

Simulation and Scheduling in the Cloud

Cloud Computing

Cloud computing relies on sharing of computer resources that are located at remote sites and are accessed and controlled by the user over the internet. The term “cloud” is used as a metaphor for “the internet.”

Cloud computing gives businesses on-demand access to a shared pool of configurable computing resources, and a variety of software. Cloud computing lets businesses do more and do it faster by letting them leverage the power of massive datacenters and IT services without having to build, manage, and maintain them. These datacenters can rapidly scale to provide 10,000's of processors for compute intensive applications.

The economic advantages of the cloud are driving businesses to adopt this framework for many core business functions including sales and customer relationship management, communications, and entire enterprise resource planning systems. Information technology savings are derived from the maintenance of a single version of software on the cloud installation, avoiding the need to install and maintain the software on multiple computers throughout the organization, as well as the cost of building and maintaining datacenters. In addition, the software applications can be accessed from anywhere, including mobile devices such as tablets.

In this paper we will consider the unique advantages that cloud computing offers for simulation and scheduling applications. Although cloud based simulation and scheduling applications offer many of the same economic benefits as other enterprise applications, they can also uniquely leverage the computational power of the cloud to dramatically improve the business value of these applications.

Simulation

Simulation modeling has become a critical technology for the 21st century. It is used by enterprises throughout the world to improve the design and operation of complex systems. Simulation technology is in a state of rapid change. The technology is becoming more powerful, easier to use, and useful for an expanding range of applications.

Simulation models are used to compare alternative designs, or optimize design parameters. For example, in a manufacturing application we might use a simulation model to compare different material handling concepts, size input buffers at each workstation, or determine the number of AGV's that are required to handle the expected transfers between workstations. Each possible combination of input values in the model produces a separate scenario that we want to compare to all other possible scenarios under consideration.

In a typical simulation project, we might have many different scenarios to compare. For example, we might have a number of different design parameters, where each parameter having a range of possible values, with a large number of resulting combinations. Each specific combination of values defines a particular scenario to be evaluated. It's not uncommon to have 100's or even 1000's of scenarios to consider. In addition, since simulation models typically have random variation as part of the model, each scenario must be replicated a number of times to get reliable estimates of performance. For example, we might have 100 scenarios to compare, where we replicate each scenario 50 times, requiring a total of 5000 replications. If each replication requires 10 minutes to execute, the entire experiment would require more than a month to execute on a single computer running 24 hours a day. In most cases this length is not acceptable, and so fewer scenarios are examined, and/or fewer replications are made. Evaluating fewer scenarios might result in potentially missing good solutions. Making fewer replications of each scenario might result in making a bad selection based on the sampling error in the model.

Cloud computing offers the ideal solution for this problem. A user can utilize the cloud to scale up to 5000 processors for the next 10 minutes, so that all 5000 replications can be run in parallel. Hence instead of waiting for more than a month to get the full results for the experiment, the user can get the completed set of results back in 10 minutes. The user only pays for this massive processing power for the 10-minute period where it is required.

Although simulation benefits from the other standard business advantages offered by the cloud, it is the ability to scale up to simultaneously perform multiple replications that make it ideally suited for running simulation experiments. Decision makers can now compare a large number of candidate scenarios without waiting long times for the results.

Scheduling

Although simulation has traditionally been applied to the design problem, it can also be used on an operational basis to generate production schedules for the factory floor. When used in this mode, simulation is a Finite Capacity Scheduler (FCS) and competes with other FCS methods such as optimization algorithms and job-at-a-time sequencers. However, simulation based FCS has a number of important advantages that make it a powerful solution for scheduling applications.

Simulation provides a simple yet flexible method for generating a finite capacity schedule for the factory floor. The basic approach with simulation based scheduling is to run the factory model using the starting state of the factory and the set of planned orders to be produced. Decision rules are incorporated into the model to make machine selection and routing decisions. The simulation constructs a schedule by simulating the flow of work through the facility and making "smart" decisions based on the scheduling rules specified.

In contrast to the simulation in manufacturing design, in scheduling applications we are dealing with deterministic data. All the features of a traditional modeling tool that help us interpret the results from a random process are of little value to us. We assume that we have complete information on the system, including routings, processing/setup times, material requirements, delivery schedules, etc. We assume away all of the randomness in the system.

As we run the real system random events typically do occur. We have machines breakdown, workers show up late or not at all, and material arrive late. These unplanned events will typically make our current schedule invalid, and in many cases this requires us to regenerate the schedule using our new information. At any point in time our schedule gives us a picture of what will happen if no unplanned events occur. In reality we will often end up with a schedule that is modified by unplanned events and is worse than the current schedule. Variability in the system typically degrades performance over time.

Risk-based Planning and Scheduling (RPS) addresses this issue by employing the stochastic model of our system to assess the robustness and quality of our schedule. By automatically adding random events to our scheduling model (e.g. breakdowns, shortages, etc.) and replicating the schedule generation process many times we can obtain measures on expected number of tardy jobs, average lateness, etc.

The risk assessment phase of the schedule generation requires multiple replications of the model to generate accurate risk measures. However, when reacting to an unplanned event on the factory floor, action must often be taken immediately and there is no time to wait for 50 or more replications of the simulation model to complete. However, the power of the cloud allows us to execute all 50 replications in the same time that it would normally take to execute a single replication on a desktop computer. Using the cloud, you can quickly scale up to the required number of processes for the few minutes that it takes to analyze the risk associated with the new schedule.

In planning and scheduling applications targeted results typically need to be distributed simultaneously to users throughout the organization. For example, each workstation might require a “work to” list that summarizes the expected workflow at the workstation, each line manager might require summary reports/dashboards that highlight key performance measures for the line, and the production manager might require separate reports/dashboards that highlight performance measures for the entire facility. The cloud provides an ideal mechanism for publishing and making these results available to users throughout the enterprise on any internet-enabled device including mobile tablets.

Conclusions

The convenience and economic benefits are driving the movement of many enterprise applications to the cloud. Simulation and Risk-based Planning and Scheduling share these same benefits, but also benefit from the ability to rapidly scale the number of compute nodes to run

many simulation replications in parallel. The heavy computational demands of simulation and Risk-based Planning and Scheduling, along with the ability to execute experiments by spreading replications across processors, makes these ideal applications for the cloud.