Dear Manager team

I hope this message finds you well. I am writing to proportion a few insights from our recent records first-rate evaluation and to outline the steps we are taking to resolve identified issues. Your comments and input could be beneficial as we move forward.

Questions About the Data

- Can you provide more context on the business logic behind the relationships between Users, Receipts, Items, Brands, and Cogs? Understanding these relationships in a product context will help ensure our data modeling aligns with business operations.
- Are there specific attributes within each entity that are more critical than others for business decision-making? For example, is the bonusPointsEarnedReason in Receipts as crucial as totalSpent? Prioritizing attributes will help us focus on optimizing the most impactful data.
- How frequently is the data updated, and what processes are currently in place to ensure its accuracy? This information will help us determine if the data quality issues stem from outdated information or inconsistencies in the update process.

Discovery of Data Quality Issues

We found the statistics nice problems via a sequence of systematic assessments. These covered inspecting column values for repeating groups, identifying partial and transitive dependencies, and validating foreign key constraints. By very well analyzing each desk, we have been able to pinpoint particular regions where the information structure did not adhere to exceptional practices.

Information Needed to Resolve Data Quality Issues

To remedy those problems efficiently, we want:

Documentation: Detailed documentation at the records replace strategies and any current modifications made to the records shape.
Business Logic Changes: Insights into any upcoming modifications within the business common sense that might impact data relationships. This will make sure that our solutions align with future enterprise desires and keep away from introducing new issues.
Additional Information to Optimize Data Assets
To optimize the information property we're trying to create, we would gain from:

- User Behavior Insights: Detailed insights into consumer behavior and how facts are utilized in decision-making approaches. This will assist us tailor the records structure to higher help enterprise targets.
- Performance Metrics: Current overall performance metrics and any present pain points associated with records get entry to and processing. This will guide our efforts to beautify records accessibility and performance.

Performance and Scaling Concerns

As we scale, we expect numerous overall performance and scaling concerns:

1. Query Performance: Ensuring that statistics queries remain performant because the dataset grows. To cope with this, we plan to implement indexing on key columns and optimize question systems.
2. Data Integrity: Maintaining records integrity throughout a growing dataset. Regular statistics, first-class exams and validation techniques could be essential in mitigating this threat.
3. Scalability: Managing big volumes of information efficiently. We plan to apply partitioning strategies for large tables and distribute the burden throughout a couple of servers if important.
4. I might appreciate it if we ought to agenda an assembly to speak about these findings and the following steps in element. Your insights and comments will be helpful in ensuring we deal with these problems correctly and optimize our facts property for future use.

Thank you for your time and attention to this matter. For more details, please refer to the attached file.

Best regards,

Ian Huang
yh3389@columbia.edu
646-318-4946

_____

Here is the Github link for the whole project: https://github.com/yh3389/Fetch-test-Ian/tree/main