Transcript

October 22, 2025, 1:03PM

Sajid Inayat started transcription

Sukul, Abhra 0:03

Basically we have an application, a web application hosted on Azure. From there customer will hit Azure function API, Azure function URL. OK, this is when I said this is a simple part.

For an example, this is a web app. This is handled and developed by customer site. Customer means PPT site. They have some C++ peoples who have building web app, web app OK.

So this our journey started from here. It eventually hit API which is basically a Azure function. So think it as Azure function one.

A1.

What is the purpose of this Azure function? The purpose of this Azure function take the payload and store into a service queue.

We have a bus.

Where we are storing all the messages, whatever user is uploading, the user is sending us a payload payload in the times payload have a the main value in the payload is the your PDF file, PDF file name.

Those everything payload will be stored into this matches scheme. OK, then we have another Azure function. This is Azure function 2.

So past Azure function is HTTP ticket.

Peti, Helios 1:39

Mhm.

Sukul, Abhra 1:40

OK.

This is a HTTP.

And this is on New York is a service bus QT guy.

Service bus queue ticket. This is a queue ticket.

One quick thing is here why you have developed this architecture in that way

because initially we observe suppose customer sending us 10/15/20 files. Eventually our going forward you get to know that our core is dependent on API LLM API.
So LLM API have some restriction and we have some we can't able to due to some limitation we can't process more than five or six. Sometimes we'll get token rate error and those things to just to limit those things we take all the request from customer where they are not. This is a direct post.
Get post request. So as soon as they send a file they will get a 200. TAI process has started. From there we have loaded our queue and from our queue as per our limitation we have processed the file one by one. So this is the service bus queue trigger. This service bus queue trigger is going to call.
Fast API a container instance.
Currently it is container instance. Already in development we have converted into container app because to why to container app because container instance has a limitation. You can't run multiple node of same container instance.
Contain instance can be scale up or scale up. OK, so that is a limitation. So that's why we are going to eventually going to move to container app. Same as container instance is this container instance or container app is going to run a Docker image which is present in your container registry.
Your Docker image is present post to container registry that this CI or this container instance or container is going to run this particular Docker image. It is in container registry.
So what is this Docker image? This Docker image is our core functionality. It is developed with MLflow in Database. So we have a Database where we have developed our AI model and that is in and we are maintaining the version using MLflow.
And after e-mail flow, we have built a fast API on top of that e-mail flow model.
This is a fast debate. Any question till now the overall structure?

Sajid Inayat 4:24
Just a quick question. So once the request is basically made it all the way to Databricks. In Databricks you have your model, but then model uses the fast API, is it?

SA

Sukul, Abhra 4:39
No, we have developed our model in Databix, OK.

Sajid Inayat 4:43
Oh, custom model, OK.

Sukul, Abhra 4:44
Yeah, you have developed our ML flow custom model in Databix. Why do you have use ML flow? Just to maintain the version because tomorrow you have to develop something so you can redeploy when you are building this fast API. This fast API doesn't play a role with any Databix. The first thing it will going to do.

Sajid Inayat 4:47
Oh, OK. Yeah.
Oh.

Sukul, Abhra 5:04
So we have download the ML flow model into this on Prem server. This first phase is running on on Prem server.
So we are going to download the ML flow model locally, then build a fast API top of that.

Sajid Inayat 5:13
OK.

Sukul, Abhra 5:25
Is that OK?

Peti, Helios 5:25
I had a yeah, I just wanted to clarify some things. So basically around the current architecture, firstly we have a customer which logs in into this web app, then the Azure function one is an HTTP trigger. Then we go to the server bus queue.

Sajid Inayat 5:26
Yeah.

Sukul, Abhra 5:28
Yeah.

Peti, Helios 5:42
Then the Azure function two is called by the service bus trigger, right? And then we process the files 1 by 1. Yep. OK And my question was regarding the container app and fast API. So the container app will it replace the Azure function #2 or?

Sukul, Abhra 5:46
Service bus cute ticket.

Peti, Helios 5:58
Does it run alongside it as a processing layer, an additional one? What's the plan regarding that? Will we keep it that way or just?

Sukul, Abhra 6:03
No, no, no.
Yeah.
The Azure function is Azure function two will be remain there. It is going to call the container app itself.

Peti, Helios 6:12
Mhm.
OK.

Sukul, Abhra 6:16
Your point is valid, like we can't directly replace this container Azure function itself with the container app because on the service bus can be implemented in container app itself. But there are some limitation when you get to know the actual code.

Peti, Helios 6:23
Yeah, yeah, there is something else, yeah.
Yeah.

Sukul, Abhra 6:32
It will be.
More complicated approach I can say.
So we don't want, we don't want to hamper our existing model in that much. OK, so that's why we are just going to replace the container instance which is currently running the our Docker image that with the container app.
So eventually right now from content instance I'm getting API.

Peti, Helios 6:56
OK.

Sukul, Abhra 7:00
I'm getting URL. This URL will will be going to replace with the content and app URL.

Peti, Helios 7:09
OK.

Sukul, Abhra 7:10
Is that answer your question?

Peti, Helios 7:12
Yeah, yeah, I understand. The only thing I I asked was will we have the whole implementation into one single specific usage or will we do those parallel the container, the container with Azure function #2? That's what I asked because.
If we can, we can merge them into one implementation, but if you want to, we can just let them separate. That's that was my question.

Sukul, Abhra 7:34
We will keep the, we will keep them separated and this container up eventually is going to scale up based on the service bus queue, HTTP or service bus queue. We do need this service bus queue tigger actually.
It's not. Basically it's going to call this content instance and apart from directly calling the content instance, it's going to do some other validation check. OK, it will go in to do some validation check based on that.

Peti, Helios 8:06
OK, OK, I I understand. I understand.

Sukul, Abhra 8:08
And one configuration we have something like this HTTP trigger can load any number of file till 1GB to the service bus queue. How we have managed to run only three instance at a time? We have this restricted our container Azure function scale out to three only.
So basically what will happen? And another thing is like on process a single request may take 5 minute OK and if you ever like Azure Service Bus queue has a limitation on the message locking.
So we are basically locking this message for 30 minute because by default it is one minute. After one minute you will again it it will release this log and another trigger will pick up the log. So eventually it will fall in a infinite loop. So that's how to restrict those things we have some.
Host in host we have mentioned like maintain the call trigger as 30 minute and run this Azure function 2 and till three only. So eventually you can only process one single message at a time and you can up to three times you can scale up. So basically we are going to process at the same time 3 request.

Peti, Helios 9:16
No.

Sukul, Abhra 9:22
This is the way we have developed this liability.
So basically the CI can process 3 instance as of today. Tomorrow we'll have a CA so we can scale up the CA to another three. So this core functionality will remain same like at the same time you can only process 3 file and based on that like 3 file is going to process in CA.
One instance 3 another three file will be processed in CA2 instance. Another three file process will be CA three instance. So eventually we can based on this CA we can scale up our how many file how many file we are going to process.

Peti, Helios 10:04
OK, OK, let's go.

Sukul, Abhra 10:05
Did you did you get that whole logic, this three logic one?

Peti, Helios 10:09
Yeah, yeah.
That's fine with me.

Sukul, Abhra 10:14
OK, this is the overall structure. Now today I will mainly focus on the your.
This ML flow OK which is the core.
So I think Sajid has an idea how the file is working. What are the? What is the expectation from the? Did you know anything about the overall like they have uploaded a file? What they're going to get? Do you have any idea? Helios? Is it Helios or?

Peti, Helios 10:46
Helios. Yeah, it works in the Helios.

Sukul, Abhra 10:49
OK.

Peti, Helios 10:50
Yep. So basically the question I had was regarding the what I know already is that we are going to losing open open AI for it. We are going to use it in Databricks and when we upload the document, does this document validate around the schema for the open AI? How is the extraction working?
And what is the plan regarding how we are going to improve it and regarding the hybrid approach as suggested as well with document intelligence doing the first extraction model and then doing an enhancement with an open AI? What is the logic behind that?

Sajid Inayat 11:23
Yeah. So at the moment, I don't believe this exists. This is a proposal from our side. Let's get to know how it's currently set up so we can propose what needs to be done. That's OK, go ahead.

Peti, Helios 11:28
OK.
Yep.
OK, that's perfect as well.

Sukul, Abhra 11:36
So yeah, I got your question. That is a hybrid approach we are going to. So this project is divided into two structures, structured way and unstructured way. So this is a generic way we are going to handle all of all of the file as of today.

Peti, Helios 11:47
Mhm.

Sukul, Abhra 11:52
The new request was like we'll have some more hybrid approach where I think Sajid Yas has mentioned Sajid in the last call like how we are going to classify the file and then we are going to pass it to.
Different different models for individual files. So that is a future approach. We don't have that approach right now.

Sajid Inayat 12:11
Yes.
I.

Sukul, Abhra 12:20
Current approach is.

Sajid Inayat 12:20
I'll, yeah, it will. I'll share that with you guys based on the discussion and what I think

might be the right choice and then we'll refine it based on our brainstorming sessions, if that's OK.

Sukul, Abhra 12:34
OK, so the current approach we have is basically they're going to upload a PDF file. If I show you the input and output, I think that will be much better.

Peti, Helios 12:35
OK.

Sajid Inayat 12:44
Hmm.

Sukul, Abhra 12:52
So.
This is the input file.
OK, so based on that input file they are specifically looking for particular few set of. They have a schema, the predefined schema. So predefined schema means it's under document ID. They are going to pass the document ID.
The page number page with and this is for their bounding bounding box related.
Then the main thing is that under generally you have negatives, you have client negative will have hold the value. They have provide us a data dictionary for the same also.
I.
The data dictionary is basically.
So they have kind of what are the things they are going to need and what are the description. The glossary is present in this particular file. OK, so based on that we have a approved schema structure which we are currently following. So this is our output schema structure.
General narrative, client side, we have multiple feeds and few of them as like product is there, product could be at least, medical history could be at least, event could be at least.
These are the basic information we are pulling from the file. So are you clear with the output?

Peti, Helios 14:30
OK.

Sajid Inayat 14:32
Yep, that's OK. Standard. Yep, Yep.

Peti, Helios 14:33
Yes.
Yeah.

Sukul, Abhra 14:37
OK, so now question is how we are, how we are going to how we have achieved this particular part. Let me go into the database.
I'm assuming you don't have any access as of now.

Sajid Inayat 14:50
Not yet, no.

Sukul, Abhra 14:52
Thank you.
So eventually you're going to have git access in this particular git repository. Then you can see the actual code so.
If I explain you, these are the basic and as you are using key vault for our secret management. So in Databix we can't access this key vault. So we have a config ENV file in Databix itself. This is for local running.
And when you're going to run into Azure environment, it will be eventually come from your keyboard.
OK, so this is our customer flow model.
So in custom ML flow model we have multiple function.
This is a function I can how you are.
Just one question to Sajeet and so like go line by line code or I just give you idea what the your dip function wise.

Peti, Helios 16:01
Uh.


Sajid Inayat 16:05
I would say go with the go to the high level using the class line by line might take too long and I think it may not be necessary for now.

Peti, Helios 16:12
Yeah.

Sukul, Abhra 16:12
So this this create default section Jason. So we have observed in our long running process that sometime what is happening GPD is not following the correct Jason format. Sometime we may GPD is maybe miss some of the key.

Peti, Helios 16:14
Yeah, it's just as well that.

Sukul, Abhra 16:31
Totally itself and that is creating a issue at customer side because they are always expecting a valid fixed JSON format. OK, so that's why we have built this the default JSON. Suppose our GPT failed, we encountered some issue.
So we are going to send them a simple.
Uh.
The blank Jason structure, who is they're expecting, rather not sending them, not filling, not sending anything which will fill their process. This is the part.

Peti, Helios 17:04
Like a fall back.

Sukul, Abhra 17:07
This is the purpose of this create default just section Jason.

Peti, Helios 17:11
OK.

Sajid Inayat 17:17
Makes sense, OK.

Sukul, Abhra 17:20
And this same Jason parser is something to the same purpose itself like it is going to check do you have the Jason is valid or not and those thing only. So if it is a invalid Jason then we are going to call this. So basically is the calling function of the previous function.
Save Jason person. OK, we have consolidated Jason. Why do you have consolidated Jason? Because we have called multiple. Initially we tried to fetch whole Jason from a single element part. Then we realized that is not going to possible out of context is going to happen multiple time.
We have split our segment like currently as of today we have 9 segment of data we are looking for. So we are calling seven or nine. We are calling 7 different LM. So in each and each and every individual section will be called individually. Now we have to consolidate them.
And even further like for a particular segment, segment means when I say segment before we just give you the idea another more segment means this one. In case of death drug history, these are individual segment. OK made history is individual segment.
Product reaction is individual segment, the same as product. We have event, we have patient and we have reporter and we have general. These are the segment we have. OK. So another thing is like sometime we observe for a product product segment. Even though we are calling a single error to get all the product information from from the file, we still get your out of context. Yeah, a context length error. So that time we have some predefined logic. We are not going to suppose we are getting for.
We have initially the approach was like first we call LLM with the whole PDF OK and we asked the LLM to go each and page and identify does it have any segment which segment information is present in that particular page.
So suppose you have 10 page of PDF, so eventually you are going to get 10 page of

10 page and each and every page is having like page number is having product information, page #3 is having product reaction information. Similar way even page number 123 have product information that kind of classification we are going to do. At the very first LLM call. Then based on that we have our segment wise page number. Product information is present in 123 page, where event information is present in core and five page. Then we are going to call each LLM for each and every segment and only passing those particular pages.
Is needed. There is another case. Suppose some product is having 5 or 6 pages. We have some predefined token limitation set as per the understanding as per the observation of the running and we have said like OK if it is if the product information present in five page.
And the number of token or for those 5 pages suppose 6000 then we split that then we are going to call product two times. OK, first for three page get the product information for the next three page get the again get the product information.
And those two products Jason file, we eventually have to consolidate.
Those consolidated logic has been written in the safe consolidated Jason and we have a consolidated consolidated Jason function that you you will be able to find in your I think here utils under utils folder.
Any question for this consolidate concept?

Sajid Inayat 21:10
No, for now, we're good.

SA

Sukul, Abhra 21:12
OK.

PH

Peti, Helios 21:13
It makes sense.

SA

Sukul, Abhra 21:14
OK, then then we have our init function. Don't think why unit is after four or three function because those function that come into the picture newly. So those that's why those are written in the top. OK, so unit is basically we have LLM accuracy. So this LLM accuracy is.
Basically they need a confidence score. As you know LLM will not give you any

confidence score. So we are using AWS to extract to extracting the text from the PDF and in Textract is going to give you a confidence score.
Line by line confidence code. That is not for the extraction confidence code. So based on that we have like initially we build our solution. We run it for five or six file. For 5-6 file user has given us OK, this is our expected data.
And we run our process and we get a output and we do a comparison and we created this LM accuracy list. So that time you observe we are able to pick 80% time general as 80%. So this 80 will eventually going to multiply with the confidence score of textract.
Line noise and it's going to give the confidence to the user like how much confident we are.
In future we may have to regenerate this LLM accuracy list. This is the purpose of this LLM accuracy function very variable.
And then the load key concept I just told you everything like AWS key, AWS secret key, account key, Mapbox access token. This is is going to need it. We are using this third party API for your.
Address validation, then account name, LM Azure, endpoint, LM Azure key. Those are all secret we have and we have mentioned them. We'll keep them in key vault, OK.
For local we are not going to access this keyboard, so that's why we are using this uh config file. I think it is pretty much clear till here.
Now check at this standardization like initially you may have address addressing this in a single text. You can have a single text that needs to be standardized and standardized in the sense like what is the.

PH **Peti, Helios** 23:26
Yep.

SA **Sukul, Abhra** 23:43
Street name? What is the region name? What is the district name? What is the postcode name? What is the country name? Those are need to be find out and that's why you are using this API map box token.
Need to be shaped.
So.
Maybe we are not using this function currently. Previously we are using this, then we have an issue. Then maybe how we are addressed the example.

I'm just for now it might like we have initially we have a JSON import function then it got removed. Similarly maybe at the standardizer is not is in use right now, but we have a provision. We have a provision if needed.

Sajid Inayat 24:36
OK.
Oh.

SA **Sukul, Abhra** 24:39
We can do that.

Sajid Inayat 24:41
OK.

SA **Sukul, Abhra** 24:42
I have to check like how we're doing that this and now.
Same like they have a height and weight. OK, height and weight. They ask is height should be always in KG. Weight should be always in. Sorry, my bad weight should be always in centimeter.
And height should be always in KG, even though suppose in your form the height is mentioned as a count, then that you have to change convert.
So this is the function for to convert the unit to standard unit.
Whatever the input like in the form, if I have height as a centimeter, so that time it should take the centimeter height and convert it to a sorry height as a your meter, then we have to convert it to centimeter.
Is that are you OK with that?

Sajid Inayat 25:41
Yep, Yep, it's self-explanatory. I looked at the code. That makes sense.

PH **Peti, Helios** 25:42
Yeah.

SA **Sukul, Abhra** 25:44

Yeah.
Just let me know if you have any issue in the understanding of the code, I'll be there.

Sajid Inayat 25:54
Yeah, no, actually, even though I think I don't see a lot of comments, but the code is written very clearly, so it's easy to understand, yeah.

PH **Peti, Helios** 26:04
Yeah.

SA **Sukul, Abhra** 26:07
Good, yeah.
Basically me and is there initially. So we do have a code review day-to-day basically we are developing most of the code by sitting together. So that's why.
And this is generated AWS profiler creation. We are creating this generated AWS profile. One thing you can work on this alias like there is a huge limitation of our whole project is we are totally relying upon AWS key and AWS secret key.

PH **Peti, Helios** 26:42
OK.

SA **Sukul, Abhra** 26:42
And if you know, if you knew that AWS key is getting change up every 3090 days, I think it is going to reload, reallocate. So that is huge limitation for us. We may have to work on this. So if me and you can work on this.

PH **Peti, Helios** 26:51
Yeah.
Yeah, of course.

SA **Sukul, Abhra** 27:00
Lit.
Instead of using AWS key and what are the approach we can use like it should be I am low but what are the things we have to do?
Then you know.

Sajid Inayat 27:16
So question for you, you are hosting the entire ecosystem in Azure, is it? But you're using AWS keys so you have multi cloud, is it?

Sukul, Abhra 27:23
Yes.
No, no, no. Yeah, we have multi cloud and mainly how you have used AWS. Based on our finding, we observed that AWS text stack to get the text from a file is much better.

Sajid Inayat 27:44
OK, OK.

Sukul, Abhra 27:45
So that's why to just to use AWS to extract function service we are using this AWS.

Sajid Inayat 27:52
OK.

Sukul, Abhra 27:53
And in this thermo feature we have multi cloud so we don't have any limitation. So if some even in current whatever the current thing I'm working we are going to call some model from hugging face itself.

Sajid Inayat 28:10
OK.

Sukul, Abhra 28:14
Who is there is no limitation like infrastructure is we are in Azure or a GPT model hosted in Azure AWS model, a tech sec model is coming from service and.

Sajid Inayat 28:27
Hmm, OK.

Sukul, Abhra 28:29
That's that's to take help from any AI service. You can go for it until unless it is have approved and it has a commercial license.

Sajid Inayat 28:34
OK.
OK, OK.
OK.

Sukul, Abhra 28:41
OK, so this is your generate text using Textact. So this is the service I was talking about. We are going to pass the whole PDF file to this particular service Textactor and we get the document information.
So basically the purpose of this text using text is getting the text from the file and along with the confidence code which I just mentioned like each and every page will line wise they will give you the confidence code for each and every line or.
That you will get the confidence code that is also needed. So basically if you pass image path you are going to get so text along with the confidence code dictionary.

Sajid Inayat 29:18
Oh, oh.

Sukul, Abhra 29:33
OK.

Sajid Inayat 29:33
OK, OK.

Sukul, Abhra 29:35
Then this function is written to generate the confidence code for a individual segment. So can't directly go for the confidence code. So do we did some fuzzy logic also like you are getting something value.
ABCDZY and because we don't get confidence score word as it's line wise. So is the our search extracted value is present in that line and those kind of written in this

particular fuzzy logic. That's why you have used fuzzy logic also.
And based on that we are going to get the confidence score.
So either it could be fuzzy logic, it could be exact match. Exact match means you whatever you are getting that is present in your confidence code dictionary. So you definitely get the confidence code. Suppose it is not present, we have to do a fuzzy match based on a.
You you know what is and those things.

30:31
Yeah, yeah, yeah.

Sajid Inayat 30:32
Oh, Yep, Yep, Yep.

Sukul, Abhra 30:34
And this is a search function. Why do you need search function on ask for us from user is like if you notice this particular thing, we're not only sending the value itself along with the value we are sending the page number X axis, Y axis, height and width.
They have to build a bounding box on the extracted text.

Sajid Inayat 30:59
OK.

Sukul, Abhra 31:00
So LLM is giving me this answer. OK, don't go for this long answer, just go for this answer. So LLM is giving me this is your extracted value. Now I have to go here. LLM will not give you any page number X axis, Y axis. It will not give you the bounding. So this bounding box is also coming from your text extractor also. OK, so if I then I go to my text extraction, I go for the line and there are some logic. We have predefined logic.
First is logic. How we are going to get the page number? Is it in page four or page one, page two, page 3? And then we are going by line by line. OK, particular this after list is present in which line?
Then based on that we are going to grab the by your dimension from that particular

line itself. So this logic has been written in here.
So if you check this is the exact match. First we are doing the exact match. Exact match means we are going to line by lines. We are going page, page in page. First we go by page wise. We are innovated the documents. The documents means it is coming from your text stack. Then we go for line and the line line.
So take you convert the line into text and lower and just checking does it contain the suppose on case my I do have only this information in line.
So it will be exact match. So there could be another reason. It could be a part of this another line.
Cool.
And we have this logic like to get the page number. Suppose this page number is kind of a assumption we have made like if.
A particular value was found previously, then if the same value appear in the next case then we are not going to consider the previous page.
I I don't know. Is that makes sense to you? I it'll much better. If you go just go to this cage. I think you'll find this. You'll have a question like why do we need? There'll be commenters there in case of no finding after first work, then pass the first work and then.
I will like answer this question here. First you go through the all the code, else I have to do the dry run right now because this code has been written one year back, one and one half year back.

Sajid Inayat 33:43
Hmm.
OK.
OK, OK.

PH

Peti, Helios 33:54
Yeah.

SA

Sukul, Abhra 33:55
So then I can give you a high level like this is the first is your trial is a equal match then is in like we are getting a value that could be a part of a line. So we are going for exact match on the it's in operator. So in that case also we are taking normal search but it is in operator.

Sajid Inayat 33:58

Yep.

SA

Sukul, Abhra 34:14

And then we have a fuzzy match and now we have a checkbox option like sometimes we observe they are looking for checkboxes. So if the value is coming from any checkboxes or not then we are getting the checkbox dimension itself and if all are not present then we are going for a fuzzy latch fuzzy logic.

Is something is like for this particular case, narrative is slightly different. Slightly different when I say means maybe in my actual it will be double quote. Somewhere it will be double quote. Here it is a simple text.

So that's why you are using the fuzzy logic.

To get the uh, same thing.

Then after that we'll try to find the bounding box. In the meantime you can check we are calling a format result. Format result is nothing to just to add the given first was the confidence score and you have to change the extracted value into the.

Oxed JSON format. That's it. Nothing else.

To your formatting the result.

This is some exceptional case. We have write additional page meta Jason. This is for your.

This one.

This page meta Jason, this is the only thing is needed in under document ID so it is not a generic one. So we have developed this additional page meta Jason file function.

So we have a generate uh generic people process box instead Jason is just going to the Jason D twice and calling those thing like if it is a document ID then you have to call the addition additional page meta Jason and if it is not.

Then you have to call the search function and that's it.

So it's eventually the final output Jason we have generated because eventually you get to know like.

So what I'm saying here?

Let me check to have any output or no.

The same e-mail flow function has been written here to just for debug purpose. OK, so this is your debug version PVG e-mail flow.

So because that is a huge all everything is written there. So eventually your LLM is going to return you something like this drug name, then NA, event term, then NA, casualty, then NA.
If you check our output.
Drug name. Then we have this structure value, V box and everything. So to do that, that function is basically doing the same thing, converting this drug name in a value to this.
Is that understood?

Sajid Inayat 37:40
Yep, Yep.

PH **Peti, Helios** 37:40
Yep.

SA **Sukul, Abhra** 37:44
Purpose of this people's uh process bar B box is the same thing.
This is OCR. Install OCR. We are installing few predefined test track OCR and the popular utils. That's why those are needed. Then this is for download blog from your. Azure like eventually their file will be present in a Azure BLOB storage. As soon as you get the payload you have to payload will be something like this if I show you this.
It will be something like this document ID, document name and document URL is is there but is not playing no purpose. So we are going to get the document name from here and we directly go to our predefined BLOB storage. We download the file into our local. First thing is the.
We are going to do that that.
So this is for your download BLOB. We have another for upload BLOB. In upload BLOB we have few more function like in case of they want image also web rib image also for each and every page that is going to upload there also.
Then this check image size is not needed right now. Initially we are trying to there is option like if they have sent us image. OK so not needed right now it's kind of. Absolute, but just to resize the image, nothing else. This is PDF to image converter. So basically you are going to get the PDF path. OK, you are going to get the PDF. That PDF has to is going to convert it into images and these images is going to store

into it.
This uh under uh result images there will be a result images container and at that we are going to store those images.
And along with we are going to get the image name, image width, image height, image URL. These are going to needed for this additional parameter Jason if you remember.

Sajid Inayat 40:19
Correct.

Sukul, Abhra 40:26
This is one particular part was to convert the image into base 64 because initially we thought if it is a simple small image we have a provision to pass the directly a PDF image or is to LLM.
And asking, but we are not going for that approach as of today, but we have that option.

Sajid Inayat 40:50
OK.

Sukul, Abhra 40:51
So initially it's like we observe like OK if it is on page directly send the image to PDF or if it is more than 10 page send it to don't go for the text that.

Sajid Inayat 41:02
OK, so I know we are out of time almost. Can you continue or we need to schedule a different call?

Sukul, Abhra 41:04
To those kind of.
I can go. I can no help. I till another half an hour. I don't have any issue. If you don't have any issue, I can let me finish this particular part. I think that will be much.

Sajid Inayat 41:15
OK.

Yeah.
OK.
OK.

Sukul, Abhra 41:25
It's almost half done. It will take another 20-30 minutes.

Sajid Inayat 41:29
Yeah, that's OK.

41:31
OK.

Sukul, Abhra 41:33
OK, I'm giving you more detailed idea, so that's why it's more taking more time.

Sajid Inayat 41:39
Oh yeah, I appreciate that. I just wanna make sure that we don't hold you back from your other meetings. That's why I was asking. Please continue.

Sukul, Abhra 41:47
OK OK so we are using to get the tick token count. OK we are using tick token for to get the token count. This is that function is needed for that number token form a string is self-explanatory function. This is GPT call build. This is is mostly used to. Creating the user message and the system message for each and every GPT. So if you check we have each and every GPT has the requested Jason format.
Just give me one minute.

Sajid Inayat 42:32
OK.

Sukul, Abhra 42:42
So we have predefined structure of each or every segment and we have predefined glossary of each and every segment like what is the area, what is the field, what is the description and these are coming from the file I just showed you initially and even

day by day we have comment where.
Redefine this prompt. Right now we are not going to change this prompt because it's what it is working. We don't want to hamper anything right now. If we want to fix something, we will go for the next level.
So each and every segment has its own decent structure along with the glossary then.
We are we have a provision to suppose general need some extra information OK that needs to be added with your system problem like product needs this extra information. We have a generics a common system problem. I think that is here.
If you need something else that can be added using this concept, OK, this is product reaction. I don't want to change the system form for others, but I want to add this same particular line that that is this particular equals clause in there and we do have the system prompt.
Where in the system basically passing the based on the segment, we are passing the Jason format and we are passing the data dictionary and the extra model specifically if if it is there any.

Sajid Inayat 44:09
Oh.
OK.

SA

Sukul, Abhra 44:17
Then this is the purpose of this particular GPT call build. We are calling the GPT call function that is right now I'm in GPT call build caller called build something which is building the whole message and GPT call is we are going to call the GPT itself.

Sajid Inayat 44:23
Yeah.

SA

Sukul, Abhra 44:34
Before we call the GPT, this is a simple GPT call. Nothing is there. We are taking just logging the how much token we are passing as input, how much token we are getting as output. We have a weight functionality there in case of any failure, retry, retry, then GPT 402 model. This deployment name is.
Study in our Azure. I'll show you later tomorrow or two day after tomorrow.

Then we have a recount logic like we sometimes face content filter issue. Why do we face content filter issue? Because we are dealing with the prescription where we have ethnics type, where we have sex type. Sometimes GPD thought we are asking some bias information, some racial.
Resist information. So that's why you apply content filter.

Sajid Inayat 45:23
Hmm.

SA **Sukul, Abhra** 45:25
Next time it will not. So that's why you have a logic to just to retry. OK, you got a content filter error. Go for a retry. That's it.

PH **Peti, Helios** 45:33
OK.

SA **Sukul, Abhra** 45:36
Scope of improvement. Currently we are only dealing with the GPT 402 model. What we have planned, we are going to deploy another 45 model in our Azure and we'll do a round Robin approach. We're going to use a round Robin approach to select the model name.
Like tomorrow we'll have a container app, container app can have multiple we can going to increase our processing, processing of payloads. So the limitation we had like what we have currently depending on a single model.

Sajid Inayat 45:59
Mm.

SA **Sukul, Abhra** 46:12
Single deployment. If you have multiple deployment we can use round Robin. OK, the process one will go to that model, process two will go to that model, process three will go to that model. That will have a we will reduce the mode load on from GP model. So there is another point scope of improvement.
But it first is the AWS key was the one and the second one is this one.
If we get the stop reason then we are done. If we don't get the stop reason then I

think we are going to. I will check like what is the status final reason. If it is stop then we are good. If it is not stop, if it is content filter we'll do for 2-3 times more.
I think 5 times.
Counter logic times B. Yeah, five times with the count is 5.

Peti, Helios 47:06
Yeah.

Sukul, Abhra 47:10
I told you the classifier call. This is the classifier call LTLM call.
So what is this classifier call? It's basically having all the hell don't have segment wise. We have all segment mentioned here. OK and we are passing this prompt key. You are a text classifier. Your purpose is to do the text classification go through the H and.
On every page and decide which segment of data category of data you have in this particular page and based on that it is going to give us a output of like page number one having this, page #2 having this this information, page #3 having this this information.
Is it understood that function classical function?


Sajid Inayat 47:56
Yep, Yep.

Peti, Helios 47:56
Yeah.

Sukul, Abhra 48:01
Now the main part.
So it might possible.
I think past is going to check.
This logic to split the like 1 segment have four or five pages we are going to chunk or is that I the story I have you mentioned I that is the thing like we are going to 1st check the number of tokens we are having in the total token PDF.
In the total PDF, how much token we have? OK, suppose in the whole PDF we have only token less than two 3500, then we don't want to go call this classifier.

There is there is no point of calling this classifier like which page is holding though overall PDF is very less. How do you going to call a classifier? So that time we are sending OK, it's all page content, just process all page. You don't have to classify it if it is more than 3500.
OK, then we are going to do this classification.
On the classification I'll get a expected answer like this page is having this, this, this means I'm going to get this dictionary where page number one will have this, this information, page #2 will have this number information accordingly.

Peti, Helios 49:31
Mhm.

Sukul, Abhra 49:32
We are going to check OK, is there any particular segment is having more than 3500 record? If not then we don't have to chunk it. Suppose total token in a PDF is 10,000. Product is require product is need five page for the token count is 5000. That time we have to split that 5000 into 3500. We want to try to keep our token as 3500 less than 3500.
To just to avoid the context length error. So that time we are going to call that chunk one. OK, chunk those.
Pages for the segment.
Is that understood?

Sajid Inayat 50:18
OK, Yep.

Peti, Helios 50:18
Yeah.

Sukul, Abhra 50:23
Eventually we are going to get this category wise info.
And this is a else part is not needed. This is for the image processing.
I just keep that code as it is. Uh, we're not using it because these sales never come.
We always have a document, OK, but we have the provision just for that.

Sajid Inayat 50:47
Yeah.

Sukul, Abhra 50:50
This is your get Jason from GPT. Get Jason from GPT means basically now the GPT call is there like which method you have to call. It is basically these are the method mentioned like method. There is a discrepancy between method name and actual category name.
Because if you pass general the classification model is also not working properly. So category wise we have mentioned the different name closely related but not exactly same the method name.
That's why you have this method category mapping.

Sajid Inayat 51:27
OK.

Sukul, Abhra 51:31
So basically is nothing. We are going to method wise then we are going to check does our category dict is containing all page. All page means no, there is no classification, nothing breaking, no chunking. It's simple simple process.
Just pass everything as it is to your own GPT calls. The user message will be simple. OK and suppose we have a multiple page, then each and every method we are going to build the user messages with the required pages. We are just significant in the pages, nothing else.
Then eventually we are going to call this method. I think the method I have mentioned you earlier get GPT called build. So we are passing the user dict and method so accordingly it will create the whole. It will call the GPT.

Sajid Inayat 52:25
Yeah.

Sukul, Abhra 52:27
Here also we are checking do you get a stop reason. If it is not stopped then we are

going to call the section field and section file to save Jason person. This is the validation step I mentioned at the start.

Sajid Inayat 52:40
Mm.

SA **Sukul, Abhra** 52:42
OK.

PH **Peti, Helios** 52:43
Hello.

SA **Sukul, Abhra** 52:52
You can go through this segment of code. It's self-explanatory like JSON formatting is there.

Sajid Inayat 53:01
Oh.

SA **Sukul, Abhra** 53:02
Jason March is the place where we're marging few Jason's like, uh, they have some. Pedefined structure event should be in between products. OK, those kind of structure is there, so we are using this JSON must there.
Like this is our structure structure. We are going to pass. We have individual JSON, so we have to merge and create the final JSON. That's it.

Sajid Inayat 53:35
OK.

SA **Sukul, Abhra** 54:01
Suppose everything failed, then that time we are creating a default finalization just to process their pipe.
Nothing. These are not that much of these are for validation purpose. Jason in Richard is the where we are going to call the those standardized present information height. We're converting the height like we're calling this.

Sajid Inayat 54:16
No.

Sukul, Abhra 54:26
Convert to standard unit. Previously it it also contained the address verification, but we have obsolete the address verification. We are currently most probably we are currently dependent on the element itself for address validation.
This is a validated and fixed final Jason. Is there any? Suppose sometimes GPT may say Jason validator, nothing else. Where is the final Jason file against the schema and auto fix if needed? Suppose you have missed some.
Closing bracket and those things will be done deal with this function.
Then I think we have a. This is your upload result output. Nothing else means to upload. We have to upload the final Jason file into two position. One is your our internal block and on to the customer block.
You will get to know.
Then the predict function. Predict function is pretty much we're going to get the data frame. From data frame we're going to get the request ID document and from document is mainly holding the document name and then we're going to directly call the.
Your.
So this is a temporary. We are creating a temporary directory. Download the file. This is our base path. This is our temporary directory base path. This is the blog name. It is going to download the file into our local temporary directory. From temporary directory we are passing it to PDF to image converter. It's basically converting this image.
Then all file names means list of all images names and images path and then from there we are passing all file names to this generate text using text extractor. It is giving us the document info and confidence score as I showed you then that.
Document info and all file name will going to get to will be feeded into general prompt from GPT. Basically it will do the GPT prompting and create the GPT prompt and loop wise it is going to give you the.
Individual segment Jason output that Jason output will be passed to Jason merge to get the final Jason that is your output Jason. Then output Jason will be go to enrich for enriching functionality, then the fix and final Jason for the validate the.

Output Jason. If it is valid, then we are sending this file to the upload result to output. Till process successfully processed.
And in case of any failure, suppose we are not able to process product person, that failure will be added to the JSON structure itself with the below like these are the error we are we are having right now. Currently we don't have any errors, so we don't have the error tab.
But it may have the error tag.
Then this whole function is your deployed through ML flow into our UC catalog. OK from UC catalog we are going to download into our one prime server like 103.
Then from there we are created going to get our first. So if I show you the our first code also.
This is a Docker file and we are basically going to download the whole ML flow model in here. ML runs under ML runs. This is our main API. This is the actual first API we are creating top of the table flow model. So it's basically doing nothing.
It's just calling the ML flow model by passing the. It's calling the predict method. The predict method function I have just showed you of the ML flow model and it is going to predict that output and we have another function like update keys because initially we have this update keys function.
To update the AWS key, AWS access key if needed because we can't deploy the whole function uh every 19 days no. So we have kept this function.
That's it.
Then we are going to post this Docker image. We are going to get the Docker image on this first API and we are going to pass the Docker image into our content registry. From there we have to go to Azure.
Uh, laptop.
Any question till now?
I know you will have multiple question, but as soon as you get the code access, code base access, you can go through the code and you can get to we can.

Peti, Helios 59:32
No.

Sajid Inayat 59:36
Yeah, yeah.

Sukul, Abhra 59:42
Do a no doubting session question. If you have anything else apart from that, you can ask me.

Sajid Inayat 59:49
Yep, Yep, that makes sense. No, it was very helpful. Appreciate that. So I think we need the access first. So I think Yash is already looking into it. So he's gonna help us from that perspective, right? So then once we have the access, it will be better to actually take a look at code ourselves.

59:52
Yeah.

Sukul, Abhra 1:00:08
Yeah, you have to go to the agent code. I think it's that code is not that much difficult.

Sajid Inayat 1:00:09
Off as well, yeah.

Sukul, Abhra 1:00:15
But we do have some more logic that you have to we have may have to explain you further like how you have wrote that code in that way.

Sajid Inayat 1:00:25
No, that makes sense. I appreciate the help. Thank you. All right. I think we got what we needed. I think it was very helpful and I wanted to actually thank you for starting from a higher level than going down to the trenches.
Because Helios wasn't around when Yash and I had that conversation and now we have that recorded as well. So that's another key thing here, right? So we can always go back to this recording and then consult with it so you don't have to repeat the same thing again. So these was there were like multiple reasons I wanted to actually go through.

Peti, Helios 1:00:44
Mhm.


Sajid Inayat 1:01:03
Through some of these details, so I think that is helpful. So Helios, I'll save the recording somewhere where you can have access to the SharePoint or something, right? But you know who can we can look into it later, but.

Peti, Helios 1:01:11
Mhm.
Yeah.
Of course.


Sajid Inayat 1:01:23
OK, so so Abra, your name of the pronunciation is Abra, is it?

Sukul, Abhra 1:01:29
It's all bro. I'm from Bengal.


Sajid Inayat 1:01:32
OK, Abrew.

Sukul, Abhra 1:01:34
No, no, not. It's.


Sajid Inayat 1:01:37
Abdul.

Sukul, Abhra 1:01:38
Yes, the last A should be pronounced as O.


Sajid Inayat 1:01:39
Abdul.
Oh, OK. OK. OK.

Sukul, Abhra 1:01:46
Like Anand, Anand do those things.


Sajid Inayat 1:01:47
OK.
OK. OK. OK. OK. Nice. OK. Sounds good.

Sukul, Abhra 1:01:56
Yeah.
We're actually having an issue production deployment is I don't know why that for this particular thing only that whereas working actually.

Peti, Helios 1:01:59
Well.

Sukul, Abhra 1:02:10
If Helios doesn't have anything, he can stay on this call because I'm going to. I may not explain those things, but he can see me to play.

Peti, Helios 1:02:12
OK.


Sajid Inayat 1:02:16
Yeah.

Peti, Helios 1:02:24
Yeah.

Sukul, Abhra 1:02:24
If that ends good.


Sajid Inayat 1:02:27
Yeah, that would be helpful actually.

Sukul, Abhra 1:02:30
So this is the Azure first function. We are talking about the AGF one that is Azure function one HTTP ticker.

Peti, Helios 1:02:34
Mhm.

Sukul, Abhra 1:02:40
So we have deployed it but I don't know why the problem we are having. You can help me also Elias like if you have some input like the problem I'm having I'm getting 50500 error.

Peti, Helios 1:02:50
Yep.

Sukul, Abhra 1:02:55
And no log, nothing in application inside. So if I go to test.

Peti, Helios 1:02:59
Are the environmental.
When you go to the Azure function, are the environmental variables all set and correctly set?

Sukul, Abhra 1:03:03
Sorry.

Peti, Helios 1:03:08
Because 500 errors, it means you are not reaching the server, so go to settings.
And the variables environmental variables underneath the settings.
You are using managed identity or connection string.

Sukul, Abhra 1:03:23
Connection string. We don't have anything.

Peti, Helios 1:03:23
OK no no go to up settings, up settings. There you have the up settings. Here you have all the values that are in the but here you can use manage identity or connection strings when you show values so.
It's the key.
Go be down.

Sukul, Abhra 1:03:48
Go with that.

Peti, Helios 1:03:50
A bit down, a bit Scroll down. Do you have any other? Yeah.


Sajid Inayat 1:03:51
Scroll down please.

Sukul, Abhra 1:03:54
We don't create any extra environment variable apart from this environment. So all the others are the auto generated. This is the only variable we have created.
And this is simple.
To be brought as for testing purpose as change it.

Peti, Helios 1:04:24
And where do you get the 500 euro? You get the 500 euro when the function starts or when you're trying to reach?

Sukul, Abhra 1:04:24
The main thing.
OK, let me see. No function is running. It's up and running and.

Peti, Helios 1:04:35
You get into invocations 500.

Sukul, Abhra 1:04:37

Yes, when I try to do the invoke the function, I'm getting this error and the funniest part is that I'm not getting any log itself like in invocation also you're not able to see no 00.
No 00. So basically it's not hitting the function itself.
So I'm simply trying to.
Run it without any payload. So ideally it should give me invalid payload, but it's giving me 500 HTTP this error.
Same code is working in.

Peti, Helios 1:05:30
Quest.

Sukul, Abhra 1:05:30
We.

Peti, Helios 1:05:32
OK.

Sukul, Abhra 1:05:36
It's not going to the function itself.

Sajid Inayat 1:05:42
I'm gonna jump off. Helios, you can stay.

Peti, Helios 1:05:44
Yeah, yeah, I can say. I can say.
So the function is not expecting a payload, right?

Sukul, Abhra 1:05:54
No.

Peti, Helios 1:05:56
OK.

Sukul, Abhra 1:05:56

It's expecting a payload but it's the we have our try and catch block. So at least I am expecting this log HTTP received AI processing. This is the entry point. So if the this particular function hit it will always be in this then afterwards it is going.
To check does it need do you have the payload or not? If we don't have the payload then it will fall it and we'll get error like something like this 500 invalid payload. Yeah invalid request body and see request is provided document list is provided.
Even if I got something error from my code, everything is as it enclosed on the try, I'll get error started buzzing this error. But what I'm getting, I'm 100% sure like this PVG HTTP is not getting hit itself.
Because I tried to run it now, you can see invocation has zero.

PH **Peti, Helios** 1:07:00
Yeah, that's.
Did you try with postman with get or post?

SA **Sukul, Abhra** 1:07:09
And.

PH **Peti, Helios** 1:07:12
With another just to copy the function and try to test it manually, not using the Azure.

SA **Sukul, Abhra** 1:07:19
Not from postman, not from postman, from database set, right? I'm getting there's a not getting any host name or something like that.

PH **Peti, Helios** 1:07:19
So busy.

SA **Sukul, Abhra** 1:07:30
Name of service not known.
If I'm not able to.

PH **Peti, Helios** 1:07:35

end call.
Max retry succeeded with the URL because when you can change.

Sukul, Abhra 1:07:59
Or well, it is working perfectly fine.

Peti, Helios 1:08:08
It's.
So basically we're making a post request to it and it says the connection error.

Sukul, Abhra 1:08:19
Yeah.

Peti, Helios 1:08:19
But.
But I think.
Um.
OK, I think I have an idea.

Sukul, Abhra 1:08:29
I think so.

Peti, Helios 1:08:33
So basically, yeah, I think so it's it does say 500 here, but I think it's a connection that's happening because your request even reaches the function. That's why you are not seeing invocations. So if so, yeah, so it doesn't even reach the function. If it doesn't reach the function, it means it's a DNS.

Sukul, Abhra 1:08:34
Sorry.
Yeah, yeah, yeah, seems like to me also.

Peti, Helios 1:08:52
Problem regarding it so.

So basically the function app is in Azure and I can see that and Azure is is it an internal or private network? Where is this function running?

Sukul, Abhra 1:09:00
Yeah.
He's on broad PPD.
I have to most probably I have to wait for the admin guy.

Peti, Helios 1:09:24
Yeah, so.
So basically, uh.
You don't need VPN artist, right?

Sukul, Abhra 1:09:36
Yeah, I'm with the BP.

Peti, Helios 1:09:37
You are in with VPN.
Have you configured any DNS here or?
Because The thing is, you are not.

Sukul, Abhra 1:09:48
Maybe I didn't want this IP.

Peti, Helios 1:10:04
You do it.

Sukul, Abhra 1:10:07
One problem I on doubt I have.

Peti, Helios 1:10:11
Mm-hmm.

Sukul, Abhra 1:10:14
The default domain is reachable.

Peti, Helios 1:10:23
Yeah, the function is running, but.

Sukul, Abhra 1:10:26
On Sun app is running and new is reachable.
And the access issue, host name issue, those should be.
Appears when I call it from my local and but here I'm calling it from simple from your Azure itself.

Peti, Helios 1:10:47
Yeah.

Sukul, Abhra 1:10:48
So it's suitable to call that. I don't know why.

Peti, Helios 1:10:56
OK.

Sukul, Abhra 1:11:03
No response.
The problem is.

Peti, Helios 1:11:25
Do you have all the dependencies there?

Sukul, Abhra 1:11:29
Sorry.

Peti, Helios 1:11:29
We have all the dependence here because.
A 500 year and no invocations means the function is failing when we are starting it up. So even if you try to execute, even if you try to do anything you we are always going to get 500. So I think something may be missing either an either a dependency either an.

Variable.
What about in logs? Do we get anything? Go to search type log stream.

Sukul, Abhra 1:11:57
I don't chicken.
No, don't have it. It is not getting log itself that I have also take.

Peti, Helios 1:12:04
Locus.
No looks right.

Sukul, Abhra 1:12:10
No logsuit that I have already checked.
Not even in the trace itself.

Peti, Helios 1:12:26
Mhm.

Sukul, Abhra 1:12:27
So we deployed the whole function itself.
It will revalidate with the.
So do not that much of information in host, a simple host.
Do you have only problem with things?

Peti, Helios 1:13:15
Can can you go to configuration at settings maybe?

Sukul, Abhra 1:13:15
You check this thing soon.

Peti, Helios 1:13:25
You can just type application settings if you.
That's configuration.
No at configuration underneath environmental variables.

Sukul, Abhra 1:13:41
Yeah.

Peti, Helios 1:13:42
Yep, configuration. Click it.

Sukul, Abhra 1:13:44
We don't have anything.
OK.

Peti, Helios 1:13:47
Oh.
Um.
OK, OK.

Sukul, Abhra 1:14:00
Let me check the configuration here.
10.2, 10.2.
No, in certificate mode ignored.
Please are doing ignored. Both are ignored. Stack setting is should be Python 3.11.
Function runtime setting code. It's also both.
Support.
No identity also.

Peti, Helios 1:15:48
Check it.
OK, can you go again at the environment variables?
Do you have Azure Webjobs feature flags? Can you check?
Can you go down?
No, I think I don't. OK, so.

Sukul, Abhra 1:16:36
You don't have that.

Peti, Helios 1:16:38
OK, wait.
Let me type it so just copy paste.

Sukul, Abhra 1:16:44
Even that is not present in our.
Hello also.
So.

Peti, Helios 1:16:53
It's not present here.

Sukul, Abhra 1:16:56
It is not in hell also.

Peti, Helios 1:16:59
And this is working.

Sukul, Abhra 1:17:01
No, this is working.

Peti, Helios 1:17:16
You're watching in this also the Python version #2, right? You're using the same.

Sukul, Abhra 1:17:20
Yeah, yeah, yeah, yeah, yeah, yeah, yeah. Same, same. I just check all the configuration. Watson told it so.
Same port has been deployed to here and there.
Even though let me do one thing.
It valid the code itself.
This is your bell code.
OK.
It should be same code mine.
Exactly same.

Or represent.
So now authentication. This is the last change I made. It's not going it here itself.
It should be good.
Just called.
It is the exact same name.
At least I should get this logger and logging in for now.

Peti, Helios 1:19:10
And you try to push it so that the pipeline starts, so we have the latest change.

Sukul, Abhra 1:19:15
Yeah, yeah, I just pushed it few times back from Visual Studio. It's a new deployment.
Before our call started, I just put that quote.
Let me check, do you know is the Azure function 2 is running or also that is is also free. I haven't checked that.
Some messages in queue.
They got for them. We got a couple.
I just sent few messages now.

Peti, Helios 1:21:29
Mhm.

Sukul, Abhra 1:21:30
What happened with those messages?

Peti, Helios 1:21:40
You didn't get them.

Sukul, Abhra 1:21:41
It's just a disappeared.
Let me send it.
Number of messaging.

Peti, Helios 1:22:31
But if you send a message to the service bus, then you're triggering the function by

this service bus message, right? So the function, yes, the function will process the message.

Sukul, Abhra 1:22:38
Yeah, so.

Peti, Helios 1:22:42
So go to Application Insights in your function.

Sukul, Abhra 1:22:47
Yeah, I'm checking.

Peti, Helios 1:22:48
Yeah.

Sukul, Abhra 1:23:06
They know all this online.
Yeah, it picked all the messages.

Peti, Helios 1:23:54
Yeah.

Sukul, Abhra 1:23:54
So this function is running.
What's wrong with that function?
This function is particularly running very smooth, but that one is not running.
Hawaii.

Peti, Helios 1:24:13
Everything is the same, right?

Sukul, Abhra 1:24:15
No.
Yes.

Peti, Helios 1:24:17
That's kind of weird.

Sukul, Abhra 1:24:21
One function is running no. Basically this is function #1.
Let me check and that is trigger.
Oh my God.
On morning it's around checking this one.

Peti, Helios 1:24:56
Did you use the same pipeline to to deploy them both the both functions?
Functions.

Sukul, Abhra 1:25:02
We don't have that much pipeline itself, like we have our code in store in Git and using Visual Stored we have deployed that code.

Peti, Helios 1:25:11
Oh, yeah. Oh, OK.

Sukul, Abhra 1:25:26
And definitely something wrong with this.
App Service plan itself.
Maybe something from the other side is is having the issue.

Peti, Helios 1:25:38
If if you have it done the same way.
I mean that.

Sukul, Abhra 1:25:49
Both are different function first of all.

Peti, Helios 1:25:54
What?

Sukul, Abhra 1:25:55
Both are different function that is to level prod US. This is posted this.
Is under the same function app. We don't have the different 2 function, we do have two different function app itself.
Is that make sense to you?

Peti, Helios 1:26:19
Mm.
I'm trying to understand why it's not working, but.

Sukul, Abhra 1:26:37
This is me bro.
It's same exactly same as well.
No changes, even the code is also same.
Bell is working. Prod is not working.
Only defines between this Val and prod is the environment variable value. It is PROD, there it is VAL.

Peti, Helios 1:26:55
Robeson.

Sukul, Abhra 1:27:07
If I run it now, test and run.

Peti, Helios 1:27:10
So only for pro this not working.

Sukul, Abhra 1:27:11
The same thing.
Huh.

Peti, Helios 1:27:14
Only for prod is not working for the function in prod the other one.

Sukul, Abhra 1:27:16
Yes, yes, yes, yes, yes. You can get output like.

Peti, Helios 1:27:19
And what's the?
OK, let's go. Do we get?

Sukul, Abhra 1:27:22
This this is this is the expected output.

Peti, Helios 1:27:25
OK, so what's the difference? Only the difference?

Sukul, Abhra 1:27:29
Environment variable is environment.
Only defines as this.

Peti, Helios 1:27:40
Yeah, just one.

Sukul, Abhra 1:27:45
I don't think that will make a that much of issue. Here it is environment, here the same thing is here, here we have value as prod and there we have value as well. That is the only difference.
No other difference.
Same code, so code wise it should not have any issue.
It's definitely something from.

Peti, Helios 1:28:20
Did you just out of just to try? Did you try this function which you have in VAR to just manually use manually change VAR to prod restart the function and test? Because if it will work then you are sure that this was a problem right?

Sukul, Abhra 1:28:34

I did the same thing bro. I can't change the VAL one right now because VAL maybe some user is doing a bad testing so it will eventually without informing them that I can't do that. But I did what I changed this prod to VAL and I did it.

Peti, Helios 1:28:45
Yeah, of course.

Sukul, Abhra 1:28:50
I tried. Same error. No, it's same error.

Peti, Helios 1:28:51
And it worked.
So when it has is it isn't working.

Sukul, Abhra 1:28:59
No, no, it is not working.
It is giving me the same mail.

Peti, Helios 1:29:02
Well, that's that's the, yeah, that's the problem though.

Sukul, Abhra 1:29:06
E.

Peti, Helios 1:29:07
That's the problem then. If if they're both are not working with the bar as an environment, then that's the problem.

Sukul, Abhra 1:29:14
No, no, I changed it to Val. That time it didn't work.

Peti, Helios 1:29:16
Hmm.
Mhm.
Yeah.

Sukul, Abhra 1:29:23
We have the same issue even if I put value instead of prod.

Peti, Helios 1:29:28
Yeah, but with prod it's working right. If you put it at VAR, will it work or not?

Sukul, Abhra 1:29:32
No, no, it not. It not.

Peti, Helios 1:29:34
It won't work with Lark.
So that's what I'm saying like if if you put it at var.

Sukul, Abhra 1:29:40
8.

Peti, Helios 1:29:44
It will not work.

Sukul, Abhra 1:29:44
It will. Yes, it will not work.

Peti, Helios 1:29:47
And with prod it was working.

Sukul, Abhra 1:29:50
No, with anything with VAL prod it is not working.
I just change it to hell OK for you.
So I have changed you to hell right now both are same, everything is same.

Peti, Helios 1:30:00
But.
But which was the function which was working that you told me it was the prod?

Sukul, Abhra 1:30:09
That is a deep Azure function too was working.

Peti, Helios 1:30:12
Yeah, the other one.

Sukul, Abhra 1:30:14
This is the function is working. This is the Azure function 2. If you remember the diagram Azure function 1 Azure function 2 Azure function 2 is working.

Peti, Helios 1:30:21
Yeah, yeah, yeah, yeah. This is the second.

Sukul, Abhra 1:30:32
I will check tomorrow with the admin guy. I think something is missed up from admin side so.
It's not reaching the post to itself.

Peti, Helios 1:30:53
Maybe it's not really been reaching the function.

Sukul, Abhra 1:30:56
Yeah, it's not getting the function.
One thing we can try that get a simple history function, deploy it and that I'm going to do in front of admin itself. Even if I test it, if that work, there is no no one from admin side.
See similar.
OK bro, I'll check it tomorrow.

Peti, Helios 1:31:23
OK, OK, I will.

Sukul, Abhra 1:31:25
No admin guy currently present is currently there is no admin guy.

Peti, Helios 1:31:25
Also, thanks for the story.

Sukul, Abhra 1:31:32
You're from Jane Pet.

Peti, Helios 1:31:35
Yeah, I'm from.

Sukul, Abhra 1:31:38
OK.

Peti, Helios 1:31:39
You are from where? Just to us.

Sukul, Abhra 1:31:40
Genpak.
I'm from Theseus.

Peti, Helios 1:31:44
Oh, OK.
OK, also thank you for the explanation regarding the project and because it was my first code review which I wanted to see and also wanted to learn. So thank you for that.

Sukul, Abhra 1:31:59
OK. You are from, you are from Genpact or you are from that another exponential company which you?

Peti, Helios 1:32:06
Yeah. So basically I I I am for exponential, yeah, from exponential data, but I'm representing genpact, but I'm from exponential, yeah.

Sukul, Abhra 1:32:16
OK. And uh, you?

Peti, Helios 1:32:17
I'm with. I'm with Steve.

Sukul, Abhra 1:32:21
OK, OK.

Peti, Helios 1:32:21
With Steve Previt.

Sukul, Abhra 1:32:25
Your form actually.

Peti, Helios 1:32:27
I'm from Albania.

Sukul, Abhra 1:32:30
Albania.

Peti, Helios 1:32:32
Yes.

Sukul, Abhra 1:32:35
Maybe.

Peti, Helios 1:32:36
It's it's in Europe, near Middle East and near Europe. It's near Greece. You know Greece.

Sukul, Abhra 1:32:36
Middle East.
Yeah, yeah, beside of.

Peti, Helios 1:32:44
Yeah, it's near Greece.

Sukul, Abhra 1:32:45
You never can teach Turkey. Yeah, yeah, Greece, Armenia, Turkey.

Peti, Helios 1:32:50
Yeah, so basically genpact is getting both exponential a few months ago and now they want to like make a big team here in Albania and in Kosovo for data and AI.

Sukul, Abhra 1:33:03
So you have experience in AI domain?

Peti, Helios 1:33:07
Yeah, I have experience in AI engineering using Azure Functional Logic Apps, Dock Intelligence, Open AI, prompt engineering. But it depends. It depends because I started as a machine learning engineer with neural networks, conventional neural networks, RNN.
Artificial neural networks. So also have a bit of experience in data science and machine learning as well. So kind of kind of I would say a bit more into the data and also into the AI, but it depends on the project.

Sukul, Abhra 1:33:41
Since you are all rounder.

Peti, Helios 1:33:45
Yeah, all rounder, but it.

Sukul, Abhra 1:33:45
You can play as a forward. You can play as a back. You can play as a winger. You can play as anything.

Peti, Helios 1:33:50
Only goalkeeper now.

Sukul, Abhra 1:33:54
Thank you.

Peti, Helios 1:33:59
But yeah.

Sukul, Abhra 1:33:59
OK bro I'm OK myself. I do have data engineering background and from data engineering I moved to this data science team.

Peti, Helios 1:34:11
Do you? Are you enjoying it? How does it look?

Sukul, Abhra 1:34:12
So I was.
Sorry.

Peti, Helios 1:34:16
Are you enjoying it? How is it looking to you?

Sukul, Abhra 1:34:21
I didn't get your question bro.

Peti, Helios 1:34:23
Are you enjoying it? Being as a data scientist right now, how does it look to your data science?

Sukul, Abhra 1:34:31
Me it's if I it's basically me or me has more knowledge in LLM itself because I don't have any prior educational.

Peti, Helios 1:34:31
Are you liking it? Yeah.
Mhm.

Sukul, Abhra 1:34:47
Certification in a e-mail, whatever I have learned from my hands on.

Peti, Helios 1:34:55
So you have only hands on experience, yeah.

Sukul, Abhra 1:34:55
So it's good then, yeah, no.
No certification.

Peti, Helios 1:35:03
And and I to be honest with you, to be honest with you, I think certifications at the end of the day. I've done the AI engineer one and an AI engineer two of Azure, but both of those they don't and yeah.

Sukul, Abhra 1:35:12
Those are bogus. Those are bogus. Fundamental will only teach you. OK, you have a classification problem. You should go to ADI classifier. You have a document extract. You need to go to Azure document.

Peti, Helios 1:35:17
They.
Is and.

Sukul, Abhra 1:35:28
This guy, I don't know.

Peti, Helios 1:35:28
Yeah, yeah. And it's also only alternative questions. It doesn't help you. It doesn't help you. If you have a real world scenario in error, then it will. It's it's not even helpful for that. So I I agree with that the certifications are just to sell yourselves, but to be honest with you.
If you have experience in projects and yeah.

Sukul, Abhra 1:35:47
Yeah, even I'm I'm also planning to. I'm also planning to do the Azure certification.

Peti, Helios 1:35:53
Let me know, yeah, let me know if you want to plan it or if you have anything. I just did Azure AI engineer tour a couple of months ago, so it's just alternative questions.

Sukul, Abhra 1:36:02
Do you have any question answer set prepared by you?

Peti, Helios 1:36:08
Uh, sorry.

Sukul, Abhra 1:36:10
Do you have any preparation like any preparation doc which is prepared by you?

Peti, Helios 1:36:16
I don't know. I don't know this. What does this mean?

Sukul, Abhra 1:36:20
No, no. Did you write it down somewhere? Like these are the question important. Take some notes, you know?

Peti, Helios 1:36:22
Uh, uh huh. No, I just.
I just to watch some videos on YouTube about it. I had some I checked some documents from Azure which they upload online. So basically for each certificate they provide also some documents which you need to learn. All the answers are with alternative and I just go to the office. It was in this office in Albania which they have some.
People there, so they see if you use your phone, if you use anything and you can do the exam there if you that's what you asked.

Sukul, Abhra 1:36:54

No, I do have Azure knowledge like is also working in multiple Azure PIS services. I'm planning to do it because to save myself anyway.

Peti, Helios 1:37:07
Which, which, which is your stronger, your stronger point, Azure or AWS?

Sukul, Abhra 1:37:14
Um, no, no, nothing that strong. I'm not that strong in any way.
Kind of all around that, including goalkeeper also.

Peti, Helios 1:37:25
That's good.
I have I have to be honest, I have more experience with Azure, but also AWS has the same logic, just if names change at the end of the day, both are cloud architecture so.

Sukul, Abhra 1:37:42
You know, it doesn't like, yeah, at the end of the day, based on the customer, you have to play it like tomorrow your customer only have AWS. I do have more knowledge in Azure itself.

Peti, Helios 1:37:54
Yeah, you.
Hmm, have you have you been working for too long? Also for for for the thermal future client or?

Sukul, Abhra 1:38:05
Yes, I've been here for like five years.

Peti, Helios 1:38:09
For five years working for them.

Sukul, Abhra 1:38:12
Yes, yes.

Peti, Helios 1:38:16
That's good to hear.

Sukul, Abhra 1:38:22
OK, this whole POC was started me by me and a citizen.

Peti, Helios 1:38:28
Yes, that was good POCSPOC is I think the most important part, right? If you make the POC from the beginning from scratch, it takes a lot of also ideas to provide a clear POC, not just to have it done.

Sukul, Abhra 1:38:47
OK, anyway, um, have to drop.

Peti, Helios 1:38:50
Yeah, of course. Also, what about your time? Just yeah, just to have, just to have a question, what about your time? What's your appropriate time for texting or calling or so?

Sukul, Abhra 1:38:51
It's nice talking to you.
Just takes me. I I'll be available pretty much. IST awards Arbenia is. I follow mainly ISTRS, but I do stay at night. It's mostly.

Peti, Helios 1:39:04
Oh.
Oh.
Some.
We have like 3 hour and 30 minutes difference. So basically now in Albania it's 4:42 PM.

Sukul, Abhra 1:39:22
In India it's 812. So in from India perspective I'll be most of the time I'll be present till 10:11 PM means your 10-6 or 7:00 PM of your time.

Peti, Helios 1:39:25

That's it, alright.

OK.

OK, OK. OK then.

OK. Thank you and hope you have a good day.

Sukul, Abhra 1:39:44

OK mate, same to you.

Peti, Helios 1:39:47

Thank you. Goodbye.

Sukul, Abhra 1:39:48

When?

Sajid Inayat stopped transcription