

APLICAÇÃO DE BIG DATA PARA VISIBILIDADE NA OPERAÇÃO INDUSTRIAL*

André Dias Leite¹
Giuliano Felipe Soares²
Julio César de Oliveira Melo³

Resumo

A resolução de problemas na operação industrial é algo complexo e precisa ser baseada em dados, muitas vezes diversos e não estruturados. Este trabalho traz uma abordagem em *big data* para dar visibilidade dos problemas em sistemas industriais, trazendo insights que permitam sair da intuição para assertividade, contribuindo, assim, para redução do tempo de detecção e resolução de problemas.

Palavras-chave: Big Data; TI Industrial; Gestão da Operação.

APPLYING BIG DATA TO IMPROVE VISIBILITY IN INDUSTRIAL OPERATIONS

Abstract

Troubleshooting in industrial operations is a complex task that needs to be based on data, many times diverse and not structured. This work shows how to use big data to gain insights and move from intuition to a structured approach, contributing to reduce time to detect and resolve issues.

Keywords: Big Data; Industrial IT; Operations Management.

¹ Gerente, Accenture, Belo Horizonte, MG, Brasil.

² Gerente Sênior, Accenture, Belo Horizonte, MG, Brasil.

³ Analista, Accenture, São Paulo, SP, Brasil..

1 INTRODUÇÃO

Neste artigo será apresentada uma solução de *big data* e sua aplicação com o objetivo de melhorar a operação industrial usando os dados gerados por sistemas de informação e automação.

Big data é um termo da tecnologia da informação que se trata de processar e armazenar grandes volumes de dados, estruturados ou não estruturados, de diversas origens e em alta velocidade. As três principais características que definem um sistema de *big data* são: volume, velocidade e variedade. O volume representa a quantidade massiva de dados que são coletados e armazenados em um único local, e uma das tecnologias mais conhecidas para esse tipo de trabalho é o Hadoop. A velocidade representa a boa performance que o sistema de *big data* deve ter para processar altos volumes de dados. Já a variedade representa as diferentes origens que os dados podem ter, desde dados de bancos relacionais tais como Oracle ou SQL Server, dados de bancos não relacionais tais como MongoDB ou Cassandra e até logs de sistemas operacionais, registros de sensores de máquinas, imagens, vídeos, sons, entre outros.

O software escolhido para esse trabalho foi o Splunk. Ecossistemas tradicionais de *big data* tipicamente podem utilizar mais de um software: um encarregado do framework e armazenamento, outro para streaming de dados, um terceiro para visualização e mais. O Splunk apresenta-se como um software completo que atua desde a coleta e armazenamento até a visualização. Passa também por streaming de dados, monitoramento, tem sua própria linguagem para pesquisa, segurança, inteligência artificial (machine learning e deep learning) e capacidade para ser estendido com softwares *add-on*. Não é objetivo deste trabalho contudo fazer uma análise extensiva do Splunk; além disso é bom lembrar que outros softwares disponíveis possuem funções semelhantes e podem ser usados para a mesma finalidade como Elastic Stack¹, por exemplo.

Neste trabalho, o Splunk foi usado para ingerir, processar, armazenar e analisar os dados dos sistemas de nível 2, MES, barramento de mensagens corporativo e ERP, cobrindo os níveis 3 e 4 da hierarquia funcional presente na norma ANSI/ISA 95 conforme na figura (Figura 1).

¹ <https://www.elastic.co>

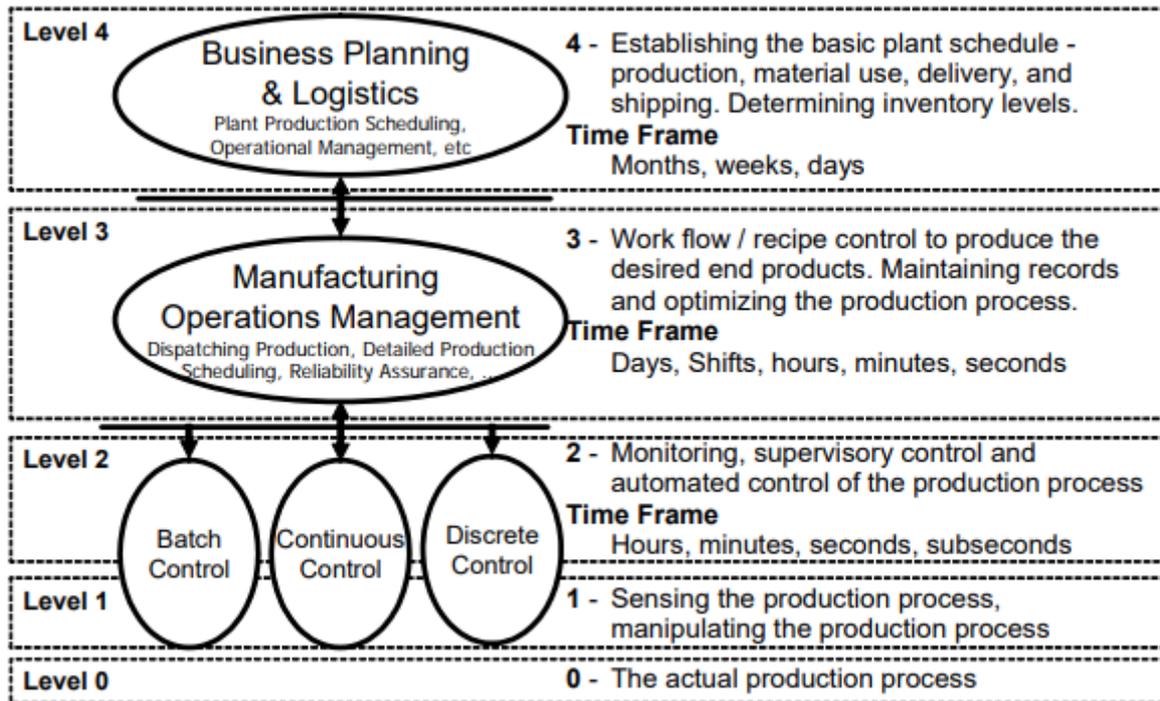


Figura 1. Hierarquia funcional multinível de atividades

Fonte: Norma ANSI/ISA - 95.00.03 — 2005

2 DESENVOLVIMENTO

Para ilustrar a relevância e aplicação deste trabalho é interessante imaginar um cenário típico na indústria. Uma área de produção pode ter uma variedade de sistemas de automação e informação, incluindo desde software executando em controladores lógicos programáveis (CLPs) até softwares ERP², passando por SCADA³, MES⁴, LIMS⁵, PIMS⁶, WM⁷, sistemas especialistas e barramentos de mensagens. Diante dessa malha de sistemas que deve trabalhar em conjunto, não é incomum haver falhas e identificá-las pode ser uma tarefa árdua. Responder a simples perguntas de usuários ou administradores como "por que a transação está demorando a completar?" ou "por que os dados não aparecem na tela?" pode requerer identificar os pontos de conexão entre os vários sistemas e a análise de logs cada um deles. Isso seguido de um processamento manual dos logs e a tentativa de reestabelecer os eventos dos vários sistemas em uma sequência temporal. Em resumo, uma atividade executada de forma *ad hoc* que pode levar

² ERP – Enterprise Resource Planning, ou Planejamento de Recurso Corporativo

³ SCADA – Supervisory Control and Data Acquisition, ou Sistemas de Supervisão e Aquisição de Dados

⁴ MES – Manufacturing Execution System, ou Sistema de Execução da Manufatura

⁵ LIMS – Laboratory Information Management System, ou Sistema de Gerenciamento de Informações Laboratoriais

⁶ PIMS – Plant Information Management System, ou Sistema de Informação de Gerenciamento de Planta

⁷ WMS – Warehouse Management System, ou Sistema de Gerenciamento de Armazém,

horas e cujo resultado frequentemente contém uma visão incompleta dos dados e limitada no tempo.

Diante desse cenário, cabe ressaltar que as indústrias modernas contam com sistemas de informação e automação que já se encontram conectados – ainda que com as medidas de segregação necessárias para segurança. Esses sistemas, portanto, já geram um determinado volume de dados, voluntária ou mesmo involuntariamente, que podem ser lidos, processados e transformados em informação e conhecimento.

Os dados gerados de forma voluntária tratam-se de logs de aplicação, escritas em bancos de dados ou outros sinais com o objetivo de retratar o comportamento de um sistema.

Já exemplos de dados gerados de forma involuntária são mensagens necessárias ao funcionamento do sistema que podem de alguma forma ser entendidas por um sistema externo, tais como: escritas em banco de dados necessárias ao funcionamento da aplicação, escritas em arquivo, trocas de mensagens com outras aplicações, registros de servidores web, de aplicação ou de banco de dados, informações sobre utilização de hardware ou de rede, entre outros.

Esses dados, quando entendidos de forma relacionada, revelam informações importantes sobre o comportamento das aplicações, seja de forma individual ou, mais valiosamente, em conjunto.

Vale a pena destacar que é cada vez mais frequente o caso em que os sistemas no chão de fábrica trabalham em conjunto. Algumas ações disparadas por usuários em um software são refletidas em toda a rede de sistemas de informação e automação. Como exemplo pode-se visualizar como a confirmação do planejamento da produção nos sistemas ERP desce para os sistema de operação como MES, PIMS, SCADA e WM. Já no sentido inverso, o sinal de um sensor no equipamento pode rodar um código no CLP que culmina com um apontamento de produção automático no ERP. Essa intrincada cadeia de eventos constitui o que chamamos de “transação distribuída”, onde uma ação é composta por várias transações individuais em cada um dos sistemas.

A análise desse tipo de dados para gerar as informações e conhecimentos relacionados possuem desafios de natureza técnica e funcional. No âmbito técnico podemos separar os desafios em 2 aspectos, (1) ingestão e processamento e (2) recuperação dos dados. É importante ressaltar que os dados se apresentam nas mais variadas formas e fontes, possuem um grande volume e são gerados em alta velocidade – os conhecidos 3Vs de *big data*. Já do ponto de vista funcional é relevante mencionar que é necessário agregar os dados para apresentá-los de uma forma visual que seja humanamente inteligível – nos casos em que usuários devem fazer o monitoramento e tomar ações.

Por esses motivos é importante utilizar ferramentas adequadas para processamento e visualização de *big data*. No caso do presente trabalho optou-se por utilizar Splunk como já foi mencionado anteriormente.

A Figura 2 a seguir ilustra uma rede de sistemas de automação informação onde é possível ver como eles se relacionam por meio de mensagens. Uma ferramenta de big *data* adequadamente configurada é capaz de ingerir dados das aplicações industriais em conjunto com dados das aplicações de suporte (p. ex.: servidores web, servidores de BD) e informações sobre utilização de hardware e elementos de rede.

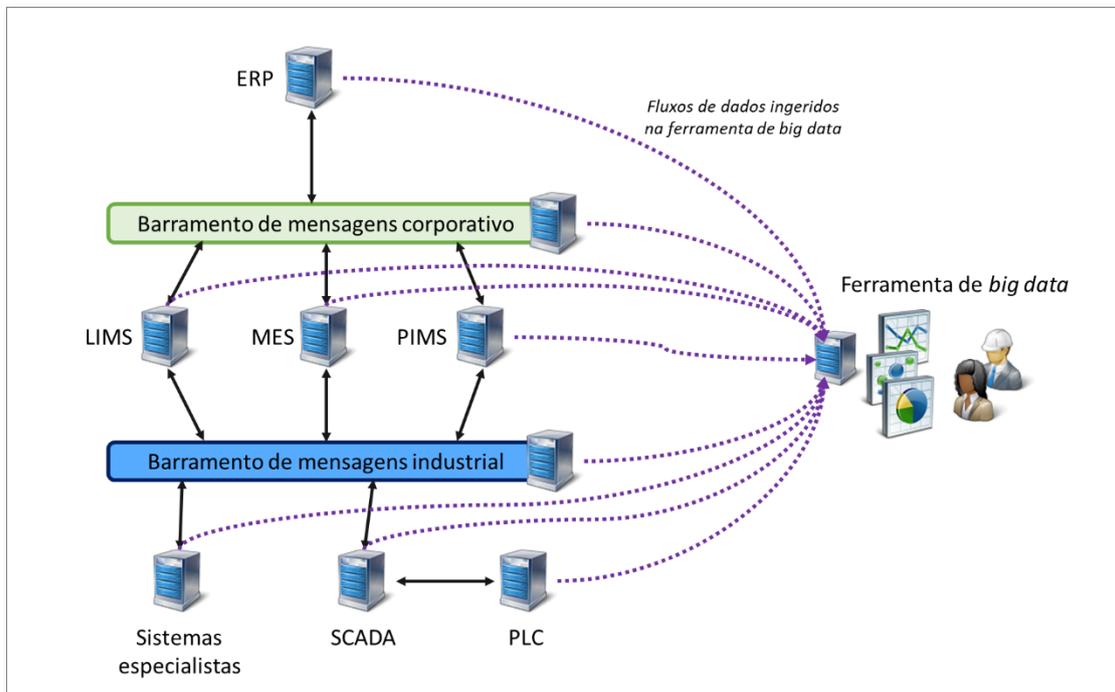


Figura 2. Uma ferramenta de big data adequadamente configurada é capaz de ingerir dados de toda a rede de sistemas industriais, incluindo aplicações, serviços de apoio (como servidores web e de bancos de dados) e hardware (servidores e elementos de rede).

Fonte: Accenture, 2019

Diante do cenário ilustrado é possível enumerar alguns benefícios do uso de uma ferramenta de *big data* aplicada à operação da indústria:

- Análise de logs em um único sistema, reduzindo tempo médio de análise e geração de *insights* mais completos sobre operação;
- Notificação de problemas em tempo real, reduzindo tempo médio de detecção (MTTD⁸), investigação (MTTI⁹) e resposta aos problemas, o que aumenta a disponibilidade;
- Análise de infraestrutura e funcional na mesma plataforma, possibilitando visão mais completa *end-to-end* da operação; e
- Simplificação da arquitetura para registro de logs e análise, liberando os sistemas de operação para atividades de negócio enquanto a carga de ingestão e análise está concentrada num único sistema especializado nesse assunto.

⁸ MTTD: *Mean Time To Detect*, ou tempo médio para detecção.

⁹ MTTI: *Mean Time To Investigate*, ou tempo médio para investigação.

A seguir serão detalhados alguns casos de uso para aplicação de big data no ambiente industrial.

2.1 Análise de desempenho integrado (*bottlenecks*)

A ferramenta de *big data*, por conter dados de medição de desempenho dos sistemas de informação e automação, permite o detalhamento do tempo gasto em cada sistema participante em uma transação distribuída.

Na **Erro! Fonte de referência não encontrada.** a seguir é possível ver, para uma determinada ação que tem início no MES, passa pelo barramento corporativo e chega no ERP, os tempos gastos em cada sistema ao longo dos dias. Dessa forma, tem-se uma análise bem objetiva a respeito de qual sistema está consumindo mais tempo e precisa ser analisado e otimizado dentro um contexto integrado de transação de negócio.

É possível notar que 2 eventos, relativamente próximos (cerca de 2 dias), indicam problemas de desempenho em uma mesma transação tem causas e sistemas diferentes. Essa visibilidade em tempo real é possível de ser alcançada quando se integra dados dos três sistemas na ferramenta de *Big Data*, permitindo a correlação dos dados em uma visão global de desempenho.



Figura 3. Divisão de tempos gastos por sistema na execução de uma operação de negócio.
Fonte: Accenture, 2019

2.2 Análise de pontos de interface entre sistemas (análise *end-to-end*)

A análise de uso de interfaces entre os sistemas permite dimensionar o impacto de alteração nas interfaces com uma visão global, *end-to-end*. Também permite estabelecer de forma objetiva quais são as interfaces mais críticas entre os sistemas do ponto de vista de usuários afetados.

A Figura 4 mostra um dashboard de análise de integração de sistemas. Na primeira coluna da esquerda estão listadas ações do sistema MES. Na coluna do meio estão listadas as interfaces de integração do barramento corporativo. Na coluna da direita

estão listadas as funções do ERP. A altura das barras corresponde ao volume de chamadas.

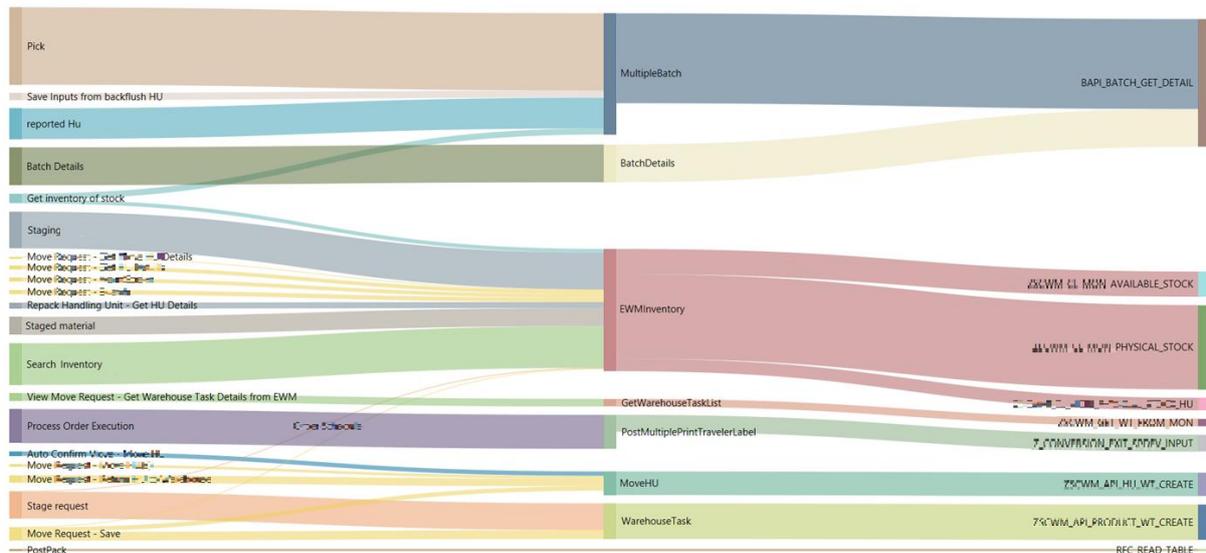


Figura 4: Dashboard de análise de integração de sistemas.

Fonte: Accenture, 2019

É possível notar que a função *EWMInventory* é usada por mais de uma dezena de funcionalidades do MES, portanto qualquer alteração nesse serviço tem um grande impacto. Além disso também é possível ver que, do lado do ERP, a função central (*Physical Stock*) é a que mais consome tempo. Nesse caso a conclusão possível é que uma otimização nessa função específica do ERP traria um grande impacto positivo para vários usuários do MES, em várias funcionalidades.

2.3 Detecção de problemas em tempo real com SMS, aplicativo e criação automática de tickets de suporte

A ferramenta de *big data*, por meio da ingestão de dados de monitoramento de sistemas, pode ser configurada para detectar problemas no ambiente e agir automaticamente diante disso.

Nesse trabalho mostramos um exemplo de problema de comunicação entre MES e ERP. Assim que a ferramenta percebe que as comunicações entre o MES e o ERP demoram mais tempo que o usual, um alerta é enviado aos times de suporte via aplicativo celular e mensagem de texto (SMS) indicando a situação e onde ela está ocorrendo. Em caso de normalização da comunicação, nova mensagem de alerta é enviada informando aos times que a situação está normalizada.

Caso o problema perdure, a ferramenta de big data cria automaticamente um ticket para equipe de suporte e notifica o time de suporte do MES. Enquanto o time de infra recupera para o estado normal de operação, o MES é colocado em modo off-

line. Essas ações são tomadas muitas vezes antes mesmo que o problema seja percebido pela operação das plantas.

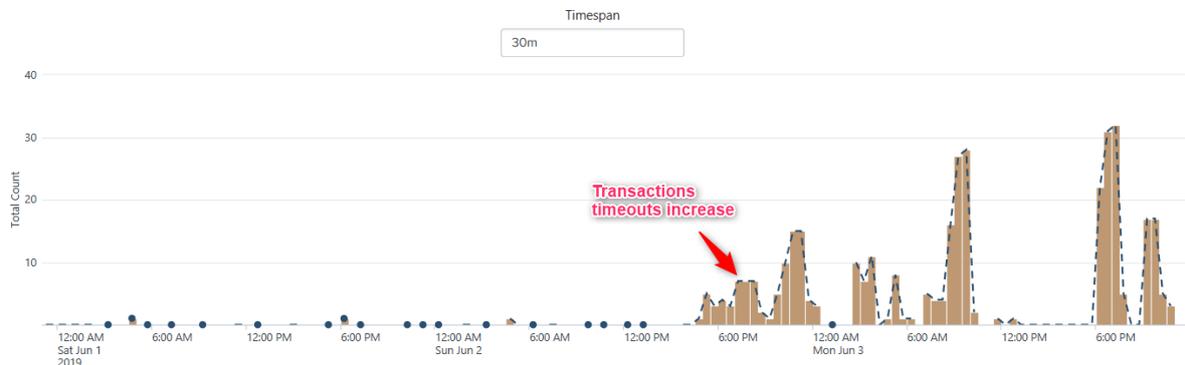


Figura 5: Dashboard de Timeouts de transações indica um aumento progressivo no volume, indicando algum problema na conectividade do barramento corporativo e ERP.

Fonte: Accenture, 2019

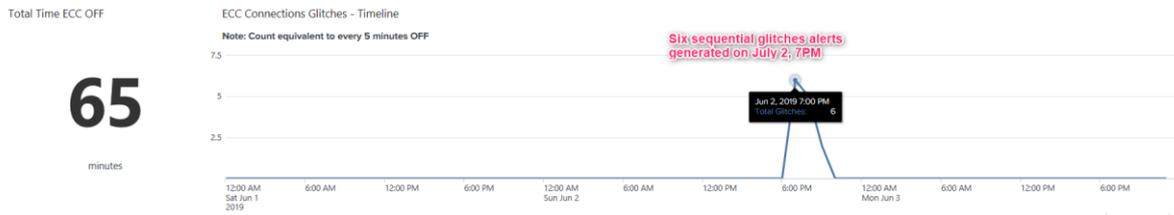


Figura 6: Dashboard de monitoramento de *glitches* na instância ERP de produção mostra um salto repentino, o que corrobora com os timeouts nas transações.

Fonte: Accenture, 2019

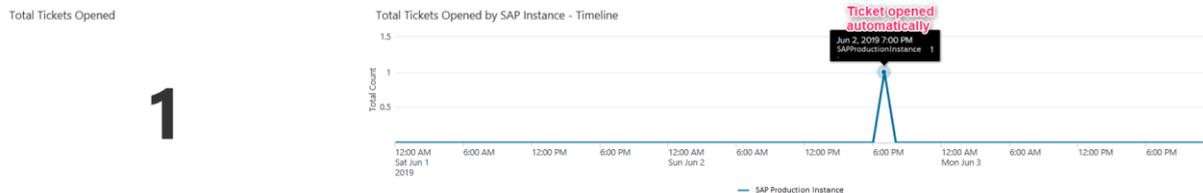


Figura 7: Dashboard de geração automática de ticket mostra que a lógica foi acionada no mesmo momento em que o número de *glitches* e timeouts das transações ocorreram.

Fonte: Accenture, 2019

Esse processo automatizado permite a economia de centenas de horas anualmente na análise de problemas, suporte aos sistemas e operação industrial que segue sem interrupção. Ao prover uma visão global de como está funcionando a comunicação e integração dos sistemas por meio da coleta integrada de dados distribuídos, a assertividade na resolução e antecipação dos problemas se torna fator chave de sucesso na redução de custos operacionais.

3 CONCLUSÃO

Diante do que foi exposto é possível ver como o uso de uma ferramenta de análise de *big data* pode ajudar no processo de identificação e solução de problemas nos sistemas de informação e automação industrial.

Ao aplicar tais ferramentas é possível passar de um processo *ad hoc*, manual e oneroso para um processo sistematizado e ágil onde dados de todo o ambiente de informação podem ser usados em conjunto.

Esse processo tem impacto direto na redução de MTTD e MTTI, como foi discutido. Isso pode ser visto dado que os problemas ainda iminentes são detectados por meio dos seus sintomas antes mesmo que ocorram; e a investigação desses problemas é bastante facilitada na medida em que os dados são agregados e exibidos em contexto, facilitando e potencializando sua análise.

Por fim cabe dizer que a maior parte dos dados necessários a este tipo de operação já existem dado que são naturalmente gerados, em grande parte dos casos basta que sejam coletados e processados. A informação e o conhecimento gerados por esses dados beneficia diretamente a operação industrial na medida em que representam diretamente redução de paradas de operação por indisponibilidade de sistema e simplificação do processo de manutenção de sistemas.

E uma vez que um sistema de processamento de *big data* esteja em funcionamento ele abre outras possibilidades além das mencionadas tais como otimização de sistemas e processos, análise de comportamento de usuário e uma vasta aplicação em cyber segurança.

REFERÊNCIAS

- 1 Big Data – O que é e qual sua importância? Disponível em: https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html Acesso em: 14 de Junho de 2019
- 2 Why Splunk. Disponível em: https://www.splunk.com/en_us/about-us/why-splunk.html Acesso em: 14 de Junho de 2019
- 3 ISA. (Enterprise-Control System Integration: Part 3: Activity Models of Manufacturing Operations Management. ANSI/ISA-95.00.03-2013.