

Data Engineer Aufgabe

Kontext:

Plattformdaten, die an den verschiedensten Stellen des Unternehmens eventbasiert generiert werden, sollen in einen DataLake für Auswertungen abgelegt werden.

Aufgabe:

Erstelle einen Architekturentwurf für einen generischen Prozess, der die Daten in einen AWS DataLake auf s3 und ggf. in die Redshift Datenbank überführt.

Rahmenbedingungen:

- Ein grober Entwurf als Diskussionsgrundlage ist ausreichend
- Die Lösung soll in der AWS Cloud realisiert werden
- Die Daten sollen im SelfService von Interessenten über den DataLake auswertbar sein
- Die Redshift Datenbank ist nicht über das öffentliche Internet erreichbar
- Data-at-rest soll verschlüsselt sein