

## Producer

globoplay



## Security Group



Amazon  
EC2

Colector

Enricher

Storage



Mensagens utilizando  
tópicos de streaming  
Kafka

Dados armazenados em GOOD files e  
BAD files para análise caso necessário



Camada RAW – dados  
crus sem formatação  
específica



Camada TRUSTED – dados  
sem duplicidade,  
organizados e  
padronizados.



Camada Service – Dados  
com maior grau de  
confiabilidade.  
Estruturados de acordo  
com a demanda



Airflow como scheduler e  
orquestrator para  
recorrência de atividades



Linhagem de dados  
utilizando DBT Docs



Grafana

Jobs de qualidades de dados utilizando  
spark, com acompanhamento pelo  
grafana



Apache superset  
como ferramenta de  
BI Opensource



Análise dos dados por  
parte de outras áreas /  
equipes



Dashboards de  
acompanhamento  
sincronizados com as  
camadas e tabelas



AWS Glue

Catálogo  
utilizando  
AWS Glue



Amazon Athena



Amazon  
S3

EC 2

Query

Persistência

mensageria