

Bootcamp Kickoff

Transcript:

0:00 uh welcome everyone uh we are currently 0:02 uh November 15th this is the day I 0:04 promise that we are going to start the 0:06 data expert free data engineering uh 0:08 boot camp kickoff and uh we will be 0:10 launching be we will be publishing the 0:12 first video in about five hours from now 0:15 and every video going forward should be 0:18 expected at around 5:00 pm Pacific each 0:20 day there's a schedule at data expert. 0:22 that you can check out but let's get 0:24 into 0:25 this so welcome uh to an intense six 0:28 weeks you're going to learn a lot over 0:30 the next six weeks especially if you 0:32 commit and dedicate yourself and 0:33 actually give your full effort to all of 0:35 this if you can uh really focus and put 0:39 in the one to two hours a day of time 0:42 that I'm asking from you and treat this 0:44 like the data engineering gym if you can 0:46 do that you will find that you will be 0:49 able to just grow immensely technically 0:53 and by the end of the year you are going 0:54 to be like wow I am really good at this 0:56 stuff now so uh let's dig into like what 0:59 we're going to cover today so we're 1:01 going to cover cover a couple different 1:02 things here we're going to cover uh why 1:04 you made a great decision joining this 1:06 program uh it's a really good one talk 1:08 about how to get the most out of this 1:09 boot camp uh it's going to come I know 1:11 six weeks sounds like a lot of time but 1:13 uh you're G to Blink and it's going to 1:15 be over and you're going to be like wow 1:16 that was a lot then we're going to be 1:18 covering the curricula and the boot camp 1:20 structure then we're going to be going 1:22 over uh the difference between this boot 1:24 camp and the paid one because I have 1:26 been getting an an obscene number of DMs 1:30 about that because there are very big 1:33 differences between this boot camp and 1:34 the one I'm doing in January they do not 1:36 replace each other if anything they 1:39 complement each other and if you do both 1:41 you would in February be very well set 1:45 up for an amazing opportunity uh we'll 1:48 talk more about study groups and how 1:50 they work and then uh we'll talk a 1:51 little bit about policies and structures 1:53 there's not that many policies and 1:54 structures here because y'all didn't pay 1:57 me anything so let's get into it 2:00 uh so why you made a great decision so 2:02 um I've been uh doing boot camps for 2:04 about 18 months now I've have trained 2:06 over, 1500 data engineers and data 2:08 engineering I've helped hundreds of data 2:10 Engineers land jobs or get promoted 2:12 across the um Spectra not just 2:15 entry-level jobs talking about people 2:16 getting promoted to senior engineer 2:18 people getting promoted to staff 2:19 engineer I've helped a lot of different 2:21 Engineers along the way and we take 2:23 engineering very seriously here and uh 2:27 keeping in mind that like all all this 2:29 content is going to be on YouTube for at 2:31 least the next year so make sure to take 2:33 advantage of it then I might take it 2:35 down uh December 2025 we'll see uh but 2:39 make sure that you know that it's not 2:41 just going to be up here forever you 2:43 want to act with a sense of urgency with 2:45 this content um I have a goal here to 2:48 help a thousand data engineers get jobs 2:51 by February 15 whether whether that be 2:54 through uh teaching them on YouTube here 2:57 or through my paid boot camp in January 2:59 uh those two things together will be 3:01 amazing and I really do think that we'll 3:03 be able to get there because our cohort 3:05 is over 10,000 people and I I know that 3:08 there is going to be a lot of very 3:09 hungry people looking to learn data 3:11 engineering here

uh and yeah get ready 3:14 this is going to be a really exciting 3:16 time so how to get the most out of this 3:19 boot camp uh keeping in mind that if 3:23 you're just here to learn information uh 3:26 like I kind of don't want you here I 3:27 kind of want you uh to be thinking about 3:30 how to make friends how to engage how to 3:32 be part of the community how to uh do 3:35 all that stuff because it's like uh boot 3:38 camps are not really about the 3:40 information as much as you would think 3:42 not to sound cliché but it's really 3:44 about the friends that you make along 3:45 the way so uh we have uh different 3:48 channels in Discord for the different 3:50 regions of the world that you can uh go 3:52 into and we're going to be adding a 3:54 little feature to the Discord bot pretty 3:56 soon that will also allow you to and to 3:59 create and join groups with people so 4:01 that then you can uh talk with people 4:03 maybe either over email or over LinkedIn 4:06 or whatever there's all those different 4:08 uh options that you can potentially do 4:11 uh so definitely uh check out that uh 4:13 and so um you want to be social make 4:17 some friends watch all the videos do all 4:19 the homework that's going to be like the 4:21 basic the basic expectation here if you 4:24 want to like get value out of this right 4:26 and like obviously some of you have 4:27 full-time jobs some of you are parents 4:29 some of you you have like you know 4:30 there's a bajillion trillion different 4:32 things that are going on for each one of 4:34 you and you all have different 4:35 circumstances for like how you do things 4:37 so like I I don't expect all of you to 4:40 do the try hard so for the try hard like 4:42 after every video post what you learned 4:44 post take a slide or a screenshot from 4:48 uh the video post it on LinkedIn post it 4:50 on Twitter post it on Instagram and uh 4:52 tag me and I can help you get some more 4:54 reach and we can grow your brand 4:55 together right uh because my like and my 4:58 comment makes your post go I further and 5:00 then uh also give your full effort on 5:02 the homework really try the homework is 5:04 not easy the homework's not easy the 5:06 homework like uh for most people if you 5:07 think about it like each week here 5:09 you're going to be spending you know 5:10 roughly speaking about 8 to 10 hours 5:12 watching videos and then it's going to 5:14 be another eight to 10 hours doing 5:16 homework and then like if you are a 5:18 little bit more entry level if you do 5:19 the homework like it's not going to be 5:21 eight to 10 hours it's going to be more 5:22 like 20 hours like if you want to 5:24 actually do all the homework as well 5:26 keeping in mind that that's the goal 5:27 here this is a this is this isn't like 5:30 uh uh some people think that this boot 5:32 camp I'm doing for free is like some 5:33 sort of like halfway boot camp that's 5:36 like a teaser for the next boot camp it 5:38 is not this is the same material that I 5:40 covered a year ago in my second boot 5:42 camp uh back in July of 2023 and uh we 5:46 are going we we're republishing that 5:48 content now because no one's watching it 5:50 and so this is a full boot camp a full 5:52 boot camp so uh understand that like if 5:55 you give your full effort here you will 5:57 be rewarded 6:00 um so curricula I think this is one of 6:02 the favorite parts of everything uh 6:04 we're going to be spending a lot of time 6:06 on data modeling I think data modeling 6:07 is one of the most important things that 6:09 you need to learn as a data engineer 6:10 because your uh uh final outputs are 6:14 data right your product is data so if 6:17 you aren't uh using data in your data 6:19 models then uh or if you're not doing 6:21 data modeling to your data you're going 6:22 to produce a shipit product and that's 6:24 not good and so we're going to be 6:25 spending probably at least two weeks 6:27 there and then uh applying analytical 6:29 patterns uh this is a great week where 6:31 we talk about things like growth 6:32 accounting and advanced SQL so growth 6:35 accounting is a really powerful way that 6:37 Facebook tracks user growth and the

user 6:39 journey and then uh Advanced SQL we'll 6:41 be covering things like window functions 6:43 and uh you know all the different uh you 6:46 know kind of deeper SQL Concepts to kind 6:48 of get you to that next SQL level as 6:50 well uh we'll be talking about kpis and 6:53 experimentation this is where it's like 6:54 how to define a good experiment how to 6:56 define a good metric these are going to 6:58 be like really kind of like thinking 7:00 like a product manager that's a big 7:01 thing you want to be thinking about in 7:02 that space it's like those product 7:04 minded data Engineers are very very 7:06 powerful we'll also be doing data 7:08 visualization and impact this is a lot 7:10 about communication and pushing back and 7:12 when to say no to low value requests and 7:15 then we do a little bit of we bust out 7:17 Tableau just a little bit we do just we 7:18 do like one lecture on Tableau so uh get 7:21 ready for that we where we'll talk about 7:23 the differences between like exploratory 7:25 and executive dashboards and then on the 7:27 right side we have more of the 7:28 infrastructure track uh we have a unit 7:30 testing pie spark pipelines that's going 7:32 to be uh really powerful stuff we use a 7:35 library called chisa for that and then 7:38 uh we have a data pipeline maintenance 7:40 it's going to be a good one where we 7:41 cover things like runbook and on call 7:43 rotations and how to actually maintain 7:45 your pipelines at scale uh then we're 7:48 going to have a week that a lot of 7:49 people are excited for is the Apache 7:50 spark fundamentals it's going to be 7:52 about four hours of content just going 7:53 over spark and everything you need to 7:55 know about spark and how to use spark 7:57 and then uh the after that we'll be 8:00 doing data quality patterns uh the big 8:02 thing here is we we will be using 8:03 astronomer a little bit here and uh 8:05 we'll be using astronomer in the cloud 8:08 and uh we will be using uh we will be 8:10 learning a very powerful data quality 8:12 pattern called Write audit publish write 8:14 audit publish is a very great uh data 8:16 quality pattern that you can learn to uh 8:19 make sure that you don't publish bad 8:20 data in production and then you know 8:22 last but not least we will be spending 8:24 about four hours of time on realtime 8:26 pipelines with Flink and kofka uh for 8:29 this segment make sure that you have 8:31 Docker set up Docker is the most 8:33 important for uh this La this last like 8:36 realtime pipeline segment this is where 8:38 we need a lot of Docker and uh like 8:40 spark also kind need kind of needs 8:42 Docker as well so uh don't like you can 8:45 do you know for the first week of the 8:46 boot camp you can do a local setup with 8:48 like stalling postgres locally and just 8:49 doing it that way which is totally fine 8:52 and uh but like later on you will have 8:54 to use Docker because we do not have 8:56 another setup for spark and Flink we 8:58 only have one setup and it's Docker so 9:01 uh don't think that you're going to get 9:02 away with not using Docker in this boot 9:05 camp uh boot camp structure um there's 9:08 going to be uh essentially two ways that 9:11 you can uh view this content all the all 9:13 the classes are pre-recorded and one 9:15 lesson will be uploaded each day I might 9:17 I'm gonna I'm going to pull the audience 9:19 here about that a little bit because I 9:21 think there's another way that we could 9:22 do this that might be better we'll talk 9:24 about that a second uh but uh you can 9:27 find the videos you can find the YouTube 9:29 videos at youtube.com/ exactly that's 9:32 going to be uh a great place to look at 9:35 it you can also if you're taking the the 9:37 class for credit and you want to get the 9:39 certification uh you're going to want to 9:41 find the videos on uh Slash lessons 9:44 you're going to be able to find the uh 9:46 the content and that's where there's 9:48 time tracking so we'll be able to get 9:50 you uh closer to certification that way 9:53 um and we have a schedule so if you go 9:54 to data expert. or or boot camp. Tech 9:57 creator. schedu that's where you can 9:59 uh find the

content schedule of like 10:02 what what what we'll be publishing when 10:05 should be pretty great uh each module 10:07 has we'll have two lectures right so 10:10 every lecture has a lab associated with 10:13 it so there's a lecture and then a lab 10:15 this is where this the the content 10:16 schedule I'm going to pull with you guys 10:18 because like I advertise this as like 10:20 one video a day for the rest of the year 10:23 part of me feels like I should publish 10:24 the lecture and lab together and uh then 10:27 we do a video we do two videos like um 10:30 every other day essentially roughly 10:32 speaking it's the same thing like it 10:34 will be the same content it's just like 10:36 how I put it out because I think some of 10:38 y'all will want to do like the do it in 10:40 a two-hour block we'll P that in a 10:42 second uh and then the infrastructure 10:44 required for each lab will be shared in 10:45 the YouTube caption and on the um it 10:48 will also be on the lessons page if you 10:50 are missing that and uh definitely uh 10:52 check out the data engineer handbook if 10:54 you have not given that a star that's uh 10:57 that's the one thing I really ask of you 10:58 here is to star my GitHub repo like if 11:00 you don't do that if you like watch all 11:02 my content and don't star my GitHub repo 11:04 that's that's that's that's a shame 11:07 that's honestly a shame that's like that 11:08 that's literally all I ask is just click 11:10 one button and then that's all that's 11:12 all I want from you so uh definitely uh 11:15 check that stuff out you're going to 11:16 need to clone that repo and then also 11:18 pull in updates because we're going to 11:19 be publishing the homework each week as 11:21 uh as that stuff comes up more 11:24 online um a lot of you have been asking 11:27 what the difference is between the paid 11:28 and free boot camps are uh here are the 11:30 big differences uh so uh the paid or the 11:34 free boot camp we have uh six weeks of 11:36 detailed content uh AI graded um 11:39 homework assignments community and 11:41 Discord notifications via email and 11:43 WhatsApp when new content drops we do 11:45 not have dedicated ta support do not DM 11:48 me about questions I will not answer 11:50 them because there's 10,000 of you and I 11:53 cannot answer those questions don't make 11:55 me regret freaking doing this so like 11:57 ask the community don't ask me 11:59 right that is different if you're in the 12:01 paid boot camp and you want to ask me a 12:02 question guess what I will answer it 12:04 because you paid me money right but you 12:07 if you have not paid me anything I'm not 12:08 GNA answer your question that's a big 12:10 difference there's no and we also have a 12:11 set of TAs in the in the paid boot camp 12:14 who will also answer your question a lot 12:16 more dedicated support in that way uh 12:19 and another very big thing is the 12:21 curricula is actually pretty different 12:23 so let's talk about what the paid boot 12:25 camp 12:26 offers uh the paid boot camp is sign 12:29 different there's a weekly uh Q&A a 12:31 weekly small audience Q&A with Zach 12:33 there's usually maybe 30 people who show 12:35 up uh there's two industry expert guest 12:38 speakers each week I always get the most 12:40 Banger guest speakers like literally 12:42 every famous person that you can imagine 12:44 in data engineering is at these and you 12:46 can ask them a question um and then uh 12:49 the content is different so uh in the 12:52 paid boot camp we cover snowflake trino 12:54 DBT we cover spark more in depth we 12:57 cover AWS glue that's how we deploy 12:59 pipelines in the cloud with spark we 13:01 cover Apache iceberg in depth we cover 13:03 spark streaming in depth and how to 13:05 deploy real-time streaming pipelines in 13:07 the cloud we cover Co more in- depth and 13:09 we cover airf more in depth uh and all 13:12 of this stuff gives you free access to 13:15 the cloud so you can deploy your 13:16 pipelines to the cloud for free any 13:19 pipeline you want not the homework if 13:20 you want to build a

Capstone project you 13:22 want to build some funky little thing 13:24 that is on the side that's totally okay 13:26 and a lot of students are already doing 13:27 that so uh yeah you the the the free or 13:30 the paid boot camp comes with Cloud 13:33 support you get Cloud access for a year 13:36 and on top of that you get access to 13:38 paid apis so like polygon polygon is a 13:40 really powerful stock market API and 13:42 scraping dog is a very powerful social 13:44 media API that you can get access to 13:46 which are paid and then you can build 13:48 pipelines on top of that to build richer 13:50 more interesting pipelines uh it also 13:52 comes with a Capstone project that you 13:53 will get dedicated feedback on we will 13:55 tell you if your Capstone project is 13:56 going to be good enough for an employer 13:59 to be like wow this guy knows data 14:01 engineering uh like I mentioned before 14:03 dedicated ta support and then last but 14:05 not least is if you get certified in 14:07 boot camp and you give all your effort 14:09 in the paid boot camp and you get and 14:10 you do all the homework and you do the 14:12 capstone then at the end of it you will 14:14 get mentorship and job interview 14:15 training uh that will help you really 14:18 get that last mile that will get you 14:20 from I know my skills I have my skills 14:22 to I have a job so I hope that this 14:25 clears up any uh confusion from people 14:29 of like how these things are different 14:31 key differences if you want to talk 14:32 about it very succinctly is paid boot 14:35 camp has a lot more support a lot more 14:36 humans involved that's why it's paid 14:38 because I have to pay my people uh the 14:41 content is different the content is in 14:42 the cloud the content is all about 14:44 deploying stuff in cloud and using the 14:47 infrastructure that I have built up over 14:48 the last 18 months and then uh Capstone 14:51 project and uh getting more feedback on 14:54 building out like a portfolio and 14:55 Capstone project those going to be the 14:57 key differences between the paid and and 14:59 free boot camp keeping in mind the paid 15:00 boot camp we have about 25 seats left so 15:03 if you're interested in joining that I 15:05 suggest getting in fairly 15:09 soon uh how do study groups work um uh 15:13 reach out to people on the Discord 15:14 through the global regions like it's 15:16 it's best to try to find people in your 15:17 study group like or in your region so 15:20 that like because time zones and you 15:22 know some of you are in India some of 15:23 you are in Australia some of you are in 15:25 the United States some of you are in 15:26 Europe right and so it's like you want 15:28 to kind of work with people who are in 15:29 your time zone uh that's better and uh 15:32 definitely recommend doing that uh and 15:34 don't reach out to me I needed to update 15:35 this slide don't don't don't reach out 15:38 to me uh with questions or do not do 15:39 that I will not answer um uh we learn 15:43 together here on Dat expert right and um 15:46 if there are people who are spamming 15:48 right in the Discord Channel we will 15:51 just we will kick them out uh like 15:54 understand that like it's going to be P 15:56 you need patience with this stuff and 15:58 that like people and like you need to 15:59 figure out some of this stuff yourself 16:01 like this is not just on uh it's not 16:04 just on the community to help solve your 16:06 problems like you should struggle with 16:08 it and because within that struggle and 16:09 in that pain is learning uh that's how I 16:12 learn data engineering and I know that's 16:14 how you can learn it too I'm giving you 16:16 the guidance and the road map but you 16:17 still have to walk the road so know that 16:20 and do not do not spam anything else in 16:23 the community do not try to sell 16:24 anything do not if you do any of that 16:26 you will be banned immediately you'll be 16:28 done zero Tolerance on that awesome 16:30 thank you very 16:33 much okay how to submit homework um at 16:35 the end of each week so we will be 16:38 opening up the homework uh and it will 16:40 be

submitted um via uh the assignments 16:46 right so you will need to go to boot 16:48 camp. tcre doio assignments you'll see 16:51 it in the dashboard should be pretty 16:53 easy to find and uh you can uh you 16:55 essentially need to upload a zip file so 16:57 you upload a zip file 16:59 of all of the the required files and 17:02 then that will kick off an open AI job 17:06 to look at if your solution is correct 17:10 and it usually takes open AI about two 17:12 minutes to get back to you to grade your 17:14 homework uh remember don't don't submit 17:17 extra don't submit like if you 17:20 submit like if you zip your whole folder 17:22 and you zip all the files including all 17:25 the stuff that like open AI might not 17:26 need then uh like it's just not going to 17:29 work because openi has a limit on how 17:32 many characters you can upload so be 17:33 very careful about that because it will 17:35 just fail make sure that you only upload 17:37 the the files that are required and 17:39 asked of you in the GitHub repo uh the 17:42 homework will be found also in the data 17:44 engineer handbook make sure to start it 17:46 like I said before uh yeah homework is 17:49 going to be great I think we're going to 17:50 be grading hundreds of thousands of 17:51 homework assignments uh in this boot 17:53 camp so I'm really excited uh for you to 17:56 get feedback on that as well uh we will 17:58 be uh releasing the first homework 17:59 assignment uh for grading so the the 18:02 actual homework assignment is already 18:03 out there on the data engineer handbook 18:04 we will open up submissions for that 18:06 fairly 18:08 soon um so there's going to be uh two uh 18:13 certificates here um if you want the I 18:15 watched all the video certificate you 18:17 can get that uh I don't know it's I 18:20 consider that the same as like the 18:21 participation trophy I mean it's a Vibe 18:24 it's a Vibe it counts for something but 18:26 if you want to actually get like the 18:28 certificate that matters uh you're going 18:29 to need to do all the homework and so uh 18:32 that's where uh you'll see on the 18:33 dashboard page there will be two things 18:36 two criteria you'll see a homework um 18:39 criteria as well as an attendance 18:41 criteria and you're going to need to do 18:43 both of those in order to get certified 18:46 so that's going to be something to be 18:48 thinking about as you go through all 18:49 this content and you are seeing what's 18:52 out there uh like really try to go for 18:54 that uh higher level um certificate 18:57 because that 18:59 uh because most of you aren't going to 19:00 get it you know there's like what more 19:03 than 10,000 of you in here if I had a 19:05 guess um I would guess that the number 19:08 of people who actually get certified out 19:10 of the out of 10,000 right is like 300 19:14 or 400 of you right and it's like if you 19:17 get certified the thing is is like I 19:19 that's great and I'm going to put you on 19:21 a list and help you with things later on 19:24 I promise it'll be worth it if you uh uh 19:26 get certified it will help you in 19:28 different ways and maybe unexpected ways 19:30 in the future as uh I build up more of 19:33 my recruiting and hiring funnels and all 19:35 these other things that are kind of very 19:37 early on in my business model right now 19:40 but uh getting certified early will 19:42 definitely help you there so uh yeah 19:45 don't just watch the 19:47 videos so uh this free boot camp is not 19:51 just me uh we have a whole stack of 19:54 humans who are involved here so we have 19:57 me we have matali matali is awesome 20:00 she's been working for the company for 20:02 data expert almost since the beginning 20:04 since like well like July 2023 since 20:07 like since we filmed this content 20:10 essentially and then we have uh Wesley 20:12 Wesley is uh a lot of the AI grading the 20:15 the AI grading that you will see Wesley 20:18 is the owner of that and then we have uh 20:21 Wan woo Who is uh he has owned a lot of 20:24 the infrastructure pieces of this boot 20:26 camp a lot of stuff around Flink and C 20:29 and Spark stuff like that uh so this is 20:32 our team uh I'm

very grateful that you 20:34 have been here listening to this 20:36 presentation today I am very excited to 20:39 get this rolling with you it's going to 20:40 be a really great time make sure to 20:42 invite your friends uh because keeping 20:43 in mind that this is free it's free so 20:46 uh and I think that that's something 20:47 that you will find a lot of value in so 20:50 uh yeah I'm going to