GENERAL DYNAMICS
Mission Systems

Data Lake Query Tool Scenario Based User Engagement Results

Shari Benko

Human Systems Engineering

Goals of User Engagements

Capture actionable user feedback to inform feature prioritization and UI designs through

USER GENERAL USER UNDERSTANDING

Goal: **Capture a user's current role and expectations for the Data Lake Query Tool.**

SCEN SCENARIO DEVELOPMENT

Goal: **Refine/define operational scenarios for the Data Lake Query Tool.**

TEST UI USER TESTING

Goal: **Capture UI specific feedback on how to improve the existing user interface to meet the various operational scenarios identified.**

SECTION
01

USER

General User Understanding

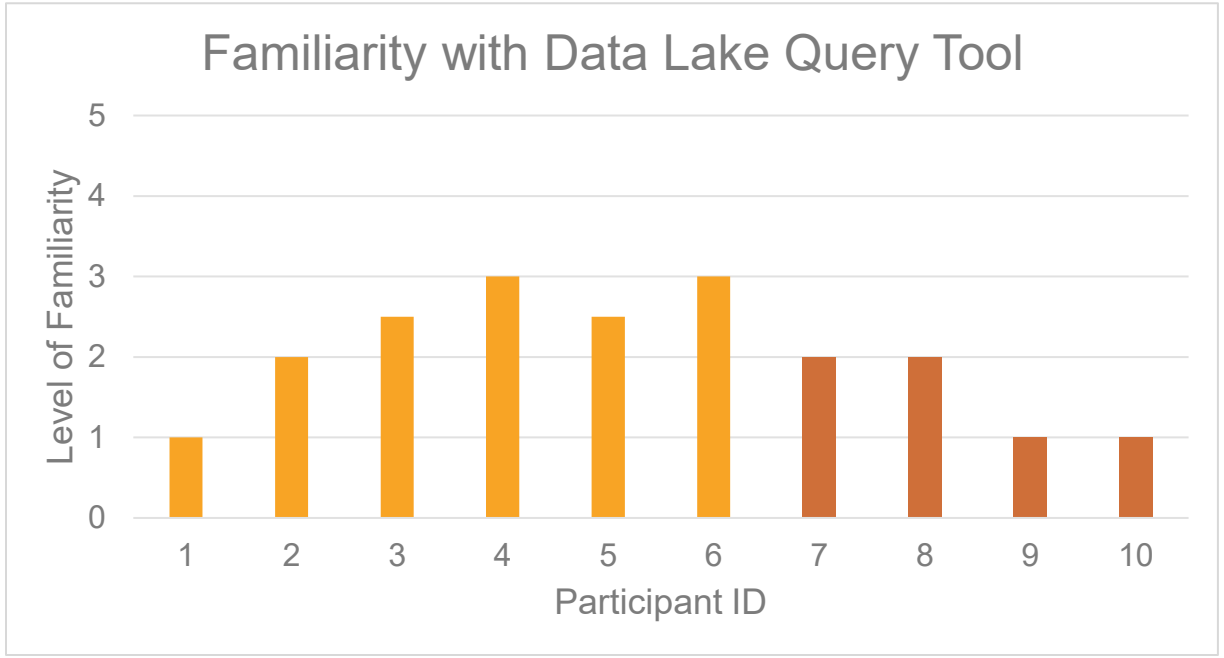
Intro and role-based questions were asked to capture a user's current role and expectations for the Data Lake Query Tool.

Overview of Participants in Training

USER

GOAL: Capture user's role and expectations.

- Participants were asked to rank their level of familiarity with the Data Lake Query Tool and if they support or work with operations data directly.
- ***Round 1:*** Only **1 of 6 Participants** answered that they do support operations directly, or work with operations data.
- ***Round 2:*** **3 of 4 Participants** were with LOCATION 2 and support operations directly.



ROUND	EXPECTATIONS OF TOOL VALUE
1	"...you're building this so that we have an ability to interrogate the data, build meaningful charts, etc. but so that we can do it in such a way that that it is not possible to manipulate the data in any way."
1	"For reporting and other purposes like you can look at XXXXXX data."
2	"It queries different information for something, you are able to go in and search information and get information back, but I don't know exactly what data will be provided"
2	"As far as I know, this is going to be a repository of information archive archiving system."
2	"Originally I understood this was going to be a raw data archive, but then I found out today that in one of our later tranches we may actually be using this to collect data and then move it to the appropriate user at a certain point."

User Expectations: Roles of Highest Use

USER

GOAL: Capture user's role and expectations.

- Users were asked to choose from a list of roles, which of the roles might have the highest need of use for the Data Lake Query Tool based on their expectations.

Warfighter Government
Authorizing Official Government
Security Control Assessor Government
LOCATION 1 Operator Contractor
LOCATION 2 Operator Contractor
LOCATION 3 Analysts Contractor
LOCATION 4 Analysts Contractor
Government Mission Director / Deputy Mission Director
Mission Ops Government

Round 1 From the **LOCATION 1** perspective **LOCATION 3 Analysts** and **LOCATION 1 Operators** have the highest need of use for the Data Lake Query Tool.

Round 2: *This question was eliminated from Round 2 to focus more on the scenario-based line of questions.*

Recommendations from Data

USER

GOAL: Capture user's role and expectations.

1. *Round 1:* Focus on "Operators" as the primary target users for future rounds of interviews.
2. *Round 2:* No further recommendations from this data.

SECTION
02

SCEN

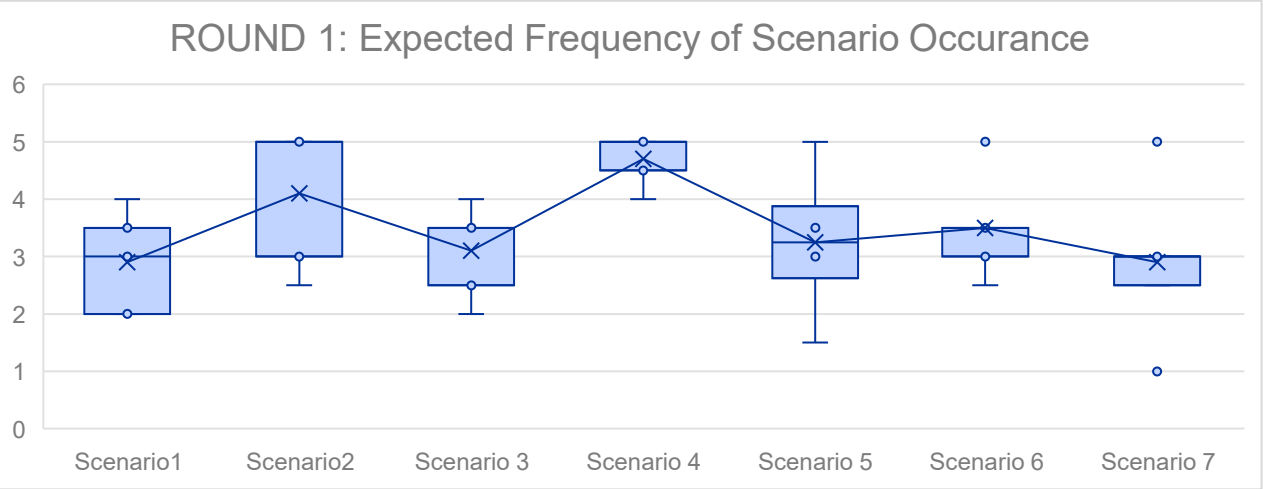
Scenario Development

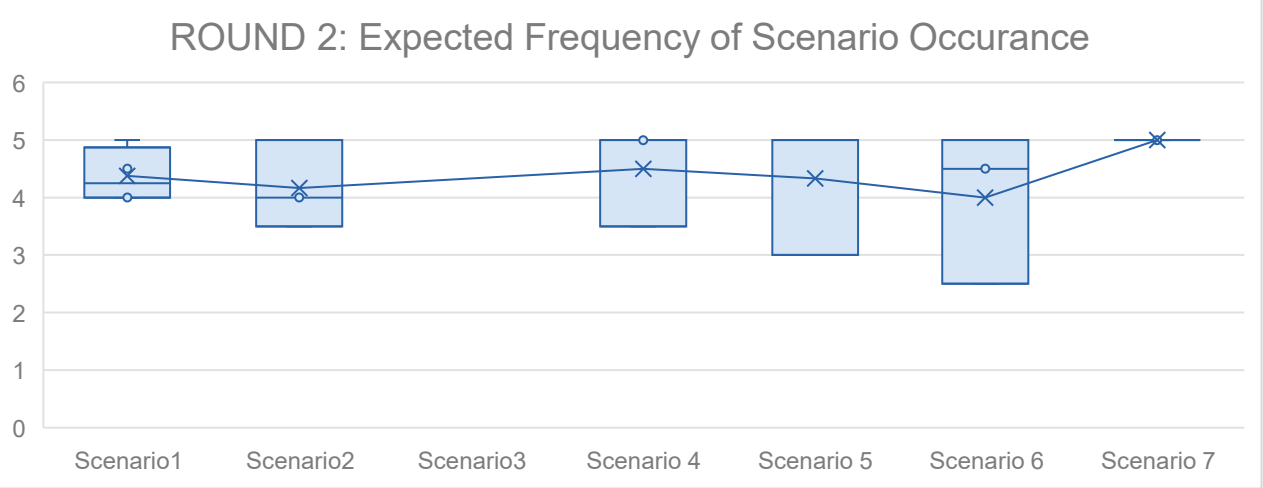
To help refine/define operational scenarios for the Data Lake Query Tool, users were asked to rank scenarios terms of how often they expect the scenario might occur.

User Expectations: Scenario Frequency

SCEN **GOAL:** Refine/define operational scenarios for the Data Lake Query Tool.

1	Users need to **access historical data** they didn't retain**.**
2	Users need to **access data from different environments** (Example 1, Example 2, etc.).
3	Users need to **access data they wouldn't otherwise have**.
4	Users need to **communicate data to others.**
5	Users **need to access data in** ways I don't already have available to me (e.g., **charts, map visualizations, grid views**).
6	Users need to **download data for offline analysis**.
7	Users need to **generate automated reports** from data.





ROUND 1: From the LOCATION 1 operators perspective **Scenario 4 and Scenario 2** ranked the highest in terms of expected frequency of need.

SCENARIO	MEAN	STD. DEV.
1	2.9	0.8
2	**4.1**	**1.11**
3	3.1	0.73
4	**4.7**	**0.4**
5	3.25	1.25
6	3.5	0.83
7	2.9	1.28

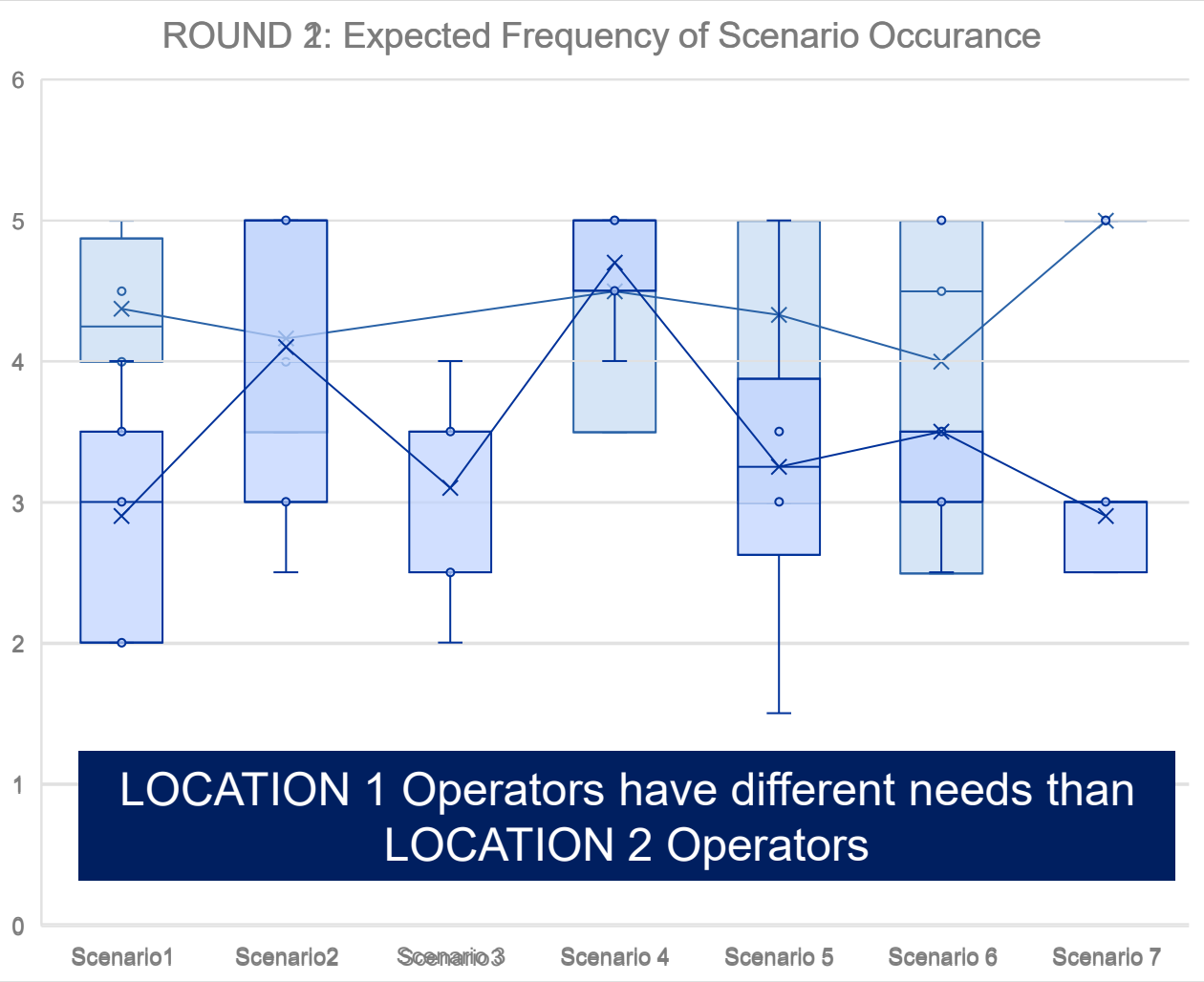
ROUND 2: From the LOCATION 3 operators perspective **Scenario 1 and Scenario 7** ranked the highest in terms of expected frequency of need.

SCENARIO	MEAN	STD. DEV.
1	**4.5**	**0.5**
2	4.1	0.76
3	N/A	N/A
4	4.5	0.86
5	4.3	1.15
6	4	1.32
7	**5**	**0**

User Expectations: Scenario Frequency

SCEN **GOAL:** Refine/define operational scenarios for the Data Lake Query Tool.

1	Users need to **access historical data** they didn't retain**.**
2	Users need to **access data from different environments** (Example 1, Example 2, etc.).
3	Users need to **access data they wouldn't otherwise have**.
4	Users need to **communicate data to others.**
5	Users **need to access data in** ways I don't already have available to me (e.g., **charts, map visualizations, grid views**).
6	Users need to **download data for offline analysis**.
7	Users need to **generate automated reports** from data.



ROUND 1: From the LOCATION 1 operators perspective **Scenario 4 and Scenario 2** ranked the highest in terms of expected frequency of need.

SCENARIO	MEAN	STD. DEV.
1	2.9	0.8
2	**4.1**	**1.11**
3	3.1	0.73
4	**4.7**	**0.4**
5	3.25	1.25
6	3.5	0.83
7	2.9	1.28

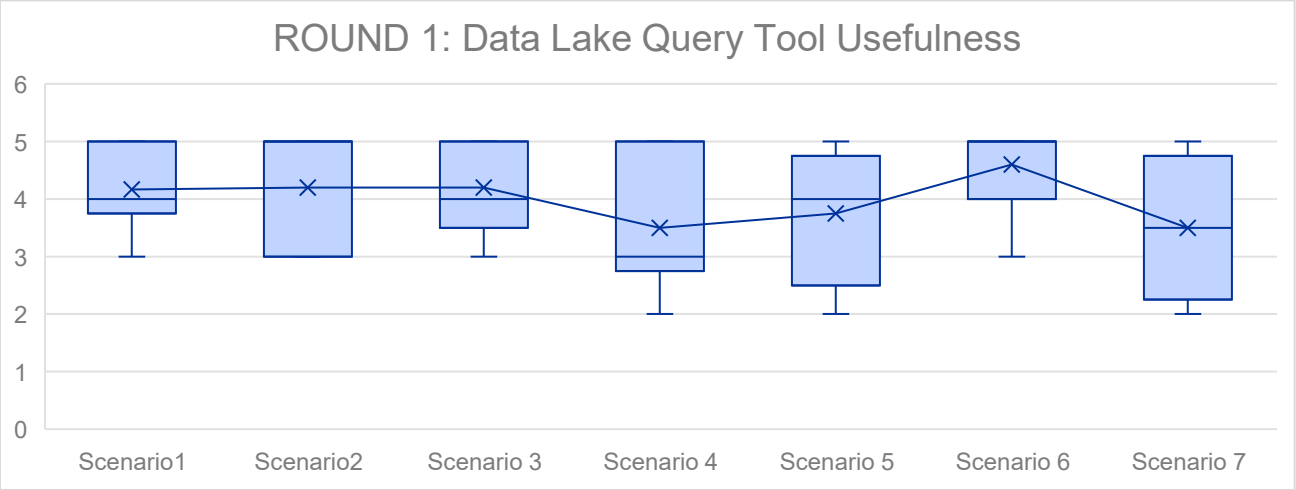
ROUND 2: From the LOCATION 3 operators perspective **Scenario 1 and Scenario 7** ranked the highest in terms of expected frequency of need.

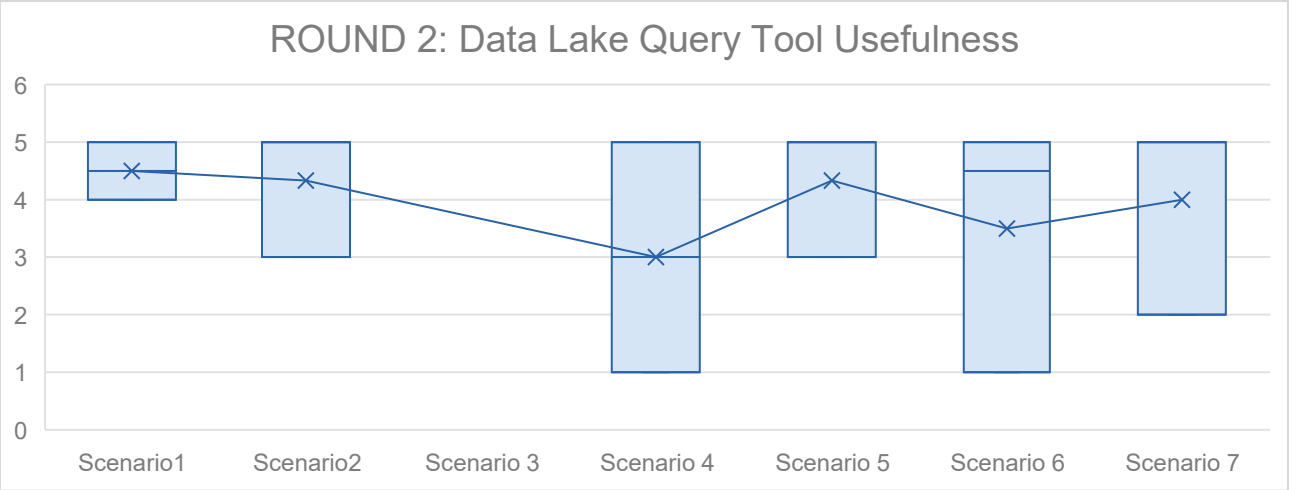
SCENARIO	MEAN	STD. DEV.
1	**4.5**	**0.5**
2	4.1	0.76
3	N/A	N/A
4	4.5	0.86
5	4.3	1.15
6	4	1.32
7	**5**	**0**

User Feedback: Tool Usefulness

SCEN **GOAL:** Refine/define operational scenarios for the Data Lake Query Tool.

1	Users need to **access historical data** they didn't retain.
2	Users need to **access data from different environments** (Example 1, Example 2, etc.).
3	Users need to **access data they wouldn't otherwise have**.
4	Users need to **communicate data to others.**
5	Users **need to access data in** ways I don't already have available to me (e.g., **charts, map visualizations, grid views**).
6	Users need to **download data for offline analysis**.
7	Users need to **generate automated reports** from data.





ROUND 1: From the LOCATION 1 Operators perspective **Scenario 4, 5, and 7** ranked the lowest in terms of the Data Lake Query Tool usefulness.

SCENARIO	MEAN	STD. DEV.
1	4.16	0.75
2	4.2	1.09
3	4.2	0.83
4	**3.5**	**1.22**
5	**3.75**	**1.25**
6	4.6	0.89
7	**3.5**	**1.29**

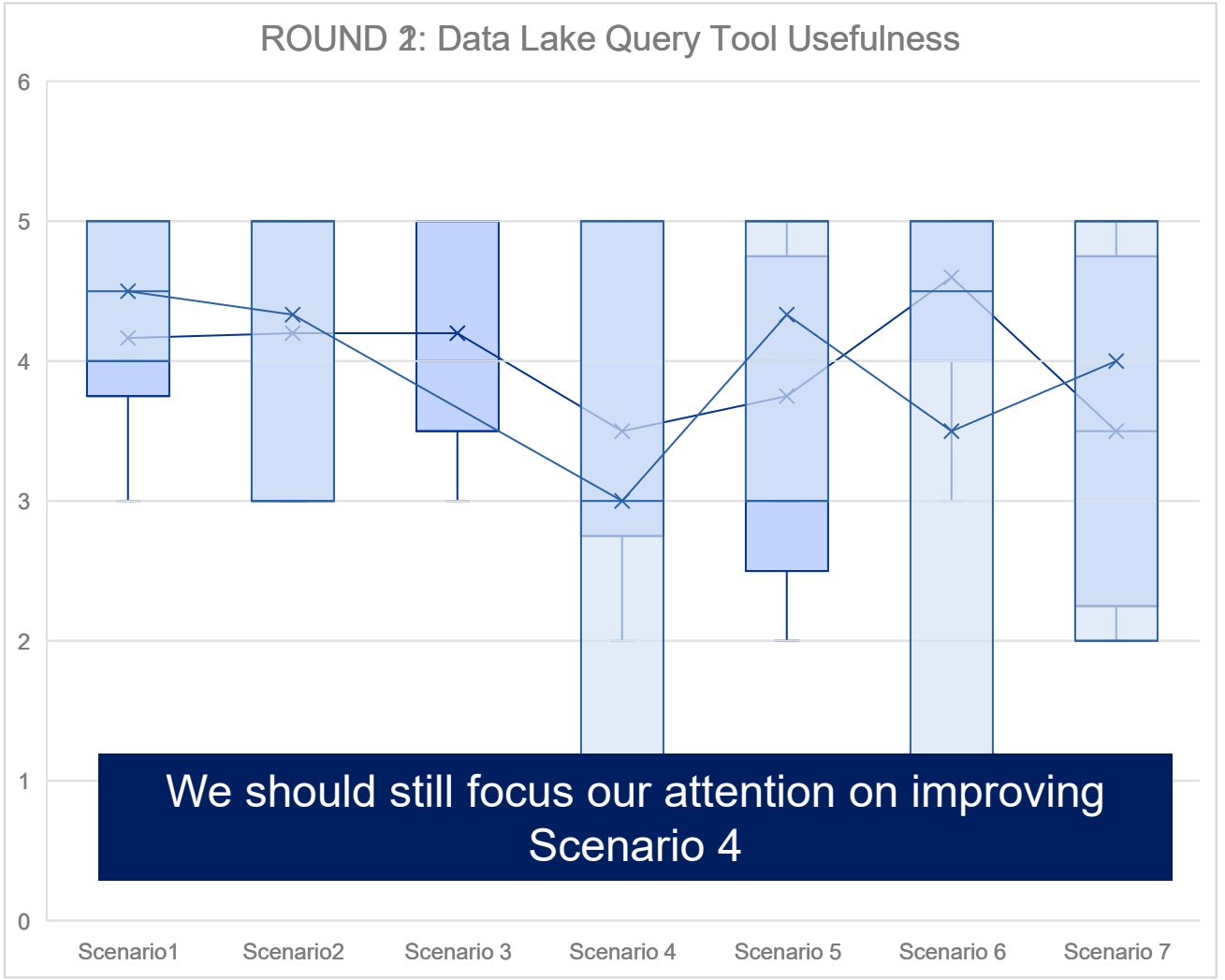
ROUND 2: From the LOCATION 2 Operators perspective **Scenario 4, 6, and 7** ranked the lowest in terms of the Data Lake Query Tool usefulness.

SCENARIO	MEAN	STD. DEV.
1	4.5	0.5
2	4.3	1.15
3	N/A	N/A
4	**3**	**2**
5	4.3	1.15
6	**3.5**	**2.18**
7	4	1.75

User Feedback: Tool Usefulness

SCEN **GOAL:** Refine/define operational scenarios for the Data Lake Query Tool.

1	Users need to **access historical data** they didn't retain.
2	Users need to **access data from different environments** (Example 1, Example 2, etc.).
3	Users need to **access data they wouldn't otherwise have**.
4	Users need to **communicate data to others.**
5	Users **need to access data in** ways I don't already have available to me (e.g., **charts, map visualizations, grid views**).
6	Users need to **download data for offline analysis**.
7	Users need to **generate automated reports** from data.



ROUND 1: From the LOCATION 1 Operators perspective **Scenario 4, 5, and 7** ranked the lowest in terms of the Data Lake Query Tool usefulness.

SCENARIO	MEAN	STD. DEV.
1	4.16	0.75
2	4.2	1.09
3	4.2	0.83
4	**3.5**	**1.22**
5	**3.75**	**1.25**
6	4.6	0.89
7	**3.5**	**1.29**

ROUND 2: From the LOCATION 2 Operators perspective **Scenario 4, 6, and 7** ranked the lowest in terms of the Data Lake Query Tool usefulness.

SCENARIO	MEAN	STD. DEV.
1	4.5	0.5
2	4.3	1.15
3	N/A	N/A
4	**3**	**2**
5	4.3	1.15
6	**3.5**	**2.18**
7	4	1.75

Recommendations from Data

SCEN **GOAL:** Refine/define operational scenarios for the Data Lake Query Tool.

1. ROUND 1: Because Scenario 4 ranked highest in terms of frequency of need, but lowest in terms of usefulness of our tool in addressing, we should prioritize the improvement of our tool in addressing scenario 4.
   - 5 of 6 participants suggested that a simple new feature that would highly improve the usefulness of the tool in addressing scenario 4 would be a screen capture feature for the map and/or charting data visualizations.
2. ROUND 2:

SECTION
03

UI

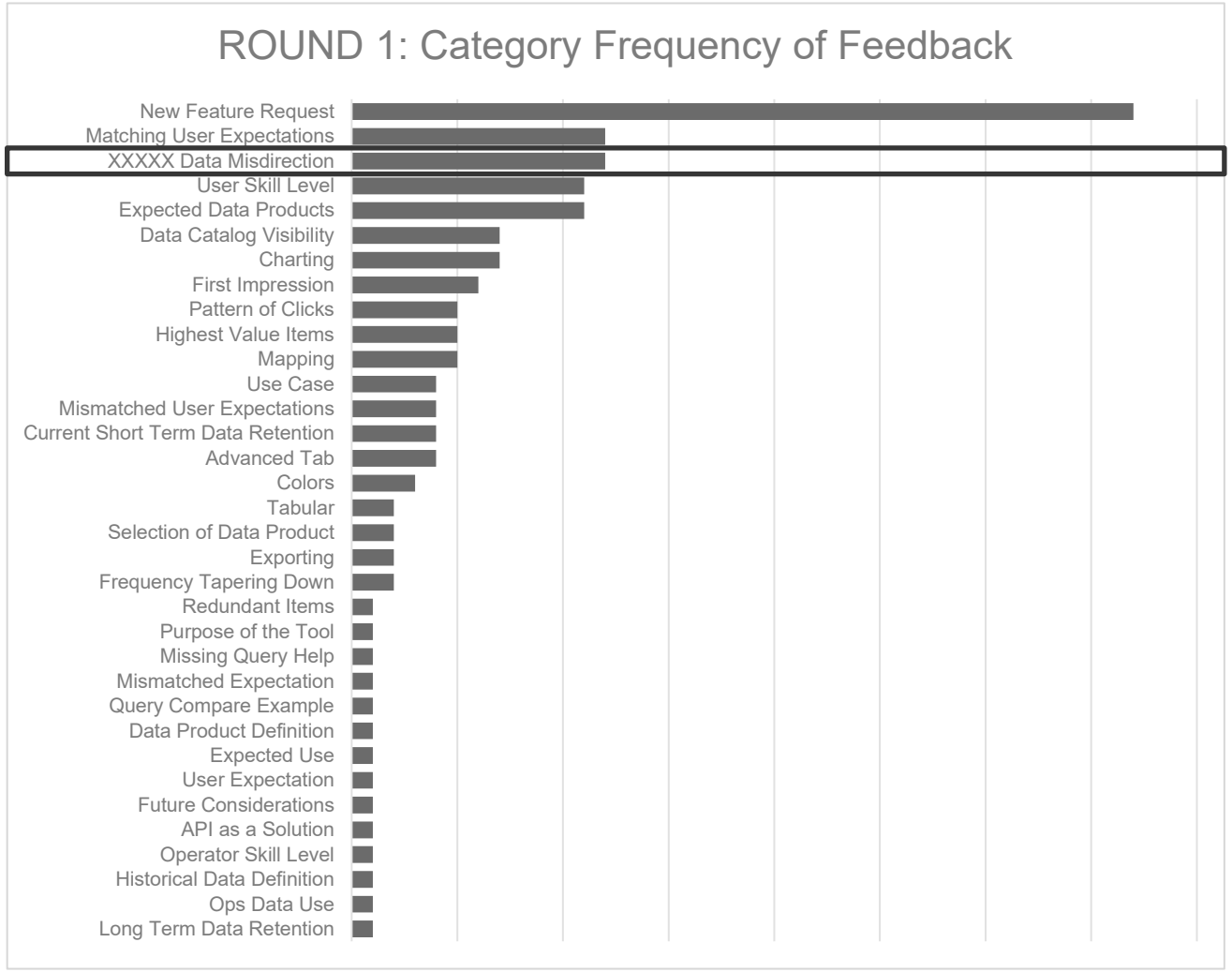
UI Feedback

The Data Lake Query Tool user interface was presented to users and questions were asked to gain valuable insights into how to improve the design of the tool to meet users' needs and expectations.

ROUND 1: UI Evaluation Overview

UI

GOAL: Capture UI specific feedback



- ROUND 1: First impressions were generally positive, however the focus on XXXX Data caused some first impressions that were biased towards negative and caused multiple negative comments throughout the interviews.

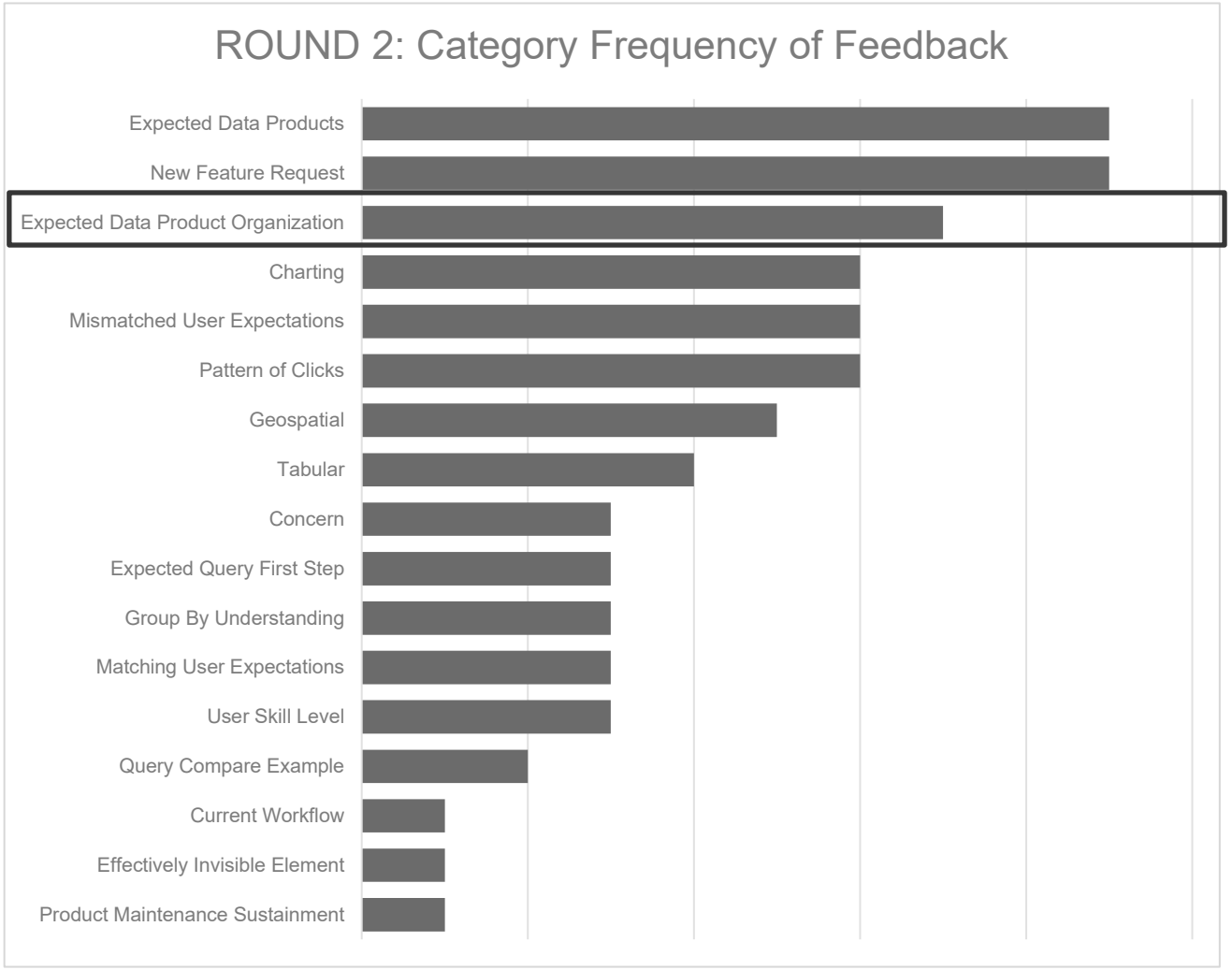
General First Impression Quotations

"I like the look and feel. It's clean, it's tidy, it's pretty."
"It seems like it is focused on data analysts more than it is on generic mission analysts, unless I am missing something."
"It's not too colorful."
"historical trending is something I know that people like to use a lot, having that begin and end date is useful."
"This is a sharp looking user interface."
"There's nothing that I see that doesn't make sense."
"I like the look and feel. It's clean, it's tidy, it's pretty."

ROUND 2: UI Evaluation Overview

UI

GOAL: Capture UI specific feedback



- ROUND 2: The XXXX data misdirection was not an issue in Round 2, however since the data catalog was front and center, we did see an increase in conversation regarding the organization of the catalog.

Concerns Quotations/Observations

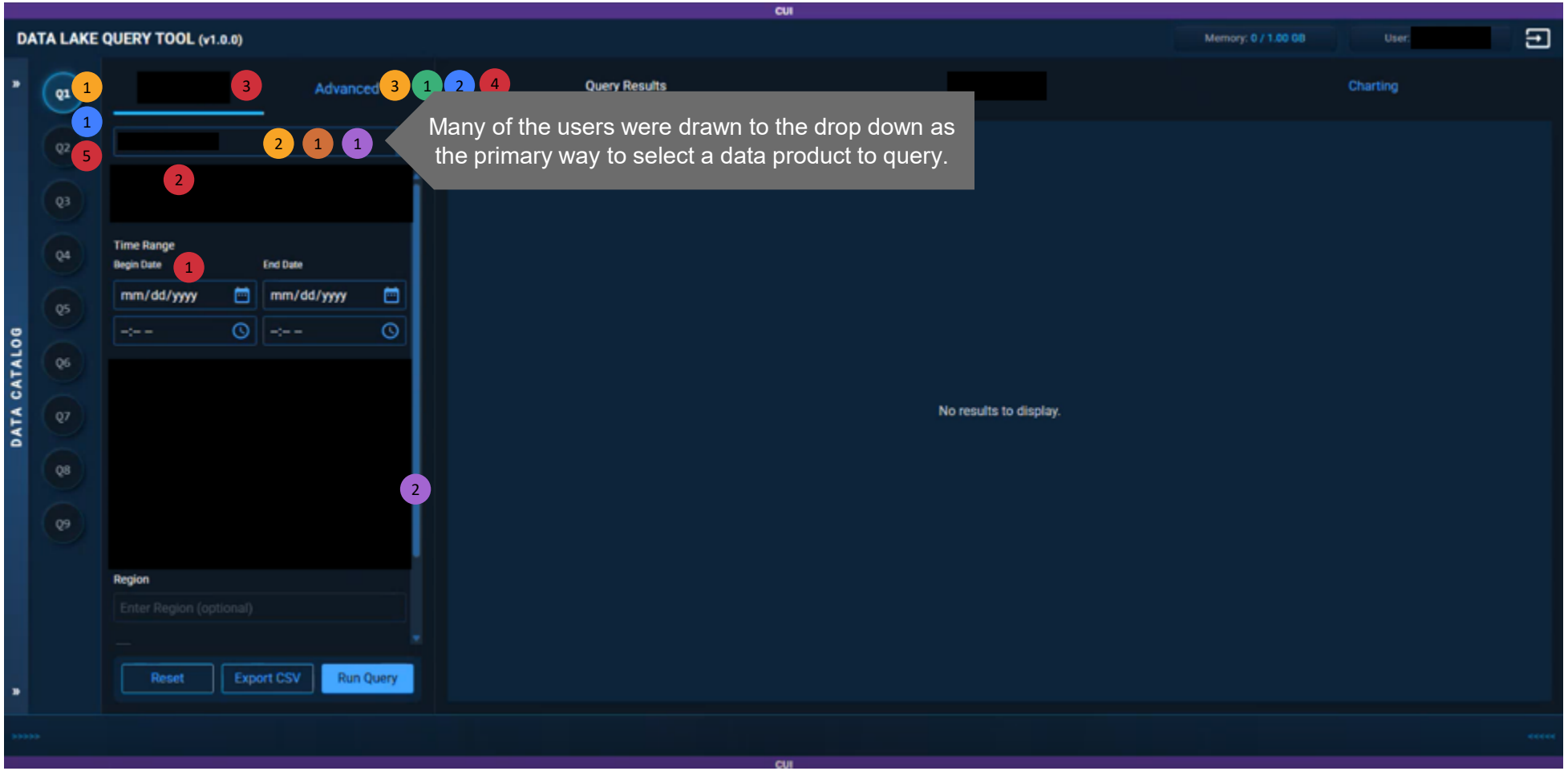
Users were concerned about the system performance of the map view in the operations environment, asking about whether this has been tested in an environment that closely reflects the operational environment.
“You are going to having people at both locations trying to access this simultaneously.”
“The charts are great, but there are limitations of hardware and the networ,k you need to test this in the actual environment we are going to use it in.”

ROUND 1: Pattern of Clicks Data Catalog

UI

GOAL: Capture UI specific feedback

- Participant 1
- Participant 2
- Participant 3
- Participant 4
- Participant 5
- Participant 6

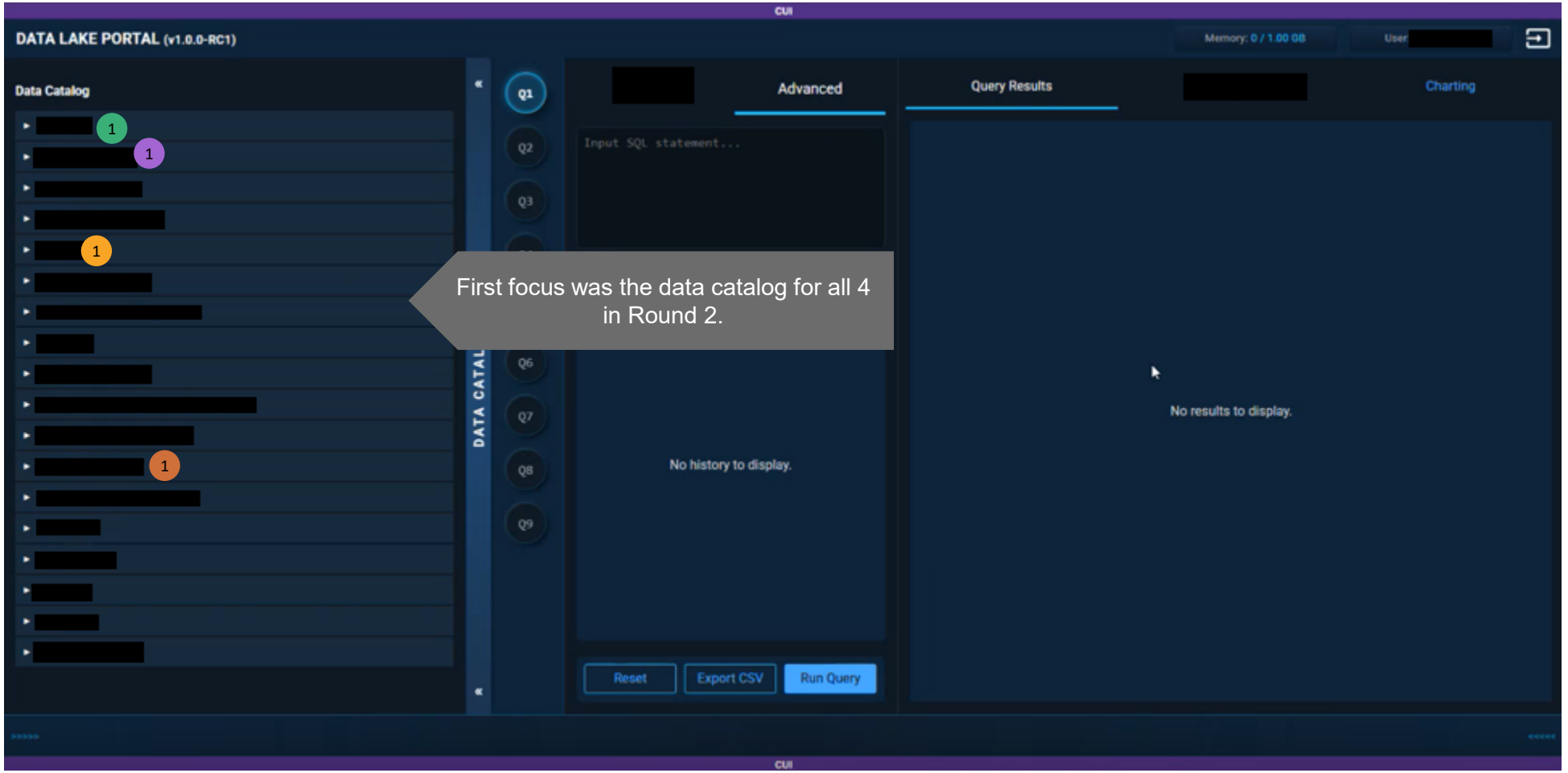


ROUND 2: Pattern of Clicks

UI

GOAL: Capture UI specific feedback

- Participant 7
- Participant 8
- Participant 9
- Participant 10



Data Catalog

UI

GOAL: Capture UI specific feedback

- ROUND 1: 6 out of 6 interviews did not see the Data Catalog when asked where to find data products.
- ROUND 2: 4 out of 4 interviews did see the Data Catalog first when asked where to find data products.

Round 1: Expected Data Products

Round 2: Expected Data Products

Querying

UI — GOAL: Capture UI specific feedback

- Round 1 and Round 2: All of the participants generally agreed that most operators would not have SQL skills that would enable easy querying using the Advanced Tab.
- Round 1 and Round 2: Most agreed that we need to provide both a basic query capability and an advanced query capability.

"I would probably say operators are probably a 2 and 3 (skill level on a scale of 1-5) or even entry level a 1 to a 2."	"My experience is that you have people that are fresh out of college or high school that just joined, as who the operators are, and maybe they don't know that stuff, maybe they aren't interested in software, so they don't want to learn it. "
"My very painful history, it tells me that most people do not want to be here (Advanced Tab SQL Query Input). So I think the vast majority of people would be in the XXXX tab (i.e., general query builder). Having said that if you don't have this tab, even if somebody doesn't know how to use it, they're probably gonna yell at you even though I think 9 times out of 10 people will not use this, I think if you don't have it, it will be a problem.	"You would have to have a basic knowledge just to understand what was happening and as you started to type out your own queries you would have to have a more in depth knowledge, so maybe they don't have Google on the ops floor so you can't just google it, if I couldn't do that and I didn't already know how I wouldn't have a way to produce the data I was asked to provide."
"I've worked and got really familiar with messing around with P SQL data bases and creating advanced threads, but most operators don't come with that."	Most of the operators themselves didn't have any coding background, there were a few of them but they were all civilian contractors, the XXXX folks would have that knowledgebase, but my history tells me that most of the operators don't.
"operator skill set is really important, if they operator has this skill then they will know how to retreive the data themselves."	I do think that if you wanted somebody to create their own queries to generate those reports and look at the charts that gets more complicated and requires significantly more user training.
"Some operators definitely know what they are doing, but some might be good a scripting not SQL."	How many folks would you expect to be comfortable with SQL? "Probably not very many on the operation side. Some of our engineers and people that are maybe doing analysis or things like that are probably going to be more familiar, but on the operations side of the house, not very many people."
I would say they (operators) would be just as limited as me for most of them (user ranked himself a 0 in terms of SQL skill level). Now we do have a couple with some experience. I think we have two operations support engineers who have experience with SQL, but other than that I think most of them would probably not. Will they be capable of learning? Absolutely.	"Right now I would say 90% of the people would fall within my bucket (a 1 on the skill level ranking)."

Data Visualizations

UI

GOAL: Capture UI specific feedback

- ROUND 1: Opinions were varied on which data visualization was most useful, this could be since these interview subjects were not working with ops data or supporting operations directly.

ROUND 1: Data Visualization Quotes

Quotes Redacted

Data Visualizations

UI

GOAL: Capture UI specific feedback

- ROUND 2: These users found the tabular and chart view the most useful for various use cases, map was less useful to LOCATION 2 operators.

ROUND 2: Data Visualization Quotes

Quotes Redacted

Round 1: New Feature Requests

UI — GOAL: Capture UI specific feedback

1. **Data Visualizations Enhancements:**
   - Export Charts and map to images.
   - Add histograms with confidence interval lines or Box and Whisker Charts.
   - Automatically apply regression modeling to trending charts.
   - Implement affinity diagrams for reliability analysis.
2. **Query Enhancements:**
   - Intelligent prompting for SQL commands.
   - Support for Analytic SQL.
   - Dynamic tailoring of query tabs to frequently used queries (i.e. basic query UI).
   - Maintaining and accessing (persistent) query history, including user and role details.
3. **Data Catalog Improvements:**
   - Search by app ID, generally expand search functionality.
   - Submit new data products through the UI.
   - Pin or prioritize frequently used data.
   - Provide larger viewing area for data catalog entries.

Round 2: New Feature Requests

UI — GOAL: Capture UI specific feedback

Green text indicates repeated requests from Round 1 to Round 2

1. **Data Visualizations Enhancements:**
   - Export Charts and map to images.
   - Add min and max average of the data points on the charts.
2. **Query Enhancements:**
   - Maintaining and accessing (persistent) query history, including user and role details.
3. **Maintenance & Troubleshooting:**
   - There were concerns that the map view with its animation might not work well in the operational environment, it was suggested that we make sure to test this in an environment that most closely represented the operational environment.
   - We were asked if there was a dedicated administrator for this tool after deployment, we should work to create a maintenance and troubleshooting plan for the tool itself (who is the main POC for users, where can users submit bugs, etc.)

Round 1: UI Feedback Recommendations

UI

GOAL: Capture UI specific feedback

1. First Impressions:
   - Expand our marketing of the tool to include higher frequency scenarios to combat the impression that our tool is mostly for long term data storage.
   - Take advantage of the upper left drop down interaction under the XXXXX Tab as a high visibility area for future designs.
2. Data Catalog:
   - Discover a more noticeable place to access the data catalog.
   - Incorporate a search in the data catalog.
   - Potentially investigate the use of the data catalog as a jumping off point for the Basic Query builder tool.
   - Investigate a “Expand to full screen” feature for the data catalog.
3. Querying:
   - Create a basic query builder tool to accommodate novice SQL users.
   - Investigate possible add-ons to support SQL query predictive or prompting for novice SQL users.
4. Data Visualizations
   - Prioritize adding feature for exporting Charting and map to images.

Round 2: UI Feedback Recommendations

UI

GOAL: Capture UI specific feedback

Green text indicates repeated requests from Round 1 to Round 2

1. First Impressions:
   - With the advanced tab as the default view users struggled to figure out where to start with querying, many of the users defaulting to the data catalog as the starting point rendering the advanced tab as essentially invisible. Recommendation is to build a basic query builder and take advantage of users first instincts to look for a starting point in the data catalog.
2. Data Catalog:
   - Users had conflicting recommendations when it came to what their expectations were the organization of the data; recommendation is to investigate the organization of the data products further in round 3.
3. Querying:
   - Create a basic query builder tool to accommodate novice SQL users.
   - Investigate possible add-ons to support SQL query predictive or prompting for novice SQL users.
   - Investigate the possibility to export query history as a data product.
4. Data Visualizations
   - Prioritize adding feature for exporting Charting and map to images.
   - Most of the Round 2 users did not understand what “Group by” allowed, recommendation is to relabel it or provide some help or hint text to make it clear what it does.

SECTION
04

Recommendations

The following is a list of recommendations based on the results of this user engagement.

ROUND 1: UI Feedback Recommendations

1. Focus on “Operators” as the primary target users for future rounds of interviews.
2. Because Scenario 4 ranked highest in terms of frequency of need, but lowest in terms of usefulness of our tool in addressing, we should prioritize the improvement of our tool in addressing scenario 4.
3. First Impressions:
   - Expand our marketing of the tool to include higher frequency scenarios to combat the impression that our tool is mostly for long term data storage.
   - Take advantage of the upper left drop down interaction under the XXXX Tab as a high visibility area for future designs.
4. Data Catalog:
   - Discover a more noticeable place to access the data catalog.
   - Incorporate a search in the data catalog.
   - Potentially investigate the use of the data catalog as a jumping off point for the Basic Query builder tool.
   - Investigate a “Expand to full screen” feature for the data catalog.
5. Querying:
   - Create a basic query builder tool to accommodate novice SQL users.
   - Investigate possible add-ons to support SQL query predictive or prompting for novice SQL users.
   - Add Help text to Advanced Query Input box
6. Data Visualizations
   - Prioritize adding feature for exporting Charting and map to images.

ROUND 2: UI Feedback Recommendations

Green text indicates repeated requests from Round 1 to Round 2

1. Continue to focus on “Operators” as the primary target users for future rounds of interviews.
2. Because Scenario 4 ranked highest in terms of frequency of need, but lowest in terms of usefulness of our tool in addressing, we should prioritize the improvement of our tool in addressing scenario 4.
3. First Impressions:
   - Expand our marketing of the tool to include higher frequency scenarios to combat the impression that our tool is mostly for long term data storage.
   - Take advantage of the upper left drop down interaction under the Tracking Tab as a high visibility area for future designs.
4. Data Catalog:
   - Incorporate a search in the data catalog.
   - Potentially investigate the use of the data catalog as a jumping off point for the Basic Query builder tool.
   - Investigate the organization of the data catalog in round 3
5. Querying:
   - Create a basic query builder tool to accommodate novice SQL users.
   - Investigate possible add-ons to support SQL query predictive or prompting for novice SQL users.
   - Investigate the possibility of turning a query into a data product for storage and even querying.
6. Data Visualizations
   - Prioritize adding feature for exporting Charting and map to images.
   - Most of the Round 2 users did not understand what “Group by” allowed, recommendation is to relabel it or provide some help or hint text to make it clear what it does.
7. Maintenance and Troubleshooting
   - Ensure test environment for this tool matched an operational environment before deploying
   - Create a troubleshooting and maintenance plan for this tool before deploying.