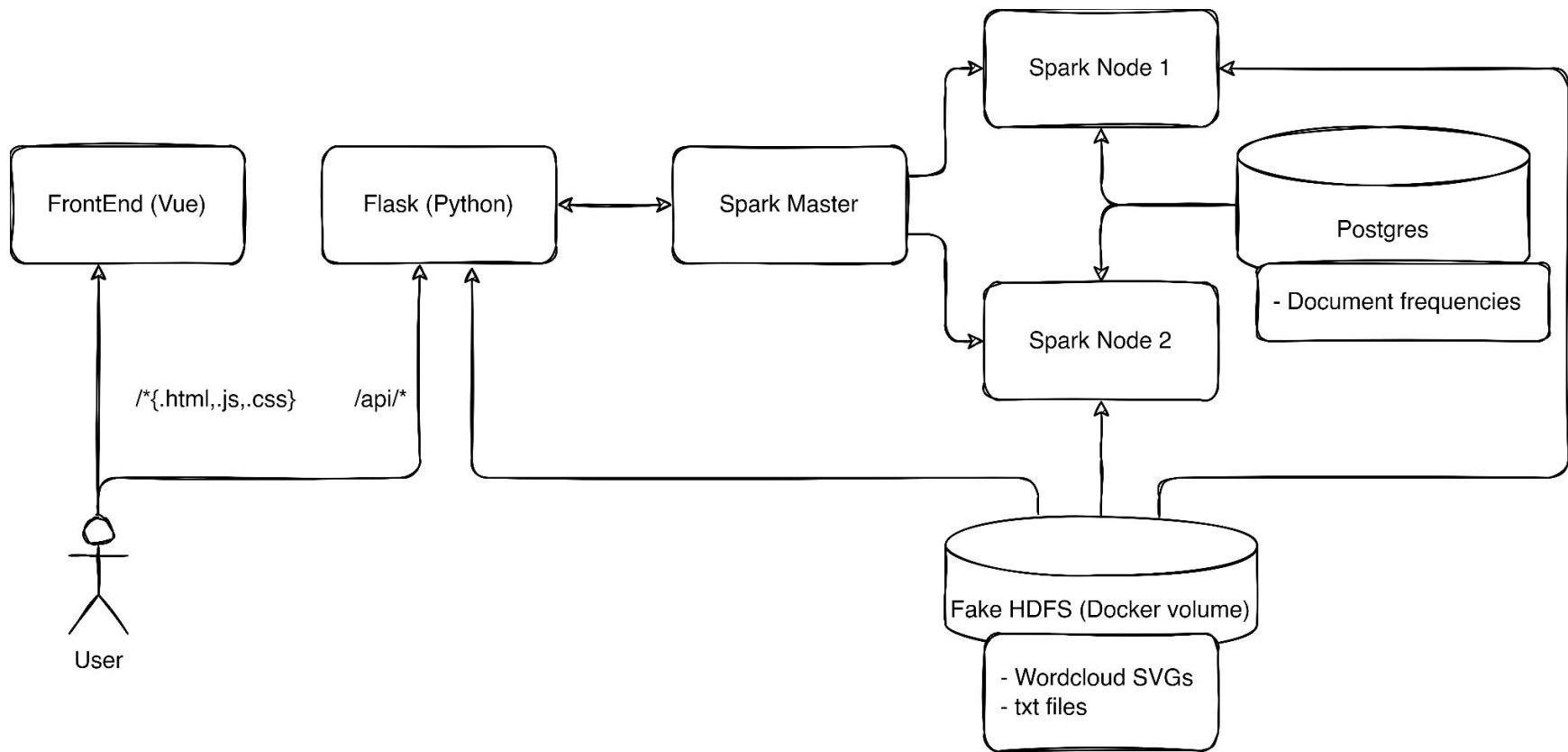


WORDCLOUD

CA 2 - BDEA

Julian Becker, Anton Bracke & Jan Mayer



LESSONS LEARNED 👍

- Setup mit Docker & Gitpod 🐳
- Normierung der TFiDF zur Anzeige der Wordcloud ÷
- “Long-Polling” zum Erstellen der Bilder 🖼️
- Docker Volume als “Fake” HDFS 📁

Hello

Ä

ريال

Ö

Kon'nichiwa

Bonjour

INTERNATIONALISIERUNG

□

は

Ø


Ciao

ß

Hallo

Unterstützung von Unicode Wörtern aktueller Regex: '\W+'

Diskussion

- Internationalisierung  
 - Stoppworte
 - Unterstützung von Unicode Wörtern aktueller Regex: 'W+'
- Geeignete Möglichkeit Daten ins Spark Cluster zu bekommen
 - HDFS?
 - Datenbank?
- Skalierbarkeit für z.B. 1000 Texte einfach durch weitere Spark Nodes?
 - Einsatz von HDFS
 - Postgres Cluster skalieren

also da kann man ruhig
mal klatschen



Merkel Memes