

FLHSMV-RFI-001-20 DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

February 27, 2020

Item No.1

Questions and Answers

The Department is seeking a solution with the initial capacity to host 30 Terabytes of structured data. The storage is expected to be adjustable on the fly. Storage options such as glacial storage will be considered.

Current Data Warehouse:

- Stages 300GB on average to support daily and weekly loads
- Ingests data into 314 tables
- Data is pulled from around 10 difference sources
- Data is stored in 860 reporting structures and has been growing 750MB per day

OLTP Business Data:

- Business Data Transactions: Average Daily # Transactions- 440,000
- Source Systems: Average Daily Growth 800 MB
- About 10 data sources and over 1000 tables
- Some larger source tables contain between 70-150 million rows.

Note: All written questions are reproduced in the same format as submitted by the Vendor.

Question(s) received on 1/29/2020 from Deltek:		
1.	Question:	<ul style="list-style-type: none"> • The VBS page here, states the responses due date as February 28, 2020. However the RFI document itself states the responses due date as February 27, 2020. Would you please clarify the correct date?
	Answer:	The VBS advertisement indicates the advertisement end date. The RFI document includes a calendar of events. Please use the calendar of events.
2.	Question:	<ul style="list-style-type: none"> • Do you happen to know the estimated funding source, or budget for this project? <ul style="list-style-type: none"> ○ If so, has funding for an RFP been secured?
	Answer:	No funding has been requested and/or secured at this time. This RFI will help the Department estimate the costs to further plan for potential project requirements and an estimated budget.
3.	Question:	<ul style="list-style-type: none"> • Is this a new requirement? Or is there an incumbent vendor providing these services? <ul style="list-style-type: none"> ○ If there is an incumbent, would you be able to provide the contract number, vendor name, and term of the contract?
	Answer:	The Department has no contract in place for the services described in this RFI.
4.	Question:	<ul style="list-style-type: none"> • Is there a timeline for an RFP to be released? If not, what steps are expected to make a decision on releasing an RFP?
	Answer:	There is no anticipated timeline for an RFP. After the options are evaluated, the Department will further plan for potential project requirements and an

**FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION**

Questions and Answers

		estimated budget. At that time, potential funding and procurement options will be evaluated.
Question(s) received on 2/5/2020 from Snowflake:		
5.	Question:	Section 6.0 Response Submission: 3. One (1) electronic copy of the entire "Original" response, and one (1) redacted electronic response, if applicable, in Adobe (.pdf) on a USB flash drive. Question: Is the agency amenable to receiving an electronic copy via other means such as email, a secure document collaboration tool i.e. Google Drive, or cloud file sharing?
		No, all interested vendors are required to submit their responses in accordance with Section 6.0 RESPONSE SUBMISSION.
Question(s) received on 2/5/2020 from Teradata:		
6.	Question:	1. The RFI has listed out a few areas of data sources such as driver license, motor vehicle, crash, citation, customer service and computer aided dispatch (CAD). When you move to the new environment do you envisage additional data sets to be ingested such as road sensor data or other large data sets? If so, do you have a sense of the potential growth sizes for the additional data to be ingested?
	Answer:	The solution is considered a data sharing and integration platform for analytic purposes. The intent of the Department is to make the solution available to the Department's data partners.. This approach will be directed by the price, options and performance of the solution and will assist with defining the additional structured data sets to be considered for ingestion into the new environment.
Question(s) received on 2/6/2020 from World Wide Technology:		
7.	Question:	1. What is the expected delay between a change in a source system being reflected in the data warehouse? • What is the delay or cadence for those changes being reflected in the APIs?
	Answer:	It varies from nightly to weekly. The Department prefers at least nightly. Please indicate in your response the offerings, how do they affect performance and what are the estimated costs?
8.	Question:	2. Does FLHSMV already have cloud connectivity from the Northwest Regional DC or Southwood Shared Resource DC? • What type of connections are there, and to what clouds?
	Answer:	The Department has ExpressRoute from the State DC. The purpose of this RFI is to determine the requirements for a Data Management and Science Solution that should be indicated in your response to this RFI.
9.	Question:	3. What latency in response time is expected in the APIs (we assume these would be fast-responding (sub-200ms) APIs, but want to ensure

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

		we are not driving to a goal that is not desired or not needed by FLHSMV)?
	Answer:	Please indicate in your RFI response your recommendations for the following: a) Latency vs. performance options and estimated costs. b) Recommended environment that is to initially be loaded nightly. c) Scalable options.
10.	Question:	4. Does FLHSMV already have a well-defined Master Data Model? <ul style="list-style-type: none"> • What percentage of the data between source systems needs to be interrelated and is not already linked? • Approximately how many source tables/collections would be involved in the initial full delivery of this solution?
	Answer:	The Department is modernizing its systems. Refining and/or modernizing the Data Model is part of the modernization effort. The Data Model is built for transaction processing. The Department would like to understand any tools available to transform the transaction processing model into more of a data lake/warehouse type repository.
11.	Question:	5. In Tab D, bidders are instructed to provide general pricing only. However, certain bulleted requirements in this tab request specific pricing (such as number of each product or line item required, cost per each product or line item, and the extended cost by product or line item; and license fees). Please provide additional clarity regarding how general or specific pricing should be for the bulleted requirements in this tab.
	Answer:	The Department is trying to understand the significant impact on pricing that items such as user licenses, Processing Units or additional quantities of a feature will increase the estimated budget.
12.	Question:	6. Is FLHSMV considering a hybrid or multi-cloud solution, or is there any preference (or otherwise) to deploy a single cloud vendor?
	Answer:	The Department expects a wide variety of solutions to be proposed. The Department needs options to demonstrate a seamless managed environment.
13.	Question:	7. With what existing tools or platforms does FLHSMV interact?
	Answer:	The Department interacts with the following: Oracle database; SQL Server database; SQL Server Integration Services; SQL Server Analytic Services, Symantec - Data Loss Prevention; Tableau; Web Services; SFTP; and etc.
14.	Question:	8. What existing staff at FLHSMV are dedicated to data management and data science?
	Answer:	The Department has a small team of professionals at this time.
15.	Question:	9. Has FLHSMV studied the potential use case and impacts of applying data science to its business/mission? <ul style="list-style-type: none"> • If so, can you share those use cases and/or priorities?

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

	Answer:	The Department has not performed an impact assessment of data science to the Department's business. Data analytics has become very successful and is growing in demand and complexity within the Department.
Questions received on 2/6/2020 from Siemens Mobility, Inc.:		
16.	Question:	1. Are the vendors expected to meet every category (solution and service) covered in the RFI?
	Answer:	The RFI is asking what solutions and/or features each vendor offers within the managed service and what will have to be procured separately when the need arises.
17.	Question:	2. Is it acceptable to redact the pricing information in order to protect the confidentiality of such information?
	Answer:	Please see Section 6.0 RESPONSE SUBMISSION of this RFI.
18.	Question:	3. Can FLHSMV provide a high list of its data sources, open and private, which data management and science services could be used against?
	Answer:	The list would include in general, the five (5) big databases with multiple tables: Crash, Citations, Motor Vehicle, Driver License/Identification and Computer Aided Dispatch.
19.	Question:	4. Can FLHSMV specify whether they are willing to consider cloud-based applications?
	Answer:	Please see the answer in Question No.12.
Questions received on 2/6/2020 from Google:		
20.	Question:	Page 2: Is the Department looking to replace the existing MS SQL server Data Warehouse?
	Answer:	Yes, the Department is seeking to move the data warehouse to a managed data management/warehousing solution.
21.	Question:	Page 2: How many transactions are produced from transaction processing, issuance and traffic records applications? How much data is generated per day?
	Answer:	The Department averages 440,000 transactions daily across key business data tables, creating an average of 800 MB of data growth. The transaction mix includes updates and deletions, so the change rate is higher.
22.	Question:	Page 2: Is a development environment envisioned?
	Answer:	A development environment or equivalent would be beneficial but is not a strict requirement. Please provide options and estimated cost for a combined small Dev/Test warehouse environment.
23.	Question:	Page 2: Is data archiving or backup required for the Data Warehouse (6 TB) of the application data (30 TB)?
	Answer:	Yes, a backup or replication across the platform is desired. Please provide options, performance and estimated cost.
24.	Question:	Page 5: Is Tableau currently in place?
	Answer:	Yes, the Department currently has a Tableau server, Tableau Desktop/Public, and creator licenses. .

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

25.	Question:	Page 5: Is Apache Spark currently used? Is there a Hadoop cluster?
	Answer:	No, Apache Spark and Hadoop are not in use in the Department.
26.	Question:	Page 5: What is the size of the data growth over time for existing data warehouses?
	Answer:	The existing data warehouse has been growing 750 MB per day on average.
27.	Question:	Page 5: What is the current lag for data going into the existing system?
	Answer:	The Department prefers at least nightly but it currently varies from nightly to weekly.
28.	Question:	Page 5: The data within the environment require specific certifications to process?
	Answer:	The Department has extensive sensitive data and it would be dependent on what the Department decides to store. The Department would be interested in hearing solutions that accommodate elevated security and audit requirements for Driver's Privacy Protection Act (DPPA), Health Insurance Portability and Accountability Act (HIPPA), and Payment Card Industry (PCI) data as well as tools that might be able to obfuscate data.
29.	Question:	How many staff users are expected to access the data lake?
	Answer:	To clarify, the Department is not looking for a data lake. There will be less than 100 internal users with access to the data warehouse. Data Marts may have public users as well as targeted users from other agencies.
30.	Question:	Is there a requirement for data from the analytics environment to be sent back to the originating applications?
	Answer:	Warehouse data will remain in the reporting environment with no transfer back to the Online Transaction Processing (OLTP) environment.
31.	Question:	Tab C.6: Can the department further define "built-in" Machine Learning & Artificial Intelligence capabilities?
	Answer:	Please indicate in your response what capabilities and typical use cases are currently utilized. The Department is interested in methods to crawl data to detect potential fraud and develop methods to mitigate such issues.
32.	Question:	Tab C.2.n: Is there an existing Enterprise Content Management system in place?
	Answer:	No, currently there is no Enterprise Content Management system in place.
33.	Question:	Tab C.3. c: Does the Department have a preference for Customer Managed Encryption Keys vs Vendor Managed keys?
	Answer:	The Department requires the ability to manage encryption keys used for the proposed solution. Vendors are encouraged to respond with a solution that would allow the Department to securely manage encryption keys.
34.	Question:	Tab C. Are there any ETL/Data Quality tools already in place?
	Answer:	The Department is looking for a complete solution and not looking to leverage any current tools.

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

35.	Question:	The RFI mentions "Much of the Department's data and information are distributed across multiple operational systems, data stores, data types and a data warehouse (Microsoft SQL server)". What type of data is in the operations systems and data stores (not the SQL Server system? What are the "data types" mentioned?
	Answer:	Please see the answers to Questions No. 13, 18, 39, and initial paragraph above. The data to be ingested currently is structured data from databases.
36.	Question:	Will the data in the operational systems and data stores stay in place to support those systems but then need replicated into the new EDW? What latency?
	Answer:	The Department envisions that the solution will be separate from the operational systems and data stores. The operational data stores will not be affected and will continue to be maintained by the Department. Latency between the operational data stores and the new EDW is optional and depends on business need.
37.	Question:	Will unstructured data (such as Waze or GPS data) ever be in scope for the FLHSMV EDW in the future?
	Answer:	In the future, there is the possibility that geospatial data will be used to map information for the public.
Questions received on 2/6/2020 from SAS Institute, Inc.:		
38.	Question:	Are data storage amounts provided in the RFI compressed or uncompressed?
	Answer:	The data storage amounts are uncompressed.
39.	Question:	Can the Department provide additional detail about the source data such as the following: 1) How is the source data currently stored, is it in relational data bases, flat files, something else? a) What is/are the file types of the source data (e.g., Oracle)? 2) Roughly, what is the number of data sources? a) Number of business units providing data b) Number of source tables 3) What is largest table size of the source data?
	Answer:	1) Most of the source data is stored in relational databases. a. Mostly in Oracle and some in SQL Server. This RFI is focused on structured data from the databases. 2) There are more than ten data sources. a. Data is provided from national, state, local and internal customers, users and partners through flat files and web services to the OLTP environment. b. The Department estimates that there are over a thousand tables. The current data warehouse ingests 314 tables. 3) Some of the larger source tables contain between 70- 150 million rows.
40.	Question:	How frequently will new data need to be ingested into the data warehouse? a) Will new data be entered in batch only, or will live

**FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION**

Questions and Answers

		pulls need to be performed? b) What is average file size for data that will need to be ingested into the data warehouse? c) What is the largest file size for data that will need to be ingested into the data warehouse?
	Answer:	The data will be ingested into the data warehouse in a manner that does not impact production transaction systems. Please also see the answers to Questions No. 13, 18, 39 and initial paragraph above.
41.	Question:	Can the Department modify the hard copy submission requirement to that of an electronic response?
	Answer:	No, please see Section 6.0. RESPONSE SUBMISSION of this RFI.
42.	Question:	What is the total number of named users who will need access to the data warehouse?
	Answer:	Please see the answer to Question No. 29.
43.	Question:	What is the total number of concurrent users (people who will be accessing the data warehouse at the same time) who will need access to the data warehouse?
	Answer:	Please see the answer to Question No. 29.
44.	Question:	The Department's RFI document cites a total data size of 30 TB and the need for the warehouse to have duplicated storage capacity of 60 TB. Is the 60 TB needed to provide a 30 TB production environment and another 30 TB for Development and test environments? Or is it needed for something else?
	Answer:	30 Terabytes is the base line for a production warehouse. The Department is seeking options, estimated price and performance for a Disaster Recovery (DR) environment and a test environment (with significantly less storage and data).
Questions received on 2/6/2020 from Microsoft.com:		
45.	Question:	1. What are the data sources in scope for this data warehouse implementation? Can you provide details such as what database technologies are used today.
	Answer:	Please see the answer to Questions No. 13, 18, 39 and initial paragraph above.
46.	Question:	2. Does the Department foresee that this data warehouse solution will store CJI at any given time?
	Answer:	The Department currently does not foresee the data warehouse solution storing CJI data.
47.	Question:	3. Will the Department host the new solution in either the SSRC or NWRDC?
	Answer:	The Department is considering all options for hosting solutions.
48.	Question:	4. Will there be any data migration requirements? (ie Oracle to SQL)

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

	Answer:	Yes, the Department is seeking information on estimated pricing, options and performance to migrate from the Online Transaction Processing (OLTP) environment to the new solution, with updates on a nightly basis.
Questions received on 2/6/2020 from Tallan:		
49.	Question:	1. Does the Department envision a strategy for migrating existing data warehouse and analytics workloads to the future state environment? Is there a preference to maintain these assets in their current form vs re-evaluate and/or redesign from the ground up?
	Answer:	a) Yes, proposed solutions should consider migration of existing systems and integration with the future environment. b) The Department is interested in understanding estimated price, options and performance for evaluation and redesign.
50.	Question:	2. Does the Department maintain a staff that is familiar with the existing ETL process that feed the current data warehouse? How does the department envision this staff's contributions to this modernization effort?
	Answer:	The Department has staff that are familiar with the SSIS process and Department staff will be expected to maintain future ETL processes. Please include estimated pricing, and considerations for options, performance.
51.	Question:	3. Can the Department provide more detail about current platform technologies including: a. Data warehouse platform b. ETL tools c. Reporting tools d. Source management
	Answer:	Please see the answers to Questions No. 13, 18, 39 and initial paragraph above.
52.	Question:	4. Does the department employ any data quality tools/measures for data loaded to the data warehouse?
	Answer:	No, there are no data quality tools or measures in use for data loaded to the warehouse.
53.	Question:	5. Can the Department provide a listing of data types that require special attention and care in handling (e.g. CJIS, PHI, etc)?
	Answer:	Please see the answer to Question No. 28.
54.	Question:	6. Can the Department provide more detail about the "limited capabilities" of the current data warehouse? Are there limitations the Department sees as a priority for modernization effort?
	Answer:	The current warehouse is hosted at the Florida Division of State Technology (DST) and is constrained by Storage, Performance, age, scalability, transformation tools and transparency.
55.	Question:	7. The RFI calls for sizing information using 60 Tb (10x current data warehouse size) as a target number. Can the reasoning for arriving at this sizing be shared?
	Answer:	The department is seeking an initial 30 Terabytes to ingest our structured data sets. The current data warehouse only retains a small portion of production

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

		data and the Department would like to improve the ability to access data for analysis and reporting.
56.	Question:	8. Are Disaster Recovery objectives such as Recovery Time Objective (RTO) and Recovery Point Objectives (RPO) known and can they be shared?
		RTOs and RPOs have not been determined for the data warehouse. It would be preferable if solutions were provided to determine those objectives.
57.	Question:	9. With regard to the current data warehouse, does the Department employ any data marts and if so, can more information be provided on how such data marts are structured in relation to the data warehouse?
	Answer:	No, currently the Department does not utilize data marts. Please indicate the available options in your response.
58.	Question:	Can detail be provided regarding the current and projected user base and usage patterns? This might be in terms of user counts, user types and report consumption metrics, among others.
	Answer:	The Department does not currently have usage metrics available.
Questions received on 2/6/2020 from ISF:		
59.	Question:	1. What is the structure & platform for the current data warehouse?
	Answer:	Please see the answers to Questions No. 13, 18, 39 and initial paragraph above.
60.	Question:	2. What is the timeline the Department is anticipating for implementation or a warehouse solution?
	Answer:	Please see the answer to Question No. 4.
61.	Question:	3. What is the anticipated budget the Department will allocate for this effort?
	Answer:	Please see the answers to Questions No. 2 and 4.
62.	Question:	4. Does the Department have any current staff skilled or trained in data sciences? What are their area's of expertise and preferred toolchains?
	Answer:	The Department is just beginning to build this expertise.
63.	Question:	5. What is the mix of current department systems that would integrate with a warehouse? Languages, platforms, etc..
	Answer:	Please see the answers to Question No. 13, 18, 39 and initial paragraph above.
64.	Question:	6. What is the growth rate of the data intended for warehouse ingestion?
	Answer:	Please see the answer to Question No. 26.
65.	Question:	7. Is the Department intending to pursue cloud-enabled providers, or is there a preference for maintaining data on-premise?
	Answer:	Please see the answer to Question No. 12.
66.	Question:	8. Have any data quality analyses been performed on the existing warehouse data?
	Answer:	Please see the answers to Questions No. 34 and 52.

FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION

Questions and Answers

Questions received on 2/6/2020 from Mirmiran & Thompson (JMT Technology):		
67.	Question:	Tab C, 2a Of the 30-60TB of data, how much is expected to be ingested, available for processing, at any given time and what are the performance requirements?
	Answer:	Please see the answers to Questions No. 13, 18, 39 and initial paragraph above.
68.	Question:	Tab C, 2a Of the 30-60TB of data, how much is expected to be processed at any given time and what are the performance requirements?
	Answer:	Please see the answers to Questions No.13, 18, 39 and initial paragraph above.
69.	Question:	Tab B, 6; Tab C, 12 Does the requestor have an ongoing PaaS partner such as AWS, MS Azure, or Google GCP?
	Answer:	No, the only platform the Department uses are the state data centers: Northwest regional and Southwood data centers.
70.	Question:	Tab C, 2F-S, 6, 7, 8, 22, 24, 25 Does the requestor have experience with Hadoop/Spark SaaS implementations?
	Answer:	No.
71.	Question:	Tab C, 2F-S, 6, 7, 8, 22, 24, 25 Does the requestor have experience with Databricks/Delta Lakes implementations?
	Answer:	No.
72.	Question:	Tab C, 2F-S Is the requestor prepared to manage the life cycle of machine learning models with a product like MLFlow? How large is the data science group?
	Answer:	Please indicate estimated pricing, options and resources needed to manage MLFlow.
73.	Question:	Tab C Has the requestor migrated any of the existing 30TB of data into a 'big data' solution such as Databricks or other Spark/Hadoop dependent solution? What was the result? Is information about that effort available?
	Answer:	No, none of the data has been migrated to a 'big data'; platform.
74.	Question:	4.0 Background What is the current daily data transfer size/throughput? Is there a projected daily throughput size?
	Answer:	Please see the answers to Questions No. 26, 39 and initial paragraph above.
75.	Question:	4.0 Background The RFI list the data sources, but what are the data source types?
	Answer:	Please see the answers to Questions No.13, 18 and 39 and initial paragraph above.
76.	Question:	Tab C, 3 What are the network security standards required for this project?

**FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION**

Questions and Answers

	Answer:	The network should enforce encryption at rest and in-transit. Reference Administrative Rule 60GG-2 "Information Technology Standards".
77.	Question:	Tab C, 3 Can we work with you to define new security standards/processes?
	Answer:	The Department is not interested at this time.
78.	Question:	Tab C, 2.c. Can you provide your existing PII controls in place?
	Answer:	Data is encrypted at rest and in-transit. Role based access controls are in place.
79.	Question:	4.0 Background How quickly is the data growing?
	Answer:	Please see the answers to Questions No. 26 and initial paragraph above.
80.	Question:	Tab C, 2.f. Can the solution utilize Florida's existing Esri cloud platform for geospatial mapping components?
	Answer:	Please provide estimated pricing, performance and options to utilize the existing Florida ESRI cloud platform.
Questions received on 2/6/2020 from Mythics:		
81.	Question:	In section 4.0 BACKGROUND: Which of the 30 TB of transaction applications are contained in the 6 TB small SQL Server data warehouse?
	Answer:	The server indicated in your question includes a small subset of each of the five (5) big databases with multiple tables: Crash, Citations, Motor Vehicle, Driver License/Identification and Computer Aided Dispatch.
82.	Question:	In section 4.0 BACKGROUND: What tool set(s) are used to prepare, ingest, cleanse, extract, and transform the data for the existing small data warehouse?
	Answer:	Please see the answer to Question No. 13.
83.	Question:	In section 4.0 BACKGROUND: Is it the desire of the Department to create an enterprise data warehouse for all 30 TB of multiple transaction processing, issuance and traffic records applications?
	Answer:	Yes, the Department is seeking a solution with the initial capacity to host 30 Terabytes of structured data. Please also see initial paragraph above.
84.	Question:	In section 4.0 BACKGROUND: What tool set(s) are currently in use to analyze exiting data sets?
	Answer:	Please see the answer to Question No. 13.
85.	Question:	In section 4.0 BACKGROUND: Does the Department desire to capture unstructured data in the proposed enterprise data warehouse? If so, what types of unstructured data (reports, sensors, photographs, traffic crash diagrams, customer service related social media, etc.) are desired to be included?
	Answer:	No, currently the Department does not intend to capture unstructured data in the proposed enterprise data warehouse.

**FLHSMV-RFI-001-20
DATA MANAGEMENT AND SCIENCE SOLUTION**

Questions and Answers

86.	Question:	In section 1.0 INTRODUCTION, subsection D: Does the Department currently use any data enrichment, Artificial Intelligence, or Machine Learning tools? If so, which tools are being utilized?
	Answer:	No, currently the Department does not use any data enrichment, Artificial Intelligence, or Machine Learning tools.