Algorithms.io

B Box Algorithm Platform

Phase I Proposal

9 September 2013

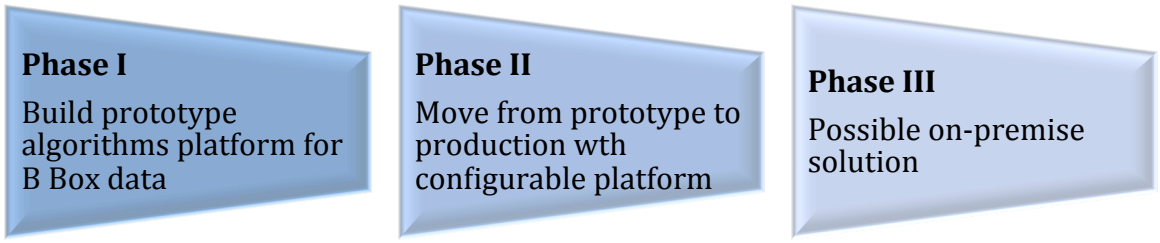
800 W El Camino Real, Suite 180
Mountain View, CA
Phone: (650) 224-6023
E-Mail: andy@algorithms.io
Web: www.algorithms.io

Project Overview

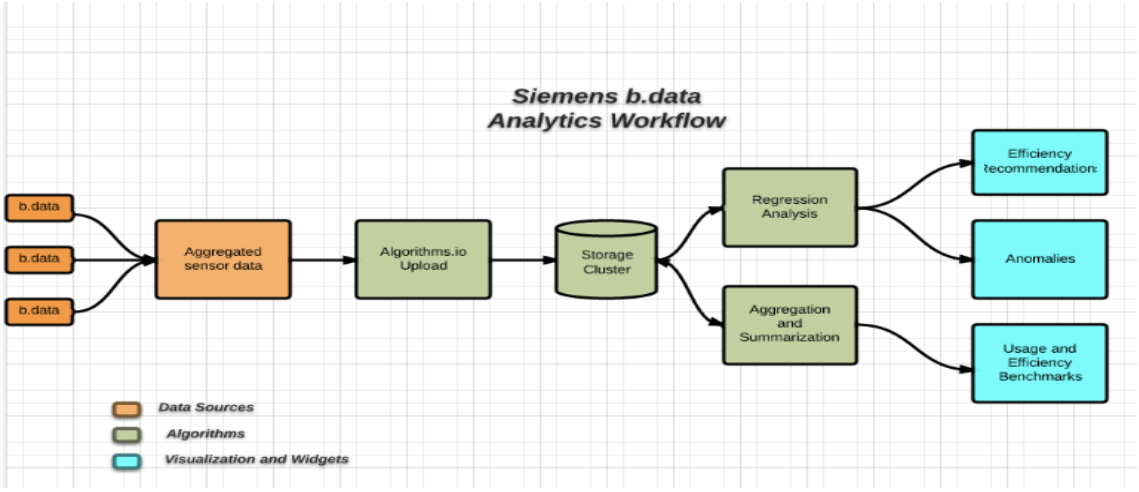
Algorithms.io will be developing a cloud algorithm platform for the ingestion and analysis of data from Siemen's B Box data collection product. The core of this platform will be a configurable algorithm engine that can be tailored to specific customers needs. This catalog will provide various algorithms including regression, anomaly detection, and machine learning for providing intelligent analytics. This algorithm catalog will allow Siemens sales and support to tailor B Box analytics for different customers needs. Data from the B. Box system is broken down into three tiers:

1. Tier 1 – Energy usage aggregate for a manufacturing plant
2. Tier 2 – Energy usage for units inside the plant
3. Tier 3 – Energy usage for substations inside a unit

This project will be completed in three phases:



During Phase I, Siemens will provide data to Algorithms.io as .txt files. As the project moves on from Phase I, the workflow for data passed from the B. Box platform to Algorithms.io will follow the outline below:



Phase I Objectives

1. **Answer Questions Specified By Siemens**
   The Phase I system will follow the deliverable outline included in Appendix 1. Questions highlighted in green will be answered in Phase I. Those highlighted in red will not be answered in Phase I. Preliminary suggestions have been provided on what additional information is needed to answer. A more detailed report will be included at the end of Phase I.

2. **Build Algorithms Engine for Analyzing Plant Manufacturing Data**
   Algorithms.io will architect a catalog of algorithms that can be used for the analysis of B Box manufacturing data. This catalog will be architected to enable addition of new algorithms, and customization for the specific algorithms that are provided to customers. The initial catalog will include regression, classification, and anomaly detection. Anomaly detection analysis will be included for Tier 2 data.

3. **Build Dashboards to Visualize Results of Analytics**
   Complete dashboards to visualize plant-level data (Tier 1) for all energy resources and unit-level data (Tier 2) for only power consumption. Mock-ups of these dashboards are included as Appendix 2 and 3. Dashboards will show the analytics defined in Appendix 1.

Phase I Timeline and Cost

	Deliverables	Timeline	Budget
Phase I Prototype Development	1. Create process for ingesting Tier 1 and Tier 2 data for multiple plant formats. 2. Develop algorithms to deliver the analytics outlined in Appendix 1 3. Build prototype visualization dashboard according to Appendices 2 and 3	**Start Date**: Upon approval of project proposal. **Time to complete**: 25 business days	**Project Cost**: $12,000 **Payment:** 50% due upon project initiation. Balance due upon project completion. Net 15.

Appendix

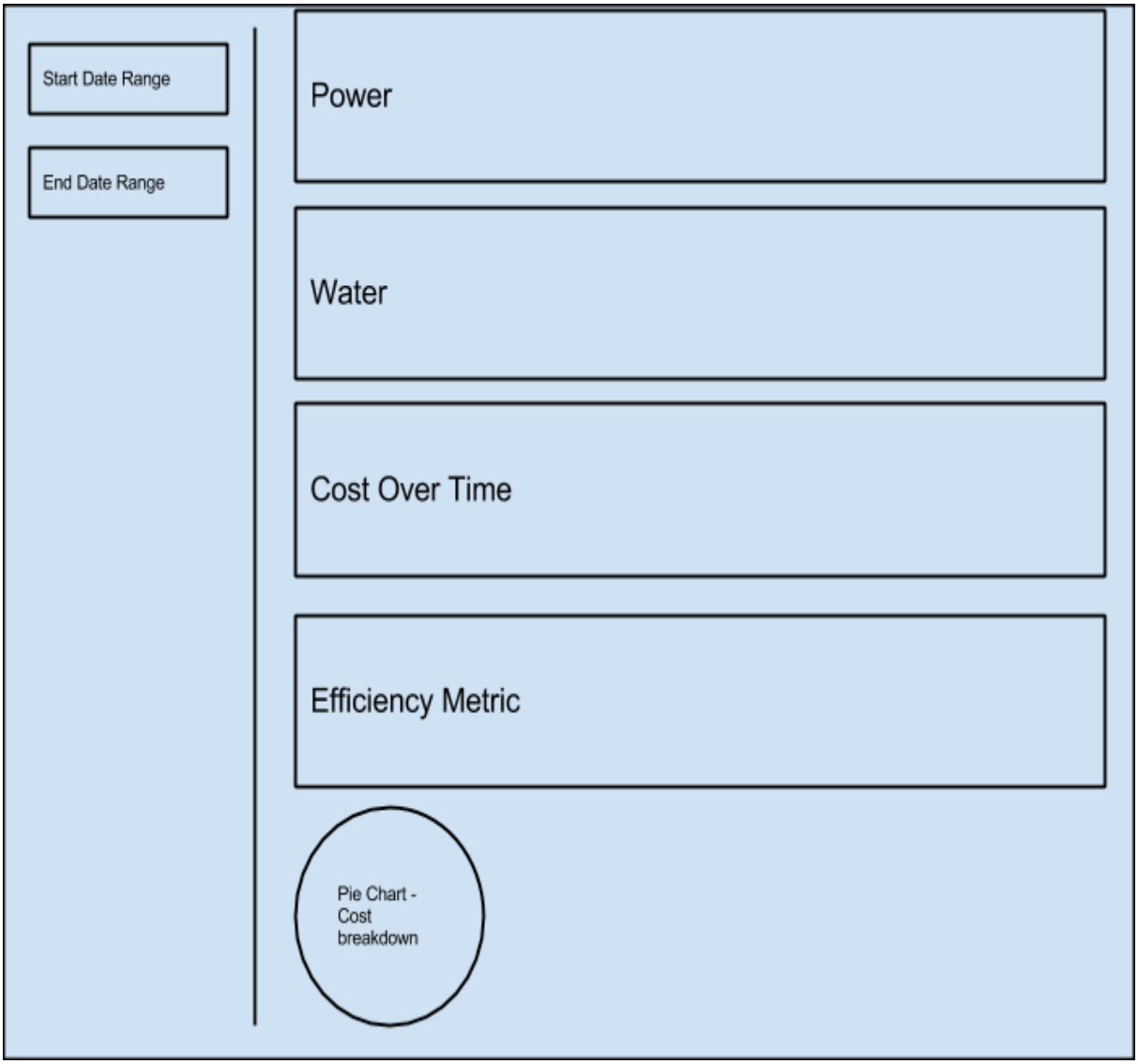
Appendix 1 – Phase I Deliverables Descriptions and Gap Analysis

Questions To Answer in Phase I	***Algorithms.io Phase I Description***
How can I identify weekends looking at my data?	Visualization to be provided for Tier 1 and Tier 2 data will show energy usage over time. Periods of low energy usage including weekends, holidays, etc, will be visible.
When is the best period for production? (electricity price/tariff needed)	Tier 1 data dashboard will include an "energy efficiency" metric that will be calculated by dividing total cost of energy over a specified time period by the amount of product in tonnage produced over the same period. The final value will show $ per ton of energy consumed.
How much energy is used at which cost center (not for cost centers, but for “areas”)	A breakdown of cost usage for power at the Tier 2 level will be included on the Tier 2 data dashboard.
Which of my cost centers have increased costs of energy relative to previous years and by how much? (not for cost centers, but for “areas”)	Year over year growth in power consumption/cost will be included on the Tier 2 data dashboard.
What are the primary consumers? (granularity low, especially for oil and gas)	Power consumption by manufacturing area in Tier 2 will be shown in Tier 2 data dashboard.
What are base load and load of production and what is their relationship between each other? (Production is by month, shift schedule needed)	Time series analysis of energy usage for Tier 1 and Tier 2 data will be provided for visualization and comparison of energy consumption cycles.
How can I draw the operator’s / management’s attention to an inefficient operation of the system? (suggestions and then customer feedback)	Phase I will include anomaly detection for Tier 2 data. Algorithms.io will include an options analysis outlining the various ways alerts created by our anomaly detection algorithm can be delivered to operators and managers.
Which alarms can I generate automatically from the data? (suggestions)	Included with the analysis mentioned above
Is my energy consumption correlated with weather / climate data? (Does the weather data correlate)	Phase I will include time series analysis for temperature in-line with energy usage over the same time. Correlation coefficients can be established as needed.

Siemens Phase I Questions That Require More Data To Answer	Algorithms.io Data Suggestions
When is the best period for maintenance? (electricity price/tariff needed)	There are many other variables to consider, like employee shifts, etc. Most likely, the data from sensors will be insufficient to make a determination.
What is the CO2 footprint/energy use of a product/part? (depending on timeline, oil on monthly basis, external CO2 data needed)	More information is needed on the calculations used to determine CO2 footprint. How will plant managers/operators need to report on this metric? Does this have regulatory implications for EU environmental reporting?
Are my primary consumers my energy hogs? (combine with external benchmark data)	Need historical benchmark data which can be established from a combination of customer data and manufacturer specifications. This industry benchmarking could be a competitive advantage for Siemens over time. Data for specific machine energy usage over time would be required. Data would need to be sampled multiple times/day and shift schedules would need to be clearly defined.
How can my energy consumption be split up (e.g. discipline, plant)? (by area)	Improved data granularity (machine-level) would enable asset level analysis of data consumption. This data combined with production scheduling and operator schedules could be used to determine the energy usage of specific products, energy variablility by operators, and the impact of high/low seasonality in manufacturing schedules on energy usage. This data could also be used to determine possible energy futures hedges and ROI analysis to power alternative energy projects such as solar panel installation.
Up to what level of detail do I need energy information (device / system / hall)? (suggestions and then customer feedback)	The current power sampling frequency of 15 minutes gives adequate information for reporting and predictions.
Which other data might have a correlation with energy data? (suggestions)	Phase I will include some high level information about energy usage and temperature trends. This analysis would be made far more powerful if hourly temperature data were collected for inside and outside the plant along with internal and external humidity at the same frequency. Hourly data would enable matching climate data with specific shift schedules to get improved visibility into possible shift energy optimization.

Appendix 2 - Tier 1 Dashboard

The user will goto a URL in their browser and be presented with the dashboard below showing data with a default time range of 1 year. This view shows data for “tier 1”. The user can proceed to select their own time ranges with the controls on the left. From there the user can click on one of the charts to drill down into a “tier 2” view. In the initial version to limit the scope only the “power” data will have the drill down capabilities.



Appendix 3 - Tier 2 Dashboard

This view shows power metrics. It has the same time range selection capabilities with some additional features. The “Period” dropdown menu allows the user to summarize the data into predefine buckets of: weekly, monthly, quarterly, yearly. There is also an Anomaly checkbox that will run an anomaly detection algorithm on the data and point out anomalies in the dataset.