Transcript

00:00:02 Interviewer

OK. So yeah, so now it is recording. Thank you so much for allowing me to record this session so.

00:00:11 Interviewer

I have few questions to ask so it would be nice if you just introduce yourself and describe your job role in this company. If you are working right now.

00:00:27 Interviewee

Sure. So do I need to mention the?

00:00:29 Interviewee

Company name as well.

00:00:31 Interviewer

It's fine if you don't want it. It can be confidential. It is fine if it it, it doesn't matter.

00:00:37 Interviewer

Like if you want to be if you feel like you don't need to share those things, this is totally acceptable. Like you don't need to forcefully share that. You just say, like, OK, like, yeah, fine.

00:00:51 Interviewee

It's fine. I can. I can mention the company name as well. So basically I work in a company called ” “.

00:00:57 Interviewee

It's in India, so it's a US based company. It's a startup basically. So my role is a machine learning architect wherein I prepare the solution, designing of machine learning part like we have different systems like UI, the API and the machine learning. So I designed the machine learning in.

00:01:12 Interviewee

Integration. So basically I work in natural language processing, so NLP problems. The real world problems like document classification, question answering, text classification, entity extraction, all those NLP related problems only because I am actually assigned to that documentation AI.

00:01:28 Interviewee

Ability. So in and out my role would be like starting from GTM activities like presel activities to fix the product to show our capabilities to the client that we are capable of dealing with such and such NLP things and we have this expectation starting from that pre selectivity till you know once we have the projects to create the Pocs proof of concepts for the client.

00:01:49 Interviewee

Design the architectures, the solution architecture, develop that architecture, test data at our end and deliver it to the client.

00:01:56 Interviewee

Once client client is happy with the POS then we go for production action. So this is my overall role in the in the company so.

00:01:57 Interviewer

OK.

00:02:04 Interviewer

Very interesting.

00:02:06 Interviewer

Yeah. So yeah, this is very interesting. Since how many years you are working like in this in this company or maybe over your over our experience?

00:02:19 Interviewee

So not not so far like it's it's around 5:00 to six months of experience I have in this organisation.

00:02:25 Interviewer

OK. Yeah.

00:02:26 Interviewee

So yeah, and before that actually I was part of some other organisations. I was part of some of the startups, some of the MNC's as well, where I dealt with all these machine learning projects.

00:02:37 Interviewee

So basically I was into data science role. As always I started as data scientist and senior data scientist and intensified. Did couple of internships with some of the organisations.

00:02:40 Interviewer

Oh, OK. Interesting. Yeah.

00:02:47 Interviewee

While doing my masters so and then I I am like here was a machine learning yeah.

00:02:53 Interviewer

Wow, OK. So you have, we can say you have one year experience or two years in general?

00:03:02 Interviewee

So total I have around 10 years of work experience, but into data science I have just 4.5 years experience.

00:03:05 Interviewer

Wow, OK.

00:03:09 Interviewer

4.510 years old. Yeah. Experience, but.

00:03:15 Interviewer

Their presence.

00:03:19 Interviewer

Thanks. So have you published any thesis in machine learning domain?

00:03:38 Interviewee

Sir, could you tell that again because your video just freezed I.

00:03:41 Interviewee

Couldn't hear properly.

00:03:42 Interviewer

Yeah, OK, OK, fine. So have you published any thesis in machine learning?

00:03:55 Interviewer

Can you hear me?

00:03:58 Interviewee

Yes. So I I have like in my masters I have worked on my thesis but I have not published.

00:04:07 Interviewee

Actually I have worked on flood intensity prediction, but I I wrote my thesis on that. But I didn't publish it.

00:04:16 Interviewee

I really didn't get that, so I have not.

00:04:18 Interviewer

OK.

00:04:21 Interviewer

So can you share your experience in your current position as I think you have already shared your current position, so 4th question has already been explained by you.

00:04:33 Interviewer

So do you have any experience in the previous company? This is already explained that you have 4.5.

00:04:38 Interviewer

Years of data science.

00:04:40 Interviewer

Is your company is service based or product based?

00:04:46 Interviewee

So I have always worked on service based companies. So far I didn't get any chance to work on any other product based companies.

00:04:52 Interviewer

So you have a, so your company is a service based the the startup company.

00:05:03 Interviewer

So which software development model do you practise in your company? In journal like Agile or Waterfall?

00:05:11 Interviewee

We started with waterfall like before 2016. I would say all the projects were at Waterfall that I had worked on and since then I have been working in Agile 2017. Actually 2017 always I've been.

00:05:20 Interviewer

OK.

00:05:23 Interviewee

Working in agile.

00:05:27 Interviewer

So could you please share your experience with the interesting project that you worked on recently?

00:06:29 Interviewer

Hi. So you're back.

00:06:32 Interviewer

I was thinking to send you a message.

00:06:38 Interviewer

I you are muted.

00:06:41 Interviewer

So you are muted.

00:06:45 Interviewer

Yeah. Now. Yeah, now, now, because you was muted. So it's fine. I can hear you now.

00:06:46 Interviewee

Is my voice audible?

00:06:50 Interviewee

Because I believe my very low, so I really apologise for that.

00:06:54 Interviewer

Oh, no, no, this is fine. This is totally understandable. Yeah, no problem.

00:07:00 Interviewer

So can you explain that any interesting projects that you've worked recently in machine learning?

00:07:10 Interviewee

Yes. So as I said in my introduction itself, I deal with basically documentations problems, the NLP problems. So it's like a client giving me a bunch of data where I have to identify whether a document is belongs to a particular type, whether it is your income statement, whether it is your.

00:07:16

OK.

00:07:26 Interviewee

Bank statement or it is your agreement lease agreement. So all those type of documentations you have to classify first.

00:07:33 Interviewee

And once the documentation is done, classification is done.

00:07:35 Interviewee

Let's say I have identified classified your driver's licence. That. Yeah, this document is a driving licence. So again, I have to extract the entities out of it. Like your name. Is this your driving licence expiry date is this?

00:07:46 Interviewee

So that comes under entity extraction. So these are all the text classifications that we do. So basically I very much deal with all these problems and document classification and entity extraction.

00:07:57 Interviewee

So for that we use most of the delivered Google parsers this, this, these are all the projects that back-to-back.

00:08:03 Interviewee

I have been doing.

00:08:04 Interviewer

OK.

00:08:04 Interviewer

Nice. So which common software architecture design technique you found being used in most company through your experience?

00:08:18 Interviewer

OK, so which common software architecture design techniques you found being used in most company through your experience?

00:08:37 Interviewer

Maybe you can if you have a laptop in front of you, you can open the questionnaire which I have shared with you, so you can also.

00:08:53 Interviewer

Hi you are muted.

00:08:56 Interviewer

Yeah. So I maybe it will be nice if you have. If you open up the questionnaire which I have shared with you on your laptop, so you can also read those. And because I'm asking the same questions in the questionnaire document.

00:09:15 Interviewee

OK, OK. Got it. Just open that.

00:09:23 Interviewee

Yes, so.

00:09:26 Interviewer

I am asking.

00:09:28 Interviewee

Load properly.

00:09:29 Interviewer

The question #10 right now.

00:09:31 Interviewee

Yeah, I got the.

00:09:34 Interviewee

OK, so which components, software, architecture, design, technique OK so I would say in my company basically what software design technique I use is mostly the serverless architecture. The microservices if you if you're aware of like we we we don't really deal with the monolithic.

00:09:56 Interviewee

Applications. So we we work on serverless architectures like.

00:10:00 Interviewee

If you say for a POC level we we handle around 3000 maximum you know parallel processing. So we basically use the cloud functions in GCP or even the Lambda functions and so I would say a server plus Microsoft microservice architecture that is that basically we use in in my current organisation.

00:10:18 Interviewer

OK. So you you also use this service less like microservice architecture and also the Lambda architecture.

00:10:28 Interviewee

So I'll I'll answer the 11th question, right. According to your experience, which are the best software design makes benefit for?

00:10:33 Interviewer

Yes. Yeah.

00:10:37 Interviewee

Your organisation. OK, so for this I would say the best as in basically we started using the blackboard and then we move on to all the way to micro.

00:10:48 Interviewee

Services in between. We had a lot of monolithic servers that were basically designed between 2014 to 2015, but since then we have been, you know, dealing with these microservices like every client basically wants us to, you know, focus on this microservices no matter what, whatever maybe the requirements.

00:11:08 Interviewee

And though it is, you're not achievable within within the monolithic server still then they they just want that microservice flexibility.

00:11:18 Interviewee

You know, on top of which we create the pipelines and all the philtres that we we need. So I would say, yeah, microservices were at least my organisation dealer.

00:11:26 Interviewee

I I have not seen this solutioning part of our other organisations either have a discussion with all my any other other colleagues in other organisations but the organisers, which I have been part of, we basically deal with this microservice and we do encourage everyone to, you know, use these things only.

00:11:43 Interviewer

OK. Yeah. Impressive. Yeah. Yes, yeah. Yeah, it is. Yeah. Like you. You are recommend. Like you. You found that microservice is the best architecture according to you.

00:11:45 Interviewee

Does that answer your question?

00:11:54 Interviewer

So 12th.

00:11:55 Interviewee

Yes, right. According to me. So basically when I talk this, I talk about the API like hosting a web service or basically I'm talking in terms of machine learning stuffs.

00:12:06 Interviewee

So maybe with context to that, I would encourage this, but I'm really not sure of other other technologies, anything other than machine learning.

00:12:13 Interviewer

Yeah, because yeah, we are actually in this thesis. We are focusing on the machine learning systems, not only the other one. So we have not concerned about the other systems. How does it does? Yeah.

00:12:20 Interviewee

Got it, got it. Yes, got.

00:12:23 Interviewer

So yeah, we go to the 12th question, maybe now.

00:12:28 Interviewee

Yes, yes. So the 12 question asks, do you have any recommendation of software architecture, design techniques in machine learning?

00:12:37 Interviewee

So this is this is the similar answer that I gave as part.

00:12:39 Interviewee

Of personal label.

00:12:40 Interviewee

So I would always recommend the server microservices microservice architecture rather than monolithic or any peer-to-peer filtering.

00:12:48 Interviewee

Anything, or even the the older ones like the, you know, the black, the blackboard ones or or the data centric or event driven I would.

00:12:58 Interviewee

I would like to say one thing. When I was working in ERP systems. So before you know entering into the data, the world of data science and.

00:13:08 Interviewee

Great analytics, I was working in reporting and ERP systems was one of them. So this object driven systems, so that is something there. But talking in terms of data science, I would I would so far encourage this microservice architecture only.

00:13:25 Interviewer

OK.

00:13:27 Interviewer

Yeah, so towards 13.

00:13:32 Interviewee

Yes. So which would be the best practise that could be useful in applying?

00:13:40 Interviewee

Designing machine learning.

00:13:42 Interviewee

Solutions. So the best practise I would say is like.

00:13:49 Interviewee

Like you know this component architecture we should encourage using the component architectures and you know the requirement management and iteratively development, iterative development, all those, all those things I would I would actually encourage like component architecture, software visually model you know all those things.

00:14:09 Interviewee

Uh, so I'm not quite familiar with the best answer for this. Like the best practise to software engineering or anything like that.

00:14:17 Interviewee

To be very honest, because you know, I I really have a very less experience in this architecture level, but to my knowledge I would, I would say yeah user component.

00:14:26 Interviewee

Architectures or requirement management, all those things are probably the best practises that that basically at least we we tend to follow.

00:14:34 Interviewee

You know, I may miss some of this, to be specific, but yeah, these these are the things I would I.

00:14:34 Interviewer

Yeah, yeah.

00:14:41 Interviewee

Would actually encourage.

00:14:42 Interviewer

Yeah, yeah.

00:14:44 Interviewee

I hope that actually answers your question. I'm not quite sure.

00:14:46 Interviewer

Yeah, yeah, actually, no, no, I think so. You know, you have, you have mentioned all the component architecture required.

00:14:47 Interviewee

Of this, but I do hope.

00:14:53 Interviewer

Architecture, software, visual models. So I think these are the concern that we are looking forward that which can be the best. What do you say the useful architecture design decision so.

00:15:05 Interviewer

Yeah, so towards 14.

00:15:31 Interviewee

When with them last year.

00:15:31 Interviewer

Yeah, I can hear you. I can hear you.

00:15:34 Interviewer

Sorry, you were you was lost. So maybe we can start 14th again.

00:15:39 Interviewee

So yeah, moving on to your next question is.

00:15:52 Interviewer

Yeah, 14. Yeah, you are bad, yeah.

00:15:56 Interviewee

Yeah, yeah, I'm back. I'm. I really apologise. Roger. This is a really embarrassing for Internet.

00:16:01 Interviewer

No, no, no.

00:16:02 Interviewer

No, no, no. I can understand. I can totally understand this is complete. This is completely fine for me.

00:16:08 Interviewer

Because I'm I'm recording this so I can be listening and I can.

00:16:08

OK.

00:16:13 Interviewer

Skip those poses, so this is fine. No worries.

00:16:18 Interviewee

Alright so.

00:16:22 Interviewee

Answering to your next question, what are all the most common design challenges in?

00:16:28 Interviewee

Machine learning so challenges I would say.

00:16:34 Interviewee

On my personal experience, the challenges is basically the complexity of the data. So if I talk about my current project that I mentioned regarding document classification.

00:16:46 Interviewee

So we do use all the Google parsers like form parser, invoice parser, vision API's. You know to extract the text, read the text using OCR and then you know to identify identify which are the accuracy of the text. Like whatever classification, accuracy, entity extraction accuracy. But here the the major problem that we face is.

00:17:08 Interviewee

Using choosing the parsers because each parser comes the Google charges are pretty much high amount for each. You know reading each document.

00:17:17 Interviewee

So we have to be very selective in you.

00:17:19 Interviewee

Know choosing the right parser because not necessarily every parser would work.

00:17:23 Interviewee

On a partner well on a particular page alright. Like then let's say I have a very structured data. So Google delivered form parser would work extremely well for this.

00:17:33 Interviewee

Like in extracting the key value pairs likes a name and then the name of the employee is let's XYZ or say SSN number and the digits are like that. So all those structured elements are very well.

00:17:43 Interviewee

You know, extracted using form parser. But if there is a very complex unstructured data then these delivered API's with the Google delivered APIs would not work. You have to go for specialise.

00:17:53 Interviewee

APIs, so each specialised API add to the cost. So we have to be very selective, like if a page is extracted using a simple parser then we have to use that one.

00:18:01 Interviewee

If a page is very complicated, we have to go for some specialised ones. So choosing this parsers eventually helps us in estimating the cost. So that is one of the very biggest challenges we face in in machine learning, you know.

00:18:13 Interviewee

Doing this architecture design.

00:18:16 Interviewee

This is one thing. Second thing is you know the cloud components that we use, if it is a lightweighted system, we can very well use the Google cloud functions like the Lambda functions in a double.

00:18:26 Interviewee

But if it is a very hugely populated data like millions of data, then there's lightweighted systems on work. We have to go for complex systems complex.

00:18:36 Interviewee

Components like cloud composer, we have to choose a queue flow to work on top of this. So these are these are the component level challenges we face again at a security perspective, we have to choose like if it is a internal VPC or you know.

00:18:51 Interviewee

Or a, you know component driven architecture. So these are these are some of the major challenges we faced. We do work with the platform teams, the data engineering teams and you know the subsystem teams.

00:19:02 Interviewee

Let's say you have you are integrating with a bigger system or a or a or any other system. So these are all the things.

00:19:08 Interviewee

That we close, even though that is not.

00:19:10 Interviewee

Completely directly related to machine learning, but machine learning things are built on these things. So basically these are some of the very big challenges we face as a machine learning developer.

00:19:19 Interviewee

I mean architect and we we actually closely spend a lot of time working and deciding on.

00:19:25 Interviewee

These things so.

00:19:26 Interviewee

Oh yeah, I I would say like bits and pieces might have got the.

00:19:30 Interviewee

Answer but.

00:19:31 Interviewer

Yeah, yeah, I can understand that. Yeah. You always have to save some charges for the Google parsing. So. Yeah. And the complexity, the machine learning, the, the the data complexity.

00:19:33 Interviewee

These are the basic challenges. We do test this.

00:19:42 Interviewee

So just one FY, I or whatever I'm saying, these are all available in GA like if you Google any of the any of the parsers that I mentioned or any of the components that I mentioned, those are all available in public domain. You can actually have a work around or more research on that to get a clear picture.

00:20:00 Interviewee

Or maybe if you need some other documentation I can provide you offline so that you can give evidence under whatever I'm saying to, you know, as a proof of concepts.

00:20:07 Interviewer

Ohh yeah, that would be.

00:20:09 Interviewer

Yeah, that would be nice. So I.

00:20:10 Interviewer

Can give a reference.

00:20:15 Interviewee

Am I audible or did I not step?

00:20:17 Interviewer

But I can hear you. Your video is stuck but.

00:20:19 Interviewer

I can hear you this.

00:20:19 Interviewer

Is important, so I.

00:20:21 Interviewer

Can hear you, yeah.

00:20:24 Interviewer

So yeah, so that would be nice. I think you have my e-mail address so you can share all the offline documents which you have shared about the parsing.

00:20:34 Interviewer

And the perks of using the cloud functions and more perks of using the cloud functions because of because of security.

00:20:42 Interviewer

And other issues.

00:21:06 Interviewee

No, Sir. I'm really sorry. We had a big. I mean, a very bad weather. So this this is, like, very embarrassing for me. My Internet is not in my hands.

00:21:16 Interviewer

Yeah, yeah, this is fine.

00:21:18 Interviewee

I'm really sorry for that. If if interview doesn't help you then maybe we can schedule it sometime like that.

00:21:24 Interviewee

Well, I can give a more clear picture on all my.

00:21:26 Interviewer

OK.

00:21:26 Interviewee

Answers and justifications.

00:21:28 Interviewee

If you find it is.

00:21:28 Interviewer

OK. OK. Yeah. Thanks for the consideration. Yeah.

00:21:29 Interviewee

Not enough.

00:21:32 Interviewer

Yeah. So maybe we move on the 15th then.

00:21:35 Interviewee

OK. Yeah. So the 15th is, what are the main architectural decisions on software architecture design, different from machine learning?

00:21:44 Interviewee

Yeah, so this is this is a generic question that that we like we face as a machine learning architect. So basically if you design A software software level architecture, maybe a back end.

00:21:56 Interviewee

Picture so it is more like a generic thing like you you create an UI where your UI would be capable of, you know dealing with a number of you know documents or should be active till and seconds. But in machine learning what we are.

00:22:11 Interviewee

Is, you know to always, always build a generic thing like like. It should not be a specific thing. We do have a range of data.

00:22:19 Interviewee

Each data differs from each other entirely, and let's say I'm talking about a documentation thing that I.

00:22:25 Interviewee

I deal with.

00:22:25 Interviewee

You know.

00:22:27 Interviewee

If I use a particular parser for extracting entities of a particular document.

00:22:32 Interviewee

That's a structured document. I'll be done within, let's say 30 seconds or 10 seconds. But if I'm dealing with some unstructured data, it.

00:22:39 Interviewee

Would it would literally take double or triple the amount of time, but always we are encouraged to, you know build a generic generic a temp.

00:22:47 Interviewee

Right. That should overall surprise things. OK, so this is the major challenge we face when we design the machine learning architecture compared to anything else.

00:22:55 Interviewee

The software team, they have a very clear architecture. They they they expect the data to fetch there and then they extract. They expect the data from machine learning to get in the form of Jason file so that they can put it on their website and so.

00:23:07 Interviewee

So that, yeah, this is the input and this is the output, but that doesn't doesn't really work. You know that one go in in a machine learning level.

00:23:17 Interviewee

What I mean to say essentially is that each document that we receive on a motional learning side, we have to, you know, deal separately.

00:23:23 Interviewee

Like each, each type of data comes up with a different challenge. So we have to like.

00:23:28 Interviewee

It it it?

00:23:28 Interviewee

It really becomes very difficult for us to, you know, make a generalised model, you know, sometimes to train our classification models, you know, to to, to give a generalised result, but not necessarily that our data is.

00:23:44 Interviewee

Always trained on that train to handle every of the things we do have to regularly, you know, update our model retraining part, you know to deal with.

00:23:54 Interviewee

The newer data.

00:23:55 Interviewee

So these are these are all the challenges that we face, which is entirely different from the software development architecture.

00:24:00 Interviewee

Side so other than that, basically these, these these are the important things on a very high level if I.

00:24:06 Interviewee

Say and on a minute label, yes, there are always post processing and pre processing items that are required for each of the entities or each of the each of the things that we deal with. Machine learning that does change time to time. That is fair enough, but yeah.

00:24:21 Interviewee

Choosing this generalised model and you know retraining of the model, these are these are quite quite the.

00:24:27 Interviewee

Things that we we, you know face.

00:24:29 Interviewee

As so as a as a solution architecture.

00:24:32 Interviewer

So these are the main architecture decisions like the general model and the retrain the model like all the time are the more common and the main attached decisions you take.

00:24:42 Interviewer

So these are the. So these men are directed decisions you.

00:24:45 Interviewer

Are saying sorry, yes.

00:24:48 Interviewee

Yeah, yeah, sorry. One point. I want to say is that whatever I have said or explained you as a machine learning solution.

00:24:56 Interviewee

Lecture is basically my experience is limited to the documentations, the NLP part only because beyond that we also have some chat bot things.

00:25:04 Interviewee

The communication part, all those things. So there maybe there are some other range of challenges, but I would say whatever answers that I have given as part of a machine learning architecture, my knowledge is actually limited to.

00:25:16 Interviewee

Than natural language processing documentation, part like mostly documentation classification and entity extraction. Question answering all those things. So I might limit myself to, you know, my thought process is limited to that only because these are the real time problems I have faced as a machine learning architecture. Before that I was not a architecture.

00:25:35 Interviewee

So I I didn't deal with such problems, so maybe you kind of find a monotonous answer, but yeah, my experience goes to this. So yeah, this is this is something I wanted you to you.

00:25:46 Interviewee

Know keep a note on.

00:25:46 Interviewer

Yeah, yeah, yeah, I know. Yeah. Yeah. Thank you for notifying this. Yeah. This this really explains that, like, like, for example, if we work in a specific domain, then you can't.

00:25:58 Interviewer

You can't be the all the time work on the website mobile development. You can't be the in two different domains, so you have to at least work in something.

00:26:07 Interviewer

In machine learning, so you're working in the national.

00:26:12 Interviewee

That's right.

00:26:12 Interviewer

Yeah. So, I would suggest if you have.

00:26:15 Interviewer

Any kind of?

00:26:16 Interviewer

Documentation for the Google parsing that would be if you if you have.

00:26:20 Interviewer

If this is not confidential, maybe if you publicly available then you can share that and you have my e-mail address so.

00:26:27 Interviewer

That would be nice. Maybe I think I will stop recording. I think this is enough for recording.