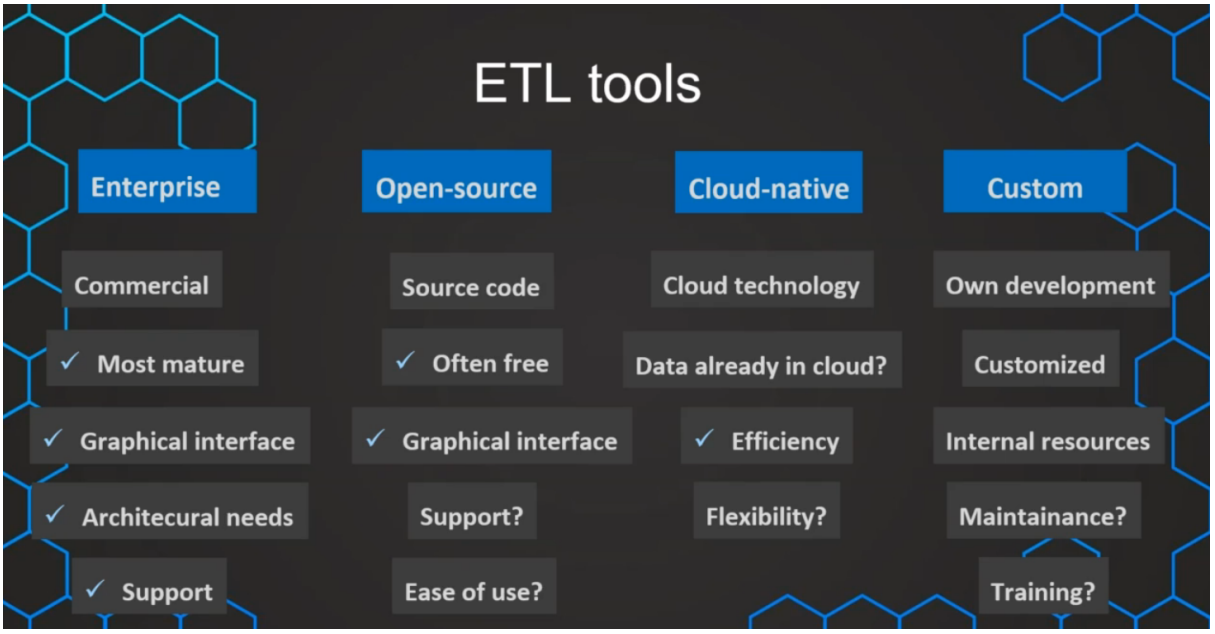
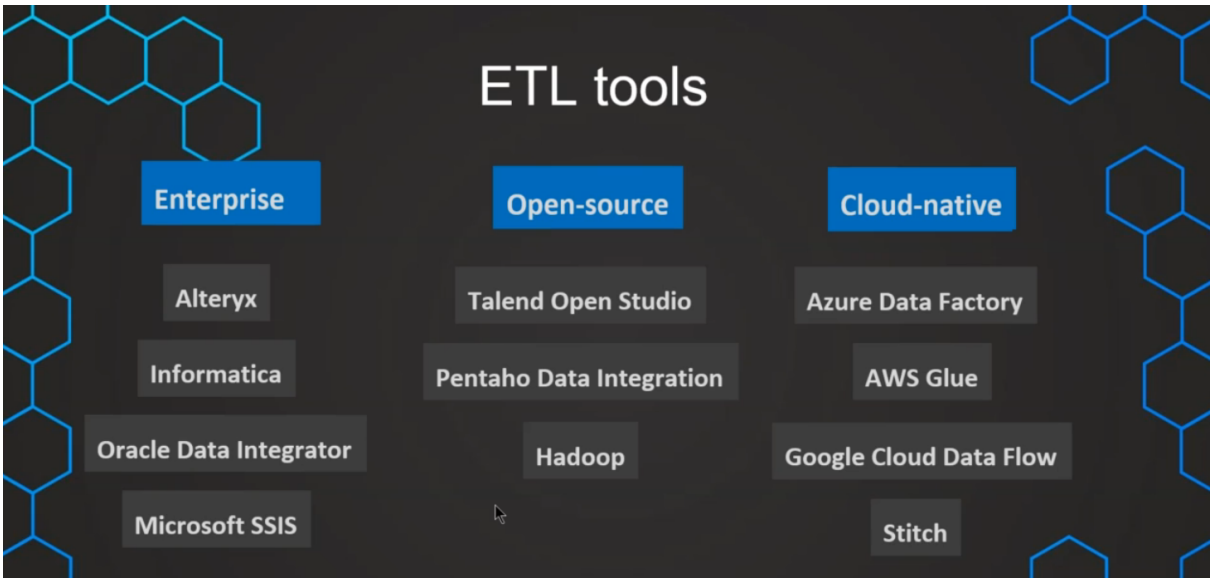
09. ETL Tools



- At one point in time, we might be faced with the question what ETL tool should we go with in our company? And that's why we now quickly want to discuss the different types of ETL tools and what is different about them.
- And then afterwards we also quickly understand how we can go about evaluating those tools and finding the right tool for us. So which is the best fit for us in our company?
- But now we want to start with the different types or categories of those ETL tools.
- So of course in the beginning we have the enterprise tools. Those are commercial tools that are offered and developed by companies, and they offer it, of course under a certain price. And they, we have to say, are the most mature tools. They have very nice interfaces. They are usually very user-friendly and they can fulfill all of the requirements usually so they can connect to all of the different data sources. And of course, also they offer a support. So this is of course very important if we are working in a large company.
- And if the things and the ETLs that we are developing are very crucial to the success of our company, then it's very crucial to have someone responsible that we can go to if some things are not working in this tool.

- And then we of course also have open-source solutions and they just make the code publicly accessible. This can just help to understand how this tool works exactly, and this just can give additional trust that the source code is publicly available. So everyone knows and understands how this tool works. Are there any security issues?
- And also I want to mention that open-source does not necessarily mean that it's free but oftentimes those open-source solutions are indeed free. And on the other side, of course also enterprise tools can have open-source solutions so they can also make their code publicly available.
- And also those enterprise tools can of course also offer free versions. So this is also important to understand. And also I want to say that those open-source tools also nowadays are quite mature and have usually a graphical interface.
- But if we are using a free tool, we have to understand that usually there's no support that's coming along with that. And also the ease of use in those free tools. This can vary quite a bit. So this is something that we also should keep in mind.
- And then of course, nowadays also everything is going to the cloud. And all of these big cloud providers, they also offer different solutions for that. So we have to ask the question if our data is already in the cloud with a big cloud service provider.
- So if we are working with AWS then also of course they offer cloud-native solutions. So this can make things then much more efficient. But of course, if we now have other cloud service providers maybe we also work on Azure then we have to also ask the question how flexible are we to also get data and work with data from other cloud service providers?
- So this is something that we also should then keep in mind. And in the end, I also want to mention that some companies can also develop their own solutions. But of course this is usually grown from history that someone needed a connector and then it was just developed and they continue to develop it.
- But this is, in my opinion not really the solution that we should aim for. Because first it takes a lot of resources. The solution is maybe not so mature and also someone has to maintain this solution and people should be also trained.

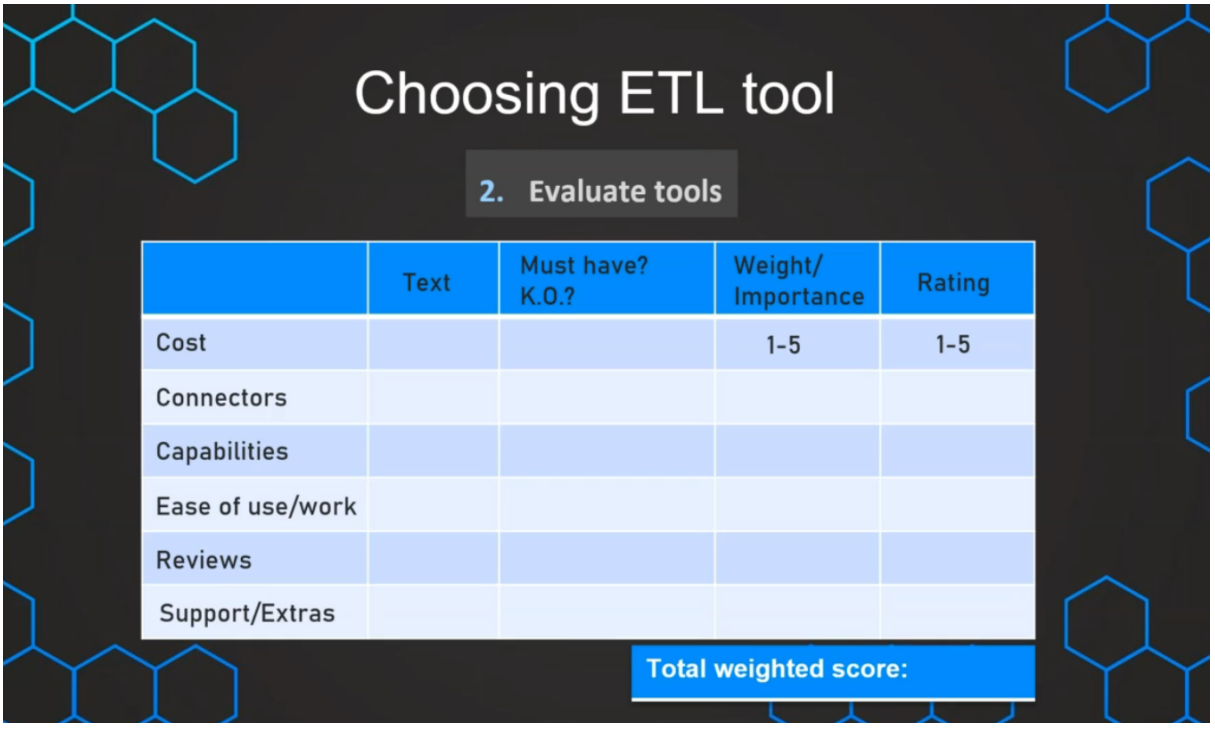
Choosing right ETL tool



- So of course in the beginning, if we are not using an ETL tool before, we should evaluate the current situation in our company. What are the needs and why do we actually want to use an ETL tool?
- So what are the use cases that we have? What are the things that we need? We can ask very general questions. What things do we want to improve? So what are things that are important for us? Also, what are some must-haves?
- So what data sources do we need to connect to and with what other tools our ETL tool needs to integrate for sure? So we should define really our requirements. And then we can also define in the company some responsible people that are responsible for defining those requirements,

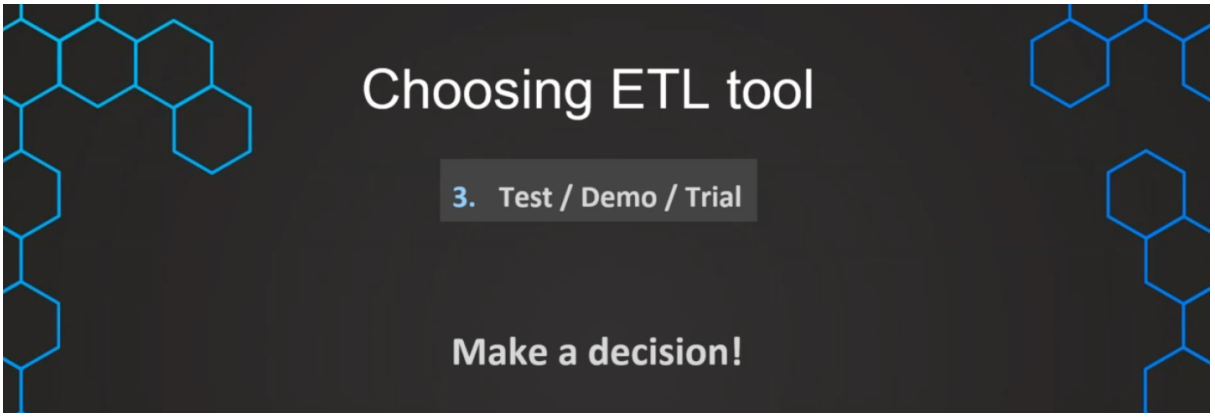
and then also for evaluating the ETL tool and of course it can be one person, but it can also be a team of people that are then evaluating the ETL.

- Of course, what we should also keep in mind for the evaluation of our situation is, who are the users of this ETL tool? Do they have experience with it or are they maybe more business users? And things need to be a little bit easier to understand and to work with. And then once we have evaluated our situation, we can focus on the evaluation of the tool and we can create a list.



- So these are some things that are in general important. So we can create a matrix, look at all of these tools, and evaluate these different categories.
- And again, these are some categories that are important, but of course, this is not an exhaustive list. But of course oftentimes the cost, the budget plays a role and also what are the must-have connectors and what is the range of the connectors in these different ETL tools.
- Maybe sometimes there are also some additional capabilities that we must have. And there we can also just rate these different tools for what they are able to do.
- One important factor, of course, is also, how easy is it to work with this tool?

- Because we spend probably a lot of time with this tool and we also have to keep in mind what are the users and how technically experienced they are.
- Once we have done that, we can also of course have a look at the reviews. So what do customers say about these tools? How are they rated? And we can also look at other companies, for example, there's Gartner the Magic Quadrant.
- They also review and evaluate those tools. So we can also have a look at their reviews and their evaluations. So this is also something that can help. And of course, there are sometimes are some additional benefits.
- How is the support for these tools? And are there maybe some special extras like a dedicated person that is helping with the training, and so on so forth?
- Of course then we can evaluate all of these categories. And maybe some of the tools, if they don't fulfill those must-haves,we can immediately decide to not consider them anymore.
- And then, with the left tools that are all fulfilling our must-haves, we can now of course create different importance levels.So for example, we can define a weight of importance. So maybe five is super important factor, and one is maybe a factor that is important but we don't want to give it a super high weight.
- And now we can then rate all of those categories. And again, here, a rating between one and five is a good compromise of having enough rating options, but not making it too complex with a rating of between zero and 100, for example. And with that weight and rating, we can then rate all of those categories, and in the end, come to a total weighted score.
- And this might now just help us with finding the best tool for us that has the best fit for us. Of course, again, there can be different approaches to evaluate these tools, but in my opinion and my experience, this is a good strategy of evaluating a tool.
- And then once we have evaluated the tools, the next stage is of course to contact the providers and these companies with these ETL tools and talk about them for testing, getting a demonstration, and maybe also to get a proof of concept, some free trial.



- And then we should of course test those tools. And then only after we have tested those tools, we should make our decision.
- Because this is of course quite some commitment then in terms of cost, time, and energy, so all of the resources that we spend. And that's why it's good that we also have a proof of concept or a trial so that we can really be sure to commit to our ETL tool. And then hopefully we have found the right tool for us.