SDAT Interview 1 Transcript

Minuets 42:00-55:59

[Dr. Gates] I’m ah I’m ah researcher I’ve set up a sensor I may be I may be, so here’s a scenario I could be grabbing them and putting them in there and if I’m grabbing and reusing somebody else’s you wanna know that this is being monitored but its reused I may go and change it and change the parameters. Then I think that’s what Dr. Pennington was talking about you want to then store that information I reuse it but I changed it.

[Dr. Pennington] Yeah and you don’t want to override because, this property and the way it was developed worked well for that camera right there and now I’m going somewhere, else maybe way far away and I liked that property but its not quite right for this new location. So I’m gonna build on this property I’m gonna use it here, now a person three years later comes in and they’re looking at these properties and they wanna be able, one of the things they wanna do is be able to is say ‘ok they used it this way here and they used it this way over here.’ Or maybe there are some comments about how it worked and why it was changed and and why that modification was made, and I can learn from that and so now I’m gonna use this someplace else so I can say ‘ok well so here’s how it worked here and here’s how it worked over here and here’s how I might expect it to work in this place,’ and I can modify it so its important that I know the whole trace of where that property has been and how its been used. Um so we don’t lose any version so you can think of it as versioning umm and its important to keep all of that, and there’s one other thing, something that you said triggered something that I wanted to say while I’m thinking of it. Umm I mentioned that the biggest problem with the interface right now is that its in the language of computer science, you know I’m flippiedly using the word ‘property’ the very first time I met with Dr. Gates about this project that was my first question, what is a property? What are you talking about? So I’m using the word ‘property’ easily because we’ve been working on this now for two years but that’s not a term that the scientist the scientist think properties they’re thinking of the proprieties of the thing they’re measuring, and you guys are talking data properties and those the data properties are related to the properties of the thing on the ground but I had to clarify that I didn’t know which I was talking about.

[Dr. Gates] That’s a data dictionary moment.

[Student] Are the data properties like rules? Or how would you define them?

[Dr. Pennington] How would I define data properties?

[Student] Yeah are they the rules that you once you go into the sensor and you pull information, does the data properties you setup to find what’s useful in that information or how is that set up?

[Dr. Pennington] Well so what’s the I’m not sure I understand your question, so I’m just gonna try and give you an example and see if that clarifies for you. I have a sensor I’m measuring temperature of something of the air, of the ground, of something so the temperature of the actual temperature measurement is a property of whatever it is I’m measuring. The entity I’m measuring so I have the measurement and I have the actual entity and it’s a property of the entity, but when I’m doing data properties for the way we’re talking about it I’m gonna say ‘this time period at this place the temperature should range from ten to thirty.’ Now you’re talking about the property of the data stream not the property of the thing I’m measuring, and that’s a subtle difference but its one that’s not immediately obvious to the scientist when talking about it, did that help at all maybe?

[Student] Maybe maybe not.

[Student] You’re talking more data constraints versus actual data information.

[Dr. Pennington] Right.

[Dr. Gates] Anomalies come based on predicting what you think the data how you think the data will be behaving what the measurements.

[Dr. Pennington] Yeah its comparing relationship and its comparing the data that you’re looking at the data, the data is now the entity of interest as opposed to the thing you’re measuring being the entity a subtle but important distinction.

[Student] going back to a question that Adam was saying earlier about transit, anything else you wanna be able to do just constraints at the moment appropriate constrains and anomalies? (garbled cannot be fully understood)

[Dr. Pennington] Yes.

[Student] So the users are able to defining the data properties?

[Dr. Pennington] The users are the ones who would define the data properties yes.

[Student] And when defining them you have to take the weather in to consideration?

[Dr. Pennington] uhh sorry I’m having a moment of cognitive dissonance here, something is not quire right about your question. Um the weather is what you’re measuring in most cases you’re measuring temperature, precipitation, and wind velocity and those are what you’re measuring and you’re collecting data about that’s what the properties are about what you expect in that.

[Dr. Gates] What you expected how do you expect weather to behave? And when you see a spike you wanna know when that spike occurs like maybe the temp raises and you’re not anticipating that, so its uh April and you don’t expect to have a snowstorm and when that happens you wanna be able to capture that moment, because it could tell you something.

[Dr. Pennington] I mean it would look like a snowstorm or maybe maybe it is a snowstorm or maybe its one of the instruments has gone haywire.

[Student] So I’m sorry, but if its not defined in the properties within the limits then its considered an anomaly?

[Dr. Pennington] Right.

[Student] Ok so do you think it would be helpful in that case to be able to somehow synchronize weather data?

[Dr. Pennington] Yh yeah yeah that’s a good idea if you knew or had some other source of information that you could use too.

[Dr. Gates] That’s number seven I think, question number seven on the second page is

‘Will we have access to database to check for predefined values historical values and time constraints?’

[Dr. Pennington] Yeah so you can imagine instead of me specifying the temperature ranges that I expect from one particular value, I can access the weather and climate databases and say ‘compare this value to that database over that data source over there,’ and I hadn’t thought about doing that but that but yeah that would be helpful.

[Student] How important is performance in terms of speed, for example would it be preferred to just have something that’s not gonna be very flashy I’m just gonna get you the very essentials in a very simplistic matter just quickly, or is it ok if it takes those extra seconds more to give you a very big but more not flahsy but more precise range of data. Like what’s more important when you’re using this thing do you get a lot even if it takes some time or do you just?

[Dr. Pennington] I think it’s always a good design decision to provide the things you can provide fast as fast as you can, and then you cordincall the others in the background. That’s always a good design decision, nobody likes to wait, for anything but somethings take time.

[Student] I’m asking like how imperative is it just a slight annoyance or is it gonna hamper you?

[Dr. Gates] How important is it? Is performance important basically?

[Dr. Pennington] Performance is always important but if I.

[Dr. Gates] Well your looking at tradeoffs, he’s looking at tradeoffs and I, that’s part of the analysis, I mean when you look at analysis and you start saying ‘if I’m gonna add these bells and whistles,’ that’s part of the analysis part so I think you wanna separate them.

[Dr. Pennington] I mean it would be easier if you had a real concrete sort of would you rather have us this information right upfront and then this is five-minuets later, or would you rather or we can provide this in two minuets and if we don’t do this if you give me something concrete like that its an easier choice to make than.

[Dr. Gates] And you’ll be doing that later on so, I’m gonna jump to question ten if an anomaly has occurred because of a faulty sensor where the sensor needs to be recalibrated how should that be recorded, I think you kinda talked little bit about anomalies is there anything else you wanna add?

[Dr. Pennington] Well I think you know I guess you should proved some choices to the user in some cases they may want instant notification on their mobile. I mean its sorta like you know, I don’t know if you guys worked with the airlines at all they have all these choices about your ‘flight is delayed do you want to be notified an hour in advance or do you wanna know a day before,’ and ‘how do you wanna know? Do you wanna know by phone? Do you wanna know by email?’ When it comes to contact that’s a very individual sort of choice you need to provide those choices.

[Student] it wasn’t a follow up I hate to go back to the data property I have been, but from what I’ve read the data property I thought the data property is also some calculations some sensors?

[Dr. Pennington] Part of understanding a failure of a sensor is knowing how if this sensor has been showing lower and lower and lower values, and maybe this sensor is suppose to be showing higher and higher and higher you know they’re correlated maybe inversely, but there’s a correlation we’re expecting that correlation if the readings are not correlated like they’re suppose to then that can be an indication that there is something wrong with a sensor. So maybe almost certainly in many use cases will be comparing across many sensors.

[Dr. Gates] That’s a good question and we’ll be digging into that more on Tuesday.

[Student] I’m sorry I have a question that builds more on what he was talking about, how would you be able to distinguish between what would be considered an anomaly and what would be considered a faulty sensor?

[Dr. Pennington] Well yeah that would have to be the scientist would to do the analysis on that, I think its gonna be hard to tell just from the data unless you know, the data flat lines then you know that’s a problem with the sensor, but if it’s a calibration problems that’s much harder to detect so I think in most cases unless its just like a flat-out failure its gonna take some analysis by the scientists to try and understand that.

[Dr. Gates] Just to give you a scenario there is so much information that’s streaming right? Depending you know if you’re doing this or if its on constantly and part of the tool is just to help the scientist to look at those places where unusual things are happening and then be able to analyze so its not sophisticated at this time to be able to say with certainty that this is a problem that has occurred.

[Dr. Pennington] I mean just think about if we were working with satellite data, I might have petabytes of satellite data coming in on a weekly basis I just want something to call my attention to the places I need to look at.

[Dr. Gates] And that’s the purpose here, ok we’re gonna go to data analysis real quick we have about ten minuets ‘Are there types of analysis needed other than transect predictions?’ I think we talked a little about that I don’t know if there’s anything else to add.

[Dr. Pennington] I would say at least initially we want to keep the analysis pretty simple, I mean I think one of the uses I see in this is that as we start developing pat, you know we can get as data starts coming in and we start analyzing how they are in this way, how they are responding across sensors and in three space. We can use data properties as a way to call our attention to patters that are occurring, so we can analyze the data that we collect about anomalies we analyze those and better understand what we’re seeing so that sort of analysis I see happening down the road but I don’t see this since this system right now you develop a different system to do that analysis I think maybe just trends and simple things to start.

Summary: The discussion in this section of the interview started with an explanation of what a data property is and how it is used and that the usage and changes to a data property is to be logged, next the topic went to defining the verbiage that the current system uses versus what the targeted users are use to. Finally the discussion concludes with an explanation of what the system should do at its core functionality.