ADDENDUM NUMBER 1: QUESTIONS AND ANSWERS

FOR

REQUEST FOR INFORMATION FOR DATA ANALYTICS PLATFORM SERVICES

November 18, 2020

Question #1: How much is the size of the historic data and incremental data?

Answer #1: The historic data is on the order of tens of billions of records, growing at 2-4 billion rows per year.

Question #2: What would be the overall approx. data volume?

Answer #2: 1.5TB to 3TB, growing at 100s of GB per year

Question #3: Please confirm whether SVCE is looking for real-time data processing or batch processing.

Answer #3: SVCE is looking for batch (daily/weekly/monthly basis) processing and near real-time data processing.

Question #4: Please let us know the frequency of ETL Jobs? How frequently will the data be changing?

Answer #4: As frequent as daily (AMI data, weather data, etc.) right now, with possibly higher frequencies for future data flows.

Question #5: Please confirm whether there is a plan for near real-time data analysis.

Answer #5: Yes, for anticipated use cases such as, but not limited to, short-term load forecasting and VPP management

Question #6: Please let us know the Analytics that SVCE is currently doing, and what are the limitations of the current architecture.

Answer #6: The current data analytics platform provides for ad-hoc analysis of demand-side data. It does not, however, provide for building end-to-end data flows. Section 6 of the RFI details the delta between DAISY 1.0 and DAISY 2.0

Question #7: Please provide details about SVCE customer, billing, interval data, where it resides, and how SVCE gets data from the source system.

Answer #7: Data is stored in BigQuery and box; sources include PG&E and Calpine; SVCE receives data via SFTP

Question #8: Are 100.0% of SVCE's customers on AMI? If not, what is the breakdown by percentage of customers and by percentage of usage in kWh?

Answer #8: Over 98% of SVCE customers have smart meters.

Question #9: Would the "public-facing dashboard" as described in Key Applications exist on SVCE's current site (ie, **www.svcleanenergy.org**) or could it be on its own (ie, as a subdomain)?

Answer #9: The public-facing dashboard can be its own subdomain.

Question #10: (From Section 5) What is the source of this data:

- Customer attribute data
- Monthly and interval electricity usage data
- Monthly gas usage data
- Public Safety Power Shutoff (PSPS) data
- Demand response program participation data
- Interconnection data
- Tax assessor parcel data
- Census tract-level data?

Answer #10: Data sources include PG&E, Calpine, CAISO, County Assessor, Census website

Question #11: Can you please describe more about the processes and means of transmission of data from Calpine going forward?

Answer #11: Calpine drops the data into the ingest fileserver for our current platform. From here, our current analytics pipeline processes it daily and loads it into our data warehouse in Google BigQuery.

Question #12: Where it says "dataset cleaning is carried out using Python", is this done by SVCE staff? Is this process desired to be maintained in DAISY 2.0?

Answer #12: Dataset cleaning is currently done by a collaborative effort between the current vendor and our staff, with some jobs taking place within DAISY 1.0 and some taking place outside of DAISY 1.0. We would like to streamline the data cleaning process by integrating the entire process into DAISY 2.0.

Question #13: Where it says, "geocoding and spatial analyses are carried out using ArcGIS Pro", who operates the ArcGIS Pro software? Will SVCE maintain ArcGIS licenses going forward?

Answer #13: SVCE staff operates the ArcGIS Pro software. SVCE will maintain ArcGIS licenses unless a more streamlined solution for geocoding and spatial analysis is implemented.

Question #14: (Section 6: Applications and Tools) Are the "Public-facing dashboards" expected to be owned and hosted by SVCE, or by a 3rd party on SVCE's behalf?

Answer #14: The dashboards can be hosted on the same platform as the DAISY 2.0 data warehouse. We would like to add/edit dashboards on an ongoing basis as described in the RFI.

Question #15: (Section 5) Last Sentence - Desire to integrate & streamline Daisy 1.0 workflows. Please provide additional information regarding the type of integration and/or streamlining that is required?

Answer #15: Automated data flows from sources in SFTP / Box / external sources through the data warehouse, including SQL or equivalent querying as well as Python or equivalent scripting that can all be scheduled to run on a regular basis. Integration of geocoding and spatial analysis functionalities that are currently being carried out outside of DAISY 1.0.

Question #16: (Section 6) Figure 1 Page 9: Could SVCE please provide the number of data pipelines, the data volumes and velocity requirements so we may determine the potential efforts associated with question #15.

Answer #16: Over a dozen data pipelines with the biggest incoming data volumes at approx. 10 million rows or 5 GB per day

Question #17: (Section 6) Figure 1. Is SVCE interested in extending the DAISY 2.0 architecture to support a ML/AI "workbench" function for developing advance ML/AI analytic services to support Tier 3 users or provide new insights to Tier 1 & 2 Users?

Answer #17: Yes

Question #18: (Section 6) Figure 1 Page 9: Is there a need to introduce a data catalog to automate and centralize metadata tagging, basic data governance and access control for Tier 2 and 3 users?

Answer #18: SVCE would be interested in learning more about a data catalog to support these functionalities.

Question #19: (Section 6) Support Page 10: Post DAISY 2.0 implementation, would SVCE only look for Dev/Ops support subscription or a potential subscription "value-add" subscription that would include DevOps, as well as DataOps (data quality management) and ongoing end-user service development support for secondary applications (Section 5 - Page 7)?

Answer #19: SVCE is interested in learning more about additional support services such as data quality management and ongoing end-user service development support for applications.

Question #20: (Section 6) Enabling Tools Page 7: what are the current mechanisms for integrating Python and R with Daisy 1.0?

Answer #20: SVCE currently does not have a way to integrate Python and R without transferring data, so that is part of the RFI.

Question #21: (Section 6) For addressing Core Functionalities, how many sources do we need to integrate? Can you categorize these sources by type (e.g. Web API call, databases, files, etc.) and velocity (Real-time vs batch)

Answer #21: Our current and expected data sources span across web API calls, databases, as well as files. Volume/velocity are addressed in questions 2 and 16.

Question #22: (Section 5) Is Calpine data going to be available in GCS or does it require ingest from another database?

Answer #22: It currently requires transferring through a file server on the cloud.

Question #23: (Section 5) How many jobs/workflows are being used outside of the Cloud environment for data cleansing? Are these all Python based or are you using other ETL tools?

Answer #23: Data cleansing outside of the Cloud environment is currently primarily conducted using Python, on an ad hoc basis.

Question #24: (Section 5) Are you using any Dataproc clusters? If so, how many and for what purpose?

Answer #24: We are not using any Dataproc clusters at this time.

Question #25: (Section 5) How many report and dashboards do you currently support?

Answer #25: We currently support around 7 reports/dashboards and expect this number to grow.

Question #26: (Section 6) How many customers do you currently have participating in VPP programs and are you using specific platforms to support these?

Answer #26: SVCE currently has dozens of customers participating in VPP programs. Participation is forecasted to grow into the hundreds by the end of the calendar year and into the thousands in 2021. We currently rely on third parties for nearly all aspects of VPP tracking and control. We expect DAISY 2.0 to be able to provide at least tracking support for VPP programs.

Question #27: (Section 6) Will the public facing dashboard be consumed via your customer portal and will it require analytics or just reporting / visualization?

Answer #27: The public-facing dashboard will be consumed via customer portal and/or SVCE's website, and will require both analytics and reporting/visualization.

Question #28: (Section 7) For the response submittal Section 5 are you looking for the distinction between the platform cost and the on-going support costs? Given the limited information we will have to provide a confidence range with a set of assumptions assuming that meets your needs at this RFI stage.

Answer #28: It will be useful for us to learn about the ranges for setup as well as ongoing costs, along with corresponding cost assumptions. We are also open to considering various pricing models.

Question #29: (Section 7) Are you open to considering a managed service for the on-going operations and support of the proposed solution / platform and associated data pipelines?

Answer #29: Yes

Question #30: (Section 7) The response submittal deadline in Section 7 indicates December 3rd as the deadline however in Section 4 it indicates December 4th as the deadline. Please clarify.

Answer #30: The response submittal deadline is December 4, 2020 5:00 p.m. Pacific Time.

Question #31: In the Tier 1 user group, are the users SVCE customers who are using the public-facing dashboard application?

Answer #31: The Tier 1 user group was scoped to be SVCE internal staff who can view and explore internal dashboards and reports. However, SVCE is interested in learning about how the proposed solution might be able to support an additional user group consisting of SVCE customers and member agencies who use the public-facing dashboard application. Furthermore, in 2021, SVCE may choose to go out to RFI/RFP for additional data services focused on a variety of externally facing applications.

Question #32: Does the application need to support on-demand creation of new reports and dashboards, or will this consist of a fixed set of pre-produced reports and dashboards?

Answer #32: Yes, the application will need to support on-demand creation of new reports and dashboards.

Question #33: Was DAISY 1.0 built internally by SVCE or with support from another vendor?

Answer #33: DAISY 1.0 was built with support from another vendor.

Question #34: Can you provide a diagram for the current state DAISY 1.0 similar to what is included on page 9 for the future state of DAISY 2.0?

Answer #34: Current state DAISY 1.0 follows a similar structure to the DAISY 2.0 schematic on page 9 of the RFI. The main difference is that DAISY 1.0 currently has no integrated applications and no Tier 2 users.

Question #35: What services in GCP are currently in use as part of DAISY?

Answer #35: GCS, BigQuery, Data Studio, Compute Engine, Cloud SQL

Question #36: Can SVCE provide a current OPEX spend for GCP?

Answer #36: Our current billing model does not delineate the GCP subscription. Please use the data volume, rate of ingest, and our expected use cases to estimate the cloud spend.

Question #37: Where does the majority of SVCE's infrastructure reside outside of DAISY?

Answer #37: We have a separate file store in the Box cloud outside of DAISY. Some of our analyst work is done in Excel.

Question #38: Is SVCE leveraging the services of any other public cloud providers other than GCP? If so, what for?

Answer #38: Box, to store flat files and documents

Question #39: Where does the source data originate?

Answer #39: Please see the response to question #10.

Question #40: Did SVCE configure the schemas in BigQuery manually, or were they developed with autodetect?

Answer #40: Primarily manual schema definition, with some autodetect

Question #41: What applications are the source of the telemetry and weather data SVCE is gathering?

Answer #41: The telemetry data source for generation assets is TBD. We currently pull weather spreadsheets from NOAA, but are open to exploring other weather data sources.

Question #42: Does SVCE have any current pain points with the current environment?

Answer #42: Please see the response to question #6.

Question #43: The RFI mentions that SVCE currently stores data on GCP. Could you tell us what service you use to store data on GCP?

Answer #43: Please see the response to question 35.

Question #44: What kind of database does Calpine use? Is it on-prem or cloud-based?

Answer #44: SVCE receives Calpine data via a file server. If your solution is dependent on details on Calpine database systems, please state the details and assumptions you are using in your response.

Question #45: Are there specific sources of weather data SVCE would like to use? Or would you like the vendor to recommend?

Answer #45: SVCE is open to vendor recommendations for weather data sources.

Question #46: Are there specific sources of CAISO market data SVCE would like to use? Or would you like the vendor to recommend?

Answer #46: SVCE currently sources CAISO market data from CAISO's data system. If there are good reasons to pull data from elsewhere, SVCE is open to considering that.

Question #47: Will SVCE be able to provide the forecasting, procurement, scheduling, and settlement data from third-party services?

Answer #47: We'd like to have automated data flows from our existing third-party data sources as well as any future ones.

Question #48: Are there specific sources of data SVCE would like to use for resource and physical asset data? If so, could SVCE specify what kind of data warehouse is being used?

Answer #48: SVCE renewable energy projects are currently being built. Once those projects achieve commercial operation, we would like to include data related to those resources in the DAISY 2.0 platform (generation, LMP price, solar insolation, etc.)

Question #49: Are there specific sources of air quality data SVCE would like to use? Or would you like the vendor to recommend?

Answer #49: SVCE is open to vendor recommendations for air quality data.

Question #50: What is the volume of data SVCE would expect transmitted and/or stored?

Answer #50: Please see the responses to questions 2 and 16.

Question #51: What tools does SVCE hope to continue using or discontinue using? For example, does SVCE hope to migrate away from using GCP, ArcGIS, etc?

Answer #51: SVCE is open to migrating to other tools/platform if the proposed solution can improve current workflows, data security, and integration ability

Question #52: How many users in each Tier does SVCE expect the application to support?

Answer #52: 2-10 users per tier for now, but this number can grow in the future.

Question #53: SVCE has mentioned that data provenance (I.e. data lineage and data processing tracking) is important. What challenges does SVCE currently face around data lineage and data processing today?

Answer #53: Current challenges around data provenance include the lack of an established framework for tracking data sources, data versions, and data processing steps. This makes analytics workflows less efficient.

Question #54: SVCE mentions data security in the RFI, for the DAISY 2.0 environment what type of data security standards are the vendors expected to adhere to (FedRAMP for example)?

Answer #54: SVCE requests the following documents to support Vendor's cybersecurity posture.

1. SOC 2 – Type 2 (preferred), SOC 2 – Type 1, ISO, or other framework-based audit report.
2. Lists of any third-parties Vendor will use as part of their offering to SVCE and proof of vendor due-diligence done by Vendor on their third-party vendors.
3. Evidence of routine internal and external vulnerability testing for the platform Vendor will use to store/process SVCE's data.
4. Information Security policy or similar internal governing document describing Vendor's cybersecurity posture.
5. Need for Cybersecurity Insurance

Question #55: Is SVCE planning on housing sensitive data on the DAISY 2.0 platform (PII for example)?

Answer #55: Yes

Question #56: Is SVCE planning on having both Internal (data they do NOT want to make public) and External (data they make public) data on the platform?

Answer #56: Yes, we would like to host both internal and public data in DAISY 2.0, with appropriate security and privacy controls in place.

Question #57: Can SVCE expand on what requirements they have for customizable user access permissions/controls – what is optimal?

Answer #57: At minimum, the user access permissions/controls should be customizable enough to support User Tiers 1 through 4 as defined on page 8 of the RFI. Optimally, the user access permissions/controls will also be customizable enough to support additional User Tiers that may be established in the future. These additional User Tiers will differ in their access to various components of DAISY 2.0 such as: data storage buckets, data pipelines, data warehousing, query libraries, data visualization tools, key applications, secondary applications, enabling tools, dashboards, and reports.