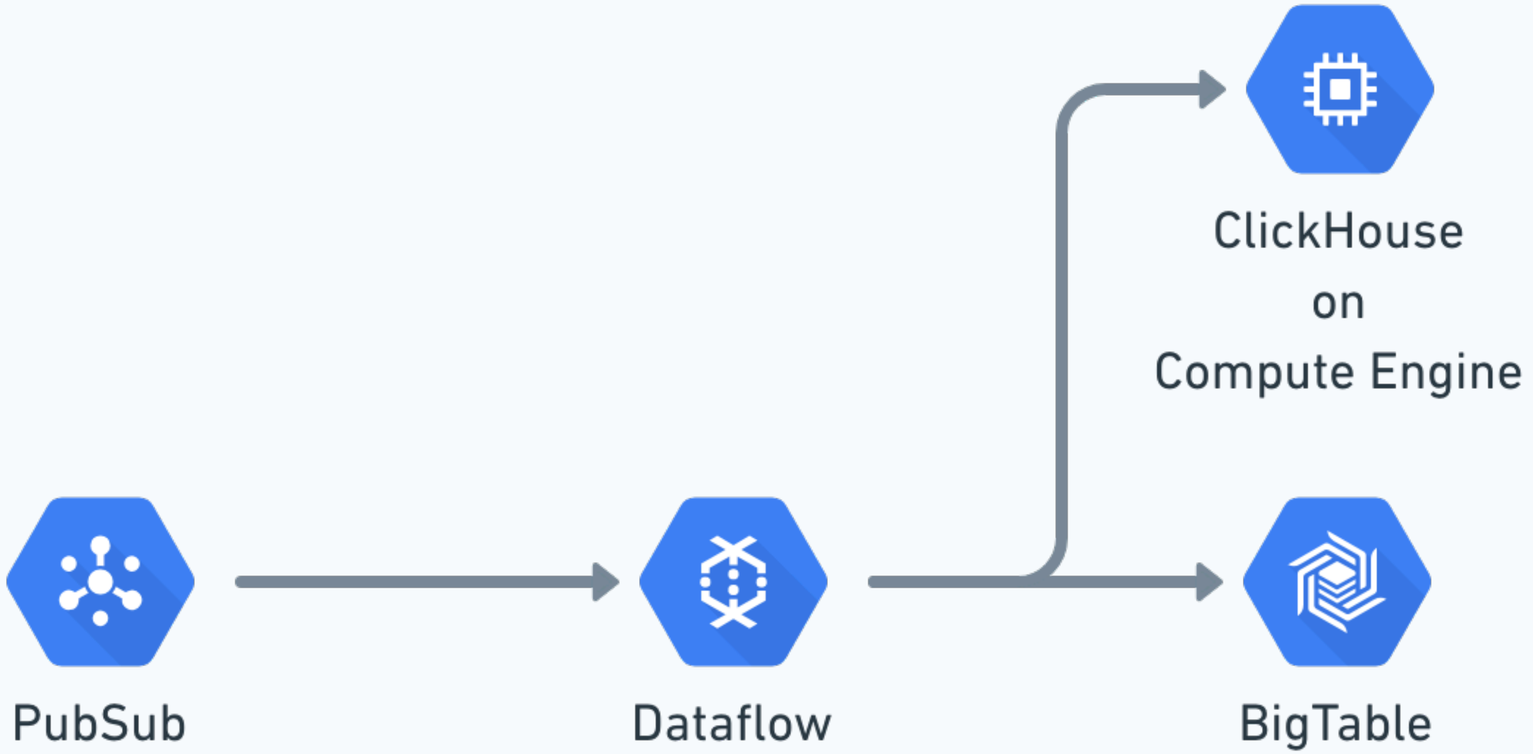
Dataflow & BigTable



Why it didn’t work?

- You have to aggregate over all dimensions, a SMS state is composed of +100 dimensions
- Multiplied by all timezones
- Over a very long window, +3 months

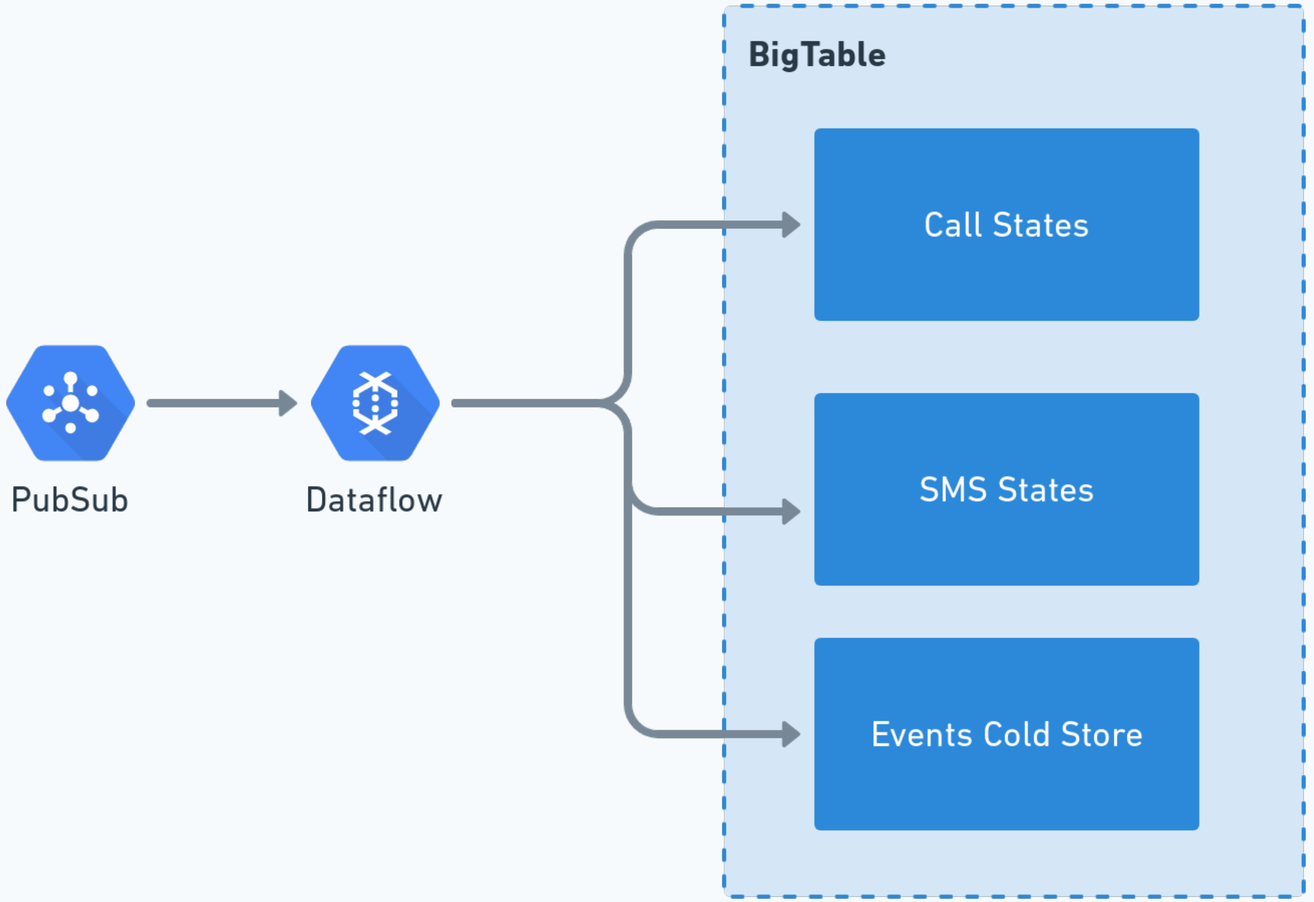
Third design, Analytical DB



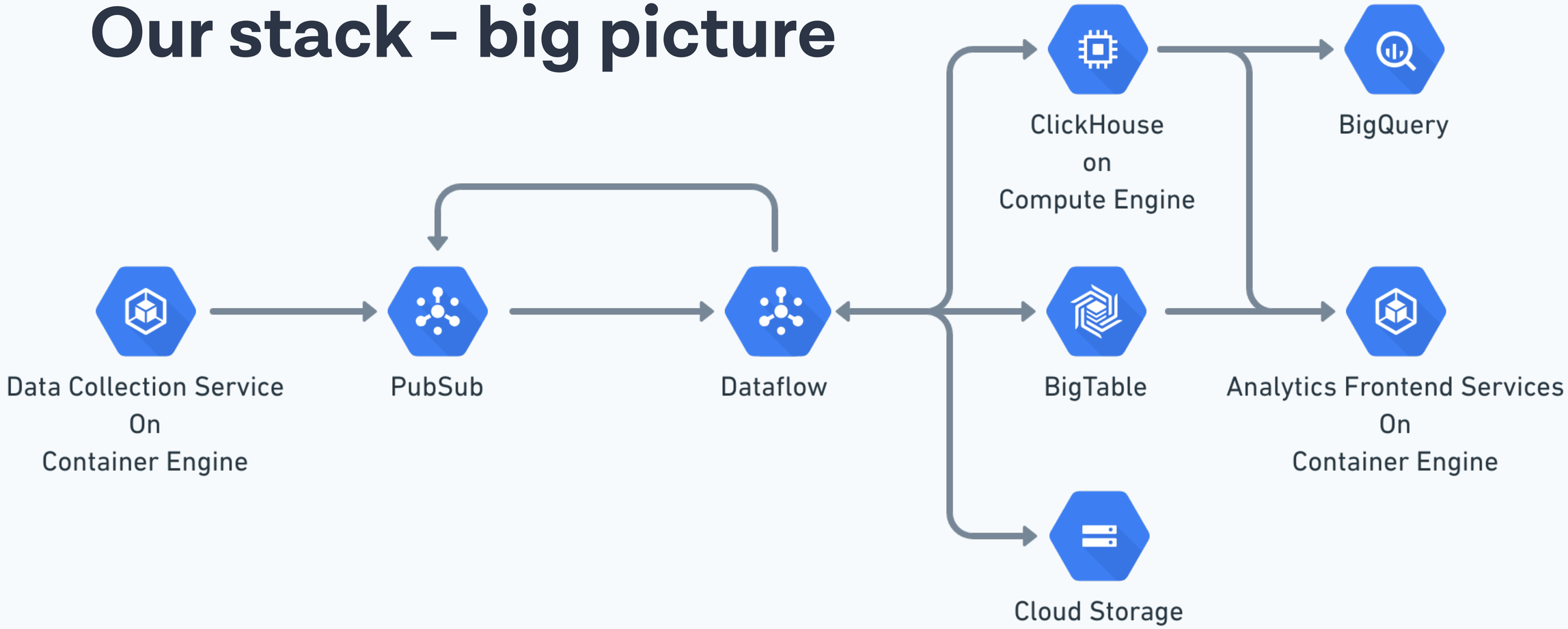
Third design, Analytical DB

- Aggregating over all dimensions - becomes feasible
- Flexible SQL
- Sub-second response time to fit our needs
- Horizontally scalable

Third design, Analytical DB



Our stack - big picture



What does GCP allow us to do?

- Scales with the growth of the business, with almost 0 effort
- Powerful real-time analytics
- Ability to react almost instantly over data streams
- High quality managed services, reducing the ops overhead

Questions

Late questions? Come say hello or drop us an email.

aleksandar@messagebird.com & felix@messagebird.com

www.messagebird.com/careers