1. Tell me about yourself.
"Hello, I am Souvik Dey, currently working as a Data Engineer with five years of experience at Cognizant, where I've had the opportunity to contribute to large-scale data solutions in the Banking and Financial Services (BFS) domain. I specialize in cloud platforms and big data technologies, with hands-on experience in Google Cloud Platform (GCP), Dataproc, BigQuery, and, most importantly, PySpark.

In my current project, I've been deeply involved in building scalable data pipelines on GCP, focusing on quality, performance, and business impact. I'm particularly proud of how we've used Apache Deequ for Data Quality checks and automated the entire pipeline using Google Cloud Functions.

I'm excited about this opportunity at EY, especially given EY's strong values around teaming, respect, and integrity, which closely align with my own. I'm eager to contribute to building a better working world by applying my technical and collaborative skills in this new role."

2. Why do you want to join EY?
"EY's reputation as a leader in providing insights and developing innovative solutions really excites me. What resonates with me the most is the strong emphasis on teaming, inclusiveness, and respect—values that I hold dear as a professional. I've always believed that a collaborative environment not only enhances productivity but also fosters creativity and innovation.

I'm especially drawn to EY's vision of 'shaping the future with confidence,' and I believe my experience in building data pipelines aligns well with EY's focus on helping clients capitalize on opportunities. I'm enthusiastic about contributing to EY's continued success in leading digital transformation, particularly in data engineering with Databricks and PySpark, where I can leverage my skills to build robust solutions."

3. How do you handle challenges or setbacks in projects?
"In my current role, I've encountered various challenges, particularly with data quality issues or tight deadlines for pipeline automation. Whenever a challenge arises, my approach is to break down the problem and understand its root cause. For example, in a recent project, we encountered issues with inconsistent data formats from our SFTP source. My team and I quickly identified the problem and implemented schema restructuring and partitioning in our Silver layer to streamline the data flow.

I believe setbacks are opportunities to learn, and as a team, we approach them with a mindset of resilience and adaptability. I've learned that by communicating transparently, seeking diverse perspectives, and being solution-focused, we can overcome even the most complex challenges. This is something I look forward to bringing to EY."

4. Can you describe a time you had to collaborate with others to achieve a goal?
"In my current project, I worked closely with data analysts, business stakeholders, and other engineers to build a scalable data pipeline on GCP. The key challenge was aligning the technical aspects of the project, such as ingestion, data quality, and automation, with the business needs.

To ensure success, I facilitated regular discussions to understand both technical and business requirements and fostered a culture of open communication and trust. For example, during the design phase, we collaborated closely on data sensitivity concerns like PII and PCI compliance. By combining my technical expertise with business insights, we were able to create a solution that met both performance and compliance standards.

I firmly believe that teamwork is about leveraging each other's strengths, and this project exemplified the importance of collaboration in achieving a shared goal. I look forward to contributing to such an environment at EY."

5. How do you ensure the quality and accuracy of the data in the pipeline?
"Data quality is crucial, especially in industries like banking and finance, where inaccurate data can lead to significant business impact. In my current role, I ensure data quality by incorporating several measures

throughout the pipeline. For instance, we use Apache Deequ to perform thorough data quality checks at the BRONZE layer, ensuring that only clean, validated data moves forward to the SILVER layer.

Additionally, we perform schema standardization and partitioning, which not only optimizes query performance but also ensures that the data adheres to a consistent structure. By building these checks and balances early in the process, we minimize downstream issues, allowing our business users to trust the data they rely on for decision-making."

6. How do you stay motivated, especially when working on long projects?
"I stay motivated by focusing on the impact of my work. Knowing that the data pipelines I build contribute to key business decisions and insights keeps me driven. In longer projects, I also break down the work into smaller milestones, celebrating progress along the way.

Collaboration also keeps me energized—working alongside a talented team and solving problems together is incredibly fulfilling. I'm especially excited about the opportunity at EY because of the company's values around energy and enthusiasm, which resonate with how I approach my work."

7. What do you bring to our team that sets you apart?
"I bring a unique combination of technical expertise in data engineering, particularly with Databricks, PySpark, and GCP, along with a strong commitment to teamwork and integrity. I've always believed in creating solutions that are not just technically sound but also aligned with business goals and stakeholder needs.

What sets me apart is my ability to bridge the gap between technology and business, ensuring that my solutions not only meet technical requirements but also deliver real business value. I'm excited to contribute my skills and collaborate with the amazing team at EY to continue building a better working world."

---
1. I am comfortable discussing only the fixed part of the compensation for now, as it forms the core of my expectations. I'm confident that this aligns with the responsibilities of the role and my experience.

2. As you can see from my track record, I stayed with CTS for more than five years, which highlights my loyalty and commitment. I am looking for a long-term opportunity where I can not only contribute as a technical expert but also grow with the organization, aligning myself with its strategic vision and long-term goals.

3. Since this role is a senior position, it comes with greater responsibilities, such as mentoring teams, driving projects, and ensuring successful delivery of business-critical tasks. Given the scope and the impact expected, I believe the compensation package reflects these additional responsibilities.

4. I have researched market rates for professionals with similar experience in BFS, GCP, and PySpark/Databricks. I believe the package I've proposed is competitive and in line with industry standards for someone with my level of expertise and the value I will bring to your organization

5. My proficiency in Big Data Engineering, Databricks and GCP Cloud services, along with certifications in these platforms, equips me to design and implement scalable, efficient data solutions. These niche skills are essential for keeping a competitive edge in today's data-driven market. Given my experience, my track record of leading complex projects, and the high demand for these skills, I believe the proposed compensation fairly reflects my value and the contributions I will make.
---

1. Why should we hire you?
I believe my experience as a Senior Data Engineer, particularly in the BFSI vertical, combined with my proficiency in tools like Databricks, GCP, and PySpark, uniquely qualifies me for this role. I've

successfully built scalable data pipelines, solved complex data engineering challenges, and led cross-functional teams. My ability to adapt quickly, lead effectively, and drive innovation while ensuring data integrity and business outcomes aligns with the demands of this role.

2. What are your strengths and weaknesses?
My strengths include strong problem-solving skills, technical proficiency in cloud platforms and big data tools, and the ability to manage and deliver data projects under tight deadlines. I have a collaborative approach, which helps me work effectively in teams. My weakness, at times, has been delegating tasks. As a data engineer, I often take personal ownership of quality, but I've learned the importance of trusting the team and am actively working on improving this.

3. Why are you looking for a change?
I'm looking for a change to take on greater challenges and responsibilities. I've reached a point where I want to leverage my skills in a more strategic role, and EY's focus on data-driven solutions and cutting-edge technologies excites me. This position will allow me to grow both technically and professionally, while contributing to large-scale projects that make a significant impact.

4. Handling Conflict with Lead Architect (Client) without Time for a POC
In such a situation, I would rely on clear and professional communication. I would present my viewpoint backed by data, explain the rationale behind my approach, and articulate the risks of alternative solutions. I would also acknowledge their concerns and suggest a follow-up or quick review to align both sides. If necessary, I'd offer to provide a detailed explanation later, emphasizing my commitment to the success of the project.

5. What will you do if another company offers more after our offer?
While compensation is important, I believe in looking beyond that and focusing on growth opportunities, work culture, and long-term career prospects. I value EY's vision, its reputation in the industry, and the exciting challenges this role offers. My decision would be based on the holistic package of career growth, learning opportunities, and alignment with my goals.

6. Tell about a time you made a mistake and had to redo work (STAR)
Situation: During a data pipeline migration to GCP, I missed validating a crucial data transformation step.
Task: This mistake caused an error in the dataset, requiring me to redo the validation while the deadline loomed.
Action: I quickly owned up to the mistake, informed my team, and prioritized the validation process. I also worked overtime to ensure that the issue was resolved and that the deadline was met.
Result: The issue was corrected without delaying the overall project, and we put stronger validation checks in place moving forward.

7. Tell about a time with competing deadlines (STAR)
Situation: I was leading two projects simultaneously: one, a client deliverable with a tight deadline, and two, an internal data infrastructure update.
Task: I had to ensure both were completed on time without compromising quality.
Action: I prioritized the client project while delegating parts of the internal task to trusted team members. I also communicated openly with stakeholders to adjust timelines where possible.
Result: Both projects were delivered on time, and we maintained our client's trust while successfully upgrading the infrastructure.

8. Why EY and what do you know about EY?
I am excited about joining EY because of its leadership in consulting and data analytics. EY's emphasis on innovation and using advanced analytics to drive decision-making aligns with my career goals. I've followed EY's growth in the data space, particularly its focus on digital transformation and emerging technologies like AI and data engineering. This resonates with my passion for implementing impactful data solutions.

9. Are you willing to relocate?
Yes, I'm open to relocation. I understand that certain roles may require being onsite or in a specific region, and I'm flexible to make such a move to contribute to EY's goals and my professional growth.

10. Managing More Responsibilities at EY
I'm excited about taking on more responsibilities, as I've been preparing for this by expanding my technical knowledge and leadership capabilities. In my current role, I've led end-to-end data projects, managed budgets, and coordinated with cross-functional teams. I'm confident that I can handle the increased scope at EY by leveraging my organizational skills, prioritization techniques, and fostering a collaborative team environment.

11. Leading a Project and Building a Team (STAR)
Situation: I was tasked with leading a critical data pipeline migration to GCP for a financial institution.
Task: Assemble a team, manage the budget, and ensure timely delivery.
Action: I carefully selected team members with the right skills, set clear milestones, and monitored the budget closely. I also facilitated regular team meetings to ensure open communication and address challenges early.
Result: The project was completed on time, within budget, and the client was highly satisfied with the performance improvements.

12. Where do you see yourself in five years?
In five years, I see myself taking on more strategic leadership roles in data engineering, perhaps leading a larger team and influencing key decisions in data strategy. I'd like to continue working with cutting-edge technologies and drive impactful projects that shape the way businesses leverage data.

13. How well do you handle change?
I handle change well and thrive in dynamic environments. In fact, one of my strengths is my adaptability. Whether it's new technology, shifting project requirements, or unexpected challenges, I approach change with a problem-solving mindset. For instance, in my current role, I've adapted to various shifts in priorities by staying organized and being proactive in communication.

14. Do you work well under pressure?
Yes, I perform well under pressure. I've been in situations where time-sensitive projects demanded quick decision-making and action. By staying calm, breaking down tasks, and staying focused on the end goal, I ensure that the pressure doesn't compromise the quality of my work.

15. How do you make important decisions?
I make important decisions by gathering all relevant information, considering the pros and cons, and consulting with key stakeholders when necessary. I also rely on data-driven insights to back my decisions and ensure they align with long-term business goals. If time permits, I seek input from the team to ensure diverse perspectives are considered.

1. A time you helped a peer or teammate struggling with their work.
In one of my recent projects, a teammate was struggling with setting up automated data ingestion from SFTP to our Data Landing Zone in GCP. Since this was a critical component of the pipeline, delays here would affect the overall project timeline. I stepped in and sat with my teammate to help debug the issues they were facing with file formatting inconsistencies.

We worked through it together, implemented schema validation steps, and finally managed to get the pipeline functioning. This experience not only resolved the immediate issue but also strengthened our team's collaborative spirit, which was vital for the project's success.

2. A time you made a mistake or missed a deadline and how you handled it.

There was a time when I underestimated the time required to optimize complex join operations in PySpark while transforming data into the GOLD layer. Despite my best efforts, I couldn't meet the deadline. I immediately communicated this to my manager and the stakeholders, explaining the reasons for the delay and proposing a new timeline.

To ensure it didn't happen again, I reviewed my estimation process and included buffer time for complex tasks. I also implemented monitoring tools to identify bottlenecks earlier in the process.

3. A time when your teammate didn't agree with you or when you disagreed with a senior but went with their suggestion.
In one of our design discussions, a colleague disagreed with my approach to handling PII data within our Silver layer, arguing that a different method of encryption would be more effective. While I initially felt my approach was simpler, I took the time to fully understand their perspective.

We had a respectful discussion, and after further review, I realized their suggestion had better long-term security benefits. I supported their solution and helped in its implementation.

4. A time when you invented or devised a novel solution to a complex problem.
In one of my projects, we faced performance issues when dealing with large datasets in BigQuery. Rather than purchasing more computational resources, I came up with a novel partitioning strategy that allowed us to optimize queries without extra cost. I proposed using dynamic partition elimination techniques along with clustering to minimize resource use.

5. A time when you had to make an important decision without enough data.
While implementing the automation triggers using Google Cloud Functions, there was a moment when we didn't have sufficient historical data to determine the most efficient trigger frequency for processing files. Based on the limited insights we had, I decided to implement a more conservative frequency with built-in flexibility to adjust based on future data.

6. A time you received negative feedback from a manager and how you handled it.
During one of our quarterly reviews, my manager pointed out that while my technical skills were strong, I sometimes focused too much on the technical solution without considering the business context fully. I took this feedback to heart and started engaging more with the business stakeholders to understand the broader picture before diving into the technical details.

7. A time you went above and beyond your job responsibilities.
When our team faced a resource crunch during the final stages of a project, I volunteered to take on additional responsibilities, such as performing extra Data Quality (DQ) checks and optimizing our monitoring setup. While these tasks were not originally part of my scope, I knew they were crucial for delivering the project on time.

8. A time when you failed or used the wrong approach.
In a previous project, I initially chose a particular method to optimize PySpark jobs, which I believed would increase performance. However, after several days of testing, it became clear that my approach was causing more overhead and was not the right fit for the workload we were handling.

9. A time when you led a project or made an important decision.
I was given the responsibility of leading the automation process of our data pipeline using Google Cloud

Functions. This included deciding the trigger mechanisms and ensuring they adhered to the defined SLAs and OLAs. I led the team through regular updates and collaboration sessions to align everyone's efforts.

10. A time when you boosted the morale of the team or resolved a group conflict.
There was a period when our team faced tight deadlines and mounting pressure. Some members were feeling overwhelmed, so I organized an informal brainstorming session, where we openly discussed challenges, delegated tasks more efficiently, and supported each other. I made it a point to recognize everyone's contributions, no matter how small.

11. Biggest failure or critical feedback from your boss.
The most critical feedback I received was that I sometimes focus too much on the technical aspects, missing the bigger picture. Since then, I've made a conscious effort to understand business goals before diving into solutions. This has improved my ability to deliver work that aligns with both technical and business objectives.

12. A time when you handled a crisis or conflict.
During a key phase in our project, we had a conflict regarding the data partitioning strategy between two teams—each proposing different methods. I stepped in to facilitate a discussion where we evaluated both options based on technical merits and business impact. In the end, we combined the strengths of both approaches and found a middle ground.

1. A time when your suggestions brought positive changes/impact to the team.
During a recent project, we faced performance issues in BigQuery due to inefficient queries over large datasets. I suggested implementing dynamic partitioning and clustering strategies to optimize query performance. After rolling out these changes, we significantly reduced query execution times and resource costs, positively impacting the project's success.

2. A situation where you had a conflict with a team member and how you resolved it.
I once had a disagreement with a colleague over the encryption mechanism for securing data in the Silver layer of our GCS pipeline. They proposed an approach that seemed overcomplicated. After careful discussion, I realized their method offered better security and longevity. We combined our ideas and implemented a hybrid approach that balanced simplicity and security.

3. A situation where you had to make a choice with limited time to explore options.
In one of our projects, we had to choose an optimal trigger frequency for our automation processes using Google Cloud Functions, but we didn't have sufficient historical data. I quickly decided on a conservative trigger frequency to prevent system overloads, knowing we could adjust later based on actual usage data.

4. A time when you failed at something and what you learned from it.
In a previous project, I designed a PySpark optimization that was meant to improve job performance. After spending several days implementing it, I realized it wasn't yielding the expected results. I reverted to a simpler approach, but the experience taught me the importance of validating assumptions early.

5. A time when you went out of your way to help someone.
A junior colleague was struggling with understanding our data pipeline workflow and the nuances of GCS. I voluntarily set aside time to give them one-on-one training, covering both technical and best-practice

aspects. We also paired on solving a few real-time issues.

6. A time when you came up with a novel solution.
While working on data aggregation for the Gold layer in BigQuery, we faced the challenge of handling large datasets efficiently. I suggested a novel strategy to use approximate aggregations, which cut down resource usage while maintaining a high level of accuracy.

7. A time you received negative feedback from a manager and how you responded.
During a performance review, my manager pointed out that I sometimes focused too much on technical details and missed the business context. I took this feedback seriously and started engaging more with business stakeholders to understand their priorities before diving into technical solutions.

8. A time when you went above and beyond your job responsibilities.
In one of my projects, when we were facing resource constraints, I voluntarily took on additional responsibilities such as performing data validation and ensuring the robustness of the monitoring system. This wasn't part of my initial role, but I knew it was crucial for the project's success.

9. A time when you didn't have enough data and had to use judgment to make a decision.
When automating the data ingestion process with Cloud Functions, we lacked sufficient data on optimal file processing times. Based on limited available data and my understanding of our system's performance, I chose a more conservative configuration that would ensure stability while giving us room to adjust as more data became available.

10. A time when you helped someone grow in their career, both when it benefited and didn't benefit you.
In a previous role, I mentored a junior developer who was eager to learn cloud technologies. Over time, I saw them transition from a beginner to a key contributor in cloud-based data pipelines. While it initially benefited the team, they eventually left for a higher position elsewhere. Despite the loss, I felt proud to have contributed to their growth.

11. A time when you were 75% through a project and realized you had the wrong goal.
During the design phase of a data pipeline, we realized late in the project that our primary metric of success didn't align with business goals. Instead of just optimizing data loading time, the focus needed to be on data quality. We shifted our efforts and implemented additional DQ checks using Apache Deequ.

12. A time when your team wasn't supporting an optimal solution but you pushed for it.
In one instance, my team was hesitant to adopt a more complex partitioning strategy in BigQuery due to its perceived difficulty. I made a strong case based on long-term performance gains and eventually convinced the team to implement the strategy.

13. A time you pushed back on management's decision for better long-term benefits.
Our management wanted to implement a quick fix for data transformation performance issues, but I pushed back, advocating for a more sustainable solution involving refactoring the PySpark job. I presented detailed projections on the long-term benefits.

14. A time you failed to meet a commitment and how you handled it.
I once failed to meet a deadline for optimizing an ETL job in Databricks. I immediately communicated the delay, explained the challenges I faced, and provided an updated timeline. I also worked late hours to

catch up and mitigate any downstream impacts.

15. Tell me about a project you're proud of.
I'm particularly proud of a recent project where I led the design and implementation of a scalable data pipeline on GCP. The pipeline automated data ingestion, validation, and transformation into Delta format in the Silver layer, and performed aggregations in BigQuery for the Gold layer.

Challenges: We faced issues with data quality and performance, but I proposed solutions involving Apache Deequ for DQ checks and partitioning strategies for BigQuery to optimize queries.

16. A time when you faced a tight deadline or time constraints.
While working on a critical project, we were given an aggressive timeline to deliver a fully operational data pipeline on GCP. I coordinated closely with my team, prioritized the most essential tasks, and delegated effectively. We also worked extended hours to meet the deadline.

17. A time you had multiple solutions and had to select an optimal one.
When deciding how to aggregate data in BigQuery for a reporting dashboard, I evaluated multiple approaches. One solution focused on speed, while the other on long-term maintainability. I chose a balanced approach, favoring a solution that offered good performance while remaining maintainable for future updates.

18. A time when you innovated and exceeded expectations.
I designed a custom solution for monitoring our GCP data pipelines by integrating Cloud Logging with a centralized dashboard. This went beyond my initial scope, but I saw an opportunity to improve how we monitored and resolved issues.

19. A time when you faced a setback but still achieved your goal.
In one project, we encountered performance bottlenecks while loading data into BigQuery due to unexpected schema changes. After diagnosing the issue, I implemented dynamic schema handling and adjusted the data loading process.