You are working as data engineer in the CAC and are member of 1 of the 8 agile teams in our Agile release train. Next to the agile teams we have a layer of product management and program architects supporting the agile teams with feature definition and architectural designs. We also have a level 2 support team, with skills for each techstack, who will take developed features into line-business operations.

We have a back-end team focusing on ETL (Talend), DWH (Oracle) setup and Batch data integration, and a front-end team focusing on building dashboards in Power-BI. The other teams are working on other components in our Ecosystem like Salesforce, Omnichannel based on Genesys cloud and middleware based on Azure.

We are working in the scaledagile framework (SAFe) in PI’s of 10 weeks with 4 sprints of 2 weeks and one innovation and planning sprint. As a Data engineer you are actively involved in the major sessions like design session, PI & Sprint planning, the daily stand-up, sprint reviews, Inspect & Adapt sessions, the system Demo’s and the retrospectives of the back-end team. The back-end team consists of a Product owner, scrum master, team architect, 4 data engineers (developers) and 2 testers. During PI planning you plan the next 4 sprints and commit to features. Also a capacity reservation is made for defect handling each PI.

1. *Could you describe how you expect a regular day as Data Engineer in the Agile team would look like?*
2. *What would you define as main tasks and responsibilities of a Data Engineer?*

The Team is preparing for the Design session, where we work out a high-level feature into detailed user stories that will be estimated including listing dependencies to other team. One of the features planned is to create a new star-schema in our DWH for the customer satisfaction data, which is gathered via surveys send to our customers via Salesforce.

1. *You are asked to prepare this feature for the design session, which steps would you take to be well prepared for the design session?*

During the design session the following requirements/input needs to be considered:

* Survey result Data is stored in Salesforce linked to the case that triggered the survey
* Survey results contains a 5 star CSAT score, 1-10 NPS score and several free text fields
* The new survey fact should be linked to the already available Case fact in our DWH
* Survey Response time is a relevant measure for the business
* Free text field answers should be available for text mining by our business analyst
* Reporting should be possible via several case dimensions like date, market, country

1. *Which design would you advice? With whom would you need to work together to get to a final design?*
2. *Could you make a small Star-schema design based on these requirements?*
3. *Design a flow from extract to load for this star-schema. Which considerations do you take in each step regarding keys, history, loading & development process?*
4. *Could you create the user stories and dependencies, you would require to finalize this feature and assign the “team role” you would expect to execute each user story and ensure a smooth handover to production.*

In the 3rd sprint of the PI, you get an urgent defect from production assigned. Seems there is an issue with duplicate data records that we loaded in Salesforce production. This will impact your progress on the committed feature. After analyzing the defect for a few hours, you realize you will need a lot more time to find the root cause and fix the defect. The time reserved for defect handling this PI is already fully consumed by the agile team.

1. *How will you handle this situation?*
2. *Which technical approaches would you use to identify the duplicate records?*