# Summary

As part of GMS project scope GE has proposed to provide the Network Connectivity Analytic to SCE in Release 2 which is tentatively scheduled to be in production in 2023. Network Connectivity application is expected to leverage increased precision in remote monitoring capabilities (OMS outage events, AMI-Reads, AMI-Events), combined with constraints identified through GIS Network Topological and GIS Geospatial analytics to expose inaccuracies in the Customer Information System (CIS) of record and the GIS Network Topological model (specifically customer-to-transformer association in CIS, and transformer-to-feeder-section and transformer-to-phase associations in GIS).

SCE is requesting GE if the Network Connectivity Analytic can be delivered sooner as this will help them correct the errors mentioned above, in anticipation of an ADMS roll-out and other distribution automation and grid planning priorities. To accommodate the request, GE is proposing to use a pre-production version of the software to integrate the required source data, and to run the analytics to obtain the mismatch predictions and recommended changes to CIS and GIS for the entire SCE distribution system.

The key features of the project will include:

1. GE will provide information to SCE detailing the required and optional data inputs to the analytic and their exact data load formats. SCE will need to first extract the required data from Hadoop (or other sources) and transform it into the reference data load files, adhering to the file format, headers, data types and data formats.
2. GE will deploy an on-prem data-fabric instance with a limited set of functionalities that will support ETL data integration and execution of the Network Connectivity analytic.
3. GE will configure and execute the analytic against the initial ETL prepared by SCE (see Data Input Summary), after which time SCE will be provided with documentation on how to re-run the analytic should they choose to correct the source CIS and GIS and re-execute against the same historic data time-series data.
4. The outputs of the Network Connectivity Analytic will be stored in the data-fabric’s S3 compatible file storage system. It is these files, the results, which will enable SCE to discover data mismatches and consider making the recommended changes in their CIS and GIS systems of record.

# Data Inputs Summary

GE will provide the data requirements from each source system, along with detailed descriptions of each column’s CIM-based semantic, and data type and format requirements to SCE. SCE will be required to perform the ETL of their data from Hadoop and/or other source systems into the data loading files that GE provides.

After the initial execution performed by GE, SCE will be able to take the results and make changes back to their CIS and GIS systems of record. After correcting changes (manually), SCE will have the option of re-executing their CIS and GIS ETL processes, and re-ingesting their corrected CIS and GIS data, and re-running the analytic against the same time-series data to check for improvements in the number of errors and convergence towards an acceptable mismatch rate.

In anticipation of the full detailed data requirements, below is a high-level summary of the data which is supported by the Network Connectivity Analytic:

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Brief Description** | **Characteristics** |
| GIS-Network (CIM based including phase details) | Substation to transformer network topology and location data. Also, mounting positions of XFMRs, and NO/NC as-built state of switchgear. | Most recent as-built snapshot of entire distribution network  (required) |
| EAM | Smart meter, XFMR, switchgear, conductors’ descriptive attributes. | Most recent snapshot  (required) |
| OMS | Event start/end time, operating device, XFMRs impacted by outage. | 1 year Historical load  (required) |
| CIS | Meter to XFMR key, usage point and service location ids and coordinates, account status history, meter type/form. | Most recent snapshot.  (required) |
| AMI-EndDeviceEvent | Outage and Power on Events | 1 year Historical load.  (required) |
| AMI-MeterRead | Daily interval Consumption (MWh) | 1 year Historical load.  (optional) |
| AMI-MeterRead | Daily interval Voltage (V) | 1 year Historical load.  (optional) |
| GIS-Landbase | Road, water, other boundaries | Most recent snapshot  (optional) |

# Description of Analytic Modules

The Network Connectivity analytic performs a series of separate algorithms against the data collection. Each algorithm provides its own output when it detects evidence of a mismatch in the Customer or GIS data (mismatch records). A subsequent phase of analysis collects the set of all mismatch meta data from these separate algorithms, de-duplicates and performs hypothesis testing to discover recommended changes (recommendation records) for the meter associations and transformer connections which better fit the data.

Below is a brief description of some of the algorithms which operate against the data described above:

## Distance based Analysis

Distance based analysis detects meters which are too far away from their reported service transformer and flags them for mismatch analysis.

## Transformer Outage Algorithm

The transformer outage algorithm evaluates each historic OMS outage event to determine if the impacted transformers can be found to have a high proportion of consistent AMI-Outage AMI-Power-on events and AMI-Consumption data. If not, the transformer is flagged for mismatch analysis. If a small proportion of meters disagree, then they are flagged for mismatch analysis.

## Landbase Analysis

Land base analysis detects smart meter locations, and distribution asset locations which are reported to be contained in invalid regions, and flags for mismatch. It also will detect if secondary service drops would likely intersect with invalid regions like highways, ditches, waterways or other boundaries.

## Recommendation Algorithms

The recommendation algorithms consider alternative connections for smart meters to transformers where they better match the other meter population in terms of available data. For transformer mismatches a connectivity model service is used to discover potential locally based alternative phase(s) for the transformer, as well as different connection locations including on different feeders nearby.

# Project Scope

For this project **GE will provide the following deliverables:**

1. W1: Complete specification of the data requirements to run the Network Connectivity analytic in its non-continuous mode (see Data Input Summary for high level data description).
2. W1: Sample columnar data file(s) for each data type (GIS, EAM, OMS etc.) in the supported ingestion format (file type, column headers, column data type, data format)
3. On-premises installation of a data fabric historian on SCE servers.
4. Installation of the Network Connectivity analytic solution (without that Application’s User Interface).
5. Ingestion of SCE’s data once GE receives it in the supported ingestion format (as in (2) above).
6. Configuration of the Network Connectivity analytic’s settings to optimize results based on SCE’s distribution standards, geography and data.
7. Analytic results, specifically: mismatches and recommended changes to Customer-to-Transformer associations, and to Transformer-to-feeder/phase connections for the portion of the network provided to GE by SCE in the initial ETL (up to the entire network)
8. User instructions on how to perform subsequent CIS and GIS data-load jobs of the corrected data, so that SCE can re-run the analytics against the same historic data sets.

For this project **SCE will provide the following deliverables**:

1. Complete set of the required and any elected optional ETL data files, populated, and complying with all aspects of file format, column headers, data types and foreign-key relationships across data files.
2. Each individual populated ETL data file will be provided to a secure GE Box Folder at the schedule shown below:

* W4 – GIS-Network and EAM
* W7 – CIS
* W10 – OMS
* W13 – AMI-Events
* W16 – AMI-Voltage
* W19 – AMI-Consumption
* W22 – GIS-Landbase

# H/W Requirements

SCE to provide the below needed hardware for this pilot:

|  |  |
| --- | --- |
| VM’s | 12 |
| Total vCPU | 136 |
| Total Memory (GB) | 1024 |
| Total Storage (TB) | 77 |

# Timelines

Below is proposed timeline. Actual schedule of milestones will be provided at the project kickoff.

|  |  |  |
| --- | --- | --- |
| **Id** | **Tasks** | **Relative Calendar Week** |
| 1 | Project Kickoff | W1-W2 |
| 2 | Data Requirements Workout | W3-W4 |
| 3 | Project Plans, Design Documents | W5-W8 |
| 4 | Development/Decoupling NC from Grid Director | W9-W20 |
| 5 | Data Review/Accept Workout | W21-W22 |
| 6 | Deployment at SCE | W23-W26 |
| 7 | Analytic Configuration | W26-W28 |
| 8 | Run Analytics and optimize | W29-W36 |
| 9 | Handover Analytic results and user instructions to SCE | W37-W38 |

# Assumptions

* GE Delivery Engineers will have access to the SCE IT infrastructure such that the data fabric and analytic code can be deployed.
* GE Delivery Engineers will be granted access to the deployed system, and the ETL data files, so that they can be ingested into the data fabric, and so that the analytic can be configured and executed.
* For this pilot, data-fabric/historian functionalities will be limited to support Network Connectivity analytics’ analysis and reporting functions and is not intended and/or will not be configured to support any other applications, data ingestion, analytic development, integrations, use-cases or APIs.
* GE is currently developing the new user interface of the Network Connectivity Analytic, and so no UI will be used for this project. The new user interface will however be available in the subsequent production release of the Network Connectivity analytic, but for this project the analytic will be configurable and executable in a headless mode a via command line interface.
* Data inputs will be loaded only as batch files for this project, in lieu of the future periodic or real-time data ingested through the OSB interface as part of the broader GMS project.
* This version of analytic is pre-production release. While GE will provide all the necessary support and critical fixes necessary for the intended usage of the application, GE is not committing to support this deployment for longer term.
* GE will perform configuration of the algorithm settings to adjust for SCE service territory and data sets, but GE will not re-design the algorithmic approaches themselves and provides no guarantees to the quality of analysis results.
* SCE will be getting the production release as part of the GMS scope at the agreed upon schedule. GE provides no commitment to port the ingested data from the pilot application to the production release version.
* SCE will comply with the ETL data specifications and, including the creation or extraction of those columns which are shown to be common identifiers across data-sets (e.g. foreign-keys, or MRID fields).
* SCE’s data sources may not have uniform coverage across the entire distribution system or continuous over the historic time-frame used. Therefore, the Network Connectivity analytic’s results are expected to be non-uniform in terms of confidence and recommendations consequently.
* GE will not provide the source code of the Data Fabric or the Network Connectivity analytic, unlike the future GMS project and associated source code.
* GE and SCE will collaborate to identify the most accurate source of location data related to their customer service locations (often redundant sources with variable accuracy).