Intuitive and Scalable Operational Simulation of PSP

Eric Hoarau and Jun Zeng; Hewlett-Packard Laboratories; Palo Alto, CA

Abstract

Medium-sized commercial Print Service Providers (PSP) are adopting digital technologies to improve their capabilities and to manage the production more efficiently. A workflow solution that encompasses the entire system from pre-press to shipping is therefore needed. An important component of this future workflow is a mobile platform that enables anyone from the management down to the operators to visualize, understand and interact with the collected information or with the operational intelligence engine. We present a software application solution built on open source software that can be used to simulate what-if scenarios for pre-sale of large digital printing presses and for manufacturing operations monitoring and control. The end-user creates models of his PSP and scenarios using an intuitive web-based application. The application interacts with the back-end cloud based SaaS to run user generated Ptolemy models, optimization engines and data storage. The user can then compare results and make more informed decision on his print operation. The solution was built with the end-user in mind by being both intuitive and scalable.

Introduction

With annual retail sales of over US$700B, the commercial print industry transformation to digital presents many opportunities and challenges. The general commercial print segment in particular is experiencing a secular trend of shorter job run-length as the demand grows for mass customization and personalization of prints. The highly variable and dynamic job mix and highly personalized customer requirements leads to the trends that medium-sized commercial Print Service Providers are adopting digital technologies to improve their capabilities and to manage the production more efficiently.

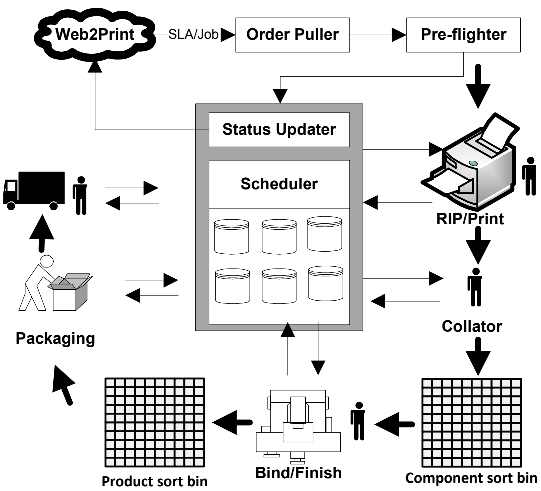

Figure 1: A typical end-to-end print workflow within a print service provider.

As a pioneer and leader of digital print, HP's strategic position in driving up print growth is through accelerating the analog to digital conversion and by improving the unit profitability. While the digital printing technologies have made great stride in the last decade, the print operations management tools and methods are now coming into the spotlight. In order to take full advantage of what digital technologies has to offer and to stay competitive in the future, Print Service Providers (PSP) need to take a complete operational viewpoint instead of the piecemeal approach that is prevalent today. The integration of embedded sensing, actuation and computing power throughout the fulfillment workflow will enable automated monitoring, real-time control, and dynamic reconfiguration of the production of the print artifacts. The sketch in Fig.1 outlines the various workflow steps in the print manufacturing process. The customers' requests for prints are submitted electronically or physically through the store front. Once accepted, a job ticket containing the job's various parameters is assigned. The submitted electronic files are pre-flighted, imposed, and rasterized into bitmaps. The printer-ready electronic files are then sent to shop floor to be printed. The sheets or web are singulated into pages which are then folded, collated, and bound into book blocks. The book blocks are joined to book covers. The finished books are then sorted, labeled, and shipped [5].

PSP must look into the entire print order acquisition and fulfillment cycle for operation efficiency. In the commercial print segment, the print volume is in systemic decline. PSPs are pushed to look beyond the print step to develop their competitive edge: operation efficiency through the entire service fulfillment chain, ancillary services to deliver the print where and when needed, and more. Currently PSPs are struggling to put their arms around the volatile and diverse print demand and dynamic workflow. Similarly, in the enterprise print segment, today's tough budget climate forces enterprises to better leverage their printing dollars. It is the operation efficiency that drives CIOs to look into different ways to fulfill the enterprises' print needs [2]. However, 79% of CIOs cannot quantify their printing costs and efficiency.

Operations simulation for print services [4,6,7] can provide key performance metrics such as cost and lead time accounting for the dynamic demand stream and high faulty rates in the manufacturing processes. Operations simulation can help PSPs to make informed business decisions by forecasting their print operations and quantifying the alternatives. Similarly, it can help CIOs to manage the enterprise's current printing cost, eco-footprint, efficiency and outsourcing options.

Consequently a workflow solution that encompasses the entire system from pre-press to shipping is therefore needed, in particular, a mobile platform that enables anyone from the management down to the operators to visualize, understand and interact with the collected information or with the operational intelligence engine. The solution needs to be cost effective, easy to deploy, scalable, and extensible to allow new functionalities.

In this paper, we describe how we leveraged our research into print factory simulation [6,7] to build a comprehensive solution for print workflows which we call ***p**rint **m**anufacturing **o**perations **s**imulator* (PMOS) [1,3,4,6,7]. PMOS includes two essential components: SimOp, the front-end client and the web-based user interface, and SimCloud, the back-end cloud-based SaaS.

Our solution

PMOS can be used for predicting the consequences of various what-if scenarios and for manufacturing operations monitoring and control. The first part of the application is the predicting tool where the user who can be a solution architect or a production supervisor will interface mainly with SimOp. He logins in the PMOS service to create a model of the PSP operation, submit different operational scenarios through SimOp to SimCloud and compare results and analysis. The solution architect can use the solution during pre-sale to show more accurately to his customer the potential improvements to his factory. The production supervisor on the other hand can apply PMOS to assist in making data-driven operational decision in the ever-changing dynamics of the production floor. A key element of SimOp is to enable the user to easily and intuitively create his own models (*virtualPrintFactory* [4,6,7]). "Master-templates" representing a range of industries are pre-loaded in the system to encapsulate the complex system dynamics in v*irtualPrintFactory*.

PMOS uses *virtualPrintFactory* as the core of the simulation engine. *virtualPrintFactory* is an operations simulator for print production developed by HPL/CPA based on an open-source simulation engine Ptolemy in partnership with RPI, one of the largest digital PSPs in the United States [4,6,7]. *virtualPrintFactory* defines digital print operations as a four-dimensional problem: orders, resources, fulfillment paths, and operating policies. They are coded into an XML file as part of the simulation model that digests input order stream and drives the simulated print production.

The second part of the application is the monitoring tool where the management team uses SimOp web interface to easily access their own operational information. The user-generated floor layout provides an intuitive communication and interaction channel between the workflow controller and the end-user which is

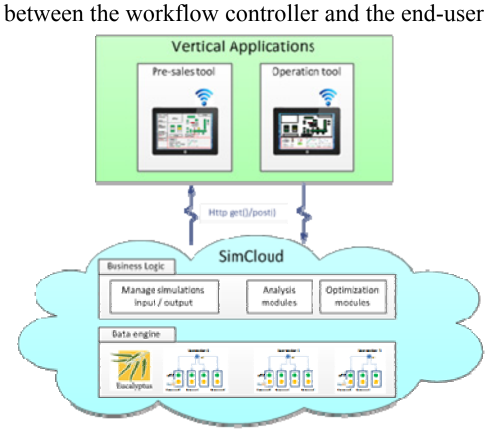

Figure 2: Architecture of the simulation platform.

not possible today. SimOp's current 2D rendering can be expanded to 3D in the future, and a multitude of data can be overlaid on it. For example, the production supervisor can interact with the 2D map of his floor to determine the health of his operation or where the bottlenecks are. He can dive deeper in the data by clicking on individual resource and accessing information such as capacity, up-time, and maintenance. By contrast, obtaining this information in current offerings if at all possible will require digging through data and log files, or by walking around the shop floor. A member of the sales team can use the tool to figure out the location of a specific job, its status and expected completion.

The solution was built with the end-user in mind to be both intuitive and scalable. In order to achieve these qualities, we built the solution using a SaaS architecture. We leveraged mobile computing standards, open-source codebases and our internal research. As shown in Fig.2, the user interacts with the web-based applications (Pre-sale or operational tool) to create models of his PSP, define and trigger multiple scenario, and retrieve the results of the simulations. These front-end application interact with the back-end cloud based SaaS using a REST API. This architecture enables all the heavy computation such as running what-if scenarios, optimization engines, and data storage to be hosted on the cloud server and not the user hardware. The user can focus on reviewing the results and consequently making more informed decision on his print operation.

Front-end Client and User Interface (SimOp)

The user of our solution will mainly interact with the front-end applications (Pre-sale and Monitoring) and as such a lot of effort was placed into its usability and deployment. The applications run entirely in the web browser and were built using HTML5 and JavaScript to be platform independent. The first version was designed to run offline to enable on the road

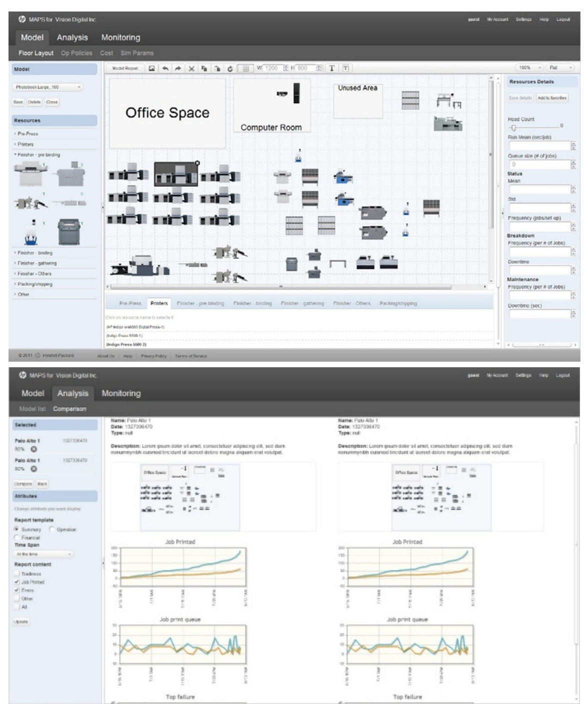

Figure 3: screen capture of the SimOp front-end to run what-if scenario.

interaction. However, as the prevalence of internet connectivity is growing, the next version will be a web application for easy deployment and supportability. The first application enables a user to derive intelligence about his specific PSP by creating a virtual model of his PSP, triggering different what-if scenario, and analyzing the results. While solutions exist on the market to create simulation of factories, they are complex and take consultants days if not weeks to create models and to produce results. One of our primary objectives was to democratize the process and enable anyone in the management chain to interact with the application. Additionally, the initial model of the PSP operation should be built in 30min and limitless scenarios could be run effortlessly afterwards. This seems to be a daunting and unmanageable task as the operation of a PSP is complex, and involves numerous resources. To solve this conundrum, we developed a two steps approach. First, we developed a standardized method of describing the operation of a PSP that consist of four dimensional spaces: orders, resources, fulfillment paths, and operating policies [7]. Second, we will have a library of models that a user can select and modify. The user starts from one such template and adds, removes, or modify any parameters to create a representation of his PSP, and any combination of what-if scenarios that interest him. Based on his need, he can choose the amount and depth of details he wants to modify. As seen in Fig.3, the user first starts by modifying the floor composition, layout, and the individual machine parameters. Second, he sets the operations policies such as the scheduling methodologies and the various possible fulfillment paths per job type. Third, the user set the cost for the resources, labor, and overhead. Fourth, he sets the input orders profile for each simulation he wishes to run. The simulations are sent to the SimCloud service where they are run in parallel. Finally, the user can retrieve and compare the results to aid him in his decision.

The second application enables the user to monitor his operations easily from a graphical interface. As shown in Fig.4, the user-generated floor layout provides an intuitive communication and interaction channel between the workflow controller and the end-user which is not possible today. A multitude of data can be overlaid on it such that the user can see in a single view the overall health of his PSP. He can determine all the machine status, and if there are issues or bottleneck without having to walk the floor. Additionally, he can track specific jobs and see their flow and location on the floor. This interface can also be used to investigate special events that were flagged by the MIS. For example, he can replay on the virtual floor layout the series of events that lead to the issue and therefore get a better understanding faster. Additionally, several of other applications that take advantage of this new interface are currently being investigated.

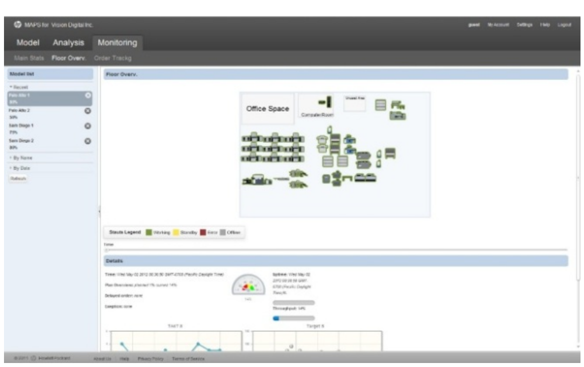

Figure 4: screen capture of the SimOp front-end for monitoring operation.

Simulation Cloud Service (SimCloud)

Operations simulation is both data-intensive and compute-intensive. A typical simulation uses order information as input; one week order data can reach several GBs [6,7]. The span of the input order data can vary from days up to months depending on simulation objectives. The outputs can be up to ~100GB depending on the simulated work span, order stream, event frequencies and output granularity. Cloud as the deployment vehicle for operations simulation eliminates the need for the PSPs and enterprise customers to own and support significant compute and storage resources required for large-scale operations simulations. Furthermore, sharing of digital assets such as simulation models and analysis results becomes much easier across the organizations.

We developed SimCloud, a research prototype that offers operations simulation as a cloud service. As a pay-for-use software service, SimCloud is designed to deliver the operation and costing metric directly to the fingertips of PSP operations managers, enterprise marketing managers and CIOs. The users (e.g., PSPs and their customers) can subscribe to SimCloud service to forecast the lead time, cost, quality of service and resource utilization before any disruption takes place (e.g., signing service contracts, admitting orders, upgrading equipments, executing new polices). SimCloud allows users to upload their own models and inputs into the shared data store, for instance, through SimOp, and execute and analyze their own simulations in the shared compute cluster. The simulation can be executed in a batch mode (“off-line”) or in a real-time mode (“on-line”) to assist real-time decision making. We have also exposed SimCloud as a web service to enable direct integration with other enterprise planning and workflow software in addition to serve SimOp, the front-end for PMOS.

SimCloud has a shared model library that includes simulation model templates that all users can leverage. Built-in model templates include digital PSPs, shipping services, and print networks that facilitate deal-making among multiple print buyers (of different budget and lead time requirements) and PSPs (of different capability, capacity and factory load). A user can also publish its own model in this shared model library to share simulation expertise, advertise service fulfillment capabilities, solicit service bid or facilitate a business engagement.

Fig. 5 illustrates the architecture of SimCloud. The user presents the simulation models and associated input data via the Simulation Design Management module. The Simulation Run

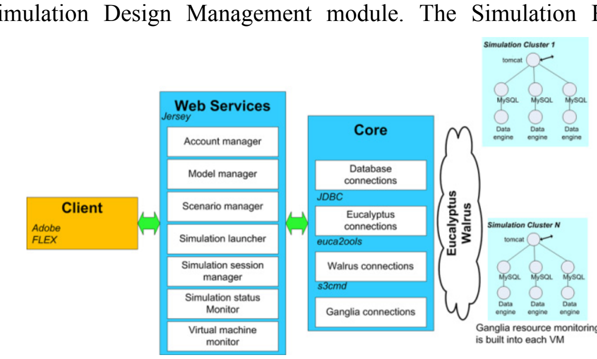

Figure 5: SimCloud system deployment diagram.

Figure 6: SimCloud computational system monitoring. At top shows the composition of simulation clusters. Each node represents a virtual machine; its IP address and role are also displayed. At bottom shows the resource and health monitoring of a virtual machine; displayed contents include CPU usage and network traffic.

Management module accepts the user simulation request and triggers the Simulation Run Scheduler module to launch scripts that communicate with the virtual machine (VM) controller and acquire VMs from the shared machine pool. It downloads input data from the scalable persistent store to populate the database engines hosted in the VMs. Each simulation run generates one simulation cluster that can execute multiple scenarios concurrently. Each simulation cluster is composed of (2N+1) VMs partitioned into three tiers: one VM as the web-based access gateway; each scenario includes one VM as a simulation engine and one VM for database. An over-provision factor is used in VM acquisition because our experiments show sizable variation in time taken for Eucalyptus to produce a VM.

When simulation is running, the Simulation Run Monitor module relies on Ganglia monitoring agents installed at each VM to report this VM's health and resource usage. The User Simulation Session Tracking module tracks all simulation runs including the simulation models, inputs, VMs used and the outputs. The Simulation Analyses Management module allows the user to generate analyses on the simulation results in both batch and real-time modes. Fig. 6 shows SimCloud's monitoring dashboard.

The User Identity and Asset Management module serves to authenticate users and restrict data access on simulation related assets. It stores and manages user credentials, including, for example, the secret access keys to access the scalable persistent store. When the simulation cluster on behalf of a user is established, the corresponding secret access key is copied to the database VMs by the Simulation Run Scheduler. The database VMs can then present this key to the persistent store to retrieve input data and store output data.

SimCloud has been developed and deployed in the 128-node Eucalyptus cluster running at a data center at HP Labs. The largest simulation cluster that we have launched successfully executes 49 simulation scenarios concurrently. This simulation cluster employs 119 VMs with 2.3GB averaged input order database.

Conclusion

Current focus is performance improvement, fault tolerance and security. On performance, we are refactoring the simulation cluster launch script to reduce the launch time, and incorporating an in-memory cache to the simulation engine so that the simulation output is first held in the cache before being flushed to the database. On fault tolerance, we are working on a recovery module that terminates and then re-launches the failed simulations reported by the Simulation Run Monitor. We are designing security improvements to enable sharing of simulation related assets among the tenant organizations.

Acknowledgement

We would like to thank the SimOp team at HP Brazil R&D (Marcelo Camozzato) and SimCloud team at HP Labs (Jun Li, Francisco Oblea and Annabelle Eseo Votaw). Our current prototype incorporates open-source software including Ptolemy II from Center for Hybrid and Embedded Software Systems (Edward A. Lee, University of California Berkeley).

References

[1] Goldratt, E. M. and Cox, J., *The Goal: a Process of Ongoing Improvement*, The North River Press, 2004.

[2] Government Technology, "Getting More from Your Printing and Imaging Investment", May 2011.

[3] Hopp, W. and Spearman, M., *Factory Physics (3rd Edition)*, McGraw-Hill/Irwin, 2007

[4] Jackson, S., Zeng, J., Mitchell, R., Lin, I-J., Comstock, S., Hoarau, E., Bellamy, R., and Dispoto, G., "On-demand digital commercial print services: a mass customization innovation", *The 2011 World Conference on Mass Customization, Personalization, and Co-Creation: Bridging Mass Customization & Open Innovation*, San Francisco, CA, November 15-19, 2011.

[5] Kipphan, H. (Ed.), *Handbook of Print Media: Technologies and Production Methods*, Springer, 2001.

[6] Zeng, J., Lin, I., Hoarau, E. and Dispoto, G. "Productivity Analysis of Print Service Providers", *Journal of Imaging Science and Technology*, 2010, 54(6), pp. 060401-(9).

[7] Zeng, J., Hoarau, E., Lin, I., Dispoto, G. and Hao, M., "Lean print manufacturing: operations simulation and analysis of digital printer", *NIP26: 26th International Conference on Digital Printing Technologies and Digital Fabrication 2010*, September 19-23, 2010, Austin, Texas, USA.

Author Biography

Eric Hoarau is a senior researcher with Hewlett-Packard Laboratories (Palo Alto, California). He received his B.S. from the University of California at Berkeley and his M.S. in Mechanical Engineering from the Massachusetts Institute of Technology. His research interests span several fields: Mechatronic systems, imaging systems, color imaging algorithms, distributed computing and system dynamics.

Jun Zeng is a senior scientist with Hewlett-Packard Laboratories. PhD and MS from Johns Hopkins University. Jun's publication includes 60 peer-reviewed papers and a co-edited book on computer aided design of integrated systems. He was a guest editor of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. He also serves as termed faculty and member of PhD dissertation committee with Duke University's Electrical and Computer Engineering Department.