Data Analysis to Optimize UPS Battery Performance and Management

Gregory W. Ratcliff
Director, Lifecycle Management
Emerson Network Power, Liebert Services
Westerville, Ohio 43082

Introduction

Failure of a data center's uninterruptible power supply (UPS) system can mean substantial losses for most businesses, and batteries are consistently a leading root cause of those failures. The key to avoiding these losses is being able to accurately identify and service battery strings that are at risk—insights that can be derived from performance data.

Utilizing hundreds of millions of data points gathered over 12 years, a team from Emerson's Liebert Services recently completed a battery analysis project in support of the company's data-driven approach to service. This project compared the operational performance of customers' individual units to the historic performance of the same model in similar environments and applications across the company's portfolio. Through data acquisition and an advanced big-data analysis process, the team accurately forecasted remaining service life and determined nine key factors that affect a battery string's service life, from a risk based approach, based upon performance seen elsewhere within the portfolio.

Portfolio Description

Our group has been preserving on-site performance details on many millions of stationary batteries installed in UPS units across North America [(1)]. As of February 2016, the portfolio consisted of more than 100,000 strings or 3.2 million active units. A subset of this portfolio primarily included 12-volt Valve-Regulated Lead-Acid (VRLA) units (jars) in enclosures, typically with doors and ventilation louvers. For this project, the team excluded all battery types which were not VRLA, but did include units which contained between one and eight cells. Our customers are particularly sensitive to risk and begin consideration of replacing strings in this size and configuration within 3-5 years. Extending this service life without increasing risk is desired. Taken over the service lifetime of battery replacements, field-derived analysis of measured performance metrics from the VRLA units was the analyzed subset of this data. All of the data referenced and utilized in this paper are from battery systems that were in operation or removed from service after February 2013. Allowing multiple years to lapse was essential to having enough recent, in-operation data to use for comparison against the machine-analyzed portfolio.

Concerning the architecture further, approximately 35% of our systems have multiple strings of batteries providing redundant DC power to those systems. Typical for UPS service and during a discharge event, battery strings with failing units are pushed into deep discharge and observable failures common for small VRLA units including vent failure, dryout, case bulge, post corrosion, leaking posts, etc. These hard failures align with high or infinite internal resistance tests during the periodic inspections described below.

Service Description

In our most common UPS configuration, 30-40 battery units, each comprised of six cells are placed in series to operate the UPS during AC mains power failures. For all that follows, we define this as a "battery string." As described above, many field configurations have multiple strings in parallel to increase capacity or redundancy, either in absolute available maximum current or to increase the runtime duration of the UPS when operating solely from battery energy.

Unlike most utility and telecom installations, there are virtually no opportunities within our customer base to measure capacity and performance through discharge testing. Often, taking a battery string "offline" for known issues is a challenge to plan and organize. Many of the strategies that are sufficient for larger stationary systems for utility, telecom, and large data centers cannot be applied when the customers perceive the battery string as a short-term consumable item with a demonstrated service life of 3-5 years. Therefore, we have chosen to utilize internal ohmic resistance as a proxy for general unit health and to communicate risk for those units that have deteriorated. This paper does not attempt to correlate internal ohmic resistance to a real capacity, but to communicate how we have observed these resistances change over time, in a unique curve for each unit that had been exposed to similar environmental effects. We define the end of service life as either when an individual unit has reached 150% of it initially observed internal ohmic resistance baseline value, or if the entire string has reached a time limit our internal experts have observed from the field portfolio.

Our goal was to determine if it was possible to forecast within a reasonable accuracy when this end of service life by internal Ohmic resistance would be reached without consideration of age. Any improvement over the historic "battery jar 6 has now failed" notification (based upon the at-risk policies described above) will be seen as an improvement and savings by data center operators with UPS systems .

The typical opportunities available for us to test and inspect small capacity batteries that fit the portfolio description above are relatively rarely, at approximately 12-15 times over their lifetime. Based on service records, more than half of our UPS operators believe quarterly inspections support their level of risk tolerance. The various methods employed today to test and measure units within strings include:

- Capacity test (2)
- Induced ripple voltage (3)
- DC float current (4)
- AC impedance (5)
- AC conductance (5, 6)
- DC conductance (5, 6)
- DC resistance (8)

The project team does not defend the validity of these methods. Within our existing customer base, all of the methods above have been or are employed. For performance comparison and this project we only reported and analyzed the following parameters from battery strings (even if additional data is known):

- Environmental temperature
- Negative post temperature of each unit
- Voltage across the battery string at full charge, in a float condition
- Voltage across each unit at full charge, in a float condition
- Portable instrument-based resistance measurement
- Baseline resistance at start of service
- Manufacturer date code
- Start of service date

- UPS model
- UPS manufacturer
- UPS load
- Fixture type (cabinet, open rack, internal)
- Cells per unit
- Measurement points on unit
- Inter-unit measurement points
- Total cells

The data utilized represents a large field portfolio that have been inspected with portable instruments. The team preserved and utilized data since 2004 as “training data” in our search for patterns. Generally, older data was given less weight and influence. Figure 1 details the relationship between the measurements taken during one of the inspections described above a “job”; the units (jars), and groups of units (strings) that exist at a customer location. Each physical visit is an opportunity, at that point in time, to correlate those measurements with previous ones that exist in the portfolio. The relationship between a site and a string are important from an analysis point of view because within the same site, the same number and duration of power failures and discharge events would likely occur. This represents to machine learning systems a historical pattern and similarity of the families of units and models at the same site. Such similarity carries more significance and is weighted by the system during analysis. As will be detailed below, these groupings fall together during big data analysis. The overall approach is based upon the premise that every failure is likely an established or emerging pattern.

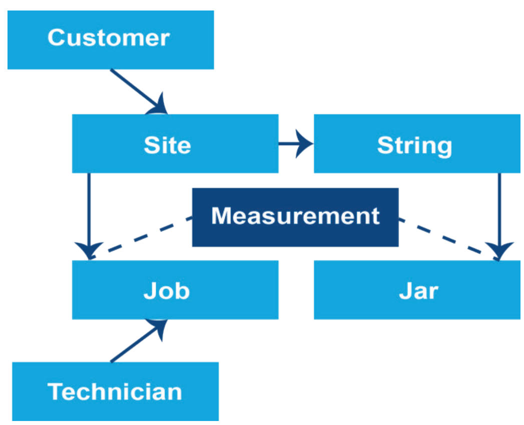

Figure 1. Data Collection Diagram

Health and Policies

As part of our best practices, UPS operators whose units are significantly above the manufacturer’s recommended temperature (typically 77 degrees Fahrenheit) are advised during each inspection that the temperature is above limits, and technicians implement remedies to bring every unit with this temperature into compliance.

Also, as part of the inspection and maintenance process for the UPS systems, policies are in place to adjust the float charge value and ensure the float voltage of battery strings is set at the manufacturer’s recommended values for the cabinet temperature. Once again, when outside these values, technicians work closely with the operator to bring all strings back to within required parameters.

The team measures and establishes an initial Ohmic resistance within the first six months of service. This measurement is preserved as the "baseline" for all subsequent inspections and measurements [1]. The team opportunistically accepts and lowers the baseline for any subsequent visits within the first six months in service for each unit, and then compares subsequent measurements against that unit's original baseline of record. According to policy, we place units into "monitor" and "critical" status at 140 and 150 percent of the baseline, respectively.

For many years, the team has felt it important to look forward—essentially to forecast the internal Ohmic resistance of each unit[8]. The team believes this achievement would allow for the highest availability and best unit performance with rare opportunities for inspection or maintenance. Therefore, the team sought to accurately forecast the **next resistance** result returned from field instrumentation and staff.

Big Data Systems

We provide a short primer on big data systems and machine learning as it will provide a basis to understand the results presented. The concepts presented below are fundamental in understanding the methods, scale and tools required.

In recent years, there has been growth of consumer-based big data analysis systems; from eBay shopping, to Facebook friends, to the well-known "recommenders" from entertainment sources like Pandora and Netflix [11]. Big data is a broad term for data sets so fast changing, large, or complex that traditional data processing applications are inadequate to return results that can be utilized. In the case of a music or movie recommender, that time may be seconds. The challenges with complex data sets include:

- Reliable and continuous data capture and flows
- Data curation
- Data analysis
- Storage
- Sharing

These issues closely aligned with those experienced by our team when developing its practices based on data-driven analysis, especially in regards to the small capacity battery portfolio we are analyzing.

The project team established a well-defined and achievable set of goals for three phases. The first phase of the project was to build a system of hardware and software robust enough to support the complex analysis that would be performed by analysts, battery experts, and data scientists. The second phase was to begin using the built system to explore the data. The team chose to build a classic Hadoop cluster (similar to what is used by eBay) to consolidate and prepare the measurement data, ingest the data to the system, and provide a discovery system for exploration. Phase three was the implementation at a commercial level, not discussed in this paper.

For us success of phase one meant that a few questions needed to be answered about the process, technology, and feasibility of putting a big data system to use. The team focused on learning to utilize the systems with existing battery data. The following success criteria were agreed upon throughout the team and among the executive sponsors:

- Are we able to adapt the tools and systems to our VRLA data set?
- From that data set, is the team able to query the data and provide **faster**, **greater**, and more **useful levels** of information than currently generated?
- Does the team believe the new insights will lead to better understanding of its **customers**, **market**, and **service offering**?

- Can these tools be used on other data sets? Is it flexible and easy to adapt?
- Is the data transformation process reusable and repeatable?

Phase One: The Data

While the historic battery data stored is excellent, the needs of a research analysis system are very different than a classic relational database used for staff and customer reporting. Previously, the data was built on a record per inspection with many fixed fields for the inspection data. Additionally, the data was stored as a continuous time series, much like telemetry historian systems (12) in which any examination of data based upon time in service would often take many hours to return results.

Reorganizing the data to support a name-value pair scheme would support a massively parallel engine to rapidly explore correlations. For example, regardless of when, in calendar time, the readings were taken, an important criteria is the unit age in days from the first day in service, not necessarily the order, job number, and date the work ticket was closed. The team chose to “age adjust” its readings before ingestion, creating a battery key that is unique for each unit, from day one in service until it was removed. This allowed for rapid queries of time in service concerning the entire portfolio with very high performance.

The System

The software and hardware system chosen for analyzing battery data included a number of very high performance computers hosted within a cloud environment and involved multiple computing technologies. Accessing the work cited will provide more information about this type of environment (13).

We constructed and deployed a system similar to that depicted in Figure 2. All systems had very high performance fiber networking between each node, from both a storage and compute resource perspective. From the software standpoint within the Hadoop system, the project team utilized Apache Sqoop™, a tool to return structured data. It also utilized PIG™, Hive™ and extensive hand-coded SQL to parse and prepare data. The team then transferred the data to each data node utilizing Hadoop Distributed File System (HDFS) for data storage and Apache Impala for interactive data analysis. At nearly every step, we used visualization tools so that subject matter experts could “see” the data as it unfolded. These systems were essential to steering discovery and exploration of the data sets.

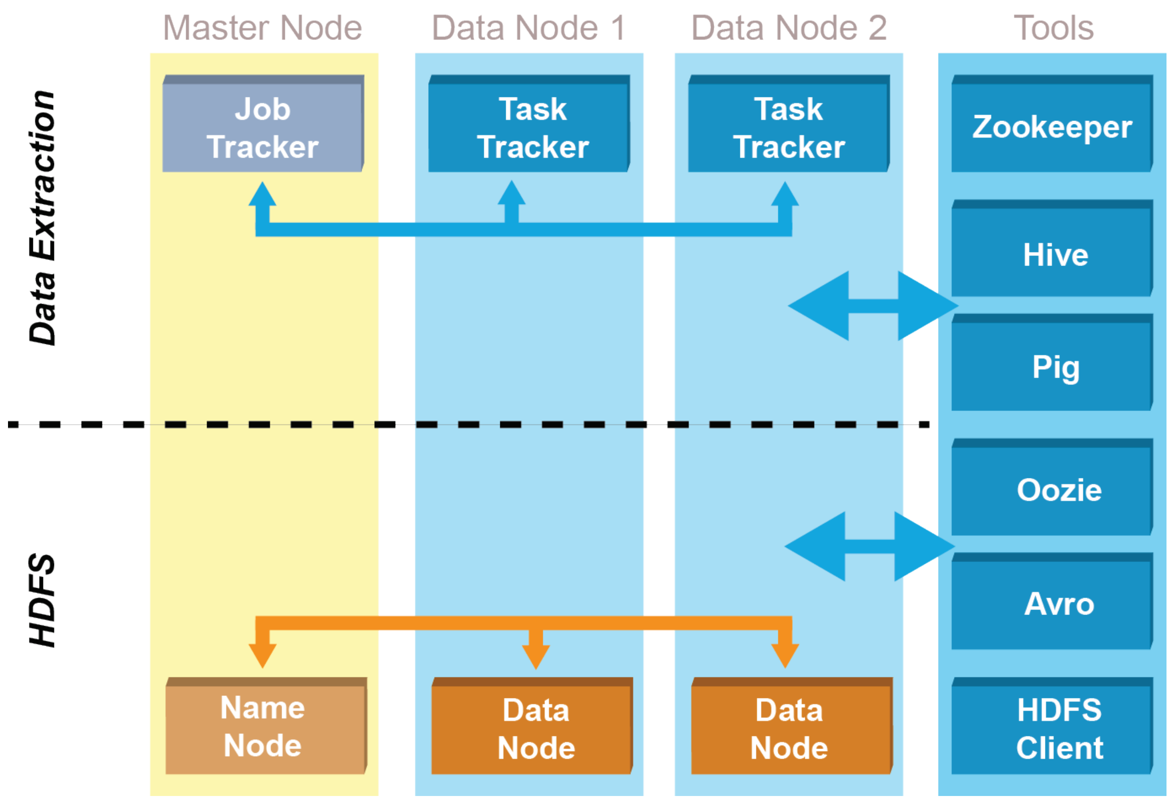

Figure 2. Typical Big Data System Architecture

Discovery Platform

We chose a primary strategy of consolidating data into one system "close" to the analysts that served as a single "source of truth". Data was previously stored across several databases, which fed into a data warehouse, as well as separate databases and tables with purchasing, work tickets, and textual notes. Once all known information was consolidated, it was optimized for HDFS results.

When all of the data was in place within the Hadoop cluster, the team wanted to ensure it had adequately invested in disk storage, random access memory (RAM), central processing unit (CPU) cores, and speed, and that the environment could serve the analysis needs. There were no budget constraints and the team could add more computing nodes as desired, but more nodes would dramatically slow down ingestion when data flow began or perhaps cause lower performance for some queries. The team wanted results in seconds; with a goal to explore hundreds of trials during the second phase.

The team was very pleased in the end to find that system-wide extractions (or queries) were returned within several seconds of request. For example, the mean resistance of every unit in the portfolio was rendered into a table in less than 20 seconds. It cannot be overstated how critical the team believed this performance to be during the discovery and exploration efforts, as many dead ends were quickly identified.

Figure 3 details some of the performance examples. Essentially, the big data analysis system was able to return results about 400 times faster than our high performance production SQL server used every day in the business.

Question	SQL Server	Hadoop Cluster
What is the model and count of jars for each string in inventory?	10:59	1.91 seconds
What is the average resistance by make and model?	not practical	1.33 seconds
What measurements do we have for a specific asset tag?	not practical	0.87 seconds

Figure 3. Timing Improvements

Other questions the team answered quickly to guide exploration included the following:

- What units are +125, +140, and +150 percent of their baselines?
- What does the distribution of internal resistance baselines look like?
- What batteries in the portfolio today are older than design life?
- By model, what is the average age when customers replaced units?
- What batteries are failing outside the normal distribution at 10/25/50/75/100 percent of their life?
- Which battery models performed above or below manufacturer's specifications?
- How did various battery models perform compared to estimated service life?
- Which strings and/or sites were operating outside of the recommended temperature range?
- Is there a correlation between single unit replacement and overall battery string service life?

The team has chosen **not to expose** the relative performance between the various manufacturers and models that support the stationary battery business. Furthermore, it is important to recognize that the specific answers to these questions are only valid for the point in time when the portfolio was analyzed (2015). Batteries being manufactured and deployed today have been demonstrated to have slightly different results. Therefore, a system of **continuous analysis** as constructed is mandatory if decisions are to be made based upon new data and comprehensive historical results. For example, the team observed within the data that one manufacturer moved its production for a specific part number (model) to a different facility. Therefore, any analysis based upon the units produced at the original facility must be completed separate from those of the current production facility; at least until a demonstration of similar results is shown in the data analysis.

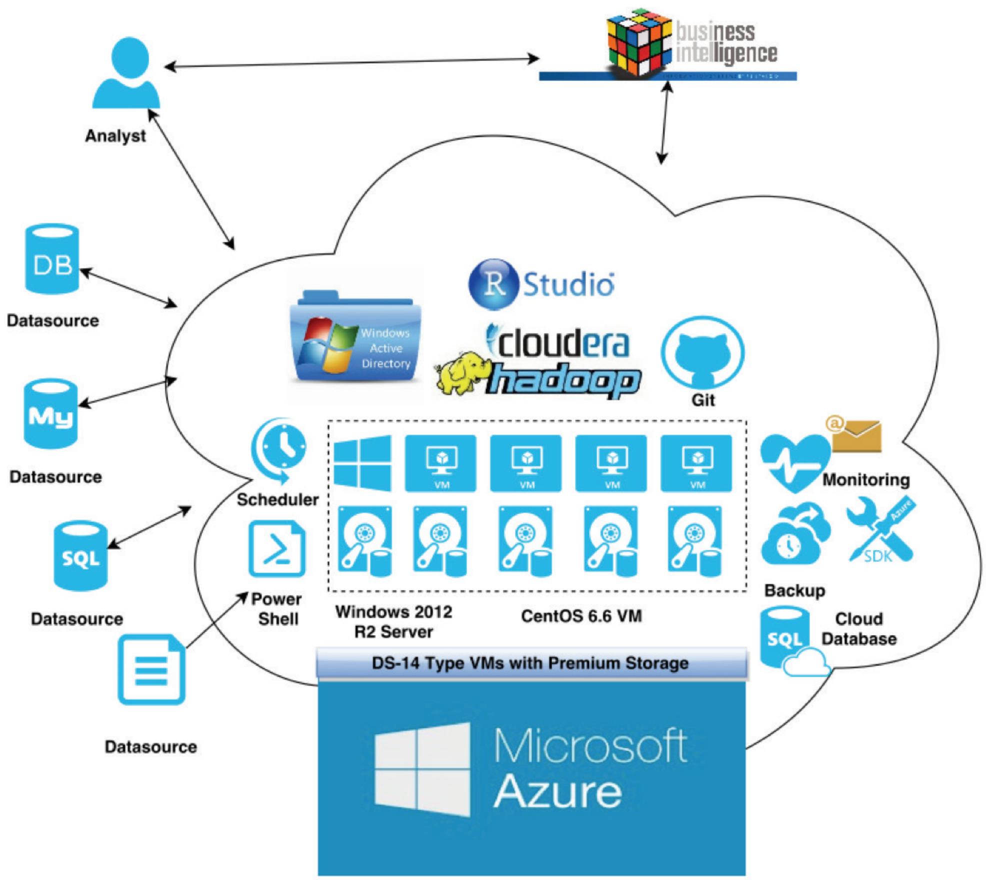

Figure 4. Typical Platform Block Diagram (Source: Cloudera)

Phase Two: Analytical Modeling

The second four weeks of the effort focused on beginning analytic iterations, evaluating results and preparing the visualization platform. For the final week of the effort, the team set aside time to develop its conclusions, transfer knowledge including time to mechanize, and define next steps. In the language of big data systems the inbound data "flows" inward through an Extract, Transform, and Load (ETL) process to permit direct analysis. Additionally, some data is very fixed and is extracted from the internal network tables. In big data systems this is often externally called the Data Store. Figure 5 details the data flow utilized. The team also put the "flow" data into the staging system for a later, production use in phase three.

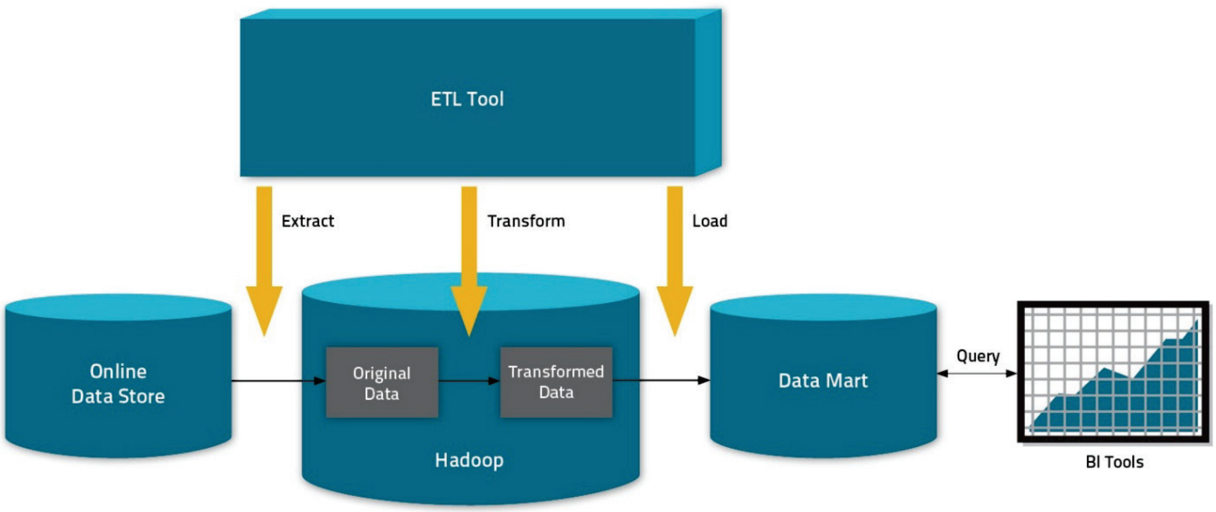

Figure 5. Data Loading Process to Business Intelligence (BI) and Visualization

Creating Holdout and Training Data

One of the challenges previously presented was validating [1] if the forecast methodology was accurate and determining how the team could best measure results within the project period. There was no time to forecast and then wait for a unit to change or fail. Although humans cannot "unlearn" something, it is very easy to cause a machine learning system to "forget" some data, improve the process and try again. Dozens of trials and ETL revisions were required during the development process in phase one. Our approach was to create two separate data sets. One set of data was withheld from the machine learning system. A second set of data, called the training set was used by the machine learning tools to create a forecasting model. Modeling approaches are discussed later. Both sets of data contained a large number of full lifecycle data elements from the portfolio, ranging from the first data measurements until the unit was returned for recycling. Figure 6 details the end-to-end process for creating each model explored.

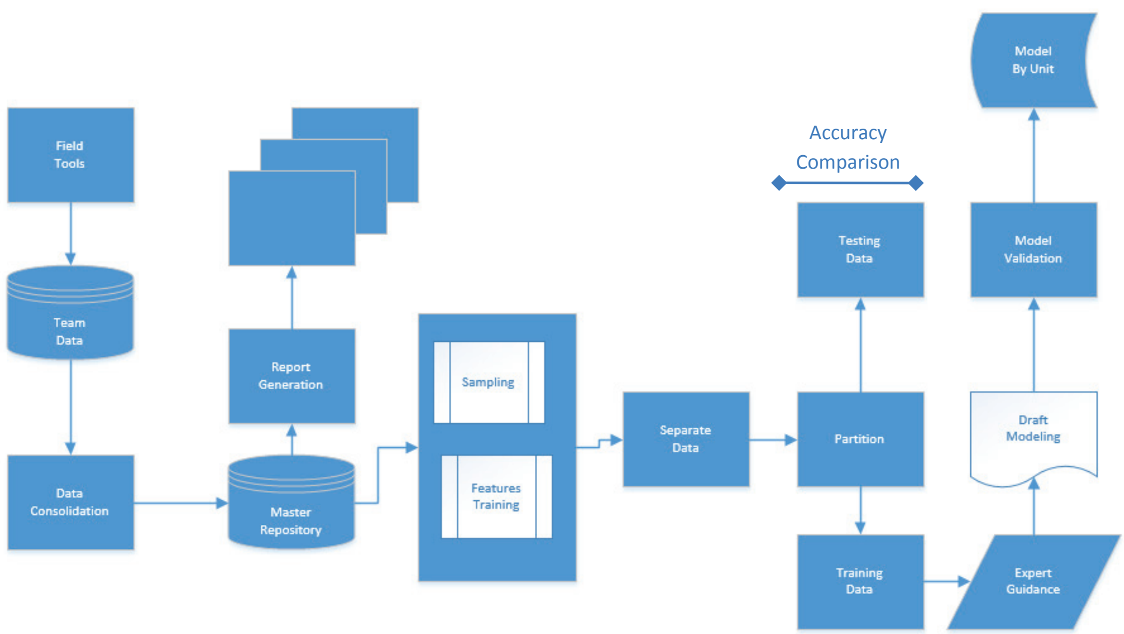

Figure 6. Data Analytics Training Modeling Process

Training Data Set

The stationary battery industry is changing with new vendors, new plants, new chemistries, differing cabinet construction, and dozens of other variations. The task of managing and understanding the implications of these changes has become overwhelming for many experts. Transferring expert knowledge into the language of the machine learning systems for the purposes of this project was a challenge. The team approached this problem by dividing the data sets into small elements that experts could understand and visualize, first by model, then by similar experiences within the data set. This resulted in tens of thousands of result "partitions." Only partial examination by battery experts are required to determine if the forecast methodology is correct, as the holdout data always grades the effective accuracy of the forecast(s).

For example, consider that within a very large set of data there are "active" units (those still in service and operational); there are failing units; there are those that served as required for years; and those that failed prematurely. Then consider if you separated and aligned all of the various similarities and differences, such that you can determine factors that have high, moderate, little, or no effect on the demonstrated service life. The number of possible outcomes for any given unit would be very large, but not infinite. Even for classic function-based mathematics, the number of possible outcomes is beyond the ability of a single polynomial to express. However, it may be possible to create groups of results that are very accurate. Figure 7 details one of the techniques used to divide and partition results into groups with automation using machine learning systems. In one classic example [14], iris flowers are divided into partitions that exhibit similar characteristics. Within the big data world, botanical studies were among the first uses of this technique.

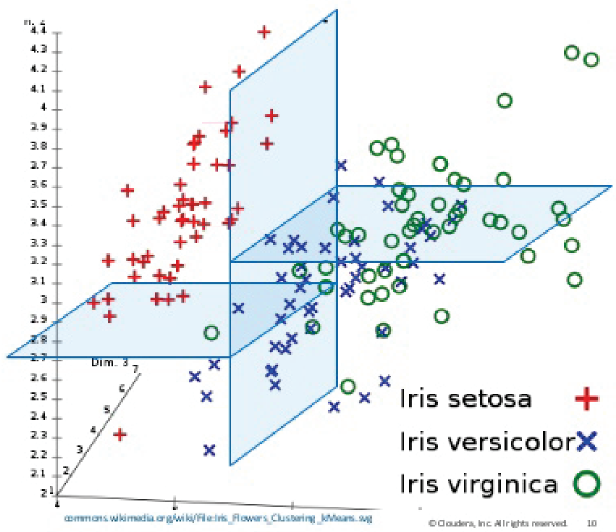

Figure 7. Result Partitioning

Once the team started to understand and visualize the data, it was possible to create and choose a probability distribution based upon the likelihood of the next value. Fifteen well known probability distributions are shown in Figure 8. For example, Bernoulli is representative of the 50 percent probability of a coin toss—heads or tails. This paper calls special attention to the Weibull probability.

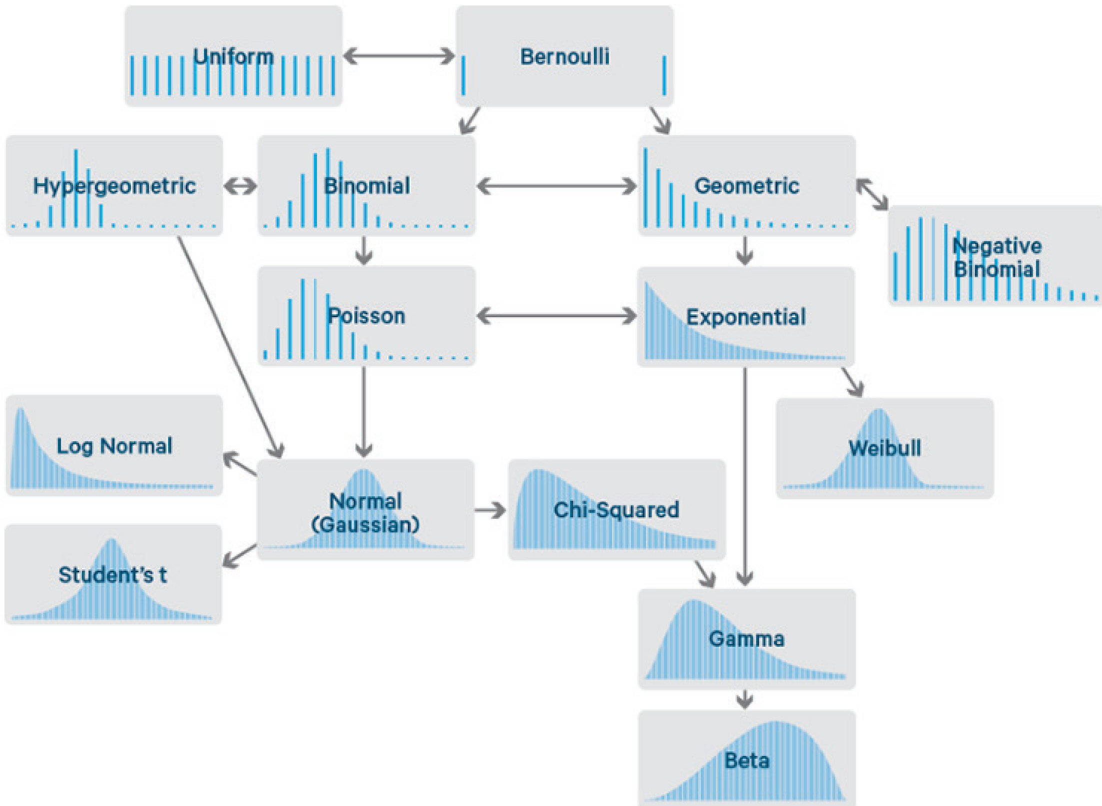

Figure 8. Probability Distributions (Source: Cloudera)

Applying Big Data to Batteries

Our analysis indicated that DC resistance followed a log function similar to mechanical wear-out. Wear-out is also very similar to the inverse of the Weibull distribution as shown in the well-known bathtub curve illustrated in Figure 9. While this curve may seem to be visualized as symmetrical, in our battery portfolio, the number of units that fail early in life is very small.

The shape of the curve in Figure 9 and how fast it "ramps up" is often called Beta. In this project, the team calculated Beta for every battery model with similar features (outside influences), and then using machine learning, predicted the short- and long-term service life of each unit within a string. As part of this analysis, the team learned that there was no single answer for Beta. However, a family of results existed based upon a particular unit, features, and their future results. The team suspected that there are unmeasured factors that affected these results.

Based upon the factors mentioned above and in our prior research [1], every unit will fail or be removed after a few years of service. For clarity, our customer base desires to operate their battery strings as "deeply" into the right side of the battery curve as possible—maximizing their investment in battery units. However, businesses must always consider their risk tolerance and weigh the cost of maintenance or replacement against the costs associated with downtime. Unfortunately, this risk tolerance isn't always fully understood or effectively communicated within the UPS market.

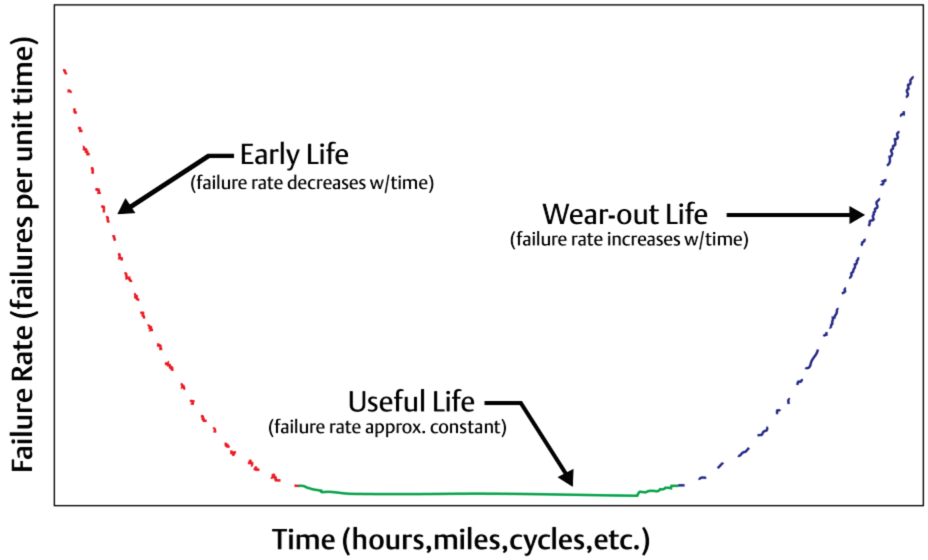

Figure 9. Well Known, Yet Not Fully Understood, Wear-out Curve

A proven and well-known machine learning technique called Random Decision Forest (RDF) ultimately (after many trials) provided an efficient method to separate large dimensional data into axis aligned stair-step decision boundaries as depicted in Figure 7. In the mathematical language of decision trees, the data consisted of 40 million examples, with tens of millions of features (type, temperature, manufacturer, position, charge voltage, etc.). Dividing these trees by manufacturer, model, place of origin, date code, temperature, and resistance produced interesting results and effect scores. Focusing on one set of results of one model, in one class of environment features produced the data which follows. Figure 10 below indicates that age is the most significant determinant of increase in DC resistance. Although self-evident, the machine had to figure this out in order to take the next steps. It is interesting to note that the next highest criteria to predict an increase in DC resistance was how long the previous units operated. Then the variables that affected resistance in order were the absolute resistance, voltage, first derivative of resistance, temperature swing, position in the cabinet, and finally cells per unit.

Variable Name	% Increase Mean Squared Error (MSE)	Increase Node Purity
First Age Difference	0.000481763	0.983969395
Previous First Age Difference	0.000891141	0.980127063
Previous Resistance	0.00003463145942	0.694333108
First Resistance	0.00002330341536	0.402220799
Previous Voltage	0.00000126429631	0.214952077
Previous First Delta of Resistance Percent	0.00004950682091	0.17245948
Previous Temperature	0.00000184046067	0.054850584
Jar Position	0.00000008614301	0.026323025
Cells Per Jar	0.00000033099722	0.005171629

Figure 10. Portfolio Subset Effect Scores

The machine learning system provided interesting “effect” scores (by weight) by unit, string, and site, for the full portfolio. By itself, the project team was excited to see these results, as they confirm what our internal battery experts have been saying for many years.

The team found it interesting that for this model voltage had more of an effect than temperature and that position and cells per unit also impacted performance. These results varied greatly depending on how many different models were included in the training set. This example utilized 12-volt VRLA units with approximately 300-400 watts/cell. Also, note that these results were somewhat self-fulfilling. Field engineers could easily optimize the overall float voltage, but could not easily change the distribution of that voltage in the string, nor individual unit temperatures.

Once the effects of the features were well understood and the RDF techniques were applied iteratively, the team explored many groups of model results. One example that produced highly accurate forecasting was RDF plus Weibull. Figure 11 details a group of models within the portfolio. The training data is shown by result count (in thousands) on the vertical axis. The holdout data (real results) is shown on the horizontal axis and cover a three-month (90 days) forecast window. The very dense dark area represents many thousands of very accurate forecasts. More than 90 percent of the forecasted results fall within +/- five percent of the measured results.

Utilizing this system to forecast resistance provided a mechanized method that improved on human analysis—one that is mathematically verifiable with the training set.

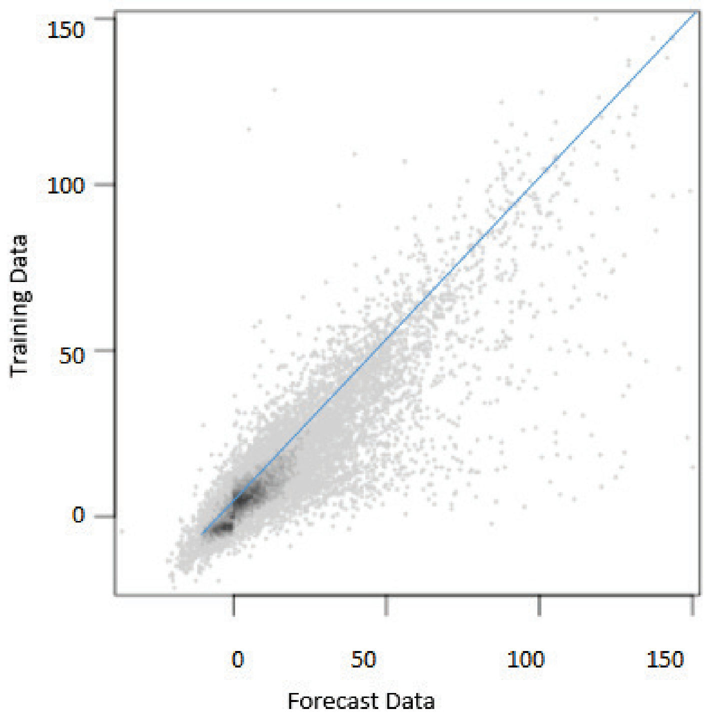

Figure 11. Forecast Accuracy within 90 Days

An additional project effort was to determine if the machine learning system could find and locate units that were not aging "normally." The team wanted to explore the early life in the first third of the bathtub curve (Figure 9). While these units are very likely under warranty if they fail, they place customers at very high risk. A string that is both new and recently tested is often trusted more than an aged string. The team was not trying to catch units in alarm, but those that would be out of policy by the next service event (90 days). Put another way, these were the units that were far from being out of policy, but were deemed very likely to be so in the coming week. To this end, the team trained the machine learning system to look at the first and second derivative of the resistance percentage change (Figure 12). The data analysis team explored units following a Weibull distribution of wear-out, essentially looking for a function that may permit early detection of failure several months in the future.

The team also trained on features pertaining to lot numbers, date codes, and the absolute value of the first baseline resistance. Our experts suspected that high initial resistance may have been an indicator of a manufacturing defect. The team analyzed how quickly the percentage of resistance change grew for a partitioned set of units. Again, utilizing the unit baseline, then creating both the training and holdout data sets, the team wanted to verify the accuracy of forecasts. These results were somewhat mixed with forecast accuracy at 80 percent one year in advance. As a reminder of the general workflow process, if the team forecasts with 80 percent accuracy one year in advance, then 85 percent six months later, then 90 percent three months after that, each visit along the lifecycle becomes more accurate as the risk increases. Work continues in this area.

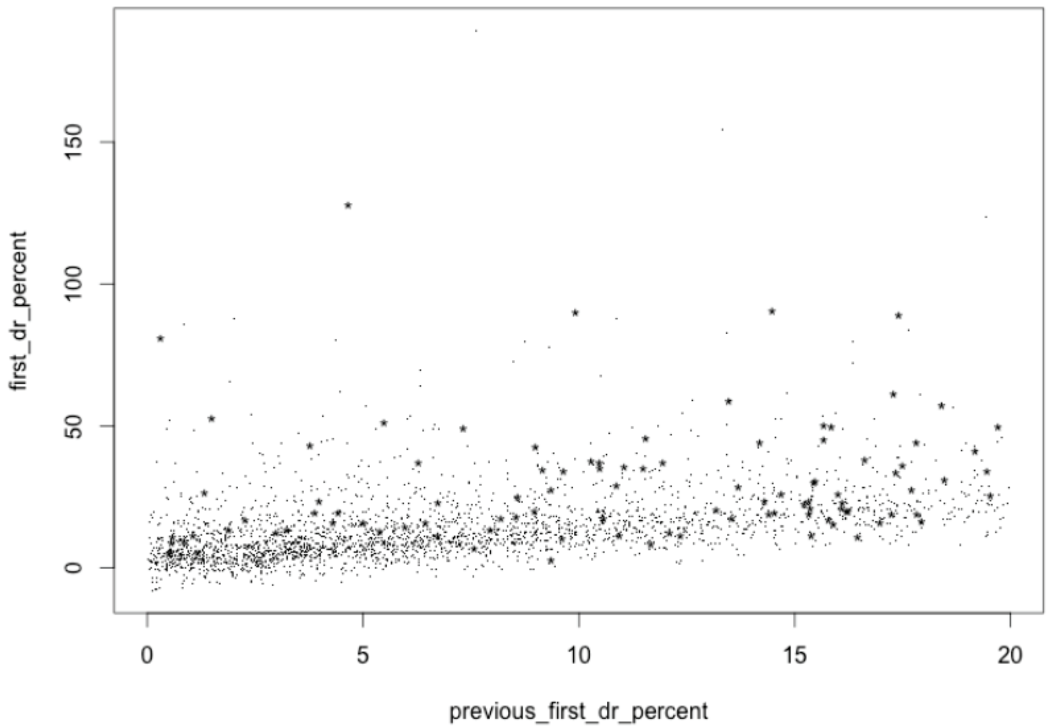

Figure 12. One Year "Rogue" Forecast (* = Error)

Machine Learning Results

Long-term (more than one year) battery health prediction appears to be out of the reach of current big data techniques for VRLA battery units with the relatively infrequent data available during periodic inspections. We were excited to prove that the approach of producing a machine learning algorithm that could forecast battery health between in-person physical inspections is entirely practical. First, the team wanted to identify the "at risk units" that would occur before the next visit, a goal well met. Next, the team demonstrated it could accurately forecast **all** of the resistances at the next scheduled service visit (approximately 90 days) for individual units, battery strings, and groups of strings within a system.

Summary

Big data systems have become tools that can be used for analysis far beyond what has been classically possible. These tools increase the productivity and enable exploration by experts that is otherwise impossible by any other method. The project team has shown that it is possible to create automated hardware and software that can help forecast internal Ohmic values in battery units, based solely on observational data within a historical record. The random decision tree method with log-based forecasting chosen produced results that exceeded 90 percent accuracy for the next visit to a UPS operator's battery string. The team has learned that additional data (perhaps from stationary systems) is required in order to forecast a battery's full lifecycle, but it is now possible perform early identification of those units that are failing prematurely 80 percent of the time. Calling out these units for additional attention during the next inspection, may increase UPS operator's availability and reduce risk from future failures.

Special thanks go to Cloudera's Sean Owen, Michael Stephenson, and Guru Medasani, as well as Emerson's Scott Baer, Chad Eiring, and Dave Saliaris for their support of this work.

Work Cited

1. Ratcliff, G. *A Comprehensive Management Approach To Maximizing UPS Availability .* Battcon Conference Proceedings 2011 paper 15-1
2. Gorlitz, T. *Battery Capacity Testing of Small, Large and Unusual Battery Systems.* Battcon Conference Proceedings 1997 paper 4
3. Noworolski, Z. *RIPPLE, NOISE, CHARGE AND DISCHARGE:TURNING UNWANTED SIGNALS FOR YOUR ADVANTAGE.* Battcon Conference Proceedings 2008 paper 17
4. Floyd, K. *A PROPOSED FLOAT CURRENT ESTIMATION TECHNIQUE AND THERMAL ALARM LIIMITS FOR VRLA BATTERIES.* Battcon Conference Proceedings 2002 paper 13
5. Lawrence, R. *THE VIRTUES OF IMPEDANCE TESTING OF BATTERIES.* Battcon Conference Proceedings 2003 paper 2
6. Cox, Daniel, *Establishing VRLA Battery Maintenance Program Using Ohmic Reference Values and Historical Test Data*
7. IEEE 1188 *Recommended Practice for Maintenance, Testing and Replacement of Valve-Regulated Lead-Acid Batteries*
8. Albér, G.. *Ohmic Measurements: The History And The Facts. Battcon* Conference Proceedings 2003, paper 1.
9. Furlong, T. *Yes...Internal Cell Resistance Measurements Are Valid.* Battcon Conference Proceedings 2000, paper 3.
10. Emerson Network Power. *The Effect of Regular, Skilled Preventive Maintenance on Critical Power System Reliability.* November 2007 Liebert.com: http://www.liebert.com/common/ViewDocument.aspx?id=852
11. Diallo, Amadou, *Pandora Radio's Dominance Built On Big Data Edge*. October 6, 2013: http://www.forbes.com/sites/amadoudiallo/2013/10/06/pandora-radios-dominance-built-on-big-data-edge/#220f18d4b332
12. Emerson Process Management. *DeltaV™ Continuous Historian*. May 2015: http://www2.emersonprocess.com/siteadmincenter/PM%20DeltaV%20Documents/ProductDataSheets/DV_COL_PDS_ContinuousHistorian.pdf
13. Cloudera. *Designing and Building Big Data Applications*. January 2015: http://www.cloudera.com/training/courses/big-data-applications.html
14. Machine Learning Mastery. *Non-Linear Classification in R with Decision Trees*. August 2014: http://machinelearningmastery.com/non-linear-classification-in-r-with-decision-trees/