Refining the Oil and Gas Pipeline with Custom Workflows

There are only a few industries as renowned for massive investments in the highest end of computing technology as the oil and gas sector. High performance computing capabilities, matched with sophisticated modeling and simulation codes, and increasingly, top of the line tools to ingest, manage and analyze unprecedented amounts of diverse data mean that infrastructure complexity mounts quickly. However, despite these bleeding edge technologies backing everyday HPC, big data, and cloud-based operations, there is still a great deal of outdated manual work required to orchestrate priority, access, and outcomes, especially at the world's largest oil and gas companies where cluttered hand-managed workflows collide, lock down silos, and hinder productivity. There are streamlining solutions on the horizon—even for the highly evolved workflows that are specific to each oil and gas organization. But where is the value, what is the return on investment in change?

Produced by Tabor Custom Publishing
in conjunction with:

HPC wire

Adaptive COMPUTING

Written by: Nicole Hemsoth
Managing Editor, HPCwire
Director of Editorial Operations, Tabor Communications

For the oil and gas industry, time to solution across a wide breadth of applications is essential to global competitiveness. These performance, data ingestion and management, and overall workflow drivers are not exclusive to this market segment, but the core value of how these elements mesh can be best highlighted by drilling into the specific requirements of oil and gas.

In this market segment, especially on the HPC modeling and simulation side of operations, the competitive edge hinges on immediately securing and deploying computational resources that are the best fit, in the shortest amount of time, and aligned with an end goal. But that is just the beginning of the story--the results may not always have a defined end point. Once a potential location is spotted, immediate reanalysis or extended visualization after the model runs might be an unplanned necessity to ensure the risk of patent securing the expensive hydrocarbon recovery process will yield resources.

To add even more complexity to the time-sensitive process, it's not uncommon for seismic researchers to quickly spot a potential hydrocarbon location, only to be met with extended queue times on busy local clusters, which requires a fast search for other remote clusters, all with their own queue times, hardware capabilities, and data transfer times. Put together, it's a complicated, ever-changing picture that shifts with conditions, findings, and a range of other subtle factors. The surprising element is capturing the best computational resource and taking priority over other jobs in the event of a very promising site is not an automatic process in many cases.

From this, several questions emerge around the manual versus automated approaches to solving these access issues with time to solution at the fore:

- How can this kind of fast-paced, evolving complexity be regimented into part of a workflow or management system?
- How can oil and gas companies meet the needs for error-free, automated, de-siloed, and high performance environments without having manual interventions each step of the rocky way?
- How do oil and gas shops, which have evolved very specific workflows over time around core technological approaches (for instance, tuning infrastructure around file systems or network architecture) find a way to carry intricate, time-honed processes into generalized "narratives" of operations via a tuned workflow?

And consider this: The above questions are only reflective of the minute-to-minute demands for the high performance computing divisions, which are but one part of the overall oil and gas datacenter picture. In other words, these are just the modeling and simulation requirements, which then must feed into a host of even larger and more complex workflows that support other elements of the discovery, production, drilling, maintenance and business operations scheme. These are often separated by invisible, unspoken walls—immovable, high barriers that restrict the flow of information due to a lack of orchestration of HPC, big data, and even cloud-based resources, tools, clusters, and people.

The needs and solutions are clear in theory; deliver an approach that helps users with highly diverse workloads unify around a comprehensive workflow that allows flexibility and agility—the capability to secure remote or local resources, to do so with an automated way to prioritize this selection based on hardware capabilities, administrative policies, and job priority. Further, there is the need to take into account a bevy of other requirements including data transfer times, where to place and send the processed data, how to retain it for other uses elsewhere in the oil and gas data pipeline and more.

Again, all of this is often handled manually, adding extraordinary weight and complexity for tasks that are reliant on arriving at the quickest and most accurate result possible. With so much potential for complication, however, it seems that this is not a process that can leverage automation. But according to Adaptive Computing's Director of Professional Services, Paul Anderson, it is possible and practical for administrators

to create an environment where users can automate cluster selection, prioritization and data movement between remote systems. It's also possible to add to the orchestration layer allowing for new things to be done with potentially valuable data on the fly (including reanalysis, enhanced resolution for promising elements, visualization, etc.). The key to all of this is contained in carefully tuned workflows that understand what oil and gas shops need, what they're missing with their existing manual operations, and a way to tailor with these and larger organizational IT needs in mind. This also means that a company's HPC, big data, and cloud or distributed environments are no longer silos meant to house more silos. They are active engines of competitiveness, working in harmony across a range of jobs, departments and more specific workflows.

"There is a strong desire to reduce all of those manual touch points," explained Anderson. "Consider an 'aware grid' of resources, not the old concept of grid, but rather a grid that consists of many disparate clusters that may or may not be tightly managed or aligned politically, so that must be built around policies. For the HPC side in particular, it solves the problem of having the disparate clusters as so many oil and gas companies do. Now, these all have different administrators, but need to be able to interoperate, automatically placing the highest priority work on the most capable resources, the next priority with the next best fit, and so on." He describes how creating custom workflows work toward this solution and extends as a concept across the entire oil and gas datacenter spectrum, integrating once-siloed elements into a swift, clean machine free from the error, time constraints, and lack of communication across the spectrum that are inherent to manual processes.

Aside from adding automation and removing manual processes to lead to faster hydrocarbon discovery, on the rare occasion that there is more compute than data available, oil and gas shops want to make sure they are leveraging all of their resources—in other words, keeping throughput high and hungry CPUs busy.

Discovering new sources of hydrocarbons is the driving force for HPC clusters in oil and gas shops but there are a multitude of

other applications and simulations pushing well management initiatives, doing work on the drilling side with computer aided engineering, as well as various chemistry, geology, computational fluid dynamics and visualization workloads. What all of this means is that oil and gas datacenters have top of the line high performance computing environments, but also have to be able to ingest, move, and analyze a lot of data after the HPC systems have done their time-critical crunching.

This opens the door for a conversation about the role of big data—a concept that is not separated from the work being done on the HPC clusters, but that has its own challenges that extend past the infrastructure divide between HPC, big data, and cloud computing. The I/O bottlenecks, varied as they are, are pushed through the chain—from simulation to reanalysis to long term storage—big data in oil and gas is another facet of the workflow model that must be considered, especially in the compute and data-intensive field of oil and gas.

A recent IDC report that drew from survey data from oil and gas industry IT end users found that their big data challenges crossed borders between different segments of their operations. On the exploration side, their concerns were around dealing with petabyte-scale seismic datasets and the HPC systems designed to tackle these issues. But added to this, were problems on the production, trading and rush to resources sides. In other words, the master challenge is mingling these various data element, often in diverse formats (structured, unstructured, etc.) and bringing real-time risk analysis, trade simulations, and seismic data acquisition into the fold as quickly as possible. The ultimate goal for these companies, of course is to boost hydrocarbon discovery, production, reduce risk and act on time-critical information using a mesh of all of these data sources.

This same report concluded that when it came to the big picture view of what challenges are ahead, a lack of sufficient technology and analytical skills, confusion about technology options, and an over-emphasis on analytics at the expense of pre- and post-analytics processes were top concerns. What's interesting here is that these are all components of a larger big data problem that can be addressed rather comprehensively through thoughtful workflow orchestration. Done correctly and with a tailored interface approach that alleviates the over-hiring and over-burdening of disparate tools, it helps address all three of the issues—and with sufficient customization around a particular oil and gas

company's internal operations on the HPC, production and business management systems, it can weave together formerly siloed components and start pushing past roadblocks.

When taking a bird's eye view over the oil and gas datacenter landscape, one would observe scattered, distinct points or "houses" of specific operations (exploration, drilling, production, risk analysis, maintenance, etc.). This is not necessarily a geographical metaphor, but rather one that reflects the siloed nature of operations. A comprehensive, finely-tuned workflow, however, has the effect of bringing these points together, forming a community of data that interacts as needed, works together to solve problems, and cooperates on a mission-critical (or even post-analytical) basis.

No matter how one considers the different processing models according to broad terms (cloud computing, big data, high performance computing), the actual challenge is its own model—and that is defined by workflow. This singular concept takes into account the combination of those processing capabilities and offers a holistic solution to navigating the many barriers, silos, and lapses in value that result from thinking inside the boxes of divisional datacenters.

On this note, consider the results of Adaptive Computing's recent survey that sought to provide unity among the processing approaches. Among the 400 respondents from across the enterprise spectrum, they found that 91 percent agreed that it would take a combined thrust of all three computing paradigms—HPC, big data and cloud computing (private/public). The only challenge that users would face then would be in software that could mesh the fine-tuned, highly optimized custom workflow around the scheduling, management and orchestration layers to reach across the divides.

Adaptive Computing also found that a staggering 84 percent of those surveyed were reliant on the time-consuming, error-prone process of analysis of their internal data using manual methods. This leads to a

continuation of the siloed approach to meshing disparate data sources and departments, in adding to detracting value, when time-sensitive, business-critical decisions could be the key to competitive advantage. However, the survey also found that respondents recognized the flaw of this approach as 90 percent said that having a customized analytical process built into a grand workflow would speed decision and add to overall value and efficiency. They accordingly predicted that for those industries they focused on (generally with a blend of big data, HPC, and cloud infrastructure requirements, including oil and gas) the demand for customized workflow automation would continue to grow.

With 72 percent of Adaptive Computing's respondents agreeing that a tuned approach to building workflow creates a bridge to competitive advantage, the question is no longer "is workflow critical" but rather becomes one around rationale, implementation, and understanding the common barriers and challenges across oil and gas silos.

"Sensing" the Big Data Picture

There is no single, unified picture of oil and gas data as the needs and mechanisms for generating the petabytes of actionable information stretch from the HPC side (the diverse models and simulations of new and existing sites, for example) to the practical operational enterprise end. What does become clear when surveying the overall oil and gas data landscape, however, is the need to pull all of the disparate dots together for an unobstructed view.

These "dots" that comprise the overall portrait of the oil and gas datacenter landscape, when viewed in high resolution, are defined data points, often from small sensors that feed a growing stream of data. This stream is so complex, alive with sensors adding specific information into the larger flow, that the real challenge is one of meshing, filtering and analyzing instead of simply gathering sheer compute horsepower. In many senses, in oil and gas and beyond, the compute is the simple part. It's this stream and its flow throughout the larger datacenter landscape that becomes the hurdle.

Although it is not a new phenomenon, the oil and gas industry is notoriously reliant on a great number of sensors and a constant stream of petabyte-level machine data that must be processed and analyzed,

oftentimes in a time-critical manner. Sensor data in this industry and many others is expected to grow unchecked, oftentimes outpacing the management and workflow tools designed to meet it. From sensors embedded in subsurface wells to those that monitor the incredible amount of equipment, oil and gas is dominated by a big data (and high performance computing) challenge that is ever more complicated due to these nearly endless data wells.

"The last decade was devoted to big data challenges. But this is the decade of sensing."

- Peter Bruening

One of the public industry practitioner voices, Peter Bruening, general manager technology management and architecture at Chevron IT, claimed earlier this year, in the oil and gas field, "the last decade was devoted to big data challenges. But this is the decade of sensing." Interestingly, these two areas overlap significantly as sensor and machine data increase datacenter complexity for a range of oil and gas operations. "Finding and extracting oil and gas from the subsurface has always been a data-driven exercise. With advances in hardware and software technologies and new sensing technologies, improving resolution within the reservoir is critical."

While one can make an argument that this new decade of sensing in oil and gas involves the need for emerging technologies across the IT spectrum, the one area of incredible value—and one that doesn't always get the attention of the larger hardware architecture, storage and other decisions—of being able to effectively manage and navigate this data in the context of a comprehensive workflow approach. Without an overarching internal IT "narrative" that orchestrates and drives actual business value from these many data sources (all of which have different prioritizations, usage patterns, and time sensitivity), data-based value can fall through the cracks. That narrative is the workflow—and that workflow and its integration into the whole is the critical piece of solving both the HPC and big data challenges inside even the largest oil and gas datacenters.

According to Adaptive CEO, Rob Clyde, oil and gas companies, as well as others across the larger spectrum of advanced enterprise, rely on

"No longer can CEOs make business decisions based on hunches and what they can physically extract from industry research. "

- Rob Clyde

collected data and simulation results for a competitive edge. "No longer can CEOs make business decisions based on hunches and what they can physically extract from industry research. Businesses are turning to their CIOs and data scientists to leverage big data to help predict likely outcomes and make data-driven decisions. Unfortunately, the increase in compute and data intensive workflows creates a logjam within the datacenter."

Finding the narrative and building workflows means being able to harness the multiple data sources and types—from petabytes of structured to unstructured information fed by the sensor networks that are so prominent, but also integrating and filtering these data with a multitude of other sources. From reservoir models and simulations to the practical business data that drives the non-HPC side of the oil and gas operations, the narrative arc must be defined, refined, and mapped into workflow—all elements that translate to the need for a sophisticated workload and workflow management strategy that offers the big picture view for the big data challenges at departmental and organizational levels.

Of course, one other aspect of meshing into flexible but tailored workflows is ease of use and manageability. "We don't want to turn our geophysicists into computer scientists," said BP's manager of high performance computing, Keith Gray, at a recent HPC event. "We want our scientists to be able to try out new ideas, but we have to hide some of the complexity and make it easy to effectively optimize and experiment."

Pipelining a Refined Workflow

While we've touched on the larger narrative around tailoring workflows that integrate the many needs on the HPC, big data, and distributed computing sides, it's helpful to put this into context.

A true, unified workflow approach, like the work of any orchestra, involves a number of independent instruments. The volume of one may be louder for certain pieces, silent for others, and operate as a swift current in the background in other cases. As in music, so it is in addressing large-scale oil and gas computational and data demands.

By pulling together onto the same stage the many engines of compute—from the traditional datacenter for operational needs, the high performance computing clusters (local and remote), the mechanisms to address big data demands (database, data warehouses, resources

for handling unstructured data, etc.), to the cloud-based environments (public, private and/or hybrid)—the symphony begins to emerge.

This act of consolidation under a comprehensive, uniquely tailored workflow means that many efficiency, performance, utilization and automatic advantages can be realized simply through shattering the silos that traditionally barricade the specific systems and data that has wider organizational value. From this single stage, intelligent allocation of resources based on time-critical findings (which themselves can be automated and set into policy as priority items for pending analysis) is possible. So too is the ability to harness automatic provisioning of the right type of resources for the right workload—sending the high probability discovery zone to the highest performing free cluster (or auto-bumping a queue for primary position), for example. This enables collaboration inside that "aware grid" concept as well with data being able to filter where it is needed for greater cross-department or cluster analysis and insight.

On the analytical platform stage this means similar efficiencies and value gains. Smart scheduling driven by customized policies might mean the slight, but winning edge to secure the hydrocarbon patent on a site. Having everything bundled onto the same platform with power management, job prioritization, remote visualization as well as the integrated hardware options (from bare metal, virtual machines to cloud), it's clear that it all works when it works together.

Adaptive Computing Puts It Together

Adaptive has over a decade of expertise in helping some of the world's leading organizations manage their infrastructure. From clouds to technical computing clusters and big data operations, Adaptive is leading the way in intelligent automation, or what they refer to as Big Workflow. The company's solutions are proven at scale and with their

core product, Moab, at the heart of their tailored workflow approach for oil and gas, the company is helping the world's energy leaders rethink what's possible with customized integration, streamlining the analysis process, and removing of the many logjams that are at the heart of mission-critical datacenter flow.

While current solutions solve big data challenges with just cloud or just HPC, Adaptive unifies all available resources – including bare metal and virtual machines, technical computing environments (e.g., HPC), cloud (public, private and hybrid) and even agnostic platforms that span multiple environments, such as OpenStack – as a single ecosystem that adapts as workloads demand.

Big Workflow orchestrates and optimizes the analysis process to increase throughput and productivity, and reduce cost, complexity and errors. Even with big data challenges, the data center can still guarantee services that ensure SLAs, maximize uptime and prove services were delivered and resources were allocated fairly.

Adaptive Computing's legacy Moab HPC Suite and Moab Cloud Suite are an integral part of the Big Workflow solution. By adding workflow capabilities to the Moab family, dynamic scheduling, provisioning, and management of multi-step/multi-application services across HPC, Cloud and Big Data environments become possible, shortening the time to discovery.