|  |  |
| --- | --- |
| The Data Team  So the first thing to know when building a data science team, is that data science is a team sport. It takes a large group of people working together to solve real,  practical data science projects. So those people include the data scientists.  The managers of those data scientists, data engineers who perform and develop the infrastructure. And then people who they have contact with outside of their data science team. So people within the organization or management of the organization,  and how do they communicate with those people?  So I just thought I'd go through the key positions in the data science team.  And so the | |
| data engineer: | So, a data engineer is a person that would maybe construct a database, or they might pull data out of that database for people to analyze.  Or they might build production level machine learning algorithms and  implement them on servers, so that they can be applied to a huge database of observations of data, and maybe even run in real time.  So they have skills in infrastructure development, both the hardware side and the software side. And they do less of the actually analysis of the day to day data. That falls to the person that's called a data scientist. |
| data scientist | So a might be somebody that would go and pull data out of a database.  Then they might analyze it, perform experiments on it, visualize it and  communicate those results to the data science manager, and  to other people in the organization to try and move things forward.  Often, a data scientist will hand off any machine learning algorithm or  prediction algorithm they develop, to the data engineer,  who will then implement it so it can run at scale. |
| data science manager | The third key person is the, so that would be you.  The data science manager, makes sure that everybody interacts with each other and  that they keep things moving. |
| And then they also interface with upper management in an organization.  As well as with collaborators that are at their same level across the organization,  to make sure that they get all the information across.  And they advertise the data science team to other people,  what their capabilities are, what they can do, and how they can help out.  So that's the data science team, and  then how do they work together?  They work together as a unit, and so often each of these people are working  on individual projects, or individual subproblems of a data science problem.  And then they come together and have joint group meetings and joint presentations,  where they discuss their ideas.  They also interact with external folks, they might interact with those people  directly through individual meetings with people outside of the data science team.  Or they might interact through the data science manager.  So, in either way they have to be able to communicate what's going on with the data  science infrastructure, what's going on with the data science team.  And how do they make sure that all the problems are getting solved | |

|  |  |
| --- | --- |
| When Do You Need Data Science? | |
| So, the way in which you'll interact with data science depends a little bit on what kind of organization you are.  To some extent, it depends a lot on the size of your organization.  So, when you're just a start up, when you're an early stage company, or  you're just one person with a very small team, maybe like one other person working for you. You may not need to worry so much yet about how to do experimentation, how to do machine learning, how to do sort of prediction and downstream calculations. The first order of business is just making sure your sort of data house is in order.  And the way to do that is to make sure you focus on infrastructure.  So the first thing that you need to do is build out the infrastructure for  storing the data, the databases and so forth. The software that's gonna be run to pull those data, the servers that are gonna serve the data to other people, and the servers that you'll interact with yourself in order to get the data out. So all that sort of requires infrastructure building up at first. So often the first people that you get to hire into a data science team, are not people that you would necessarily called data scientists in the sense that  they're not analyzing the data, they're not doing machine learning.  They might do a little bit of that but mostly they're involved on just making sure the machine is running, making sure the data's getting collected, its secure, its stored and so fourth. |  |
| So when your a mid size organization,  then hopefully you've got the basic infrastructure in place.  And you can start thinking about building out your real data science team. And so to do that you can bring on board people that are actually called data scientists. And those are the folks who will then actually use the data. So they might pull it out of the database.  They might run some experiments. They might build machine learning algorithms. They might analyze the data to see if you can identify any patterns or trends in behavior that you care about.  And so at that point, you're thinking about actually building out the data science team. You're also thinking about implementing these data science ideas and products. So again, the data scientist might build something like a machine learning algorithm that predicts, say, consumer behavior. Once you have that algorithm built out,  you might need to implement it back on to your system.  And you might need to scale it up, so that it can be run on the whole data set. You might wanna build some sort of visualization,  that people who aren't necessarily data scientists can interact with.  And so, that would be turning it back over to the data engineering team. So there's still infrastructure concerns,  because you have a large set of data that you've hopefully collected at this point. You need to be secure about it, you need to have a database, you need to be able to scale it. But, now you're sort of graduating into a more complete view of data science. |  |
| For a large organization you have all those same sorts of things.  You now have a data infrastructure,  you might have a data science team that's running experiments.  You may be using those experiments to make decisions.  But now you have one additional component which is really managing the team and keeping everybody on task and coordinated.  So at this point the data science manager role becomes a little bit more involved, in the sense that you might be coordinating multiple teams of data scientists working on different projects.  You might have a team that works exclusively on building machine  learning projects. You might have another team that works exclusively on running experiments and inferring what you can from those experiments.  And then someone has to be in charge of coordinating those activities making sure they're connected to the right people within your organization. Whether that's the marketing team, the business group or whoever else that you're collaborating with. You have to be able to connect those people. And so at that scale the full data science infrastructure is in place. |  |