

O que vem a seguir para big data em processos de manufatura

Big Data armazenado em historiadores de processo por décadas tem sido freqüentemente subutilizado - novas análises avançadas simplificam o acesso a soluções, acelerando o tempo para compreensão

Por Michael Risse

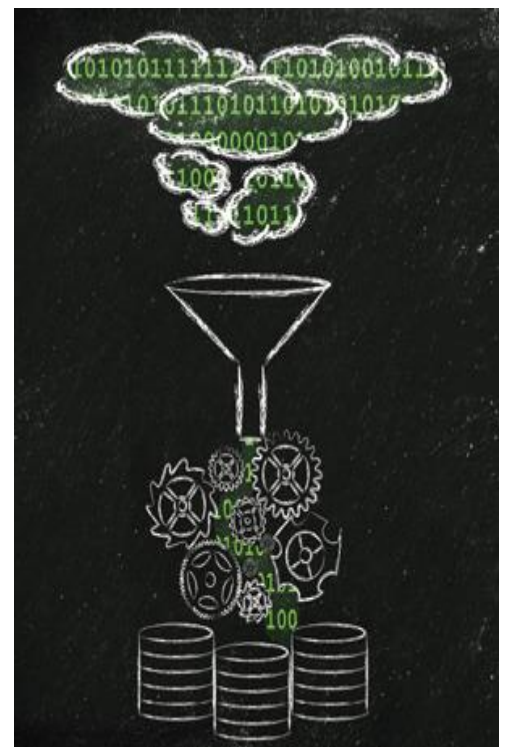
De acordo com a reportagem de capa da história do "Big Data Analytics" na edição de novembro / dezembro de 2015 da InTech, "Empresas de manufatura têm o benefício e a maldição das taxas de atualização de tecnologia, tipicamente medidas em décadas, e processos de seleção de tecnologia normalmente medidos em meses ou anos. Isso é um benefício, porque a tecnologia que compramos é construída para durar e proporciona um retorno sobre o investimento muito além do de outras indústrias. É também uma maldição, uma vez que não obtemos novas tecnologias com muita frequência, quando temos a chance, muitas vezes mordemos mais do que podemos mastigar, ficamos paralisados pelo medo do desconhecido ou procuramos o novo brinquedo brilhante. "

Como resultado, embora o Big Data possa ser consolidado e aceito em ambientes menos restritos, ele só agora está deixando seu estágio introdutório para a maioria das organizações de manufatura de processos. Os primeiros adotantes têm implantações e alcançaram sucesso, e muitas empresas estão avaliando ou incluindo novos projetos de análise em seus planejamentos, mas o termo mais popular em artigos sobre Big Data e a indústria de manufatura é "oportunidade". Então, a oportunidade está lá fora, mas ainda há muito a fazer:

- dados (armazenados principalmente em historiadores de processo) e pouca informação
- trabalho árduo, em vez de fácil acesso à inovação
- tempo de inatividade não planejado
- tempo gasto na limpeza e modelagem de dados antes que a análise possa começar

Pergunte aos engenheiros de processo qual é a ferramenta de análise usada com mais frequência e, para a maioria, a resposta é um software de 30 anos, de usuário único e um software cliente peso pesado chamado de planilha. Pergunte aos gerentes sobre seus ambientes de dados e eles responderão em termos de historiadores de processo, sistemas de gerenciamento de ativos e outras ferramentas.

Portanto, o Big Data, para muitas empresas, ainda está lá fora como uma promessa, aguardando a interseção entre inovação e oportunidade de trazer novos discernimentos e melhorar os resultados de produção para fábricas e organizações. E nenhuma indústria tem mais necessidade de criar valor a partir de big data. As fábricas produzem duas vezes mais dados do que qualquer outro mercado vertical, de acordo com a McKinsey and Company, do relatório que lançou o Big Data de

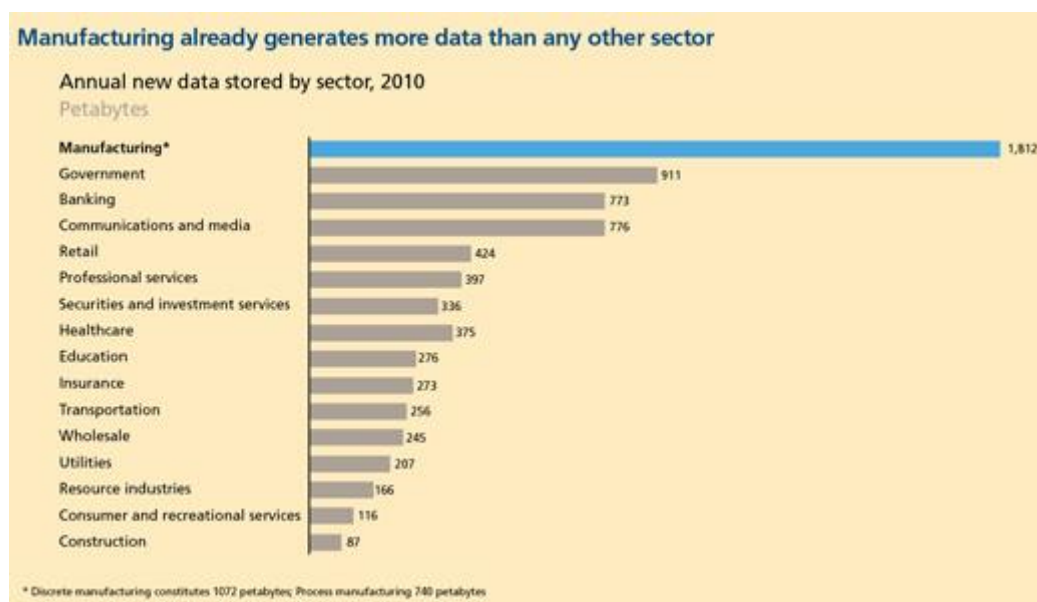


conscientização e publicidade em 2011 (figura 1). Com essa quantidade de dados, surge uma oportunidade correspondente de melhoria, no montante de US \$ 50 bilhões somente no setor de exploração, perfuração e produção de petróleo e gás (figura 2).

A ajuda está a caminho, à medida que as ofertas e expectativas de big data forem alteradas para atender melhor aos requisitos de processos e manufatura. O mercado passou de um modelo T, "qualquer cor desde que seja preto", para uma variedade de tamanhos e formas para atender às necessidades dos clientes. A interface ou a experiência do usuário com muitos aplicativos de Big Data, por exemplo, não é mais uma experiência baseada em tentativas.

Na verdade, um novo conjunto de temas relacionados a Big Data está surgindo, à medida que mais empresas estão abertas e interessadas em novas experiências de análises avançadas. Se a evolução das plantas é medida em décadas, e a conscientização e a inovação de Big Data estão se aproximando de uma década de investimento, o tempo deve estar maduro para a implementação de novas tecnologias.

Este artigo discutirá quatro das expectativas associadas a uma moderna experiência de Big Data em empresas de processos de manufatura. O cumprimento dessas expectativas resultará em uma experiência mais aprimorada e de alto nível, aproveitando os recursos de gerenciamento de dados, armazenamento e capacidade de análise agora disponíveis para melhorar os resultados de produção e de negócios.



Fonte: McKinsey

Figura 1. Nenhuma indústria produz mais Big Data do que em manufatura, criando uma oportunidade enorme de melhorias através de análises avançadas.



Source: McKinsey

Figura 2. O melhor uso do Big Data apresenta uma oportunidade de \$50 bilhões em exploração, perfuração e produção de óleo e gás, com centenas de bilhões de dólares de oportunidade em outras indústrias de processo.

Contexto através de fontes de dados

Os três "Vs" de grande velocidade de dados, variedade e volume são bem conhecidos e fazem parte da definição de big data por mais tempo que o termo big data. Mas dos três, um deles é muito mais um problema em processos de manufatura do que os outros dois.

A questão não é o volume, porque historiadores de processo e outras fontes têm muitos dados armazenados e disponíveis para análise. Da mesma forma, a velocidade tem várias soluções com redes de alta capacidade e taxas de transferência mais rápidas para historiadores. Variedade, no entanto, apresenta o maior desafio para a análise avançada, e novas soluções de Big Data estão trabalhando para resolvê-la.

O desafio com a variedade é que a maioria dos sensores de plantas existentes suporta apenas um conjunto limitado de dados de tempo, valor e talvez estado. Portanto, o tipo de dados mais típico na fabricação, a sequência do dado ao longo do tempo, é, por definição, separado de outras fontes de dados, que armazenam o contexto relacionado. Assim, antes que qualquer investigação possa ocorrer, um engenheiro tem que lidar com a questão da variedade - em particular, a integração de sinais analógicos contínuos com os conjuntos de dados relacionais ou discretos armazenados em outro banco de dados.

Essa integração, geralmente feita à mão, é um dos maiores impulsionadores do uso de planilhas nas organizações. Mesmo organizações com modelos de informação em soluções de inteligência de fabricação empresarial (EMI) têm que confiar em planilhas para análises ad hoc, porque se um conjunto de dados não é integrado e modelado no EMI, e raramente é, então ele volta à estaca zero e interpolação, alinhamento e tempo correspondente são carregados manualmente.

Existem muitos termos para o alinhamento e integração de tipos de dados diferentes do setor. Mistura de dados, harmonização de dados e fusão de dados são três exemplos - mas para empresas de processos de manufatura, o termo normalmente usado é a contextualização, que adiciona contexto ou informação sobre os dados como atributos de um intervalo de tempo.

Esses dados podem ser armazenados em outra fonte, por exemplo, os períodos de tempo definidos por um estágio em batelada ou estado do ativo em um sistema de execução de manufatura (MES) ou sistema de gerenciamento de manutenção computadorizado. O contexto pode estar dentro dos próprios dados da série temporal, definidos por quando uma leitura está acima ou abaixo de um determinado limite. Ou pode ser simplesmente períodos de tempo de interesse, por exemplo, quando

um sinal "se parece com isso", com contexto criado para definir quando uma forma ou padrão está presente em um sinal.

Em cada um desses casos, o contexto é adicionado para identificar os períodos de tempo de interesse. Uma vez identificados, esses períodos de tempo podem ser combinados para criar um novo conjunto de períodos de tempo descrevendo um conjunto de dados exato e multidimensional para análise (figura 3). Com novos recursos de big data, não é necessário haver limites para a profundidade ou o número de camadas "empilhadas" necessárias, até 15 ou mais camadas sequenciais de critérios em alguns casos. Com a maioria dos esforços de análise exigindo a integração de dados de cinco a sete fontes diferentes, essa é uma vantagem crítica em relação às abordagens atuais.

Diferentemente dos tipos de dados, em particular das séries temporais e fontes de dados relacionais, a análise avançada pode começar lentamente, exigindo mapeamento manual extenso de tipos de dados, sem mencionar a limpeza de dados e outros aspectos da preparação de dados. Mas, com inovações recentes, as tecnologias subjacentes de Big Data fornecem esse tipo de conectividade, alinhamento e mapeamento de dados para acelerar a definição e a modelagem de operações complexas. O que já foi o trabalho de um mês de programadores e interfaces de programação de aplicativos (APIs) agora pode ser recursos que qualquer engenheiro de processo pode implementar em minutos.

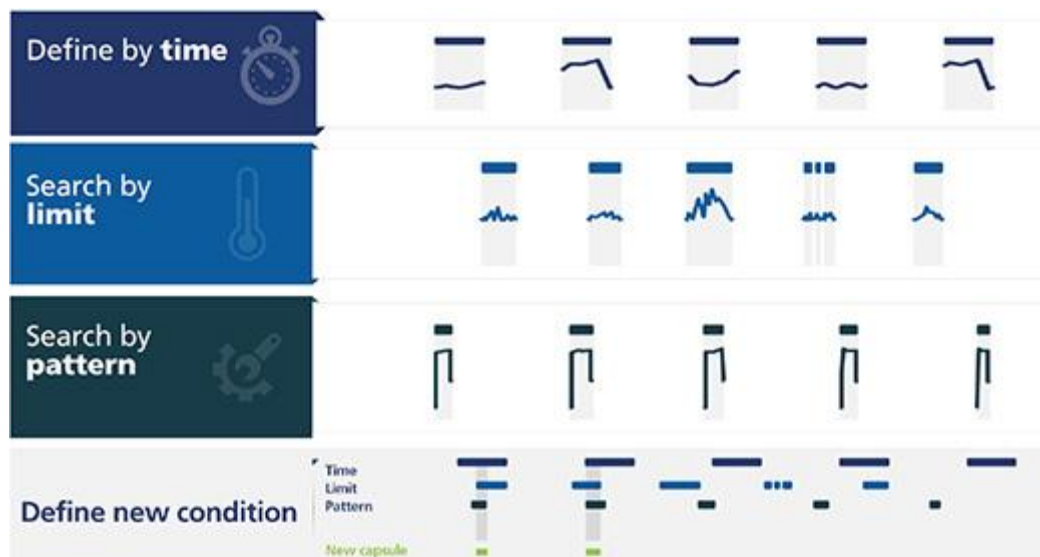


Figura 3. Usando cápsulas "Seeq", engenheiros podem combinar períodos de tempo para criar um novo conjunto de períodos de tempo descrevendo um conjunto de dados multidimensional exato para análise.

Entregando auto-atendimento(self-service)

Nos estágios iniciais, Big Data significava que os programadores escreviam código para mapear a análise de um grande conjunto de dados para um cluster de nós de computação e, em seguida, para reduzir a saída dos nós em um resumo consolidado. O algoritmo MapReduce, que definiu esse modelo de programação, foi aberto pelo Google em 2004 e se tornou a base do Hadoop, que mais tarde foi comercializado por fornecedores como a Hortonworks.

Ao mesmo tempo, o Google não expôs o API do MapReduce aos usuários como a interface do mecanismo de pesquisa. Em vez disso, eles apresentaram a funcionalidade do algoritmo em uma página da Web simples, na qual qualquer cliente poderia simplesmente pesquisar o que queria, digitando apenas dados em inglês simples.

Essa abordagem para envolver funcionalidades complexas em interfaces fáceis de usar é uma experiência comum em nossas vidas como consumidores, e essas mesmas abordagens estão sendo adotadas agora por ofertas analíticas para engenheiros em processos de fabricação.

Por exemplo, a capacidade de "pesquisar como o Google" em todas as tags de um navegador ou outro sistema de armazenamento de big data agora está disponível em algum software de análise avançada. Outros recursos que tornam as inovações de big data mais facilmente acessíveis são entregues de forma semelhante. Isso habilita que os engenheiros (nunca antes permitido) trabalhem em um nível de aplicativo com benefícios de produtividade, capacitação, interação e facilidade de uso.

A capacidade de transformar a complexa programação da ciência de dados em recursos facilmente usados pelos engenheiros é uma capacidade crítica das ofertas analíticas avançadas. Embora tenha havido muita empolgação em relação aos cientistas de dados e seu papel na melhoria dos resultados de produção, como o artigo "O mais sexy trabalho do século" da Harvard Business Review em 2012, artigos mais recentes e anedotas de usuários finais contam uma história diferente.

A questão é que, embora os cientistas de dados conheçam seus algoritmos, eles não conhecem os processos e o contexto das plantas. Tem havido uma quantidade mais recente de artigos sobre a necessidade de tradutores de dados ou ligações de dados entre as equipes de engenharia e de ciência de dados. Mas tudo isso pode ser evitado se os vendedores simplesmente fecharem a lacuna e levarem a inovação da ciência de dados aos engenheiros, criando recursos que permitam o autoatendimento, análises avançadas para engenheiros e outros especialistas no assunto (figura 4).

A estratégia não pode terminar com os engenheiros, não obstante, até porque o autoatendimento é o que os engenheiros vêm fazendo há 30 anos com planilhas. Portanto, a nova geração de analítica avançada para Big Data deve capacitar equipes e redes de funcionários que dependem de conhecimentos de produção e operações dentro da organização. Se isso soa como linguagem sofisticada para apresentações e relatórios, há uma diferença crítica.

A principal mudança é manter uma conexão entre a análise criada e o conjunto de dados subjacente, para que os usuários possam clicar e acessar os dados subjacentes. Essas ofertas avançadas de análise podem ser usadas para produzir não apenas imagens de dados em visualizações, mas também podem fornecer acesso às análises e fontes que geraram as saídas. Engenheiros, equipes, gerentes e organizações podem, portanto, usar esses novos recursos para permitir a distribuição de benefícios por toda a fábrica e uma empresa.

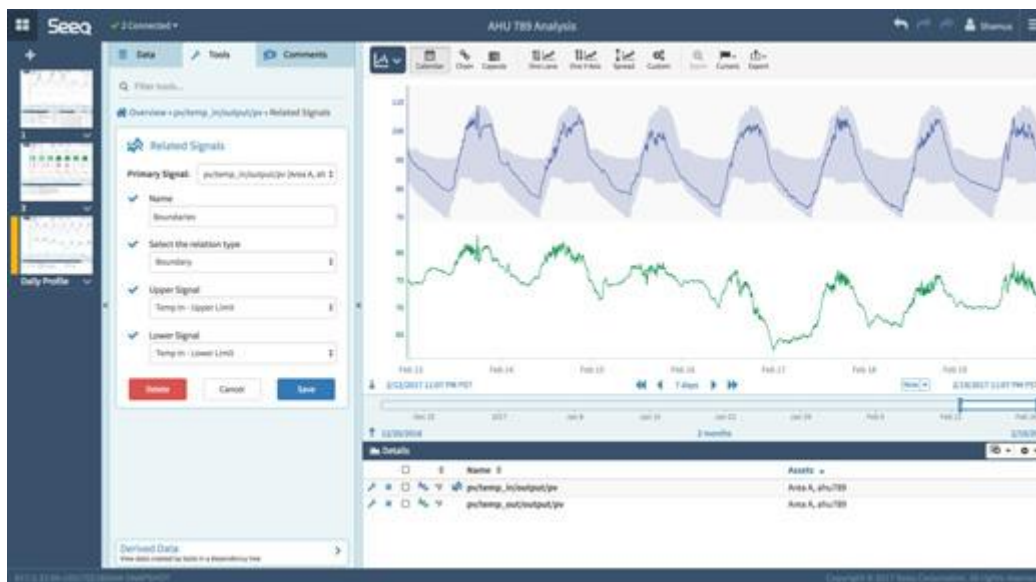


Figura 4. Software de análises avançadas disponibilizando recursos de auto-uso para engenheiros criarem varias visualizações de dados.

Revolução em desenvolvimento

"Nós tendemos a superestimar o efeito de uma tecnologia a curto prazo e subestimar o efeito a longo prazo", observou Roy Amara, ex-presidente do Instituto para o Futuro. Se Big Data não é novo, então certamente a nuvem também não é. Algumas escolhas populares para o início da computação em

nuvem incluem a introdução do primeiro grande aplicativo "SaaS" (Salesforce) em 1999; a introdução da AWS pela Amazon em 2002 e, em seguida, S3 e EC2 em 2006; e quando a competição de computação em nuvem ficou interessante com as apresentações da plataforma de nuvem da Microsoft e do Google em 2008. De modo conservador, como o Big Data, há uma década de história e inovação para alavancar a análise avançada.

Para ser claro, a nuvem não é um requisito para implementações de Big Data. Se alguém disser que uma implantação em nuvem é necessária para análises avançadas com Big Data, é provável que ele seja um vendedor de nuvem que esteja buscando atendimento de cotas. Em nossa experiência, os dados estão bem, onde quer que estejam; é a análise que precisa de atenção.

Dito isso, há muitas boas razões para alavancar a computação em nuvem, e certamente tem um impulso a seu favor, embora seja impossível generalizar os benefícios propostos em comparação com os custos específicos de cada organização. Os custos podem incluir segurança, governança de dados, tempo para obter aprovação e custo real de implantação. Se você unir a nuvem, ou não, com inovações no gerenciamento de dados de código aberto e nos investimentos em plataformas de nuvem da Internet das Coisas (IIoT), o resultado é uma série de elementos tentadores para a implantação de soluções de Big Data. Considere que em apenas 90 dias no final de 2017 e no início de 2018, oito empresas receberam mais de US \$ 250 milhões em capital de investimento para armazenamento de dados de código aberto, plataforma de nuvem IIoT e analítica da IIoT - e temos uma noção do interesse em análises avançadas.

O que isso significa é que o modelo atual de armazenamento de big data em processos de fabricação - que é local, baseado em histórico, e que o proprietário está passando por uma transição, permitirá novas alternativas para como e onde as análises avançadas são executadas. O novo modelo pode ser um "data lake" (estratégia para armazenamento de dados) para agregação de dados, no local ou na nuvem, ou uma solução IIoT abrangente, como uma plataforma de armazenamento de dados de próxima geração. No mínimo, os atuais fornecedores de historiadores de processo precisam introduzir desenvolvimentos com caminho seguro para dados mudando de locais para na nuvem.

Como fornecedores de soluções analíticas avançadas, aqui estão alguns exemplos do que isso significa para os usuários finais com os quais trabalhamos diariamente. Há três anos, tínhamos solicitações de clientes para engenheiros de vendas visitá-los no local para trabalhar com seus conjuntos de dados no local e bloqueados por acesso remoto. Hoje, em contraste, temos clientes compartilhando cinco anos de desenvolvimentos que integram ofertas baseadas em nuvem e, especificamente, pedindo contexto em algumas ofertas de código aberto, como Hortonworks e InfluxData. A suposição de que os dados nunca podem ou nunca irão migrar para a nuvem é cada vez mais incomum e mudou muito rapidamente nos processos de fabricação nos últimos anos.

Os serviços e os modelos de implantação não apenas mudarão, mas novos fornecedores entrarão no mercado de gerenciamento e análise de dados. Em particular, Microsoft, Google e Amazon possuem plataformas de nuvem e serviços de armazenamento de dados de séries temporais - Cosmos DB, Bigtable e Dynamo, respectivamente. Todos os três adquiriram empresas da plataforma IIoT (Solair, Xively e 2lemetry, respectivamente) para desenvolver suas soluções de processos de fabricação.

GE com Predix, Siemens com MindSphere e PTC e ThingWorx podem ter mais conhecimento de domínio industrial, e OSIsoft começa com a melhor base de clientes e oferta mais completa no local, mas a revolução na implantação oferece flexibilidade na implementação e níveis de serviço na forma como as empresas licenciam e executam soluções avançadas de análise.

Análises avançadas

A indústria de manufatura seria bem servida por um dicionário de marketing para definir o grande número de chavões, eras de tecnologia e "marketectures" (arquiteturas de marketing executadas no PowerPoint). Nesse dicionário de termos, o Big Data, é claro, seria incluído em "B", mas seria precedido por "análise". Análises: descritiva, preditiva, diagnóstica, interativa, prescritiva, básica, em tempo real, histórica, causa raiz e assim por diante. A análise agora é tão usada que a palavra perdeu

significado específico em um histórico de 30 anos de planilhas e em um papel de 20 anos com o termo comercializado para "desenvolvimentos úteis".

Mas agora, o papel da analítica precisa mudar para endereçar o volume, os desafios e as oportunidades associadas a volumes de dados massivos, variedade, etc. Para o resgate, surge uma nova entrada para o dicionário, "análises avançadas". Assim como adicionar "inteligente" a um substantivo denota algo com sensores para serviços de telemetria e monitoramento remoto (por exemplo, refrigerador inteligente, estacionamento inteligente), adicionar "avançadas" a "análise" traz a análise para uma estrutura moderna para os desafios de hoje.

Especificamente, as análises avançadas falam sobre a inclusão de tecnologias de computação cognitiva nas ofertas de visualização e cálculo que foram usadas durante anos para acelerar os desenvolvimentos para os usuários finais. Como a McKinsey and Company define soluções analíticas avançadas: "Essas [soluções analíticas avançadas] - que facilitam o acesso a dados de múltiplas fontes de dados, juntamente com algoritmos avançados de modelagem e abordagens de visualização fáceis de usar - poderiam finalmente dar aos fabricantes novas maneiras de controlar e otimizar todos os processos em todas as suas operações."

O que aconteceu é que os fornecedores reconheceram que há muitos dados de muitos sensores, e potencialmente de muitos tipos para uma pessoa simplesmente resolver problemas manualmente com uma planilha. Portanto, através da introdução de "machine learning" ou outras técnicas de análise, os esforços de um engenheiro precisam ser acelerados ao buscar correlações, agrupamento ou qualquer agulha dentro do palheiro de dados do processo. Com esses recursos construídos em modelos multidimensionais e habilitados pela montagem de dados de diferentes origens, os engenheiros ganham uma ordem de grandeza em recursos analíticos, como mover da caneta e do papel para a planilha.

Essas inovações avançadas de análise não substituem a caixa preta pelo expertise dos engenheiros, mas são um complemento e um acelerador de seus conhecimentos especializados, com transparência para os algoritmos subjacentes para apoiar uma abordagem de primeiros princípios às investigações. Desta forma, é um próximo passo natural na história dos processos estatísticos e de controle, ao invés de uma abordagem de ciência de dados para investigações. Ao mesmo tempo, a análise avançada reconhece o caminho para que desenvolvimentos mais rápidos utilizem inovações em áreas adjacentes para abordar o escopo de dados disponíveis para investigação.

Longo caminho a perseguir

À medida que os fabricantes de processos encontram uma oportunidade quando suas melhorias de fábrica ou investimentos de capital permitem a introdução de novos recursos avançados de análise, eles encontrarão um novo conjunto de recursos e experiências, muito distante dos primórdios da era do Big Data. A aplicação dessas soluções avançadas de análise a grandes volumes de dados aumentará a experiência do usuário acelerando o caminho para a implementação.

Contextualização, autoatendimento para organizações, novas opções de plataformas e recursos analíticos avançados se beneficiam de anos de investimento de fornecedores e realimentação de early adopters (novos desenvolvedores). A única coisa que eles não garantem, no entanto, é o sucesso na última milha de qualquer projeto de análise, grande ou pequeno, que é o pouso ou a adoção de novas percepções sobre uma cultura conservadora e questionadora. Esse, sempre, é o maior obstáculo para qualquer projeto de análise, no qual nenhuma quantidade de inovação tecnológica pode se esgotar.

Mas ao adotar uma cultura de análise e a inovação agora disponível, as organizações podem aproveitar as oportunidades para criar valor a partir de seu Big Data, permitindo que permaneçam competitivos

Considerações

- Os avanços na tecnologia permitem o uso mais eficaz de Big Data, que existe há décadas nas indústrias de processo.
- Muitas empresas do setor de processo usam ferramentas antigas, geralmente planilhas, para obter insights de seus big data, com resultados insatisfatórios.
- As soluções analíticas avançadas abordam as deficiências de uma planilha em quatro áreas: contexto, autoatendimento, implantação e facilidade de uso.

Sobre o Autor

Michael Risse é vice-presidente e CMO da Seeq Corporation, uma empresa que desenvolve aplicativos analíticos avançados para engenheiros e analistas para acelerar percepções dos dados de processos de fabricação. Anteriormente era consultor em aplicativos e plataformas de big data e, antes disso, trabalhou com a Microsoft Corporation por 20 anos. Risse é um graduado da Universidade de Wisconsin em Madison.

Artigo traduzido por Tomé Guerra para a ISA São Paulo Section e republicado com permissão da ISA, Copyright © 2018, todos os direitos reservados. Este artigo foi escrito pelo autor acima e publicado originalmente na revista InTech Online de Mai-Jun / 2018 em <https://www.isa.org/intech/20180601/>. A ISA não se responsabiliza por erros de tradução neste artigo.