



**PENINSULA  
CLEAN ENERGY**

**2020 Data Warehouse  
Request for Proposals**

**Q&A Responses**

*Peninsula Clean Energy, locally-controlled by the San Mateo County community, provides lower prices, clean energy, and community reinvestment.*

**Responses are due *September 4, 2020* at *5pm* Pacific Time**

## Table of Contents

<b>1</b>	<b>Administrative Matters.....</b>	<b>6</b>
1.1	Can we respond with a partial response to the RFP, for example, just the Data Warehouse infrastructure?.....	6
1.2	Is there room for negotiation on your insurance requirements? .....	6
1.3	Can we negotiate your standard contract ahead of time?.....	6
1.4	Keeping the COVID-19 situation in mind, what percentage of the work have to be done on-site?.....	6
1.5	What processes and systems are in place at Peninsula Clean Energy for remote work? 6	6
1.6	Can a part of the work be performed at an off-shore location?.....	6
<b>2</b>	<b>Data Warehouse Platform &amp; Security .....</b>	<b>7</b>
2.1	Is subscribing to a pre-existing warehouse tool with existing architecture, data relationships, and third-party management a consideration? .....	7
2.2	Do you have any existing Data Warehouse system, if yes: what are the various components?.....	7
2.3	Is the current data hosted On-premises or Cloud? .....	7
2.4	Are there any preferences to use on-premise or cloud applications for storage and data processing? .....	7
2.5	Do you have a preference between cloud and On-premises solutions? .....	7
2.6	Will cloud-based (Azure, AWS) solutions be considered?.....	7
2.7	Does PCE have a stance toward cloud platforms? Are there existing relationships with cloud vendors that we should be aware of for this proposal? .....	7
2.8	Are there any infrastructure/security concerns to consider when architecting the solution? i.e. is a VPN solution required if the database is kept on-prem?.....	7
2.9	Can you provide an example of a "data pipe" to a "data source"? By this definition, do you mean piping data from Snowflake to another destination? Or vice versa?.....	7
2.10	Are there any preferences to have only two environments Test & Production, or QA as well? 8	8
2.11	To which specific standard are you referring to with the statement "Secure data storage and interchange at utility-industry standard level"? Some standards (e.g. NERC) are concerned with the electric grid and some standards (e.g. ISO/IEC 27000) are more about data. 8	8
2.12	Would it be correct to say, we need to develop a centralized Enterprise Data Warehouse (EDW) with multiple Data Marts that will serve different use cases and can be leveraged across different upcoming phases ? .....	8
<b>3</b>	<b>Data Scale and Performance.....</b>	<b>9</b>

3.1 Is PCE currently processing any customer interval data collected by PG&E. If so, how many metering points?..... 9

3.2 How much history data is stored and what would be the overall data volume(approx.)? 9

3.3 How many metering points having interval data do you anticipate the system will need to store and for what period of time. .... 9

3.4 What is the size of historical data and incremental data? ..... 9

3.5 Similar to the 5 second latency for response time, do you have a desire for speed of data landing in the warehouse? (e.g. "near real time" of 15 minutes) ..... 9

3.6 How frequently are the meter reading captured (Hourly or Sub-hourly basis), is it required to display real time data or data refresh will happen based on defined schedule?..... 9

3.7 Are you looking for real time data processing or batch processing, in case of batch processing various cloud vendors have pause and resume capability thereby minimizing the overall cost ? ..... 9

3.8 What frequency would you see the data warehouse being updated with customer interval data?..... 9

3.9 What are the preferred and required refresh rates? ..... 10

3.10 What is the frequency of ETL? How frequently will the data be changing? ..... 10

3.11 Based on the Phase 1 and Subsequent phases use cases, this appears to be primarily a research and reporting system. Do you plan for this to have the capability to offer near real-time data analysis or event triggers in the near future? ..... 10

**4 Data Sources, Attributes, and Integration ..... 11**

4.1 Can you please provide the definition of "interval data"? And how it is used ..... 11

4.2 Which are the current and future data sources? ..... 11

4.3 What are the other data systems that PCE currently uses, other than Sales Force-CRM/Power Path that the Data Warehouse needs to connect to? ..... 11

4.4 Can you clarify the source of the metering interval data? Is this provided by the relationship with Calpine? Is there an existing database where this data is collected? What data volumes are expected for this data? (row count, interval frequency, # of meters) ..... 11

4.5 Can you provide example data? ..... 11

4.6 Can PCE please elaborate on the data that is being stored in the Salesforce-based CRM system? Is this customer data, program data, or utility data? What types of data attributes need to go into the Data Warehouse? Are these strings, numerics, addresses, full sentences or all of the above? ..... 11

4.7 Is the PowerPath CRM still compliant with the generalized Salesforce APIs for data integration? ..... 11

4.8 Can you share where these data (interval data, weather, rate class, customer attributes) are coming from? What are the data sources? (saas applications? databases? File storage systems?) ..... 12

4.9 Does PCE already utilize a weather data source? If so, is that source expected to be utilized by this system and can you share what it is? ..... 12

4.10 Are there any existing data transfer mechanisms in place between PCE and PG&E for load profile interval data? ..... 12

4.11 Is the data cleansed or are there issues with data cleansing and balancing? ..... 12

4.12 Can you provide the data structures and statistics for the sources you would ingest? 12

4.13 How is interval meter data retrieved and what is the update frequency of that data? .. 12

4.14 Are there any sensitive data that needs to be included in the ETL pipelines? If so, is masking required for such data? ..... 12

4.15 Will there be a need for some of the more sensitive data to be hidden from particular users of the system? ..... 12

4.16 Can you provide the exhaustive list of data sources that you would like to land into your data warehouse, both for Phase 1 and beyond? ..... 12

**5 Use Cases and Analytical Tools ..... 13**

5.1 Can you provide examples of existing reports or reports you would like to replicate?. 13

5.2 Is the development of dashboards included in the current scope? ..... 13

5.3 On page 15, in the “Load Shapes by Segment” section, please specify the granularity of segment you are targeting. Is Peninsula Clean Energy seeking geographic load shapes or “customer type” load shapes or something else more specific? ..... 13

5.4 For the forecasting and targeting use cases, does PCE anticipate users interacting with existing analytics tools specified (Tableau, Excel, Python) or should vendors propose additional analytic software for these use cases? Are we requested to quote only the data warehouse and integration to support these listed use cases or the user facing application/interface work as well? ..... 13

5.5 What specific client tools are you planning to integrate with Data Warehouse? ..... 13

5.6 Are Tableau and Ascend currently being used by PCE or are they preferred tools to begin being used as a part of this solution? ..... 14

5.7 For Forecasting Use Case, do you plan to integrate any advance analytical tool with the Data Warehouse to achieve this? ..... 14

5.8 In regards to "Analysis Tools" ... ..... 14

a.) Is Peninsula Clean Energy (PCE) only looking for guidance of which analysis tool(s) will be established as "compatible" and "integrated" (initially, in Phase 1) with the data warehouse? ..... 14

..... 14

... or ... ..... 14

b.) Is Peninsula Clean Energy (PCE) looking to partner with a firm beyond tool integration--to also take the lead (while collaborating with SMEs) in building out the entire (Phase 1) BI environment, including all required dashboard visualizations and reporting? ..... 14

**6 Staffing and Support ..... 15**

6.1 How critical is the experience of working with energy data in the evaluation process? 15

6.2 What is the expected participation level of the PCE team in creating algorithms, reports and visualizations? ..... 15

6.3 Does PCE have on staff or plan to have on staff data scientists or analysts who will take over this tool within PCE? What level of ongoing support is expected to be needed? .... 15

6.4 Does PCE have IT or development resources that are currently maintaining Cloud solutions? ..... 15

**7 Budget and Licensing Questions ..... 15**

7.1 Can you please elaborate on the suggested budget of \$100,000? Does this figure include all licensing, software, platform and implementation services? For what period should we quote licensing, software and platform fees? ..... 15

## **1 Administrative Matters**

### **1.1 *Can we respond with a partial response to the RFP, for example, just the Data Warehouse infrastructure?***

PCE is looking for a complete response and will not be making product platform decisions independent of the consulting support. Proposals should also include recommendation of a specific platform or options, including justifications for the proposed direction. Product providers may wish to partner with consulting firms to provide a complete response.

### **1.2 *Is there room for negotiation on your insurance requirements?***

Proposers may propose alternate insurance requirements. However, PCE does not guarantee alternative insurance requirements will be accepted.

### **1.3 *Can we negotiate your standard contract ahead of time?***

Proposers may propose modifications to the standard contract with a proposal. However, PCE does not guarantee that modifications to the standard contract will be accepted.

### **1.4 *Keeping the COVID-19 situation in mind, what percentage of the work have to be done on-site?***

There is no requirement for on-site work.

### **1.5 *What processes and systems are in place at Peninsula Clean Energy for remote work?***

PCE is fully remote at this time.

### **1.6 *Can a part of the work be performed at an off-shore location?***

Yes, offshore work is permitted but key team members must be able to participate in meetings as needed during Pacific Time working hours.

## **2 Data Warehouse Platform & Security**

### **2.1 Is subscribing to a pre-existing warehouse tool with existing architecture, data relationships, and third-party management a consideration?**

Peninsula Clean Energy would welcome proposals which leverage preexisting warehouse platforms and external management. Additional considerations with an existing platform likely would include:

- Alignment of existing platform with target data needs
- Extensibility across all the conceived phases
- Transferability and supportability if it should be necessary to engage alternative development resources, including potentially internal to Peninsula Clean Energy, for some or all continued work
- Long term prospects for the data tool and associated services

### **2.2 Do you have any existing Data Warehouse system, if yes: what are the various components?**

PCE does not have an existing Data Warehouse.

### **2.3 Is the current data hosted On-premises or Cloud?**

PCE does not have a data platform in place currently.

### **2.4 Are there any preferences to use on-premise or cloud applications for storage and data processing?**

The Data Warehouse must be provided through a cloud platform.

### **2.5 Do you have a preference between cloud and On-premises solutions?**

Cloud platform is required.

### **2.6 Will cloud-based (Azure, AWS) solutions be considered?**

Yes, PCE intends to utilize a cloud-based solution

### **2.7 Does PCE have a stance toward cloud platforms? Are there existing relationships with cloud vendors that we should be aware of for this proposal?**

PCE does not have a cloud platform standard. PCE is seeking recommendations on a platform, including a rationale for the recommendation. The PowerPath CRM is being integrated with an Azure based ETL process by our third-party support service for import of the PG&E 4013 and other data. PCE does not directly manage that platform however, use of Azure may (or may not) simplify some operations associated with PowerPath, PG&E or customer attributes.

### **2.8 Are there any infrastructure/security concerns to consider when architecting the solution? i.e. is a VPN solution required if the database is kept on-prem?**

Cloud solution is required.

### **2.9 Can you provide an example of a "data pipe" to a "data source"? By this definition, do you mean piping data from Snowflake to another destination? Or vice versa?**

A “data pipe” simply refers to the technical structure which provides for transferring data from a source to destination store. This is not a reference to a specific product.

**2.10 *Are there any preferences to have only two environments Test & Production, or QA as well?***

The consultant may propose two or three environments based on an assessment of need. The number of environments appropriate may depend on the complexity associated with expected development processes and structures. Regardless of the approach, the architecture must support PCE being able to adequately validate prospective releases in a test or QA environment prior to a release to production. It is also advisable to have rollback capabilities of technical structures and the integrity of the data must be retained or be recoverable.

**2.11 *To which specific standard are you referring to with the statement “Secure data storage and interchange at utility-industry standard level”? Some standards (e.g. NERC) are concerned with the electric grid and some standards (e.g. ISO/IEC 27000) are more about data.***

This reference is to standards associated with data management. PCE does not own and operate grid hardware.

**2.12 *Would it be correct to say, we need to develop a centralized Enterprise Data Warehouse (EDW) with multiple Data Marts that will serve different use cases and can be leveraged across different upcoming phases ?***

The consultant may propose specific architectures and refine them during the requirements process. However, PCE is a small organization. Needs for partitioning of data for different teams is likely to be limited. In addition, the major data set in Phase 1, interval meter data, is used by both the primary teams – Power Resources (energy supply team) and Programs (energy demand and decarbonization programs).

### **3 Data Scale and Performance**

**3.1 *Is PCE currently processing any customer interval data collected by PG&E. If so, how many metering points?***

PCE has a billing data manager, Calpine, which manages all PCE's interval data received from PG&E. PCE has approximately 300,000 metering points.

**3.2 *How much history data is stored and what would be the overall data volume(approx.)?***

PCE has approximately 300,000 accounts. Most generate data on 15-minute intervals. It is anticipated that 2 years of data storage will be the minimum storage requirement. The maximum is likely to be 5 years however, the term of the data is likely to be refined during the requirements process and is dependent on the costs associated with data environment.

**3.3 *How many metering points having interval data do you anticipate the system will need to store and for what period of time.***

All PCE accounts must be supported. We anticipate a minimum of 2 years retention. Maximum retention may depend on costs but is not likely to greater than 5 years.

**3.4 *What is the size of historical data and incremental data?***

PCE has approximately 300,000 accounts. Most operate on 15-minute intervals. It is anticipated that 2 years of data storage will be the minimum storage requirement. The maximum is likely to be 5 years however, the term of the data is likely to be refined during the requirements process and is dependent on the costs associated with data environment.

**3.5 *Similar to the 5 second latency for response time, do you have a desire for speed of data landing in the warehouse? (e.g. "near real time" of 15 minutes)***

PCE is not looking for near real-time availability of interval data. The Data Warehouse is to be used for asynchronous analytic purposes. Data is not available real-time nor near real-time from Calpine, PCE's billing data manager.

**3.6 *How frequently are the meter reading captured (Hourly or Sub-hourly basis), is it required to display real time data or data refresh will happen based on defined schedule?***

Data from PCE's billing data manager Calpine is available daily but on a delayed basis. PCE's analysis process will be asynchronous, not real-time.

**3.7 *Are you looking for real time data processing or batch processing, in case of batch processing various cloud vendors have pause and resume capability thereby minimizing the overall cost ?***

The Data Warehouse is not intended as a real-time system. Analysis work will be asynchronous. Import and processing of data may be scheduled.

**3.8 *What frequency would you see the data warehouse being updated with customer interval data?***

PCE's needs may evolve but update frequency of meter data is anticipated to be 1-4 weeks.

**3.9 What are the preferred and required refresh rates?**

Data acquisition is anticipated to be weekly at most for phase 1.

**3.10 What is the frequency of ETL? How frequently will the data be changing?**

Data acquisition is anticipated to be weekly at most for phase 1.

**3.11 Based on the Phase 1 and Subsequent phases use cases, this appears to be primarily a research and reporting system. Do you plan for this to have the capability to offer near real-time data analysis or event triggers in the near future?**

There are no plans for real-time analysis or triggers.

## 4 Data Sources, Attributes, and Integration

### 4.1 **Can you please provide the definition of "interval data"? And how it is used**

"Interval data" refers to energy meter data gathered in regular time intervals. The analysis of interval data provides for energy use analysis, most commonly through energy load shapes by segments and timeframes. A strong grasp of this core data structure and use is highly recommended.

### 4.2 **Which are the current and future data sources?**

Meter data must be drawn from our billing data manager, Calpine. Customer attribute data from our Salesforce CRM.

### 4.3 **What are the other data systems that PCE currently uses, other than Sales Force-CRM/Power Path that the Data Warehouse needs to connect to?**

The Data Warehouse will need to connect to PCE's billing data manager Calpine for transfer of meter data.

### 4.4 **Can you clarify the source of the metering interval data? Is this provided by the relationship with Calpine? Is there an existing database where this data is collected? What data volumes are expected for this data? (row count, interval frequency, # of meters)**

PCE can draw meter data from Calpine, though in principle in could be drawn from PG&E. However, note that AMI data is far cleaner from Calpine so AMI data should come from Calpine unless there is an exceptional rationale for not doing so and that the proposer can demonstrate substantial preexisting capacities for addressing AMI data issues. This data is not currently stored by PCE. PCE has just under 300,000 meters and most meters operate on hourly intervals.

### 4.5 **Can you provide example data?**

Below is a segment of what the meter data looks like:

Account Number	Service	TYPE	DATE	START TIM	END TIME	USAGE	UNITS
999999	9999999	Electric usage	7/21/2020	0:00	0:14	0.1	kWh
999999	9999999	Electric usage	7/21/2020	0:15	0:29	0.1	kWh

Data from Calpine may not exactly match this example.

### 4.6 **Can PCE please elaborate on the data that is being stored in the Salesforce-based CRM system? Is this customer data, program data, or utility data? What types of data attributes need to go into the Data Warehouse? Are these strings, numerics, addresses, full sentences or all of the above?**

Our Salesforce CRM holds customer contact, attribute, marketing events and asset information as well as incentive business process information. The principal account information is the PG&E "4013" file. All relevant data types are used. It does not include meter data.

### 4.7 **Is the PowerPath CRM still compliant with the generalized Salesforce APIs for data integration?**

PowerPath is currently on a fully-featured Enterprise Edition of Salesforce

**4.8 Can you share where these data (interval data, weather, rate class, customer attributes) are coming from? What are the data sources? (saas applications? databases? File storage systems?)**

The interval data and associated rate classes will be provisioned by Calpine, PCE's billing data manager. Weather data should be provisioned by established weather data providers at appropriate resilience and geography associated with energy analysis and PCE's territory. Most or all customer attributes are expected to be provisioned from the PCE CRM.

**4.9 Does PCE already utilize a weather data source? If so, is that source expected to be utilized by this system and can you share what it is?**

PCE does not have a preferred weather data source. The consultant may propose a source.

**4.10 Are there any existing data transfer mechanisms in place between PCE and PG&E for load profile interval data?**

A data transfer mechanism will need to be established with Calpine. Calpine has done so with other community choice energy agencies.

**4.11 Is the data cleansed or are there issues with data cleansing and balancing?**

It is anticipated that some data cleansing may be required.

**4.12 Can you provide the data structures and statistics for the sources you would ingest?**

The principal large-scale data is energy meter data. See above on scale.

**4.13 How is interval meter data retrieved and what is the update frequency of that data?**

Meter data is collected by the regional utility Pacific Gas & Electric. That data is processed for billing purposes by PCE's billing data manager Calpine. The data is available daily but with a delay of 3-4 days.

**4.14 Are there any sensitive data that needs to be included in the ETL pipelines? If so, is masking required for such data?**

The data is utility data that must adhere to strict security from external access. However, internally, all authorized users may access all data in Phase 1. Subsequent phases may have more partitioned access requirements.

**4.15 Will there be a need for some of the more sensitive data to be hidden from particular users of the system?**

Sensitive data security between system users is not anticipated in the phase 1 but may be needed in subsequent phases.

**4.16 Can you provide the exhaustive list of data sources that you would like to land into your data warehouse, both for Phase 1 and beyond?**

For phase 1, the set of attributes is expected to be narrowly defined and will only include data immediately available. Most or all of the data is anticipated to be sourced from the PCE CRM. Specific elements will be defined in the requirements stage of the project.

There is not an exhaustive list of data sources available for subsequent phases. PCE programs are evolving rapidly. It is expected that Phase 2 scoping will identify specific attribute needs and that most or all of those will be readily accessible.

## **5 Use Cases and Analytical Tools**

### **5.1 *Can you provide examples of existing reports or reports you would like to replicate?***

There are many different types of analysis that are needed for energy data analysis. The most common is presenting energy data into load shapes. Load shape analysis show energy use for given time frames and market segments, and may include showing averages (seasonal, weekday/weekend, etc), peaks and other variable and comparing them against other corresponding segments, typically within one graph or side-by-side with other graphs for different segments. The Data Warehouse must support the queries that a front-end visualization tool would use to present such data.

### **5.2 *Is the development of dashboards included in the current scope?***

Some minimal dashboards may be desirable. That can be determined in scoping but it is not the principal objective of Phase 1. Integration with business intelligence tools is required however.

### **5.3 *On page 15, in the “Load Shapes by Segment” section, please specify the granularity of segment you are targeting. Is Peninsula Clean Energy seeking geographic load shapes or “customer type” load shapes or something else more specific?***

Peninsula Clean Energy is the primary load serving entity for San Mateo County, CA. Required load shapes include our service territory with on-demand segmentation based on the analytic needs for particular programs, this includes but is not limited to: sub-geographies, rate classes, customer types, known assets (PV, EV), segmentation by load shapes, and customer attributes such as income levels.

### **5.4 *For the forecasting and targeting use cases, does PCE anticipate users interacting with existing analytics tools specified (Tableau, Excel, Python) or should vendors propose additional analytic software for these use cases? Are we requested to quote only the data warehouse and integration to support these listed use cases or the user facing application/interface work as well?***

PCE users are anticipated to continue using current tools. However, the consultant may propose alternatives (ex: business intelligence and visualization tool to replace Tableau) and may propose tools that utilize the data warehouse to extend analytic functionality. If the consultant proposes such alternatives or extensions, costs should be identified. And if the costs are for assets PCE would retain ownership of, the costs may be separate from the suggested budget.

### **5.5 *What specific client tools are you planning to integrate with Data Warehouse?***

As noted in the RFP, PCE uses some tools currently: Tableau, Excel and custom Python processes. It is anticipated that specialized analytic and business intelligence tools will likely need to be integrated for various analytic needs. Consultants may recommend specific tools.

**5.6 *Are Tableau and Ascend currently being used by PCE or are they preferred tools to begin being used as a part of this solution?***

Both Tableau and Ascend are currently used to varying degrees. Tableau is an option for Phase 1 analysis. PCE uses Ascend for supply-side risk analysis. If Ascend has tools for forecasting and analysis of AMI data, the consultant may suggest such tools. The consultant may suggest alternatives to Tableau for general business intelligence and visualization.

**5.7 *For Forecasting Use Case, do you plan to integrate any advance analytical tool with the Data Warehouse to achieve this?***

It is expected that advanced analytic tools will be desired by PCE. It is anticipated that there may be an evolving process associated with analytic tools as PCE's needs and capabilities develop over time. Consultants may recommend analytic tools as part of their proposal.

**5.8 *In regards to "Analysis Tools" ...***

***a.) Is Peninsula Clean Energy (PCE) only looking for guidance of which analysis tool(s) will be established as "compatible" and "integrated" (initially, in Phase 1) with the data warehouse?***

***... or ...***

***b.) Is Peninsula Clean Energy (PCE) looking to partner with a firm beyond tool integration--to also take the lead (while collaborating with SMEs) in building out the entire (Phase 1) BI environment, including all required dashboard visualizations and reporting?***

The RFP has been conceived to focus on the data warehouse with guidance on and integration of analytic tools. However, PCE would welcome value add services that might address the full scope of PCE's needs including a business intelligence and visualization layer. Such an offering should be specifically characterized, including associated budget, as it is anticipated that there is not a "one-size fits all" solution to the analytic layer as there are multiple needs across the organization.

## 6 Staffing and Support

### **6.1 *How critical is the experience of working with energy data in the evaluation process?***

Experience with energy data is strongly preferred. At a minimum, a proposer must demonstrate an understanding of the fundamentals – types of data and analytics required – to reduce ramp-up to implement the project.

### **6.2 *What is the expected participation level of the PCE team in creating algorithms, reports and visualizations?***

PCE team participation in developing specific analysis will be high and should be mostly self-sufficient unless there are specialized transformations that may be required to support particular analysis.

### **6.3 *Does PCE have on staff or plan to have on staff data scientists or analysts who will take over this tool within PCE? What level of ongoing support is expected to be needed?***

PCE is hiring a data manager and has analysts on staff. It is desirable for PCE to attain a measure of self-sufficiency for core administration and basic enhancement of the environment. However, it is likely there will be ongoing support needs depending on complexity of the environment and evolving needs.

### **6.4 *Does PCE have IT or development resources that are currently maintaining Cloud solutions?***

No, PCE does not have in-house IT or development. Aside from a data manager, PCE does not anticipate adding such staff.

## 7 Budget and Licensing Questions

### **7.1 *Can you please elaborate on the suggested budget of \$100,000? Does this figure include all licensing, software, platform and implementation services? For what period should we quote licensing, software and platform fees?***

The budget figure may be treated as specific to the development consulting services over a year. Additional ongoing maintenance and potential subsequent phase development should be itemized separately. Fixed costs such as licenses or product purchases may be treated separately where PCE would own those components, such as the cloud platform licenses. These must be identified as well. Where the consultant is retaining ownership of certain tools, those must be within the project budget. Terms for PCE owned asset licenses and fees should be optimized for cost over long-term use.