



Rafaella Duarte
Carolina Dias

O Desafio

- Implementar um pipeline de **extração**, **transformação** e **disponibilização** dos dados
- Responder às **perguntas**
- **Livre** utilização de ferramentas



Os Dados

- Base de **Dados Abertos do CNPJ** do Governo Federal
- De **2010 a junho de 2021**
- Estão em formato **.ZIP** e são posteriormente extraídos para **.CSVs**

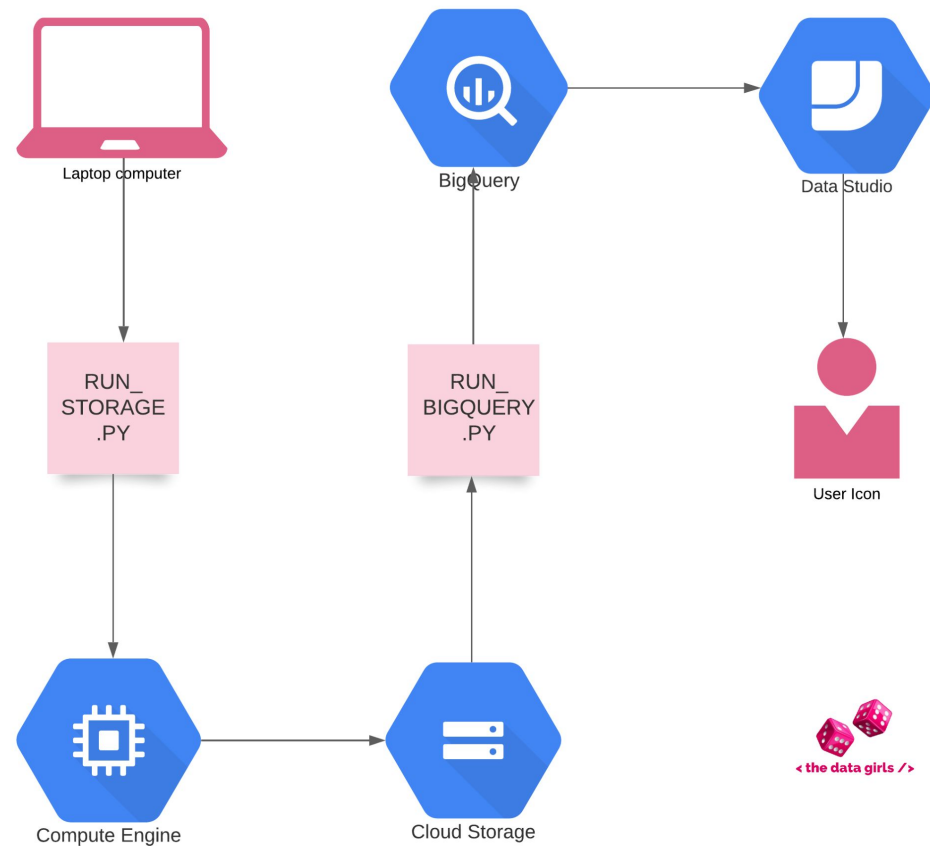


A Solução

- Google Cloud Platform
 - Alta Integração entre os Serviços
 - Facilidade de Uso
 - Baixo Custo
 - Altamente Escalável



A Arquitetura



Os Custos

- Uma Semana de Uso
- R\$ 23,01
- \$ 4,31
- Tende a Diminuir



As Respostas

- Queries em SQL no BigQuery
- Dashboards no DataStudio

