

Technical Disclosure Commons

Defensive Publications Series

February 22, 2018

Automation and orchestration of hardware and firmware data mining using a smart data analytics platform

Suhas Shivanna
Hewlett Packard Enterprise

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Shivanna, Suhas, "Automation and orchestration of hardware and firmware data mining using a smart data analytics platform", Technical Disclosure Commons, (February 22, 2018)
https://www.tdcommons.org/dpubs_series/1065



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Automation and orchestration of hardware and firmware data mining using a smart data analytics platform

Abstract:

Effective data mining is going to be important for differentiating and succeeding in the digital economy especially with increased commoditization and reduced barrier to entry for infrastructure devices like servers, storage and networking systems. There is lot of telemetry data from manufacturing facilities and customers that can be used to drive improved supportability experience, unmatched product quality and reliability of infrastructure devices like servers and storage devices. Currently data mining of hardware, firmware and platform logs is a challenging task as the domain knowledge is complex with expertise for large multinational organization distributed across the world. With increasing complexity and data mining continuing to be a very time consuming task that requires math/statistics skills, diverse programming & machine learning skills and cross domain knowledge, it is important to look at next generation analytics solution tailored to infrastructure vendors to improve supportability, quality, reliability, performance and security. In this publication we propose a smart, automated and generic data analytics platform that enables a 24/7 data mining solution using an built in platform domain modeler, an expert system for analyzing hardware and firmware logs and a policy manager that allows user defined hypothesis to be verified round the clock based on policies and configurable triggers. This smart data analytics platform will help democratize data mining of hardware and firmware logs and help improve troubleshooting complex issues, improve supportability experience, reliability and quality and reduce warranty costs.

Problem Statement:

Data is being considered as the new currency and it is important for enterprises to use data effectively to differentiate their service and product offerings. Currently there is rich set of telemetry, configuration data and failure events from millions of devices that gets stored in different organization specific data warehouses. This rich data can be used to derive new insights that can help analyze outages automatically by categorizing the root cause to software, hardware (e.g. quality and reliability issues with vendor components), user error, mis-configuration and poor quality vendor deliverables (e.g. OS drivers bug causing system crashes). The hardware and firmware logs can also be used to understand failure patterns and its relationship between data center operating environments (e.g. Temperature, Humidity, usage etc.), conformance to vendor specific operating guidelines (e.g. ASHRAE specifications) and root cause analysis of complex support cases. This rich data can be used in many other ways to reduce warranty costs by detecting patterns in data collected from customer machines and correlating it to actual failures and for validating failure and design hypothesis in rapid time using next generation data mining solutions. However, there are many challenges to achieve this end goal. Few of the major challenges are listed below

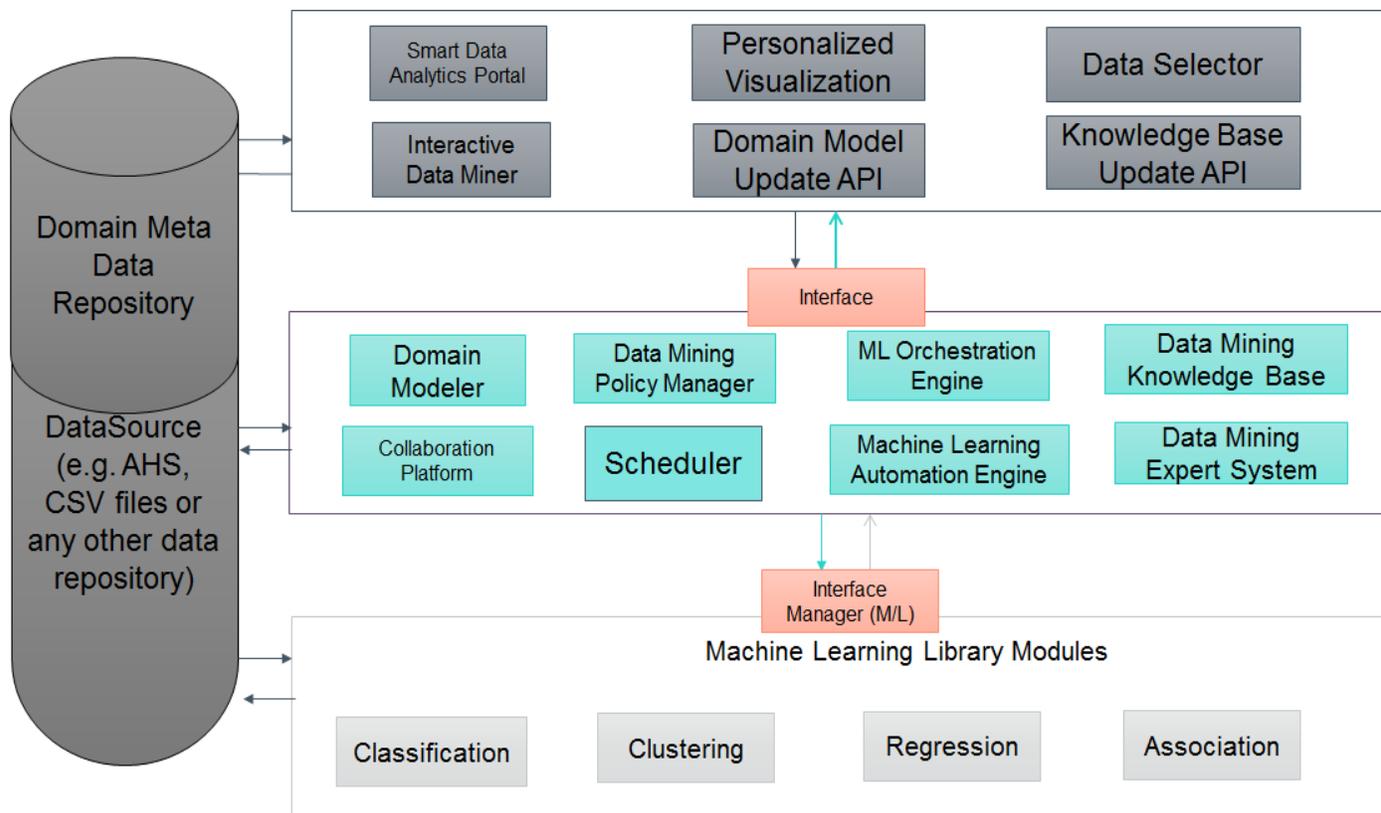
- For data mining, especially data related to hardware and firmware, we will need (i) deep technical domain knowledge of the hardware, firmware, mechanical, software to which the data relates and (ii) deep math/statistics/artificial intelligence (AI) and machine learning (ML) knowledge in order to analyze and use the data. It is also important that we understand the characteristics of the data, relationship with each other before building a data model that can be used to detect patterns, predict outcomes and perform root cause analysis. This kind of complex combination of deep domain expertise, knowledge of data and ML/AI algorithms and math/statistics knowledge is very difficult to get in one team/geography –typically the knowledge is spread over technical experts in different countries and departments and the statistics and AI/ML knowledge is held by people who do not have the technical knowledge of the hardware, firmware and software underlying the data. This is a big obstacle to deriving fast and accurate insights from any big data mining projects.
- Data mining of hardware and firmware logs is a very complex activity that requires data transformation that can be very platform specific and requires deep technical domain knowledge about each of the attributes collected from a particular device. For e.g. the inlet temperature reading in a typical server log can be a raw number from -10 to 100+ and the way to interpret this reading may vary from platform to platform depending on the cooling subsystem and hardware/mechanical design. Using hardware registers for data mining containing different error states can also be a very intensive work, and with interpretation and values changing from generation to generation, it is important to look at interpreting these values in an automated and reliable way to allow data mining on these complex attributes and find relationship with other parameters within the server.

Solution:

We propose a method and apparatus that enables automated data mining of telemetry data in a generic way using a “Smart Data Analytics Platform”. The solution will help democratize data mining by discovering knowledge and insights rapidly and the Smart data analytics platform is designed as a generic solution with an integrated Domain modeler to allow capturing domain knowledge of hardware devices and every attribute in telemetry data warehouses. Along with the domain modeler, we propose a hardware/firmware AI/ML expert system containing a knowledge base with a list of best ML/AI algorithms for analyzing platform logs and its best practices, usage policies, constraints, data affinity and performance metrics. The smart data analytics platform will support automated data transformation and data mining using the domain modeler and expert system with support for personalized and advanced visualization. The solution enables data mining to be done by any engineer with just a problem statement

along with the right data set to derive insights in an automated/semi-automated way. The Smart data analytics platform abstracts the data mining details from the user using a simple set of options/questionnaire and the domain knowledge along with ML expert system can then be used to perform 24/7 data mining using optimized math/statistics/ML libraries based on models created by either data scientists or by the expert system or both. This new solution will enable round the clock analytics using a data mining policy manager and can be configured to run the model whenever new data comes in (e.g. telemetry data from field or factory) or based on specific triggers (e.g. outage rate, failure rate etc.). The solution will also allow capturing hypothesis (e.g. Design assumptions of new hardware products) and validate them in an automated way to get rich insights and foresights

The high level architecture of the solution is shown below



The main solution components and associated details are as follows

Domain Modeler: The domain modeler helps to capture the domain knowledge in a very generic way using set of policies in XML, JSON or any data description language. The domain model will be designed in a hierarchical way containing sub-domain categories

which at the leaf level contains the attribute will all its properties that map to either a column in a telemetry data warehouse or CSV file headers. The properties of the attribute captures meta data information like

- Type of the variable (continuous, categorical etc.),
- Valid range: Valid Range for continuous variable (e.g. Temperature ranges from X-Y), for categorical, it is the number of categories, the names and associated value
- Anomaly values/range for the data set: e.g. >120 can be considered as an anomaly for temperature variable
- Platform specific continuous to categorical variable conversion rules: (e.g. raw temperature can be converted to a defined number of categories with names like high, medium, low)
- Outlier/error value handling rule: e.g. replace error values/outliers with values that map to mean/ median or mode for continuous variable type.

Data mining expert system (Dynamically configurable with Machine learning Recommendation Engine): To enable an automated and rich data mining solution, we propose a machine learning expert system with a knowledge base capturing all the different types of popular ML algorithm and recommended workflow based on input data characteristics and type. The knowledge base also contains important parameters like an overall ranking value capturing the effectiveness of the algorithm for different input types, the best practices for each algorithm that includes the minimum sample set required for algorithm to output reliable results, preconditions, performance metrics, data affinity (e.g. works for continuous variables with linear relationship or categorical inputs only) and if the algorithm is supervised learning based, the optimal ratio of training and validation data set. The knowledge base can be dynamically configurable based on research/learning and advancement in machine learning algorithms using REST APIs and can be customized to meet the needs of a particular domain and data characteristics (e.g. hardware/firmware domain vs marketing requirements). A recommendation engine inbuilt into the machine learning expert system will automatically recommend the best algorithm based on the input data set given by the users by characterizing the input parameters and output parameter type and using the ranking information of each algorithm in the knowledge base.

Suhas Shivanna (suhas.shivanna@hpe.com)