

POC DISCOVERY QUESTIONS

#	CATEGORY	QUESTION	RESPONSE	WHAT TO LISTEN FOR
1	General	What are your strategic goals and end-state in the use of cloud data warehousing?		
2	General	Please share any current state architecture diagrams that describe your existing RDBMS solutions. Please include sufficient detail on the infrastructure to be used as a comparison to Snowflake		
3	General	Please share any current state source to target data flow details. Incorporate with item 2 as appropriate. Do you have critical POC success factors and/or functionality requirements that must be met before acceptance of a new platform? What will determine if it's successful or not?		
4	General		SQL Compliancy Data Load Rates Multi-Workload Performance Integration with ETL / ELT Tools Integration with BI / Reporting / Analytic Tools Minimal Management Ability to Scale Up and Back Down Price / Performance Other	
5	General	Are their specific POC milestones and dates you've identified as part of a broader success plan?		

When will you be able to provide

6	General	DDL
		Sample Data Extract
7	General	Full Data Extract
		SQL Queries
8	General	ETL Code
		Sample
9	Source Systems / Load Files	Who will be the key POC evaluators of the Snowflake platform?
		Who will be the key POC evaluators of the Snowflake platform?
10	Source Systems / Load Files	Who will be the key POC evaluators of the Snowflake platform?
		Who will be the key POC evaluators of the Snowflake platform?
11	Source Systems / Load Files	When you would like to wrap up the Snowflake POC by?
		How many source systems do you have feeding your current environment?
12	Source Systems / Load Files	What types of sources are these (i.e., ERP, web logs, etc.)
		Where are your source systems located (on premises vs. cloud)?
13	Productivity Tools and Processes	Are there additional sources you intend to add long term?
		If so how many and what type?
14	Productivity Tools and Processes	How do these need to be loaded - full refresh or incremental?
		What BI / Reporting / Analytics / Data Science tools are in use or do you plan to use?
15	Productivity Tools and Processes	Do you have any homegrown applications you use for BI / Reporting / Analytics?

14	Productivity Tools and Processes	Do you have SQL benchmark queries, workload profiles and performance runtimes documented for this POC?
15	Productivity Tools and Processes	What ETL/ELT are in use or do you plan to use? If none, what do you plan to use for your data pipelines? Are you using any
16	Productivity Tools and Processes	replication or change data capture tools today? If so, which ones? How many ETL/ELT jobs do you currently have across your EDW / Data Marts / Analytic Data Warehouses?
17	Productivity Tools and Processes	Can you share some sample simple/mediu m/complex jobs with Snowflake for analysis?
18	Productivity Tools and Processes	What file types and formats do you desire to load into Snowflake? Delimited, JSON, XML, Parquet / Avro, Logs, other?
19	Backup and Recovery	What do you use to perform backup of your Data Warehouse / Analytic Data Stores today? How often is data backed up from each system described in the architecture?
20	Backup and Recovery	What type of backup is performed (full, incremental) How long does each backup process take?
21	Backup and Recovery	Is the system fully available or are there operational impacts caused by the backup processes? Do you have a database recovery SLA?

22	Backup and Recovery	Have there been any challenges meeting the BAR SLA's? How about other SLA's such as point in time rollbacks?	
23	Backup and Recovery	How do you provide for BI/DW data disaster recovery today? (remote replication, offsite tape, restore to alternate system, etc.)? Do you currently leverage a third-party SI to assist you in the any of the following areas?	
24	Consulting Partners	Basic training and workshops re your data warehouses Migration of data Advanced database support including query optimization, clustering, data modeling/dat a warehousing, etc.?	
25	Consulting Partners	Would you be interested in leveraging an SI to help with your Snowflake POC?	Having a clear understanding of the business value is critical in establishing a clear success criteria. Important to identify business sponsors and technical owners for any Data Science PoCs.
26	Data Science - Project Level	At a high-level, what business problem are you looking to solve?	

27	Data Science - Project Level	In terms of Data Science processes and skills, would you describe your organization and team as a beginner, intermediate or advanced Data Science practitioner?	Recommend Auto-ML tools such as DataRobot / H2O.ai for beginners. Intermediate/ Advanced: SageMaker (AWS), Azure ML (Azure), Google ML (GCP). Reach out to #datascience for more recommendations.
28	Data Science - Project Level	What cloud provider do you prefer (AWS/Azure/GCP)? Do you have any preference for any Data Science / Machine Learning libraries or frameworks?	Some customers might have a strong preference for a certain platform or library. Even if we don't have a formal partnership, Snowflake should be able to work with most platforms. Reach out to #datascience if Databricks is involved.
29	Data Science - Project Level	Are there any specific compliance & governance requirements such as ability to show lineage of ML models?	Reach out to #datascience on Slack to see if we have partners that can support this.

30	Data Science - Data Collection	What kind of data (structured, semi-structured or unstructured) would be used for training?	<p>If the use-case involved unstructured data, see if it would be possible to use Snowflake as the repository for meta-data that points to unstructured objects in cloud object storage such as S3</p> <p>This will help you further understand the use-case.</p> <p>Are there any data sets within the Snowflake Data Marketplace that can help with this?</p> <p>Important to know how data will be loaded into Snowflake?</p> <p>would be possible to position a simple dashboard built using SnowSight as a way to do simple visualizations and model evaluation dashboarding</p> <p>.</p>
31	Data Science - Data Collection	Is this first-party (owned by customer) or 3rd party data?	
32	Data Science - Data Collection	How are you getting this data? Flat files / batch-loading or streaming data?	
33	Data Science - Visualize & Understand	How do you currently visualize and explore the data? Do you use any BI tools within your Data Science or analyst teams?	

34 Data Science
- Transformation & Feature Engineering
Data Science Are you currently using any ETL tools? How is the data transformed and prepared for regular analytics and BI reporting?
- Transformation & Feature Engineering

This will inform whether the customer has a strong preference to use an existing ETL tool. If not, we should encourage them to use Streams & Tasks within Snowflake. Our strategy should be to promote SQL transformations as much as possible that can be done in Snowflake efficiently at scale. However, certain kinds of transformations and feature engineering will require Python/R coding.

35 Data Science
- Transformation & Feature Engineering
Data Science Is your team comfortable using SQL or Python for transformation?
- Transformation & Feature Engineering

36	Data Science - Training & Deployment	Is your Data Science team comfortable working in a notebook environment such as Jupyter? Are they skilled in Python or R?	<p>This will inform whether to recommend a notebook based ML platform such as AWS SageMaker or Azure ML. If the team doesn't have much experience with Data Science and/or coding in Python / R, an Auto-ML partner such as H2O.ai or DataRobot should be recommended.</p> <p>Make sure the ML platform you recommend or is chosen by the customer supports model deployment for real-time as well as batch inferences / predictions.</p>
37	Data Science - Training & Deployment	Once the model is trained, do you anticipate doing bulk predictions or use the model for real-time predictions? Or both?	

			Open ended question as some customers might give you a specific metric and number such as 90% model accuracy. Others might say that they don't initially care much about the model performance but want to see if the ML workflow can be implemented on top of Snowflake efficiently.
38	Data Science - Model Evaluation	How would you measure success for this PoC?	
39	Data Lake	What is the volume of the data set that you'd like to test in the scope of this POC?	
40	Data Lake	At what rate do you anticipate ingesting data? Streaming data or batch copy? Volume per day?	
41	Data Lake	Where does the data currently sit? Cloud blob storage (S3, Azure Blob, or GCS)? On-Prem (Hadoop)?	
42	Data Lake	Do you have a comprehensive list of the file format types for the POC and full Data Lake? Structured, semi-structured and unstructured types?	
43	Data Lake	Does you want to ingest all of the data into Snowflake as the core data lake or leverage External Tables and Materialized Views in the testing process?	
44	Data Lake	What tools, technologies and connectors do users currently leverage with your data lake? Spark, python, Hive, EMR, Apache Ranger, TensorFlow, etc., etc.?	

- 45 Data Lake What are your biggest pain points with the DL today and what are your key success criteria in this POC? Cost performance, complexity/maintenance, performance/scale of transformations or queries, governance/security, value realization, other, all of the above?
- 46 Data Lake Do you have any requirements around managing unstructured data (video, audio, image, free text, etc.) in your data lake ecosystem?
- 47 Data Lake What security controls do you currently have in place in your DL? Access controls, governance, SOX, auditability, etc. Is there anything missing that should be addressed in the scope of the POC?
- 48 Data Lake Are there any limits that you'd like to test with Snowflake in a data lake scenario with semi-structured data sets? Concurrency, throughput, ingest speed, storage volume, query performance at scale?
- 49 Data Lake Is disaster recovery/replication/business continuity in scope for this POC? If so, what is the data volume that you'd like to test for this feature?
- 50 Data Lake Do you have any requirements around rehydrating your external Data Lake? Testing of partitioned unload?
- 51 Data Lake Do you use Hive Metastore to manage your data lake today?

			Would be good to have a stack that a better integration with Snowflake (where a native connector exists). Example: Node JS is better than PHP
52	Data Applicati	What is your current/proposed development stack look like?	
53	Data Applicati	What is your expected concurrency? How many users would access the application concurrently?	This is determine the Multi-cluster Warehouse configuration Snowflake can slow down the app if there are lots of individual writes.
54	Data Applicati	Do you expect to have lots of individual writes? If yes, what is expected transactions per second (TPS) for writes/updates/deletes?	Propose that they have a Kafka like streaming service to batch up the writes. Snowflake primarily excels at analytical queries.
55	Data Applicati	Do you expect to have point lookup queries?	Inform the customer about our new Search service that can help speed up point lookup queries. They should. Inform them how zero-copy cloning can be leveraged for test/dev environments
56	Data Applicati	Do have separate dev/test/prod code pipelines and environments?	.

57	Data Applicati	How does your frontend talk to the database?	Important to understand how Snowflake will fit in that puzzle.
58	Data Applicati	(If there is a need for very fast, small queries to hit Snowflake) Do you use caching service like Redis today? Would you consider using one with Snowflake?	With a cache in front of Snowflake, we can take take workloads away from some in-memory DBs and simplify the stack.
59	Data Applicati	Is there a new product/feature that Snowflake will be supporting?	Imporant to understand this for business/pain alignment and 'who cares'
60	Data Applicati	Do you have a replica of your app/prod database that you do your work on? How is that experience like?	To uncover pain around working with multiple DBs.
61	Data Applicati	Are there any data quality issues/concerns with the current system?	
62	Data Exchang	Are there any governance requirements around the Data Exchange?	Figure out if there are any governance considerations that need to be baked into the POC
63	Data Exchang	Who will own it and be the administrator? (permissions, managing listings, adding participants)	See if you have the right exchange owner and if there is someone else that should be involved and get buy-in
64	Data Exchang	Will you be the sole data provider in the Data Exchange, or do you intend to have third parties also share?	Will the POC mainly have just one party sharing data? Or will a collaborative data provider/consumer workflow be part of the POC?

65	Data Exchange	What are the SLA's for updating shared data?	<p>Will a refresh interval on data have to be satisfied? (ie. update tables at x seconds/minutes/hours, and have shared secure view reflect changes on data consumer's inbound share)</p> <p>Will Dynamic Data Masking or user/role based access be a requirement for the outbound share? ie. sharing one view with multiple third parties where they can only see their own data, or have PII be abstracted?</p>
66	Data Exchange	Does Data need to be secured or anonymized?	

67

Data Exchang

Would you require one
exchange for all recipients?
Or multiple exchanges for
different groups?

Try to get a sense of stakeholder expectations around service and privacy (data consumer A may be uncomfortable being in the same exchange as data consumer B). Unless the customer insists on it, you may be able to "prove" the Data Exchange without creating multiple exchanges. Just be aware that this may come up as a mandatory requirement for the POC.

68

Data Exchang

Are the participants of the Data Exchange Snowflake customers? If not, would you set up an account that your company owns for them to use?

For POC purposes you could likely have Reader accounts act as "third party" Snowflake accounts, so long as the Reader accounts belong to the same Snowflake account that is the Data Exchange admin. Including actual third parties in the POC could introduce a lot of scope creep, maybe even risk. But on the flip side they could be an existing Snowflake customer, and ready and willing to be participate in the POC. Use your discretion, but I would lean towards minimizing scope by using Readers.