

---

## Data Scientist interview questions and answers

This **Data Scientist** interview profile brings together a snapshot of what to look for in candidates with a balanced sample of suitable data scientist interview questions. These questions will help you assess the skills and experience of candidates for highly technical data science roles. Similar job titles include **Certified Data Scientist**, **Big Data Scientist**, **Data Architect** and **Data Engineer**.

### Data Scientist Interview Questions (Coding)

[Data scientists](#) are more than simply data analysts, in that they understand how studying some data could lead to an important decision that can enhance a product or improve a business. Moreover, skilled data scientists can effectively gauge the complexity of a particular approach to a problem. They can propose alternative solutions to a given task depending on the available time and resources, thus allowing them to whip up a simple but functional approach with short notice and implement a more elaborate design under a larger time frame. These interview questions for data scientists will consider both a candidate's background in computer science, and their specific skills that suit them for the role.

The [data scientist role that emphasizes coding](#) targets candidates with strong software engineering skills that understand the tools, processes and exigencies of creating and maintaining software that will be deployed to production. This type of data scientist has solid programming skills in a programming language such as C++, Java or Scala, is very knowledgeable in databases, and will have worked with platforms for deploying machine learning solutions in the real world such as Azure ML or PredictionIO. In addition, a frequent requirement of the role is experience in working with big data and platforms such as Apache Spark and Hadoop. The ideal background for this type of role data scientist is computer science, but candidates with engineering and mathematical backgrounds sometimes develop strength in practical software engineering skills in order to arrive at this role.

A thorough data science interview contains a combination of data science, big data, analytics, modeling and analysis interview questions.

### Computer Science questions

*(Programming knowledge)*

- Do you contribute to any open source projects?
- With which programming languages and environments are you most comfortable working?
- Have you used any online platforms for machine learning such as Azure ML or PredictionIO?

- How would you train and deploy a logistic regression model? A recommender system?
- Describe a data science project with a substantial programming component in which you have worked?
- How would you sort a large list of numbers?
- What is hashing? Give an example of when you might want to use it
- What is dynamic programming? What is recursion?

*(Software engineering)*

- How do you test your code? What kind of tests do you write?
- How would you monitor that the performance of a model you trained does not degrade over time?
- Suppose you wanted to keep a record of some computations that your model performs while in production. How would you go about doing this?
- Are you familiar with version control? What tools and processes have you used for this?
- What are software patterns? With which patterns are you familiar? When might you use a Factory/Singleton/Memento/Builder/DAO etc. pattern?
- Have you ever worked within a developer team that followed a particular agile process?
- What is technical debt, how does one mitigate it, and how relevant is this to deploying data driven models in the real world?
- How might you deploy a model that was training in an environment such as R? Are you familiar with PMML?

## Role-specific questions

*(Big data and distributed computing)*

- In the map-reduce paradigm, what does the map function do and what does the reduce function do? What do the combiner and partitioner do?
- How would you build a search engine for a very large collection of documents?
- Are you familiar with technologies from the Hadoop stack (Hadoop, Pig, Hive etc...)?
- With what distributed environments have you worked?

**For more data science questions that emphasize technical background in machine learning and statistics, check out the interview questions for the [data scientist \(analysis\) role](#).**

**For additional technical interview questions, see our sample [coding interview questions](#).**