people historically never really thought about data as being important they thought about it as being necessary in order to perform a function and the function was always the application so data was very often designed specifically to be used within a single application

there were some sets of data that were common across a firm things like reference data or pricing data were used in common but beyond that was very enclosed in a narrow application

**how the data evolved in the past 25 years**

* we've kind of seen a few waves of that as we go through
  + I think post the financial crisis there was a big focus on data from a governance perspective for people to really understand uh do you know your data is your data of adequate quality
    - post great financial crisis there was a lot of demand for data in the risk management
  + as data science became more of a thing uh people realized that you just couldn't have the data and the applications
    - and then people wanted to access the data within their data science environments in order to do cool things with it
    - about use of data in the alpha generation side for Quant
  + the reference data function has historically been very important because without good instrument reference data nothing tends to work so people historically had teams of people very focused in getting the reference data right which is a very operationally intense area
  + within the buy side there was a big focus in the data around positions in order to do performance analytics risk analytics  so then again they had various teams that were heavily focused on bringing in all the fund accounting data processing it normalizing it before they could do their performance risk and returns
  + as we went through the crisis the focus on risk became important so then again you had people with both a risk background and a data background focused on really how do I organize the data so I can trust the numbers as regulation began to increase
  + is still a continuing effort because the importance of quality and data lineage is only going to continue to grow as people go through that Journey these areas have really been about structure data and very high levels of governance in parallel with that the research areas began to realize the value of data to some extent they always did because people built Financial models through Excel and spreadsheets where they made their assumptions Etc and that is an extensive area
  + we're at a phase where people have realized that actually the applications in the analytics can be pretty simple if one really has access to amazing data
  + the industry is really in a place where
    - it wants access to good data great quality
    - extensive depth of History
    - being able to connect and join them very simply to get to Value because ultimately this is a race for Speed
      * and the cleaner your data is the better the history
      * you've got the more joined up it is
      * the quicker an investor can get to the point of value
* the cloud has now enabled people to do things that were far harder in the past
* People have tried to unlock the data put into the cloud the types of tools and capabilities available today are dramatically different from where they were five years ago
* tools like snowflake which make data quering very easy things
* data bricks that have made machine tat have made machine learning every easy
* more quantitative approaches to find Value
* people have realized the importance of having the range of data scientists data Engineers who
  + can take the data Harvest it
  + create ideas and
  + then really work with it in order to drive Alpha
* people have seen the value of the results they are now d down in that area and in that process they are now investing in much better platforms that can take in data at scale that are much more structured where if the data is normalized once it has reusability so the next data scientist who comes around and has an idea doesn't need to go back to the basics but can benefit from the work that has been done in the past so the work is a creative data cataloging has become more of a thing where organizations want to Now understand
* their data better where it's stored and make sure that dat is accessible
* so the skills involved have grown through time it started off very much in a SQL type of area and now it's an uh SQL and databases structured databases and now we're very much in the space of data science where people use a lot more probabilistic techniques
* the data the one thing that is now common is people want data to be widely accessible across the entire organization so the same data that is being used for reference data and  operations needs to be the same data that is being used by data science is to generate Alpha
* see data span across the entire Spectrum from the front office all the way to the middle and the back office
* organizations wanting to share the same data across different functions
* it's clear that historically that probably wasn't the case but as use of data has increased whether it's at the risk management side or the alpha generation side or the trading side it's quite obvious in retrospect that they should want to speak to each other
* we look to the Future it's going to be a combination of both the structured and the unstructured data because now you can begin in the future to use large language models to work with both your actual numberss and along with textual docu

**pain points**

1. Institutional Investor their first challenge is their data comes from a multitude of sources it's not just one bank it's multiple Banks it's not just one provider it's multiple providers multiple vendors all of them deliver their data through different Technologies in different formats so the first pain point is how do I handle this vast amount of data coming to me and yet get it in the simplest possible way so I don't have to allocate my internal resources in order to ingest the data so that process needs to get really simple
2. the second aspect is once the data lands inside my organization how can I cleanse it and normalize it to make it useful because as it lands it's probably not very useful to my end user
3. that I've got the data hopefully it's been normalized how can I make it discoverable how can John know the data is present even though Jane was the one who brought it into the organization how do you really make the discovery process work then once you get past the discovery process
4. how do I give them access should this person have access to the data what's the means by which I can really entitle them so they can be able to easily access the data
5. once I've got that data then how can I easily use it within my analytical application

if you fast forward what people would really love to have is can all the data really function as an Uber Warehouse so I can then answer complex problems like give me all the portfolios that I have that are contingent on technology stocks in China but also have some dependency on raw materials in Indonesia and when one tries to answer problems like that one is really beginning to scan a vast amount of data the reference database the pricing database the research database data from vendors

 in a simple way being able to actually generate that result

* today it tends to be very hard because the data very often sits in silos and it doesn't come together

**here do you see the industry going where do you see the data management landscape**

first thing that will become increasingly relevant for us is AI itself right AI offers a massive amount of potential for investors in so many different dimensions and I won't dwell on the benefits of generative Ai and what people can do with that but what underlies Ai and unlocking AI is really the data because the difference between asking a chat GPT a question they're really going to answer

an answer based on public data whereas what we really want is you want an answer that's completely customized to all the information that we have and that is a hard problem uh that cannot be solved unless one really has full control and understanding of one's data

I think as people get excited about AI But realize its limitations Because unless You' got your data all sorted out thei tends to be less useful

industry will continue to be on a journey in order to make their structured unstructured data simple and easy to understand people to your point earlier about uh having the same data consistently they'll want to focus on that to eliminate duplication

it simplifies the connection between data producers and data consumers

investors can get clean ready to use data from any Source or domain directly into their Tech stack

allowing for analysis and insights in days and weeks instead of months and years it gives investors the ability to save 90% of their time in the data wrangling process

But what slows everybody down is really the data. The data is either dirty, it's proprietary to the organization, it's less understood, sometimes poorly structured, and then very often people just need an army of folk to get involved

it's trying to solve the complete data problem and really eliminate data as a roadblock

 And we really do this by giving institutional investors access to clean data across all sources and across all domains direct into their tech stack.

 can maximize the value out of their transaction and position data by receiving data through the Security Services Data Mesh and really being able to integrate it directly into Python Notebooks, APIs, STKs, and especially Snowflake.

you look at the data analytics journey, Snowflake really does a great job at the beginning and at the end.

It's great in bringing the data into the organization. And once the tables are finalized, it's great at people using analytics and BI data on top of that. What's really hard is the stuff in the middle where you don't just need the technologist, you need the domain experts, the analysts, the data wranglers on different domains really working together

 We have all the experts with us, we take care of all that middle stuff for our clients. So a technologist just goes into Fusion, very quickly configures this perfectly laid out physical table of data, they can have their opinions, their overrides, et cetera, and the data flows straight through inside the system.

 we not just connect with the vendors and institutions and bring the data in, we allow our clients to express their opinions on top. They can change the hierarchies, they can apply overrides, they can decide propagation rules, they can set up their own calculation criteria, streaming criteria