

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO



Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

by, David Vellante

August 2nd, 2020

According to reports, Snowflake recently filed a confidential IPO document with the U.S. Security and Exchange Commission. Sources suggest that Snowflake's value could be pegged as high as \$20B. In this week's Breaking Analysis we address five questions that we've been getting from theCUBE, Wikibon and ETR communities. ETR's Erik Bradley provides data and insights [...]

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

According to reports, Snowflake recently filed a confidential IPO document with the U.S. Security and Exchange Commission. Sources suggest that Snowflake's value could be pegged as high as \$20B.

In this week's Breaking Analysis we address five questions that we've been getting from theCUBE, Wikibon and ETR communities. ETR's Erik Bradley provides data and insights from customer roundtable discussions and colleagues David Floyer and George Gilbert contribute additional analysis and insights.

Five Questions from the Community

Investors, customers and technologists have been buzzing about Snowflake's prospects but five key questions have surfaced that we will unpack in this analysis as shown below:

Summary

Investors posing questions about Snowflake's IPO:

1. What's happening in the EDW market?
2. Whither legacy EDW players?
3. Can Snowflake compete with the big cloud players?
4. What's the viability of Snowflake's multi-cloud strategy
5. Does Snowflake's TAM justify its \$20B valuation?

To add some color to the questions. People want to know:

- Isn't the data warehouse market dead? Far from it and in many respects Snowflake did what Hadoop couldn't, bring the data warehouse back to life.
- What about the legacy EDW players like Teradata, Netezza, Oracle, Vertica, etc. Can they continue to compete and what are their prospects? How will they impact Snowflake?
- Snowflake competes with the largest, most technically capable companies in the world. Can its focus and innovation engine allow it to continue to thrive in a crowded field?
- Snowflake differentiates with a multi-cloud strategy. As a cloud-independent player, that clearly opens opportunities for Snowflake to operate within clouds but are there opportunities to provide a data layer across clouds?
- Finally - does Snowflake's total available market (TAM), justify a \$20B valuation?

Erik Bradley's [commentary on the question set](#).

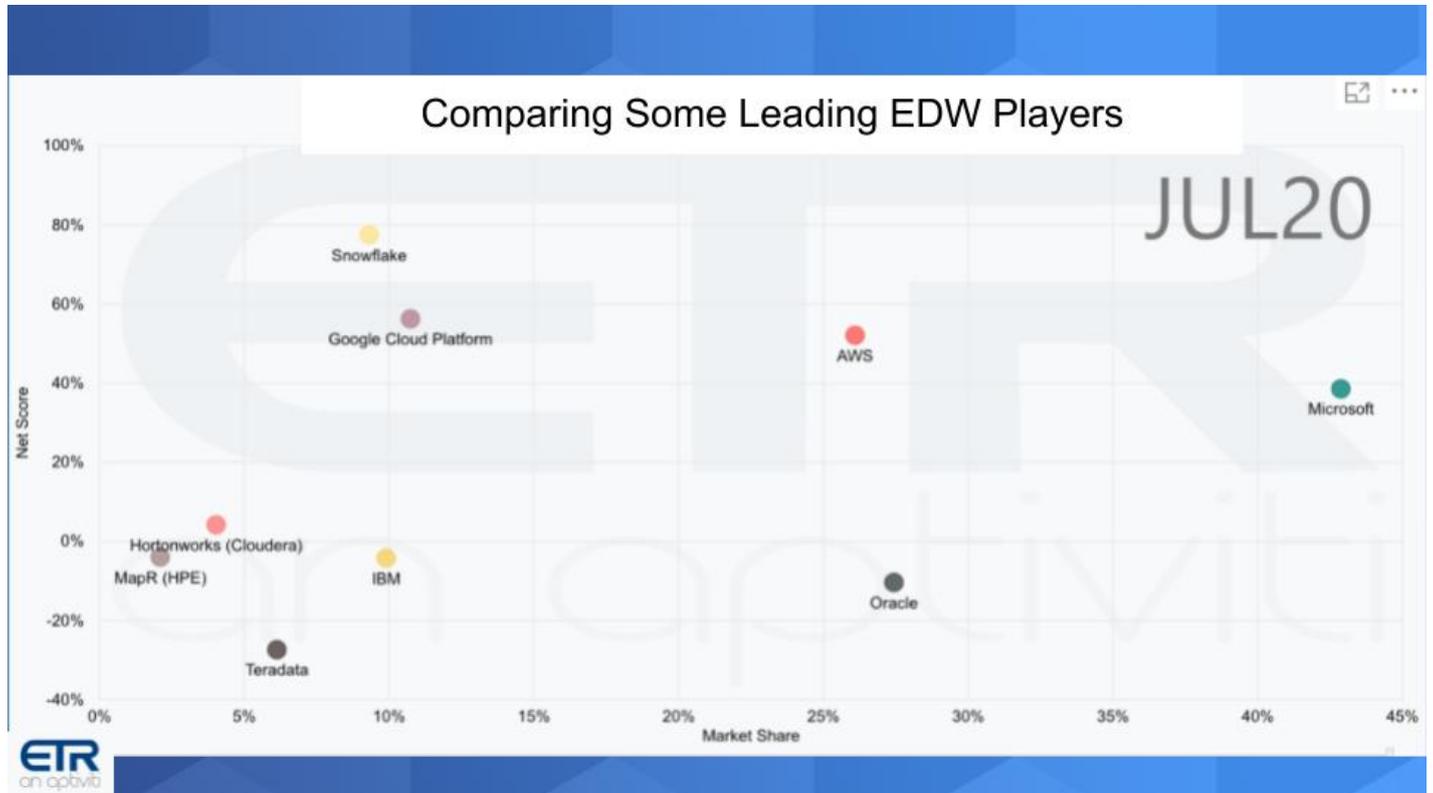
Note: Data from the July ETR survey asks respondents to compare second half spending intentions with the first half of 2020.

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

Q1. What's Happening in the EDW Market?

The enterprise data warehouse business is approximately \$20B globally, comprising traditional on-prem installations, the Hadoop data lake and the fast-growing cloud database space.

The chart below takes a snapshot of key players. The data shows survey results from nearly 1,200 IT buyers taken in the most recent July ETR survey.



The chart above shows a vendor's position on two dimensions. The Y-Axis shows Net Score or spending momentum. It is derived by asking customers each quarter are you spending more or less on a specific vendors platform. The higher the Net Score, the greater a vendor's momentum in the data set. The X-Axis shows "Market Share." Market Share is a measure of pervasiveness in the data set and is calculated by dividing the number of mentions of a vendor within a sector divided by the total overall mentions within a sector.

Let's start with some commentary on the players.

Microsoft's data comprises both its cloud and on-premises businesses. As such, Microsoft is most pervasive in the data. It has a corresponding Net Score of nearly 39%, which is respectable but not off the charts.

AWS Also stands apart from the pack in both pervasiveness and Net Score (53%). A significant portion of respondents are running Redshift.

Google lacks the presence of Microsoft and AWS but its Net Score is higher at 56%. BigQuery is a major offering of Google in the data set.

Snowflake leads all players in terms of Net Score of nearly 78%. Not surprisingly, it doesn't have the pervasiveness of the leading cloud players but there are 111 responses in the data set and Snowflake continues to increase its presence in the market.

Q2. What about the Legacy EDW Players?

Erik Bradley points out in our interview that we're noticing a widening bifurcation between the cloud native and the legacy on-prem players. The data is concrete and the gap is getting bigger. However, we stress that nobody's going to rip out their legacy applications tomorrow. It takes years to move off of these platforms.

Some additional comments on these players:

Oracle has a major presence in the data set but its Net Score is in the red at -10%. Oracle respondents are running both on-prem EDW and cloud-based solutions. Cloud is much smaller for Oracle, its on-prem business is much larger but in managed decline. Oracle invests heavily in R&D and is the #1 database on the market so it's unlikely Oracle is going anywhere soon.

Teradata is fighting to maintain its installed base. The company invented the concept of an integrated database machine in 1984 and is investing in Vantage, its location independent data platform. Teradata's market cap is only \$2.3B today and so it doesn't have to drive nearly as much revenue to hold serve. But its position in the dataset is not good with a Net Score deep in the red.

IBM - IBM acquired Netezza in 2010, marketed the hardware appliance for years, pulled the plug on Netezza and then pivoted to a software version of the database and brought back the name. The company is re-tooling its data warehouse offering with the IBM Cloud Pak for Data platform that runs on Red Hat OpenShift container platform. Not the best marketing move by IBM but technically necessary. The confusion in the market has hurt IBM's momentum as seen in its Net Score of -4%.

Cloudera/Hortonworks/MapR - Hadoop ushered in the era of big data. Cloudera/Hortonworks essentially invented the concept and catalyzed ten years of data innovation. The big 3 players of the Hadoop era are now consolidated into Cloudera (Hortonworks) and HPE (MapR). All have Net Scores that reflect maturity and Cloudera is showing a positive number. The vision of the data lake set forth during the Hadoop era is evolving to the modern era and setting up an expanded TAM opportunity for the sector.

Customer Comments - Snowflake Euphoria with Cautions

ETR runs regular customer roundtables call VENNs. Erik Bradley hosts the VENNs and he brings in CIOs, IT practitioners, CISOs, IT architects and data experts. These are open and frank conversations, but it's private to ETR clients. What follows are some curated conversations from eight technology buyers and data experts run over the past two weeks.

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

“Snowflake is synonymous with cloud-native data warehousing...there are no equals.”

“ETL is completely disrupted.”

Potential icebergs ahead for Snowflake:

“Their engineering advantage will fade over time”

“What AWS has done to separate compute from storage for Redshift is largely a bolt on”

“Similar to what happened to Hadoop.” Public cloud players *“giving away these offerings at zero cost.”*

-Curated comments from a two roundtables with eight IT practitioners and data experts | ETR Venn 7/23/20 & 7/30/20

The first comment came from the head of global architecture at a Fortune 100 company.

Snowflake is synonymous with cloud-native data warehousing...there are no equals.

The second comment was telling and requires some additional discussion. ETL stands for Extract, Transform, Load and is part of the data integration / data preparation market. It's a market that is sizable with some estimates approaching \$10B worldwide. Increasingly, ETL is being referred to as ELT as the transformation can occur inside of modern databases. According to the practitioners on the panels, Snowflake and the other cloud players are embedding many ETL, ELT and data preparation capabilities within their stacks and this is why they feel the market will be disrupted.

Another example of this disruption is seen with AWS Redshift Spectrum, a feature that allows queries to be run against S3 data with no extract or load required.

Nonetheless, as [Erik Bradley points out](#) in our discussion, Snowflake partners with [dozens of leading data integration vendors](#).

Competitive Challenges to Snowflake

There were some contrarians on the panel. One member made the following comment in regards to Snowflake and discussed how he felt that the technology would become commoditized over time:

Their (Snowflake's) engineering advantage will fade over time.

In our [June Breaking Analysis](#) we discussed this topic specifically and cited several examples of competitors like AWS innovating on roadmap features. A major topic of discussion in the community is the separation of compute and storage resources that allows for independent scaling (and pricing) of resources. This was a huge breakthrough popularized by Snowflake from its inception.

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

At re:Invent 2019, AWS announced [RA3 nodes with managed storage](#) and Andy Jassy touted this as separating storage from compute. This led everyone listening to believe that AWS was copying a Snowflake innovation with true [elasticity as the lynchpin](#) of the architecture. Essentially meaning you can turn off compute resources when not in use. A clear example of an engineering advantage fading over time.

Here's where it gets interesting. RA3, which is built on AWS Nitro compute instances, really doesn't separate compute from storage in the way we normally think about that feature. What AWS does is really a form of storage tiering to move cold data off the Redshift compute cluster on to the S3 managed instance. The RA3 node remains active but this allows you to resize the compute to accommodate less data.

This approach is why one panelist - who is a large and loyal AWS customer - said the following:

What AWS has done to separate compute from storage for Redshift is largely a bolt on.

Perhaps. But the question we have is does it matter in the long run? It's possible AWS chose this approach because it's technically simpler and was faster to market. Or maybe there are Snowflake patent issues that AWS doesn't want to touch. Regardless, it's an enhancement that AWS can market as a way to lower your bill. And we have no doubt that AWS can provide dozens of customer reference examples using and loving Redshift generally and this new capability specifically.

Q3. Can Snowflake Compete with the Big Cloud Players?

This is one of the most frequent questions we get from investors, customers and the technology community in general. Snowflake has a strange and curious relationship with the cloud players. The fact is that Snowflake needs the cloud and the cloud players need ecosystem partners like Snowflake. So they will continue to co-exist, partner and compete. Each will choose its own path and the innovation roadmap will continue to benefit customers. This dynamic will no doubt be addressed in Snowflake's [Red Herring](#) filing.

There are two key risks to Snowflake that we want to highlight:

1. Cloud players will leverage their proprietary hardware and platforms to compete.
2. Cloud players could bundle and price to compete aggressively.

On point #1...AWS and other cloud players will over time develop proprietary systems - hardware and software - to become as efficient as possible and reduce their reliance on Intel. As an example, AWS uses Nitro for its RA3 instances. AWS uses specialized hardware, Arm-based processors and proprietary hypervisors to drive efficiency and low costs. While these innovations won't necessarily be fenced off for use by ISVs, AWS and other cloud players will have early access to these designs and will optimize their services to exploit these innovations.

Is this a Hadoop Replay?

Regarding #2 above, this comment from one panelist caught our attention:

Similar to what happened to Hadoop." Public cloud players ***"giving away these offerings at zero cost.***

This makes great sound bites but we don't buy the direct Hadoop comparison. Hadoop for all its groundbreaking good was difficult to scale but it set forth a wave of data innovation. Suppliers like Hortonworks and Cloudera provided committers and essentially funded critical open source projects and that have led to the modern data era. While this effort taxed resources it catalyzed innovations within the

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

technology community that have set the groundwork for the modern AI/data era.

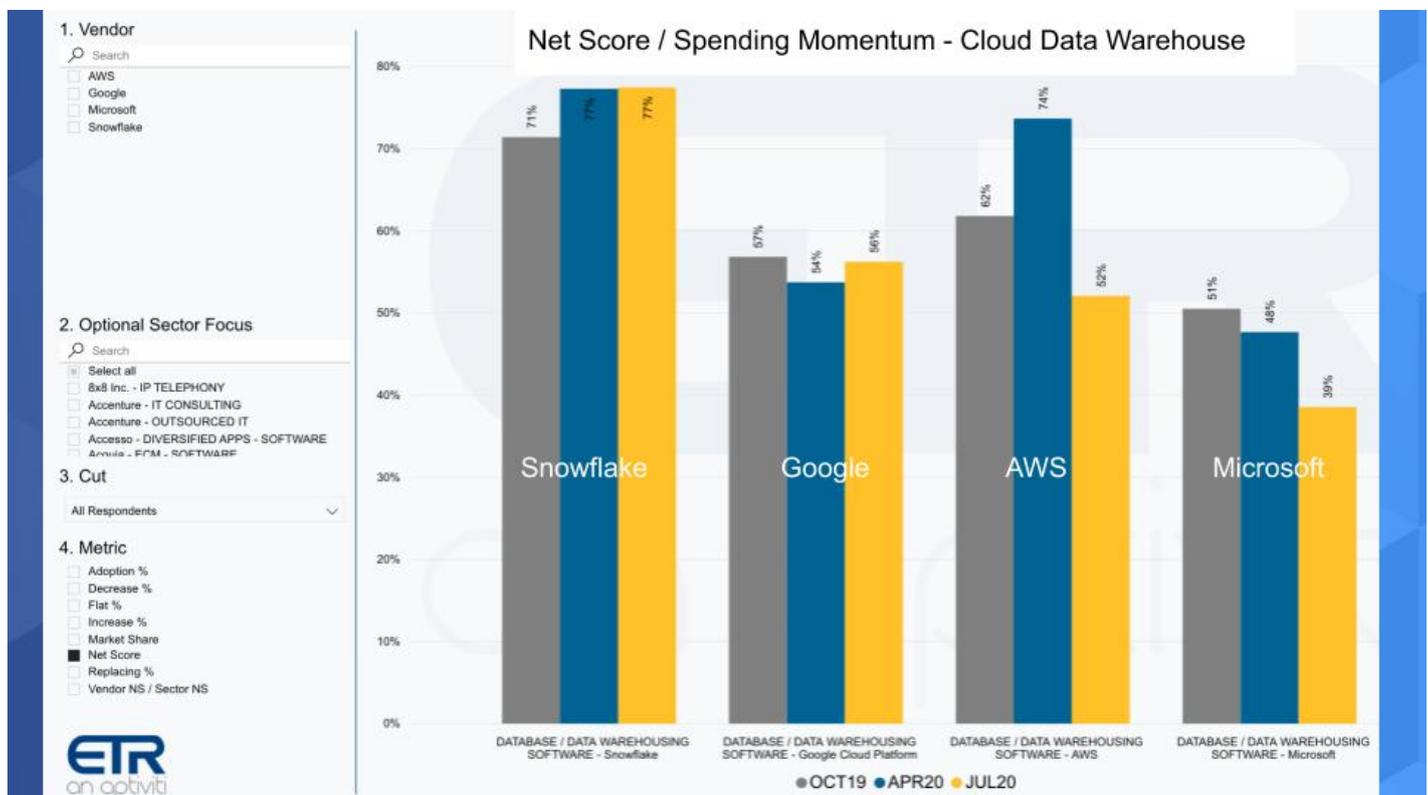
However this panelist's second point, while perhaps over stated (i.e. the cloud vendors aren't giving away data warehouse software - far from it), is important. Cloud vendors are working to automate and bundle services around the end-to-end analytic data pipeline from ingest to prepare to analyze to visualize and serve to lines of business. Snowflake plays an important role in this lifecycle but the cloud players are thinking end-to-end.

We see this playing out as the best-of-breed against the integrated suite argument that carried on for years in the technology business. Snowflake's challenge will be to continue to innovate faster in its core areas, while expanding its TAM to justify its valuation.

[Erik Bradley comments on this issue of bundling and Snowflake's valuation.](#)

What Does the Survey Data Tell Us?

As always in these segments, we like to anchor our analysis in data. The chart below shows Net Score or spending momentum and compares Snowflake to the Big 3 cloud players.



The chart above shows the last three ETR survey results for Net Score in the data warehouse sector. This survey data comes from about 1,200 respondents for each survey from last October, April 2020 taken at the height of the U.S. lockdown and the most recent July survey.

As you can see Snowflake is maintaining very high scores. Actually improving from the last October survey. Google shows lower net scores, but still very strong and increasing slightly from the April survey. AWS' Net Score dips notably from April. We think what's happening here is that AWS is exposed to so many industries that have been hard hit and, at its size, is a reflection of tighter IT spending. AWS reported results last week with nearly 30% growth, which is awesome but still slightly below expectations. Finally Microsoft is showing some notable softness as well in the EDW space, we think for reasons similar to what we mentioned with AWS - a very large exposure to broader IT spending trends.

Snowflake has the advantage of being a smaller disruptor and clear share gainer and the data clearly

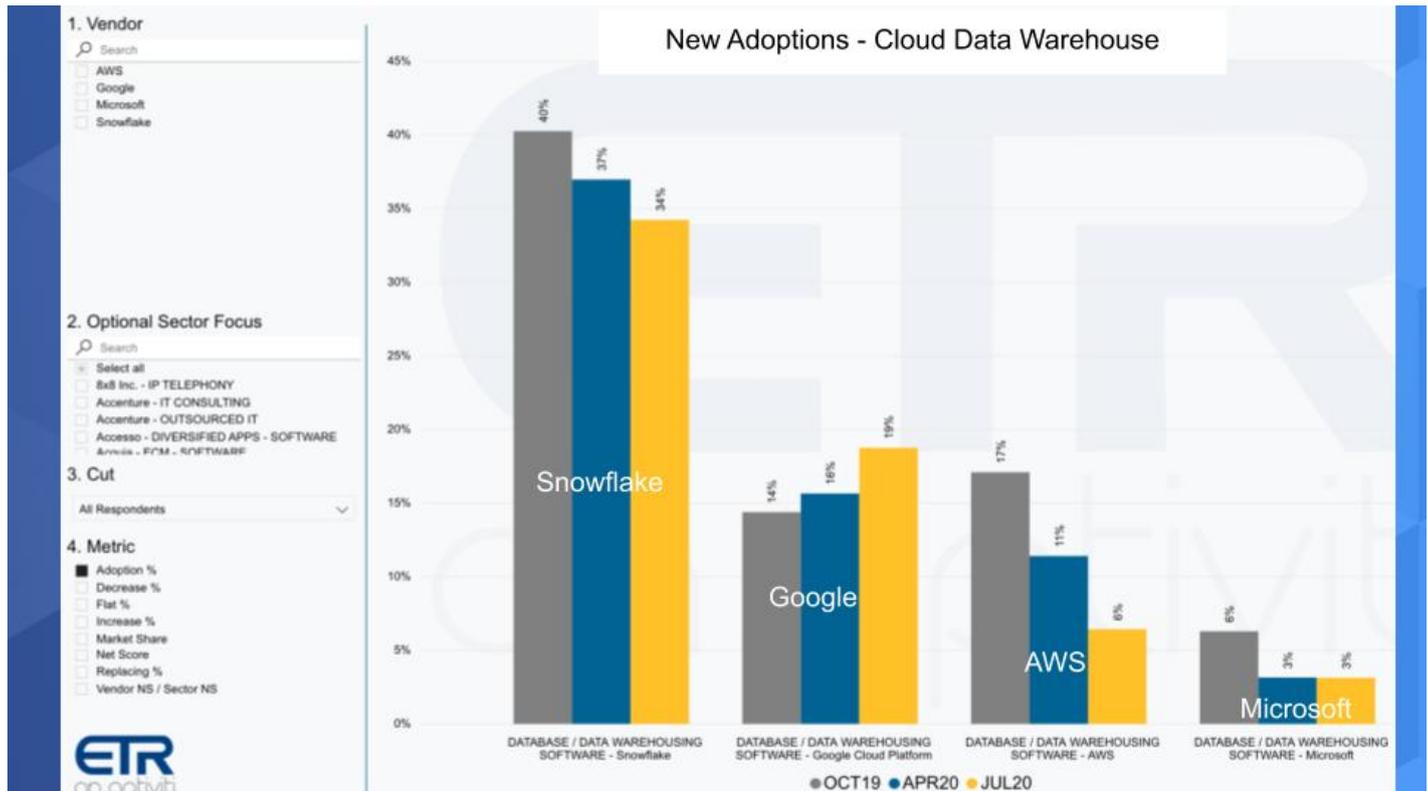
Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

supports that fact.

Snowflake is out outpacing everyone across our entire survey universe. - Erik Bradley, ETR

New Adoptions in the EDW Space are Even More Telling

Net Score has five components: New Adoptions, Spending More (6%+), Spending Flat (+ or - 5%), Spending Less (-6%+) or Replacing. The chart below isolates on the New Adoptions component of Net Score over the past three surveys.



You can see above Snowflake with 34% in the yellow, new adoptions, down somewhat from previous surveys, but still significantly higher than the other cloud players. Notably, Google is showing momentum on new adoptions, while AWS is down in a meaningful way. And again, AWS like Microsoft is exposed to a lot of industries that have been hard hit by the pandemic. Microsoft actually quite low on new adoption.

Bottom Line: Snowflake's smaller base (and Google to a certain extent) gives it a growth advantage relative to the larger players.

Q4. What's the Viability of Snowflake's Multi-cloud Strategy?

Practitioners in the ETR panel and theCUBE community cite three main factors when commenting on Snowflake's appeal:

1. Flexibility
2. Security
3. Multi-cloud

Snowflake is touted as easy to use and can connect to multiple platforms. Such commentary is not surprise.

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

However the comments on security were very surprising and we believe related to #1 - i.e. the ability to easily establish roles accelerates setting up systems and ensuring compliance with the security edicts of an organization.

Multi-cloud is consistently touted as an advantage for Snowflake and one we've reported on in the past. We've said many times that we felt multi-cloud was more a symptom of multi-vendor to data but increasingly is becoming a strategy. As Erik Bradley said in our interview, one of the things Snowflake has going for them is they can connect to all the major clouds. It's that multi-cloud ability and portability that they bring to you is such an important piece for today's modern CIO and data architects. They don't want vendor lock-in. They are afraid of vendor lock-in. And this ability to make their data portable and to do that with ease and the flexibility that they offer is a huge advantage right now.

So if you're a customer that is concerned about lock-in and/or you have different use cases running on different clouds and want to use a single cloud data warehouse across platforms you're likely to choose an independent like Snowflake. There are others entering the space as we've reported. For example, Yellowbrick Data is a startup company that has raised more than \$170M and is going hard after a hybrid EDW model with a modern architecture. Google as well has taken a more friendly multi-cloud stance and with its "ML out of the box" capability with BigQuery is winning favor in data science circles.

Guidelines for Assessing the Multi-cloud Opportunity

We see three major areas that investors should focus on as it relates to multi-cloud:

1. The degree to which Snowflake's multi-cloud strategy differentiates it from the Big 3 cloud players;
2. The level of value that comes from cloud native compatibility - i.e. I can run Snowflake on any cloud, versus;
3. Actually doing complex analytics and data joins across clouds and the degree to which this is a TAM expansion opportunity for Snowflake.

The third point requires more research and time to bake in our opinion. It's early days in the cycle and Snowflake is emphasizing its ability to support multiple workloads with one copy of data as a key advantage. We'll discuss this further in our TAM section but it's unclear how this will evolve.

In particular we see a massive opportunity for ingesting real-time data streams from machines at the edge and other data sources across clouds and on premises. Snowflake doesn't address real-time data streams today. As with AI/ML, it leaves many of these capabilities to its partners but over time could add these capabilities organically or through acquisition.

Bottom line is multi-cloud "compatibility" today is a clear advantage for Snowflake relative to the Big 3 cloud players. Whether or not cross cloud data and real-time data analytics can be a viable and differentiable strategy is a TBD.

Which brings us to the discussion of Snowflake's TAM.

Q5. Does Snowflake's TAM Justify a \$15-\$20B Valuation?

Former Wikibon and Gartner analyst George Gilbert submitted the following context for this research, which lays out the fundamentals of modern data analytic pipelines. This informs our starting point for understanding Snowflake's TAM and complements the previous discussion on the competitive dynamics of Snowflake's market, where the cloud vendors offer services in each layer of the data lifecycle. These are possible adjacent opportunities for Snowflake through acquisition or organic R&D.

The analytic data pipeline that manage today's data flows in order to automate or inform decisions has some basic requirements, no matter what products are involved:

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

Data Acquisition: At the very minimum, adapters that know how to talk to applications, databases, and other sources. MuleSoft and SnapLogic are modern examples.

Ingestion: These products move data from all the sources to some central location(s) where it can be sent to all the destinations that need particular sources. This process can operate like a PBX. Kafka is the best example.

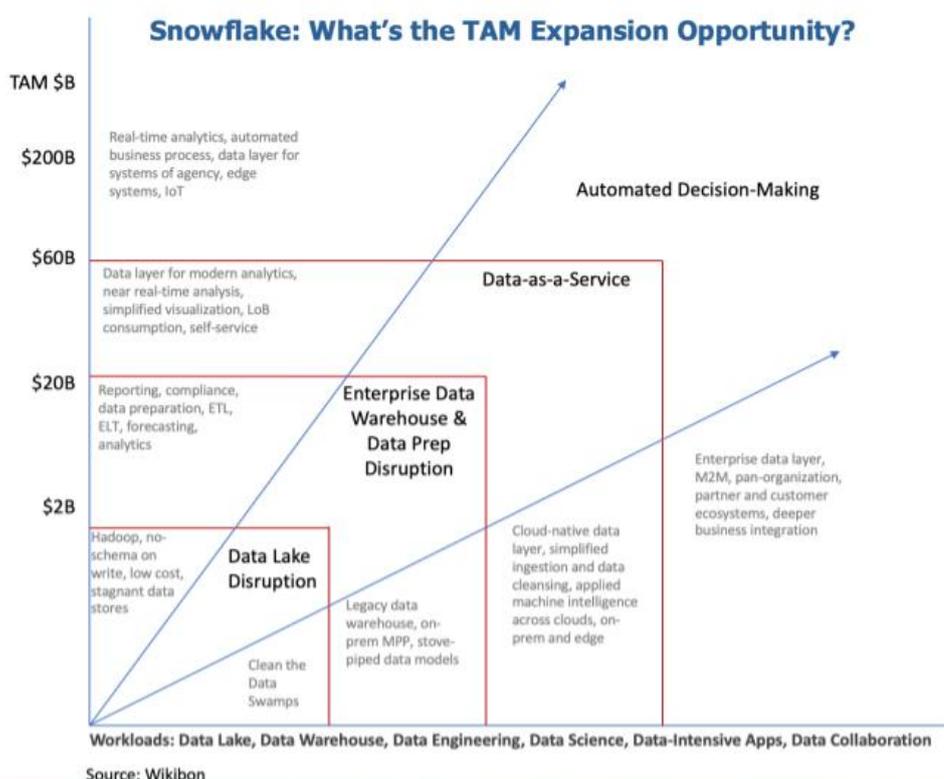
Organize: This used to be called ETL but has become far more complex with the explosion in data sources and the complexity of the data. Now, this stage has to prepare data for exploration, analysis, and AI/ML. Spark is the best example.

Analyze: This used to be EDW. Snowflake is much more than EDW, as we've stressed. EDWs were about historical performance reporting. Increasingly, they need to support forward-looking projections. At design-time, this requires data science tools. At run-time, it requires looking up data to pass to a service that does inference, which leads to the serving layer.

Serve: This is the run-time part of AI/ML, where the serving engine can be as sophisticated as a DBMS or as lightweight as a Web service. The serving engine gets the data values that either directly drive an inference or drive an analysis engine to do more lookups before getting an inference. Fraud detection requires looking up a lot of data in a short time before delivering an inference, for example.

Evolving the Data Pipeline to Support Automation and Deeper Business Integration

Imagine these components of a data pipeline evolving to support the following vision of a TAM expansion strategy for Snowflake:



We put together the above TAM expansion graphic to document and quantify our thinking.

Data Lake Disruption: Data lakes were born out of the Hadoop era and with no schema on write, data lakes became data swamps and ultimately just cheap data stores that were hard to leverage. Snowflake is bringing ease of use to the concept of data lakes and as [John Furrier often says - setting up the opportunity to create "Data Oceans."](#)

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

But cleaning up and growing data lakes into oceans doesn't really help with the advanced analytics that drives decision making.

EDW & Data Prep Disruption: Snowflake's entry in the market has breathed new life into the idea that data warehouses are not dead yet. But as we've stated, Snowflake is more than a data warehouse, which is largely used to describe what happened last quarter, not what's happening now. Or what will happen. But simplifying EDW, lowering its cost, making it more flexible, elastic and agile; and cloud-based allows Snowflake to take share in this space and grow.

Data-as-a-Service: We think of this as essentially a PaaS layer, or data data layer across data sources, across clouds, on-prem and the edge. Making it easy to ingest, prepare and analyze data from these sources and then serve the business with insights.

Automated Decision-Making: And then ultimately we have this huge TAM around automated decision making, real-time analytics, automated business processes. We peg this at hundreds of billions of dollars where data is being used in real-time to make decisions and take unattended actions. Of course attended automation will be part of this segment as well but the key here is automation and deeper business integration.

The bottom line is Snowflake's TAM is huge. Now of course Snowflake won't serve all of its markets. It will partner with an ecosystem and vertically integrate where innovation is required that drives new customer value and gives Snowflake a unique position in the market.

[Erik Bradley shares his thoughts on the Snowflake TAM.](#)

[Erik Bradley on some of Snowflake's M&A opportunities.](#)

Factors to Watch

Of course we'll read the documentation when it becomes public to see what else we can learn from the S3.

In addition we think there are four things to watch as this IPO progresses:

1. If good enough is good enough. In other words will customers gravitate to best-of-breed or will they prefer a more bundled offering from cloud players. The cloud giants will claim best-of-breed and we'll watch that closely but at the moment we feel confident that Snowflake has a position of product strength in its niche. The one exception to watch is Google with its out-of-the-box AI/ML capabilities.
2. Price Transparency. We haven't discussed pricing much but customer feedback clearly indicates some difficulties sizing and predicting Snowflake costs. Some customers have selected alternative cloud offerings because they feel more confident that they can transparently predict the monthly bill. We've asked Snowflake about this and the company believes it understands the issue and has tooling today that can help. In part this comes down to education and additional tooling to evolve price transparency. But in organizations where procurement has more influence this could be a blocker for Snowflake.
3. Ecosystem Development. With hot companies like Snowflake that are platform plays, ecosystem partners are critical and the enterprise giants always prioritize and excel at growing their ecosystems. This comes down to bottom line revenue growth for ecosystem partners. If they can make money with and Snowflake invests to support them the ecosystem will grow.
4. Pace of Innovation. Snowflake must stay ahead of the cloud giants to thrive. We believe Snowflake can continue its disruptive innovator status for quite some time but it's an important angle to watch, particularly with regard to its M&A and corp dev strategy.

On balance our data, feedback from customers and our community suggests that Snowflake is well-

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

positioned to make a run at becoming one of the next great enterprise software players.

As always, look for updates on the [ETR Website](#) and make sure to check out [SiliconANGLE](#) for all the news and analysis.

Remember these episodes are all available as [podcasts wherever you listen](#).

Ways to get in touch: Email david.vellante@siliconangle.com | DM [@dvellante on Twitter](#) | Comment on [our linkedin posts](#).

Check out this [ETR Tutorial we created](#), which explains the spending methodology in more detail.

Watch this week's full video analysis:

Breaking Analysis: Five Questions Investors are Asking about Snowflake's IPO

David Vellante is co-CEO of SiliconANGLE Media, as well as co-founder and Chief Analyst at The Wikibon Project, the world's leading open source technology research community. Dave is a long-time tech industry analyst, entrepreneur, writer and speaker. As co-host of theCUBE - "The ESPN of Tech," Vellante has interviewed over 5,000 experts since 2010. He is also a co-founder of CrowdChat, an angel funded startup based in Palo Alto using big data techniques to extract business value from social data. Prior to these exploits, Dave founded a CIO consultancy and spent a decade growing and managing IDC's largest business unit. He lives in Massachusetts with his wife and four children where he is active in town activities including serving as the president of his town's local "Kiddie Sports" association. Dave holds a B.S. in Applied Mathematics from Union College.



David Vellante

@dvellante

david.vellante@wikibon.org