

# How to Prepare for Your Data Engineering Interview

Got a callback from your dream company, but not sure what to expect and how to prepare for the next steps? We are here to help!



[Masha Danilenko](#)

[Follow](#)

[Aug 20, 2019](#) · 7 min read

At [Insight](#), we have successfully prepared hundreds of our Fellows to transition to Data Engineer and Machine Learning Engineer positions at their dream companies. In an interview for any Engineering role, the interviewer wants to understand if you have good analytical skills, problem-solving ability, communication, work culture and ability to build technical solutions. Specific to Data Engineering, they also want to understand if you have the skills to handle large data and build scalable and robust systems. In this article, we will cover how to best prepare and perform at each type of Data Engineering interview, ranging from algorithms, system design, SQL questions, to the essential behavioral component.

Phone screen	Take-home exam	Onsite
<ul style="list-style-type: none"><li>• HR screen</li><li>• Technical screen</li></ul>	<ul style="list-style-type: none"><li>• Timed hackerrank</li><li>• Coding challenge</li></ul>	<ul style="list-style-type: none"><li>• Algorithms and data structures - whiteboarding</li><li>• System design</li><li>• RDBMS optimization</li><li>• Cultural fit</li></ul>

The typical Data Engineering interview process.

## Phone Screens

There are two types of phone screens: HR, which is generally all behavioral questions, and technical phone screens.

The HR phone screen is usually 15–30 minutes and conducted by non-technical staff at the company, such as a recruiter. You'll be asked soft questions such as *Why do you want to be a Data Engineer? Where do you see yourself in 5 years? Why do you want to work at our company?* And importantly, *what salary are you expecting?* These questions can seem boring or odd if you don't know the real reason for them behind the scenes: HR wants to find the right fit for their team. They want a candidate who will be communicating well with their peers and managers and stay at the company for a long time because hiring and onboarding are expensive!

Just as the HR phone screen is a filter for basic communication ability, the technical phone screen is a filter for basic technical ability. On-site interviews are very costly in terms of time and team resources, so companies don't want to spend hours on a candidate who can't code well. An assessment of basic SWE knowledge and the ability to break down complicated ideas to smaller understandable pieces are the most essential reasons for technical phone screens.

## HR Phone Screen Summary

**Expect** a 15–30 minute teleconference call discussing your background, goals, and interest in their company.

**They are looking for** clear communication, a pleasant person to work with, someone who is enthusiastic about the company

and has done their research, ideally translating into a loyal employee willing to stay and be happy at their company.

**Example questions include** tell me about your background. Why do you want to be a Data Engineer at [company]? What is your desired salary range?

### **To prepare:**

1. Write and practice a script for your background.
2. Do a deep dive into company values and tweak your answer accordingly.
3. Practice with your peers over the phone (we know it can be awkward).
4. Settle in a quiet place with a good Internet connection at least 10 minutes before the interview.

## **Technical Phone Screen Summary**

**Expect** a 30–60 minute teleconference call answering basic DE concepts or classic SWE questions, usually from a member of the engineering team.

**They are looking for** people with basic knowledge in SWE and DE, problem-solving skills, and ability to communicate technical information.

**Example questions include** what are linked lists? How would you code them in your language of choice? Find all duplicates in a list. When would you use SQL vs. NoSQL databases?

### **To prepare:**

1. Read [Data Engineering Cookbook](#) and answer at least 50 questions.
2. Practice random questions from the book with your peers.

3. Do 50 easy [LeetCode](#) problems.
  4. Settle in a quiet place with good Internet connection at least 10 minutes before the interview.
- 

## Take-Home Exams

Your resume says you have many years of experience, leading multiple projects and being a rock star. How do companies know if you're really that good? In most cases, there is no access to your old company GitHub repository, and it takes time to read and understand personal GitHub projects — not to mention they won't know for sure that you wrote the code. A take-home coding challenge is the easiest and fastest way to assess how production-ready your code is, how you account for edge cases and exception handling, and whether you can solve a given problem in an optimal way. There are two main types of exams:

### Timed Hackerrank

**Expect** 1.5–2 hours exam with 3–5 easy-medium HackerRank questions including SQL, regular expressions, algorithms, and data structures

**They are looking for** engineers who know efficient algorithms and data structures for solving standard computer science questions, take edge cases into account, and provide the solution quickly

### To prepare:

1. Solve at least 100 LeetCode/HackerRank problems
2. Practice with [Virtual Leetcode Contests](#) — all free past contest that you can take any time, and try to solve problems quickly and correctly on the first try

3. Block off a chunk of time where you'll be in a comfortable environment where you usually do technical work and make sure you won't be interrupted and have plenty of water and snacks (if needed).

## Coding Challenge

**Expect** 1–7 days to write code to answer 1–10 questions on 1–3 datasets, push it to your GitHub repository and submit the link.

**They are looking for** clean and modular code, good README with clear delivered ideas, unit tests, and exception handling.

**Example question** Clean and analyze a dataset of employee salaries and locations. What is the distribution of salaries at different locations? Write a SQL query to do the equivalent task.

### To prepare

1. Read and internalize the [Google style guide](#).
  2. Practice using the [unittest](#) library in Python or equivalent.
  3. Read [GitHub best practices](#).
- 

## On-Site Interview

You should feel very accomplished if you get to the on-site interview, but the hardest part is yet to come! On-sites can be grueling affairs of interviewing with 4–10 people in 3–6 hours, especially if you're not prepared. Knowing what to expect and doing realistic preparation beforehand go a long way toward reducing fear and nervousness.

## Whiteboard Algorithms and Data Structures

This is the most common type of interview, because of knowledge of algorithms and data structures is crucial for cost- and time-efficient code. It's usually done on the whiteboard to evaluate your coding skills with no IDE/Stack Overflow and your technical communication skills.

**Expect** 30–45 minutes interview with 1–2 medium-hard questions to solve on the fly on a whiteboard, constantly communicating requirements and solutions with the interviewer.

**They are looking for** your preparation before the interview, knowledge of basics and great communication. Don't work on the problem in silence — make it a conversation between you and the interviewer.

### **To prepare**

1. Solve 80–150 LeetCode problems on paper/whiteboard
2. Get at least 20 practice sessions as an interviewee with peers or professionals.
3. Practice writing clean, readable code on a whiteboard.

## **Whiteboard System Design**

As a Data Engineer, on a day to day basis you are going to design entire systems from scratch or add small features to giant existing pipelines. Even you mention these skills on your resume, it's crucial for companies to check your ability in a real-life.

**Expect** 30–45 minutes interview to design a data engineering system to spec.

**Example questions:** Design Twitter — what are the system blocks needed? What database and schema would you use?

What about caching and load balancing? What are the tradeoffs of a system?

**They are looking for** your ability to clearly communicate and scope down requirements, design a concept-level pipeline, and knowledge of distributed systems basics.

### **To prepare**

1. Read [Data Engineering Cookbook](#), the [Data Engineering Ecosystem](#) and [Grokking the System Design Interview](#).
2. Practice at least 10 different questions on a whiteboard with peers or mentors.
3. Practice drawing clean, readable systems diagrams on the whiteboard.

## **SQL**

Most companies are language agnostic. You can transition from Scala to Python to Java, but a deep understanding of SQL is fundamental and irreplaceable to database work, even NoSQL databases. That's why more than 50% of companies have this type of interview as a part of the Data Engineering on-site.

**Expect** a 30–45 minutes interview with 1–3 hard [HackerRank SQL](#) questions + normalization, indexing, ANALYZE-EXPLAIN for queries

**Example questions include** print the nth largest entry of a table column.

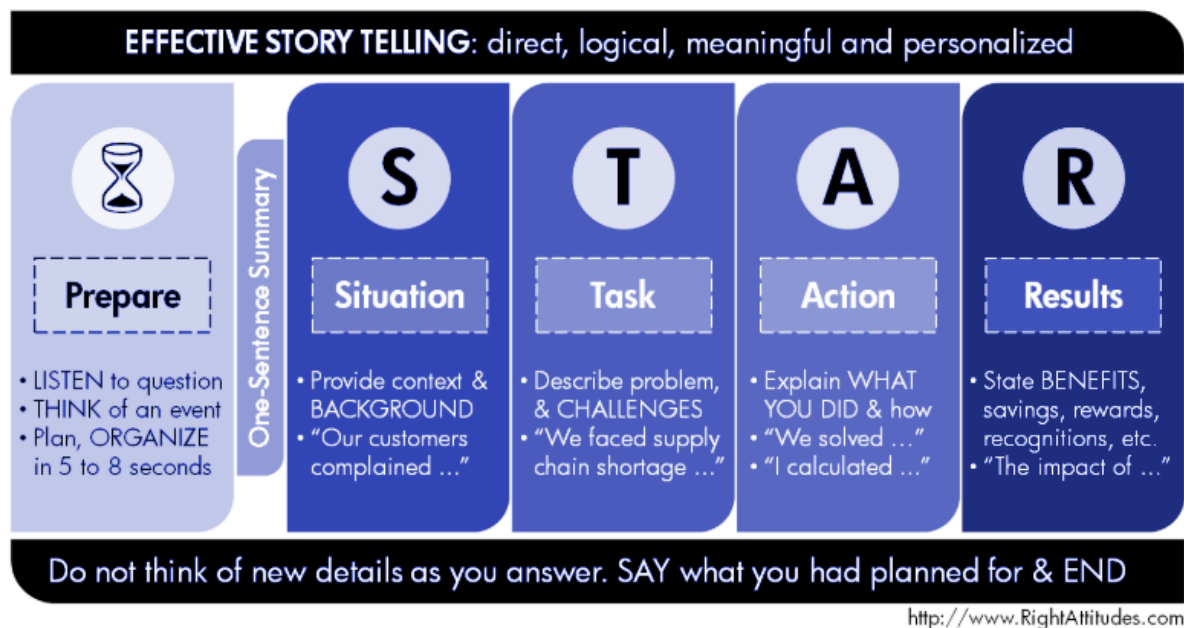
**They are looking for** your ability to write queries and optimize their existing RDBMS.

### **To prepare**

1. Practice the [57 SQL questions](#) book.
2. Google and learn about *Query Optimization*.

## Cultural fit

It's very important to be technically strong and knowledgeable, but it's not enough! If you can't deliver your brilliant ideas, then no one else can understand and use them. Behavioral types of interviews, such as cultural fit, are meant to show how you can tell your story and explain how you've handled tough workplace situations.



The STAR method.

**Expect** a 30–45 minutes interview with 2–4 questions about your past situations.

**They are looking for** consistent and complete answers using the STAR (situation, task, action, result) method.

**Example questions include** tell us about a time at work where you had a big deadline. Tell us about a time when you had a conflict with another team member. How did you handle these situations?

### To prepare

1. Practice at least 30 [cultural fit interview questions](#) alone,



writing scripts and recording yourself if needed.

2. Practice at least 10 questions with peers, following the STAR method.

## TLDR

If you prefer watching videos, here is the webinar we gave on the same topic:

<https://youtu.be/2XppoicnUIg>

### Towards Data Science

A Medium publication sharing concepts, ideas, and codes.

[Follow](#)

480

- [Interview](#)
- [Data Engineering](#)
- [Data Science](#)
- [Coding Interviews](#)
- [Interview Questions](#)

480 claps



WRITTEN BY

**[Masha Danilen](#)**