Corona-Meldung GbR

Technology Whitepaper

Vision

Certain Blog Posts[1] and research papers have shown that Big Data and Simulation techniques are very effective in order to get control over the corona crisis. They can help to decide which actions should be taken by the government, companies and individuals to react properly to current situations.

Every Big Data application needs data. Corona-Meldung GbR was founded to enable persons to provide data about their health status to research centers, government and non-government organisations.

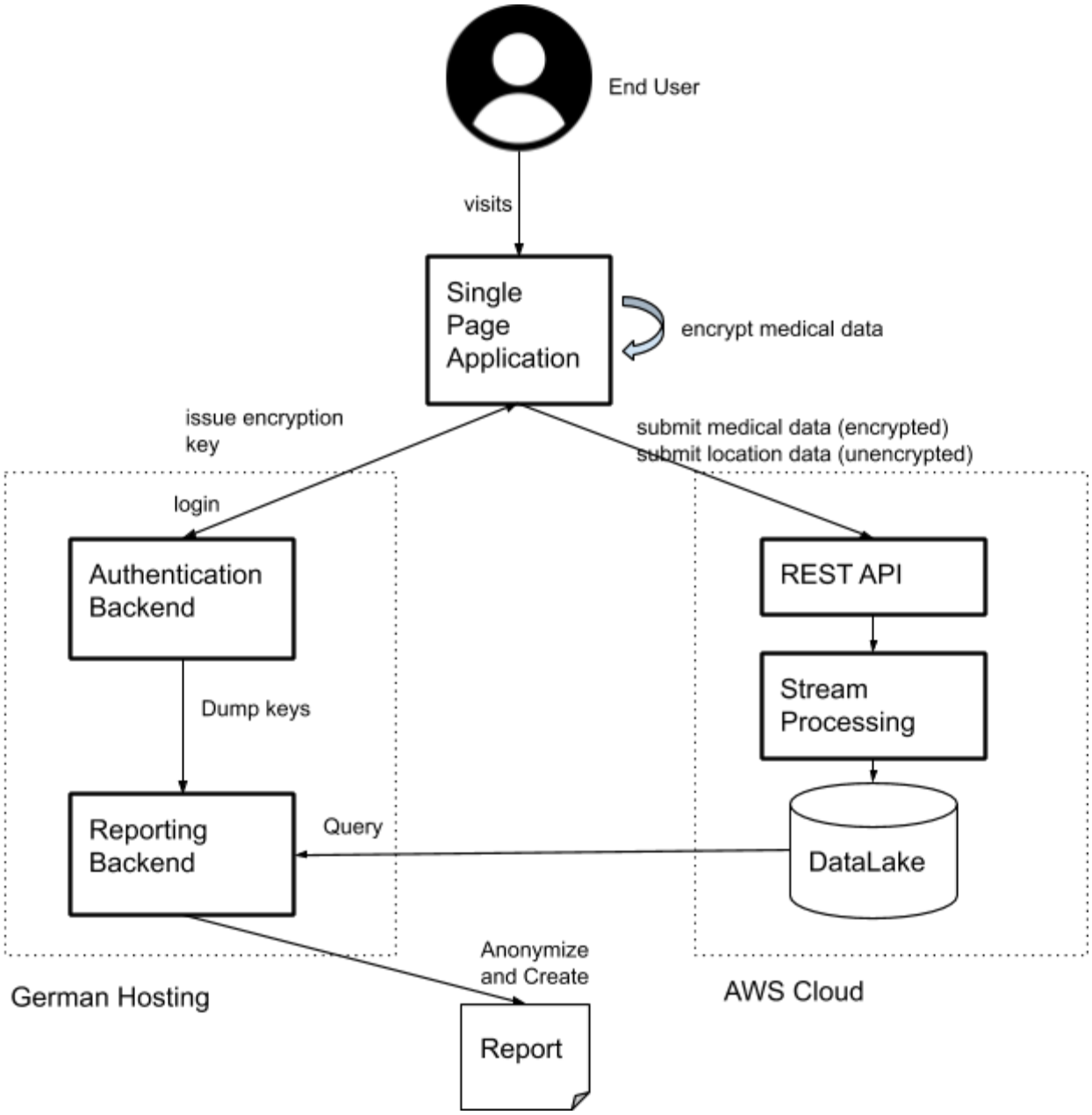
Philosophy

Medical and Location data of users is very sensitive and worthy of protection. For this case we have the ethical rules in our company:

1. The use case for the Data Gathering and Storage is strongly bound to BigData Analysis which helps to fight the corona crisis.
2. Published data is completely anonymized. It's neither possible to see nor to reconstruct a user relation from the data.
3. We use the latest possibilities in technology (encryption and authentication) to protect user data.

[1] https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca

Software Architecture



Process description

The user journey and handling of the entered data is described in the following 6 steps:

1. **Landing/frontend access**
   The user frontend[2] is realized with a Single Page Application (SPA), where a user is informed about our policies. If he/she decides to volunteer and share his/her data, he/she is requested to sign up to our service.
2. **Authentication & Key assignment**
   After the standard registration process a user can login via our Authentication backend[3], which is realized as a REST API written in python. After the login process the SPA is able to obtain an encryption key from the Authentication backend. This key must be unique for the user and of at least 256bit length and provided in JWK[4] (Json Web Key Format).
3. **Survey**
   A user form is provided where he/she will be requested to answer questions about his/her health status. To protect this data, it is encrypted with the key provided from auth backend using the javascript encryption library[5] provided by the browser.
4. **Storing of the survey**
   After encryption and pseudonymization (only user id from auth backend is referring to the user) this data will be sent via a second REST endpoint to the AWS Cloud.
5. **Sharing of location data**
   After the medical form the user will be able to share his/her location history. For this purpose the will be a guide how to download their location history from google takeout[6] and upload it again to the SPA. This data is also sent via REST Upload to the Cloud Endpoint.
6. **Data Processing & Report generation**
   Data Processing is done one AWS Cloud side in stream processing components and written to a private S3 Data Lake. For creating reports the reporting backend uses the encryption keys from the authentication backend and the data from the S3 Data Lake.

Security Concepts

1. Sensitive data is encrypted on row/user level.
2. To create a link between location and medical data it is required to have access to the data in the Data Lake and to the JWK Keys in the User Database

---

[2] https://github.com/dennisjay/corona-meldung
[3] https://github.com/dennisjay/corona-meldung-auth
[4] https://tools.ietf.org/html/rfc7517
[5] https://developer.mozilla.org/de/docs/Web/API/Crypto
[6] http://takeout.google.com

3. Data to be published is always anonymized