Journal of Manufacturing Systems 45 (2017) 236–247

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Full length article

A flexible data schema and system architecture for the virtualization of manufacturing machines (VMM)

Atin Angrish, Binil Starly*, Yuan-Shin Lee, Paul H. Cohen

Edward P. Fitts Department of Industrial and Systems Engineering, Carolina State University, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 2 June 2017
Received in revised form 3 October 2017
Accepted 7 October 2017

Keywords:
Cyber-physical manufacturing
Smart manufacturing
Digital twins
MongoDB
NoSQL

ABSTRACT

Future factories will feature strong integration of physical machines and cyber-enabled software, working seamlessly to improve manufacturing production efficiency. In these digitally enabled and network connected factories, each physical machine on the shop floor can have its 'virtual twin' available in cyberspace. This 'virtual twin' is populated with data streaming in from the physical machines to represent a near real-time as-is state of the machine in cyberspace. This results in the virtualization of a machine resource to external factory manufacturing systems. This paper describes how streaming data can be stored in a scalable and flexible document schema based database such as MongoDB, a data store that makes up the virtual twin system. We present an architecture, which allows third-party integration of software apps to interface with the virtual manufacturing machines. We evaluate our database schema against query statements and provide examples of how third-party apps can interface with manufacturing machines using the VMM middleware. Finally, we discuss an operating system architecture for VMMs across the manufacturing cyberspace, which necessitates command and control of various virtualized manufacturing machines, opening new possibilities in cyber-physical systems in manufacturing.

1. Introduction

The variety and velocity of machine data on a factory floor and cheaper methods to store, compute and analyze this data for real-time technical and business decision making are major driving forces to improve factory productivity. These developments are motivating the manufacturing information technology industry to re-examine traditional machine control and networking architectures present in manufacturing shop-floors. In discrete manufacturing job shops, such as those in machining services, networking multiple manufacturing machines can be difficult to achieve due to a variety of reasons. This include, interoperability concerns between machine controllers from various vendors, outdated hardware controllers and a legacy infrastructure that is incapable of handling latest network communication protocols. This hinders the transformation of physical factory floors to the digital era. Advancements made in machine communications standards (example – MT-CONNECT [1], OPC/UA [2], MQTT [3]) have certainly lowered the barriers of machines communicating with centralized information systems (such as MES and ERP systems).

The era of digitalization of manufacturing processes and its networking for efficient utilization of manufacturing resources requires a shift in the way data from manufacturing machines is handled, stored, retrieved and computed for actionable insights. Key manufacturing trends that are driving the need to rethink how data from factory floors and its extended enterprise are organized, include:

- *The need for enterprises to quickly respond to evolving market demands*: Democratization of manufacturing is leading to increasing demands for personalization or customization of products. Reconfiguration of physical infrastructure rapidly based on changing market needs and demands is necessary. Production in quantities of one, such as in personalized medical products, to quantities less than one hundred will require new ways in which data from manufacturing systems is 'pulled' to satisfy market demands. This pull of information is necessary to assess distributed capability and dynamic capacity across job-shop and production floors regionally and globally.
- *The shift to connecting product lifecycle data with the manufacturing processes*: The digital thread concept linking product information throughout its lifecycle from conception to production, then use and final disposable requires that we have fundamentally new ways in which information is linked across its various points of

* Corresponding author.
E-mail addresses: bstarly@ncsu.edu, starlyb@gmail.com (B. Starly).

https://doi.org/10.1016/j.jmsy.2017.10.003
0278-6125/© 2017 Published by Elsevier Ltd on behalf of The Society of Manufacturing Engineers.

use [4]. Today, data generated during a product lifecycle is stored in siloed information systems – from PDM/PLM, to ERP/MES, to unique databases maintained by end customers. Integrating data across these disparate systems and linking across them in meaningful contexts across multiple enterprises in the supply chain can be challenging and expensive to implement.

- *Increasing granularity of data collected from manufacturing machines*: With the ubiquity of sensors (RFID and machine mounted sensors), wireless networking and cloud storage, collecting and storing large amounts of data is not a bottleneck. Analytics conducted on manufacturing data can be beneficial beyond just monitoring, but useful in predictive and prescriptive maintenance [5]. It can also be used to inform key shop-floor decisions such as production scheduling and supply chain performance [6–8].
- *The scrutiny placed on machine-to-machine connections in a network*: With all this connectivity, machine generated data requires context and meaning. Semantic models must include both explicit (raw signals, feature attributes) and implicit (operator name, part name, description), compounded with a time-reference to make sense of large machine generated data. The consumer of this data is presented with customized visualization and context to make data-driven decisions.

In a key technical position paper by Lee and Bagheri et al., the authors proposed a high-level cyber-physical system (CPS) architecture towards digital factories [9]. The primary intended purpose of CPS in factory floors is to manage Big Data from factory floors and leverage the network connection ability of machines to create the goal of resilient, intelligent and self-aware machines. Five tiered elements are proposed in the CPS architecture – 1) Connection to the physical machines; 2) Conversion of Data to Information for each individual machine; 3) Aggregating this information across a fleet of machines through 'digital twins' of machines; 4) Cognitive architecture which help synthesize the information for decision making both within and beyond an enterprise, and finally, 5) Configuration, where data insights help decision making support by individual machines themselves or by humans within the loop. This paper relates to a system architecture and a flexible data storage schema surrounding Levels 2 and 3, to enable higher order elements of a manufacturing focused CPS. At these levels, we focus on a data architecture that supports the concept of a 'digital twin' for a machine or more specifically, 'Virtualized Manufacturing Machines'.

The virtual manufacturing machine (VMM) of a physical machine encapsulates its physical capabilities (static), in-process data (dynamic) generated by controllers and any external sensors attached to the machine. The core function of a VMM is to present a middleware architecture that abstracts hardware level specifics of machines on the shop-floor and then provide a programmatic interface to allow higher order information systems to feed into service applications. The community has seen a plethora of technical phrases utilized to signify the data generated from physical machines. These terms range from 'digital twins', 'Cyber-twins', 'digital machine analogs', 'digital machine shadows' etc. At its technical core, the concept of a virtualized version of the physical machine signifies a data model that encapsulates technical specifications, machine data, and information relationship about a physical machine and its environment which then represents the machine in near real-time states within cyberspace.

The cyber-twin of a physical machine can reside within the machine's own computing system, at a server with close proximity to the machine or at a remote external cloud location. It is intended to tap in between the control/communications systems of a machine and higher order execution systems. This streaming data from a machine populates a data structure, which allows it to be analyzed within the context of manufacturing. The VMM for a machine must also provide its version of the data for cross-analysis among various other types of machines on a factory floor. We focus this study to the first part, a strategy to store structured and unstructured data from machines and have that data be made available to third-party applications.

This paper is organized as follows. In the following section, we describe components of the proposed system architecture for the virtualization of manufacturing machines. Further, we provide details on the design of a document based database schema and then critically evaluate the schema structure by testing it against two query types on streaming machine data from a metal based additive manufacturing machine. This paper will discuss the use of an unstructured database (MongoDB) as the core backbone infrastructure that instantiates the virtual manufacturing machine. The paper also proposes a high-level operating system architecture when multiple VMMs for the various manufacturing machines on a shop-floor are operating in a cyber-physical manufacturing space. We demonstrate two app cases for the VMM, one each for a low end, partially open sourced 3D printer and a high-end closed source metal 3D printer.

2. Components of the virtual manufacturing machine (VMM)

The ultimate goal of any cyber-physical manufacturing system is the ability for global enterprises to quickly respond to business and customer demands while containing costs and maintaining operational flexibility. Wang et al. [10]. discusses various types of cyber-physical systems in manufacturing. Their work expands on the examples of CPS in manufacturing such as multi-agent systems [11–13] adaptive manufacturing systems [14,15], model driven manufacturing systems [16] and cloud manufacturing [17,18]. To this end, there is an imminent need to consider advanced database systems which allow large scale storage and retrieval needs for cyber-physical systems.

In recent work, Kang et al. [19]. developed a NoSQL data store using MongoDB system for supply chains in manufacturing and performed an assessment of their traceability system with a simulation using streaming RFID data. Several papers have discussed the feasibility of deploying NoSQL type data stores for storing large volumes of streaming data from small sensors used in IoT applications [20–22]. In their studies, they have demonstrated various aspects of NoSQL databases and analyzed their performance with respect to relational databases. Boicea et al. [23] demonstrated the performance of MongoDB vs relational databases where they compared the performance of both the database systems in terms of insertion, deletion and update speeds and recommended the use of non-relational databases where high speeds are needed. Liu [24] discussed the auto-sharding capabilities of MongoDB and developed strategies to improve the concurrent read/write performance of the data structure. Several organizations have implemented MongoDB as back-end IT systems for crunching through various schema-less data and as a way to store and present data in web compatible formats [25]. Most recently, NIST has started a "Materials Genome initiative" which uses MongoDB at the backend with RESTful services to enable third-party software integration [26,27].

NoSQL databases are a distinct class of databases that do not have a prescribed schema when compared to conventional SQL based databases. The most common of such flexible schema data type stores include graph and document based stores. In document based stores, data is stored in the form of key-value pairs known as the JSON format (javascript object notation). The documents form the atomic units for a NoSQL database. This enables the NoSQL databases to efficiently handle unstructured data gener-