

# Habilitando mineração de processos no domínio de operações logísticas

Extração de dados para um log de eventos centrado em objetos

Mateus Alex dos Santos Luna, Elio Ribeiro Faria Junior, Thais Rodrigues Neubauer

Marcelo Fantinato, Sarajane Marques Peres

{theew, elioribeirofaria, thais.neubauer, m.fantinato, sarajane}@usp.br

Escola de Artes, Ciências e Humanidades, Universidade de São Paulo

São Paulo, SP, Brasil

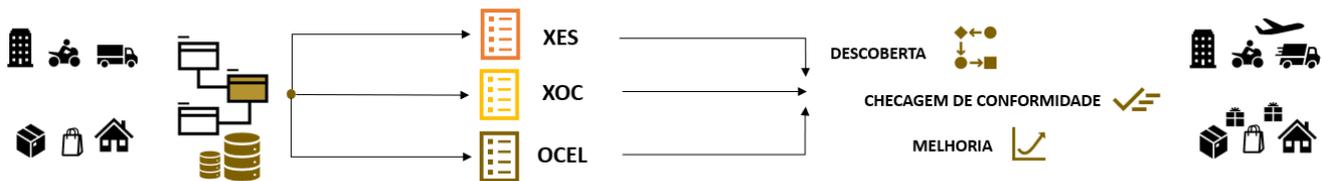


Figura 1: Engenharia de dados para mineração de processos no domínio de operações logísticas

## RESUMO

Atualmente, pesquisadores em mineração de processos têm se preocupado em ir além da análise de processos isolados, uma vez que a complexidade dos processos organizacionais exige análise de múltiplos processos inter-relacionados. Mineração de processos centrada em objetos traz uma proposta conceitual para lidar com esse tipo de situação. Contudo, trata-se de um assunto do estado da arte na área, ainda com poucos exemplos de aplicação em contextos reais. Esse artigo relata um trabalho em andamento para aproximação do estado da prática ao estado da arte no contexto de mineração de processos centrada em objetos. A proposta é elaborar e instanciar um *framework* de engenharia de dados para construção de logs de eventos centrados em objetos no contexto prático de organizações que lidam com processos logísticos. Assim, neste artigo, nós apresentamos os primeiros resultados representados pelo modelo conceitual de dados orientado a um processo logístico, pela instância desse modelo, e pela modelagem e instanciação de uma relação de mudanças cujo conteúdo reflete a operação de um sistema de informação sobre o modelo conceitual apresentado. A fase final de execução do trabalho aqui relatado inclui a proposição de procedimentos para uso da relação de mudanças para construção de um log de eventos centrado em objetos.

## CCS CONCEPTS

- **Information systems** → **Information lifecycle management**;
- **Software and its engineering** → **Application specific development environments**;
- **General and reference** → **Design**.

## KEYWORDS

process mining, data engineering, object-centric event log

### ACM Reference Format:

Mateus Alex dos Santos Luna, Elio Ribeiro Faria Junior, Thais Rodrigues Neubauer, Marcelo Fantinato, Sarajane Marques Peres. 2023. Habilitando mineração de processos no domínio de operações logísticas: Extração de dados para um log de eventos centrado em objetos. In *Workshop Virtual Loggi: Resultados do Programa de Bolsas de Pesquisa (PBP-LOGGI)*, 8 de dezembro, 2021, São Paulo, SP. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUÇÃO

Modelos de processos são ferramentas essenciais para o alcance de sucesso em gestão de processos em todos os tipos de organizações. Contudo, por razões culturais ou por falta de recursos humanos e materiais, é comum que os modelos de processos não sejam formalizados. Há situações em que os gestores dos processos não estão conscientes do fluxo real de atividades executadas no dia a dia das entidades que gerenciam. Nesse contexto, a análise automática de processos tem um papel crucial, pois fornece meios para que modelos de processos organizacionais sejam automaticamente descobertos, verificados e melhorados. Descoberta de modelos de processos e análises com objetivos de checagem de conformidade e de proposição de melhorias são tarefas estudadas na área de mineração de processos [19]. Em mineração de processos, a ciência de processos e a ciência de dados estabelecem um compromisso que permite o desenvolvimento de ferramentas para apoio à gestão de processos organizacionais e tomadas de decisão associadas. Entretanto, a complexidade inerente aos processos organizacionais impede que a automação dessa tarefa alcance pleno sucesso. Duas

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

PBP-LOGGI 2021, 8 de dezembro, 2021, São Paulo, SP

© 2023 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

questões podem ser colocadas como fatores importantes na definição de ações em mineração de processos: o domínio do negócio e a natureza dos dados que fornecem as informações básicas para a mineração de processos dentro desse domínio de negócio.

Definir os métodos, as estratégias e as técnicas de mineração de processos adequados para cada domínio é um trabalho *ad hoc*. Não há um procedimento padronizado para isso, e muito do sucesso de iniciativas na área dependem do olhar que é dado ao problema de mineração dentro do negócio em questão. Especificamente na área de logística, há a expectativa de que exista alta complexidade, devido a pelo menos dois fatores: (a) processos diferentes que se inter-relacionam e (b) escalabilidade. Esse artigo tem como ponto de atenção o trabalho de engenharia de dados para suporte à mineração de processos inter-relacionados. Assim, aqui são apresentados os primeiros esforços no sentido de explorar modelagens de dados relacionais comumente usadas em softwares que suportam a execução de processos logísticos, e atrelar essa exploração a um levantamento de requisitos para construção de logs de eventos capazes de representar a dinâmica desses processos. Nessa primeira fase, o ponto de atenção está no primeiro fator citado.

Embora o conceito de mineração de múltiplos processos esteja em foco no estado da arte, ainda não faz parte do estado da prática em mineração de processos. Em geral, as ferramentas e consultorias adaptam as técnicas clássicas de mineração de processos para atender as especificidades de organizações. Organizações que trabalham com logística trazem um cenário adequado para o estudo da aplicação prática desses conceitos e esse projeto tem a intenção de explorar tal cenário. O objetivo da pesquisa é propor uma abordagem para o tratamento de dados no domínio da logística, que apoie a engenharia de dados referente à construção de logs de eventos adequados a esse tipo de negócio. Como primeira fase da pesquisa, é proposto o desenvolvimento de um *framework* para associar a construção de logs de eventos tradicionais e logs de eventos centrados em objetos à exploração de bases de dados relacionais comumente incorporadas a softwares de gestão organizacional.

Neste artigo, um modelo de dados conceitual, sua instanciação e associação com uma relação de mudanças (ou uma relação de auditoria) são apresentados no contexto de um sistema de gerenciamento de entregas fictício. Esses artefatos representam a base para o desenvolvimento de procedimentos de engenharia de dados que serão propostos para a construção de logs de eventos aderentes às necessidades da mineração de processos para contextos nos quais diferentes processos organizacionais se inter-relacionam.

## 2 REFERENCIAL TEÓRICO

Esta seção apresenta os conceitos básicos necessários para compreensão da proposição discutida neste artigo.

### 2.1 Mineração de processos - conceitos básicos

De acordo com [21], um processo de negócio pode ser definido como um conjunto de tarefas (ou atividades) executadas de forma coordenada dentro de ambientes organizacionais ou técnicos, com o objetivo de alcançar uma meta de negócio. Processos de negócio são submetidos a procedimentos de análise, monitoramento e melhoria considerando expectativas do gerenciamento de processos de negócio e seus ciclo de vida. Quando efetivo, o gerenciamento

de processos de negócios garante resultados consistentes e agrega valor à organização e seus clientes [4].

A execução do ciclo de vida da gestão de um processo de negócio é primariamente feita com base em conhecimento tácito de gestores e especialistas ligados à organização, com vistas a procedimentos e regulações internas e externas à organização. Atualmente, esse contexto pode ser enriquecido se algum tipo de automação é adotado no suporte das tarefas do ciclo de vida do gerenciamento, considerando dados factuais e históricos sobre a execução dos processos. Tais dados estão registrados em log de eventos e são a base de uma disciplina denominada mineração de processos. A meta da mineração de processos de negócio é melhorar a compreensão sobre tais processos por meio da extração de conhecimento dos logs de eventos [18]. Logs de eventos são arquivos sequenciais nos quais eventos relacionados à execução das tarefas de um processo de negócio são registradas.

De acordo com [19], um evento  $e \in \mathcal{E}$ , em que  $\mathcal{E}$  é o universo de eventos, corresponde à execução de uma tarefa  $\tau \in \mathcal{T}$ , em que  $\mathcal{T}$  é o conjunto de atividades que ocorrem em um determinado processo de negócio. Eventos são comumente associados com atributos não-mandatários (por exemplo, atividade, selo de tempo, recurso, custo e tipo de transação). Tipos de transação se referem à tarefas do ciclo de vida associado a um evento, e podem assumir valores como *START* and *COMPLETE*.

Um caso  $c \in \mathcal{C}$ , em que  $\mathcal{C}$  é o universo de casos, engloba uma sequência de eventos  $e^c$ . Um caso refere a uma instância de processo que é uma execução real do processo. Todos os casos possuem um atributo mandatário chamado *trace* ( $\sigma$ ). Um *trace* é uma sequência finita de eventos  $\sigma \in \mathcal{E}^*$  e  $1 \leq i < j \leq |\sigma| : \sigma(i) \neq \sigma(j)$ . Um log de eventos  $L$  é um conjunto de casos  $L \subset \mathcal{C}$  e para qualquer  $c_1, c_2 \in L$ , se  $c_1 \neq c_2$  então  $\delta(e^{c_1}) \cap \delta(e^{c_2}) = \emptyset$ , em que  $\delta$  é um operador para converter uma sequência em um multi-conjunto. Um classificador é uma função que mapeia os atributos de um evento em um rótulo ou valor. Por exemplo, se  $e.[attribute\_notation]$  é adotado para indicar uma função classificadora, tem-se que  $e.\tau$  indica o nome da tarefa associada a um evento  $e$ .

Mineração de processos usa logs de eventos como entrada para algoritmos que descobrem modelos de processos (descoberta), contrastam dados do log de eventos com informações de modelos de processo (checagem de conformidade) e extraem informações que permitem realizar melhorias no processo (melhoria de processos) [19]. Nesse artigo, estamos interessados em suportar a criação de logs de eventos centrados em objetos com o intuito de gerar informação referente a processos inter-relacionados para motivar a elaboração de estratégias de descoberta, checagem de conformidade e melhoria de processos inter-relacionados.

### 2.2 Mineração de processos inter-relacionados

Um dos primeiros trabalhos a ilustrar o problema de mineração de processos inter-relacionados foi apresentado em [6]. No seu estudo, os autores ilustraram a complexidade de processos inter-relacionados de vendas, entregas e cobranças usando conceitos de modelos de processos centrados em artefatos e referentes a Proclerts [15]. Artefatos, nesse contexto, dizem respeito aos modelos de processos independentes criados para pontos de atenção específicos, por exemplo: o modelo de processo de vendas constitui um artefato,

o modelo de processo de entregas é um segundo artefato, e o modelo de processo de cobranças é um terceiro artefato. A lógica de execução de cada processo pode ser apresentada usando três redes de Petri e dependências comportamentais entre os processos podem ser modeladas usando portas e canais assíncronos provenientes da teoria de Proclats. Como uma alternativa, [14] aplicaram a teoria de *Guard-Stage-Milestones* [11]. Nessa iniciativa, o relacionamento entre as lógicas de execução dos processos de vendas, entregas e cobranças foi modelado usando uma linguagem declarativa capaz de expressar condições de verificação e de parada que controlam a dinâmica entrelaçada dos diferentes processos.

Uma estratégia simples e mais aderente aos conceitos básicos usados em mineração de processos é apresentada em [13] para permitir fluxos de controle entre processos independentes. Essa estratégia considera dependências comportamentais e a frequência com que elas ocorrem. A estratégia semiautomática apresentada nesse trabalho inicia na identificação de artefatos (pontos de atenção) e da fonte de dados (p.ex. modelos relacionais) e finaliza na apresentação do modelo de processo AS-IS. Em [5], o conceito de modelos de processos centrados em artefatos é usado junto a estratégias de descoberta de modelos de processo expressos em máquinas de estados. Essa pesquisa usa log de eventos que integram os eventos dos processos referentes a cada artefato e permitem descobrir um modelo de processo único. A origem dos eventos é preservada no modelo de forma que é possível explorar as interações entre os múltiplos artefatos.

Também em 2017, o conceito de mineração de dados orientado a objetos é apresentado em [17]. Esse conceito representou uma evolução no trabalho com múltiplos processos (agora denominados objetos), pois incorpora modelagem de dados relacionada ao log de eventos, o que permite trabalhar com relações de múltiplas cardinalidades ( $m : n$ ) e implementar algoritmos de checagem de conformidade e análise de desempenho. Esse conceito sofreu evoluções de forma que modelos de múltiplos pontos de vistas puderam ser desenvolvidos [2], culminando em 2020 na apresentação das redes de Petri centradas em objetos [16]. Essas redes de Petri são redes coloridas nas quais é possível visualizar todas as relações existentes entre elementos de diferentes processos.

A evolução dos estudos em mineração de processos gerou uma série de padrões para logs de eventos, cada qual apresentando, no decorrer do tempo, características para acomodar novos conceitos de representação de informação. A partir da identificação da necessidade de trabalhar com processos inter-relacionados, os padrões para log de eventos passaram a incluir características especializadas, dando origem a uma família de padrões que, atualmente, estão no centro das atenções da engenharia de dados da mineração de processos. A adoção desses novos padrões ainda não é unanimidade entre pesquisadores e práticos da área.

## 2.3 Padrões para log de eventos

A figura 2 resume a evolução dos padrões de log de eventos. O primeiro padrão a ser estabelecido foi o MXML, um padrão no qual as informações dizem respeito a um único processo. O último padrão, o OCEL, permite organizar informações sobre múltiplos processos relacionados. A figura ainda posiciona nessa linha do tempo, conceito e padrões de propósito geral que foram usados

pela comunidade de mineração de processos, ou como base ou como uma alternativa aos padrões específicos da área. Os padrões especificamente criados para log de eventos seguem descritos no restante desta seção.

**2.3.1 O padrão MXML: Mining eXtensible Markup Language.** Mineração de processos é uma disciplina relativamente nova que teve seus primeiros conceitos e métodos estabelecidos no início dos anos 2000, e que rapidamente se desenvolveu dentro das comunidades acadêmica e, posteriormente, industrial. Devido a esse crescimento, a área de mineração de processos precisou estabelecer um metamodelo para padronizar a organização da informação referente ao log de eventos [3]. O primeiro padrão para log de eventos estabelecido foi o MXML. Um grande problema encontrado no padrão MXML foi a dificuldade em lidar com a padronização de atributos adicionais àqueles já previstos em sua sintaxe [3, 9]. Uma versão semanticamente anotada do MXML foi então desenvolvida, o SA-MXML (*Semantically Annotated Mining eXtensible Markup Language*), porém o nível de complexidade inerente ao padrão impediu que ele fosse amplamente adotado.

**2.3.2 O padrão XES: eXtensible Event Stream.** Para resolver as limitações do MXML, um novo padrão foi desenvolvido [12]: o XES. Nesse padrão, um atributo é um par chave-valor, cujo tipo pode ser um texto, um número inteiro ou real, uma data ou hora, um tipo lógico, um identificador ou uma lista de atributos. Para fornecer semântica aos dados, o XES permite a criação de extensões. Tais extensões definem chaves e fornecem um significado para os valores associados a essas chaves. Segundo o exposto em [10], o padrão XES foi estabelecido para atingir quatro objetivos:

- Simplicidade: representar as informações da forma mais simples, com o propósito de facilitar o entendimento por parte do usuário.
- Flexibilidade: poder ser utilizado em qualquer domínio de aplicação além da mineração de processos.
- Extensibilidade: realizar atualizações no padrão sem perder a compatibilidade com versões anteriores.
- Expressividade: empregar um formato genérico no padrão, mas com o mínimo de perda de especificidade de informação.

A figura 3 ilustra o metamodelo para o padrão XES. Nele, o log de eventos é composto por um ou mais traços, associados a casos, que descrevem a ordem de execução das atividades relacionadas ao processo. Um traço pode conter eventos, os quais representam as atividades do processo. Essa estrutura de log de eventos, traço e evento possuem atributos que são utilizados para armazenar as informações, por exemplo, um evento pode conter um atributo do tipo *string* responsável por armazenar o nome da atividade associada ao evento. Há também a composição de um classificador de atributos, pois os atributos não são predefinidos. Por fim, o log de eventos contém uma extensão para determinar o conjunto de atributos que será utilizado em um domínio de aplicação específico [10].

**2.3.3 O padrão XOC: eXtensible Object-Centric.** Logs de eventos XES organizam os eventos com uma noção de caso, ou seja, o log de eventos é orientado a um processo. Considerando que em muitos cenários reais há a necessidade de analisar processos inter-relacionados, diferentes noções de casos passam a ser importantes. O padrão XES, ao assumir apenas uma única noção de caso, exige

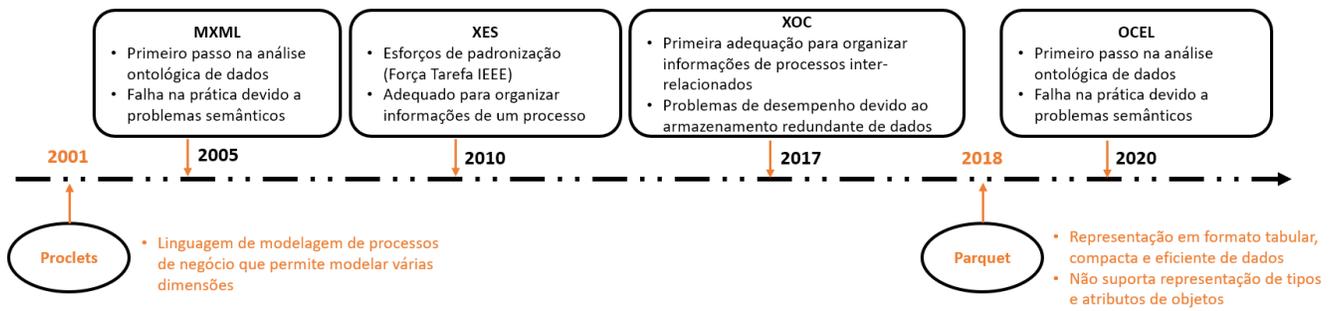


Figura 2: Evolução de padrões para log de eventos

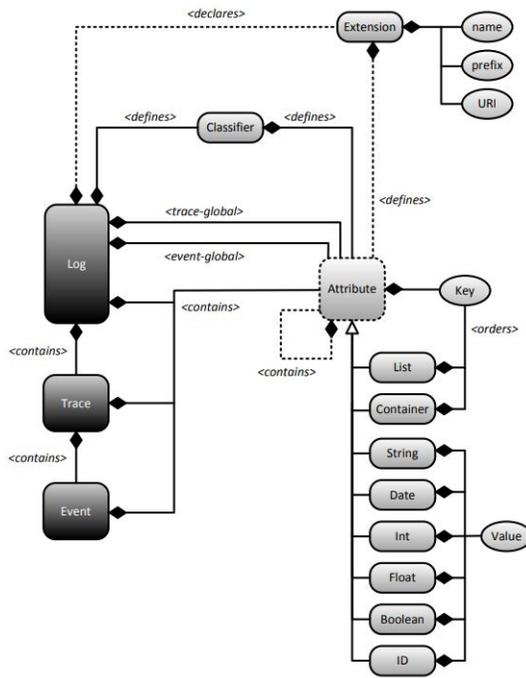


Figura 3: Metamodelo do padrão XES [10]

que a informação seja “planificada”, gerando os problemas conhecidos como deficiência, convergência e divergência de informação no log de eventos [20]. O padrão XOC foi criado para resolver o problema de uso de múltiplas noções de casos. Nesse padrão, eventos são descritos com um identificador próprio (nome da atividade associada a ele), com uma lista de objetos ao qual o evento se refere e uma estrutura que representa a relação entre objetos no momento em que o evento ocorre.

**2.3.4 O padrão OCEL: Object Centric Event Log.** O log de eventos centrado em objeto organizado no padrão XOC pode gerar problemas de desempenho [8]. A fim de superar essa dificuldade, o padrão OCEL foi estabelecido [7]. O padrão OCEL suporta informação sobre eventos relacionados a múltiplos processos e permite que a inter-relação entre os processos seja observada. A tabela 1 mostra representações informais da informação presente em um

log de eventos organizado no padrão OCEL. A primeira parte da tabela representa os eventos, por exemplo, o *evento 1* está associado a uma *atividade*, ao *dia e horário* da sua realização e aos *objetos* do tipo: **pedido**, **item** e **cliente**. A segunda parte da tabela mostra os objetos. Nela, o valor *i1* é um objeto do tipo **item** e possui os atributos **produto** e **peso**; isso significa que o *evento 1* é referente a *atividade* “realizar pedido”, realizada em “10-09-2021 às 08:05:03”, pelo colaborador “Pedro”, e essa atividade faz referências aos objetos *p1*, *i1*, *i2* e *c1*.

Tabela 1: Representações no padrão OCEL, adaptado de [8]

Eventos	Atividades	Timestamp	Pedido	Item	Cliente
evento 1	Realizar pedido	10-09-2021 08:05:03	p1	i1, i2	c1
evento 2	Confirmar pedido	10-09-2021 08:15:55	p1	-	-
evento 3	Verificar disponibilidade	10-09-2021 08:30:01	p1	i1	-
Objetos	Tipo	Produto	Peso	Idade	ID Conta
i1	item	TV	18,00	-	-
i2	item	Geladeira	37,00	-	-
c1	cliente	-	-	30	9874

A figura 4 mostra o metamodelo do padrão OCEL. Nesse padrão, o log de eventos é composto por eventos e objetos. Eventos possuem um ou mais objetos e o log de eventos pode assumir atributos globais. Os eventos e objetos são compostos por elementos que podem ser do tipo texto, data e hora, número inteiro ou real e do tipo lógico. Para suportar o trabalho com o padrão OCEL está disponível uma ferramenta<sup>1</sup> e um plugin para a ferramenta ProM<sup>2</sup>. Esses recursos são capazes de importar e exportar nas serializações JSON e XML, o conteúdo de um log de eventos organizado no padrão OCEL [7].

### 3 MÉTODO

O *framework* de engenharia de dados proposto nesse trabalho pressupõe a criação de um contexto sintético para execução da extração de informação para os logs de eventos. Para a criação do contexto sintético, um sistema simplificado referente ao contexto de operações logísticas foi projetado. Para isso, três artefatos foram criados:

<sup>1</sup> <https://github.com/OCEL-standard/ocel-support>

<sup>2</sup> Pacote: OCEL Standard (ProM 6.10).

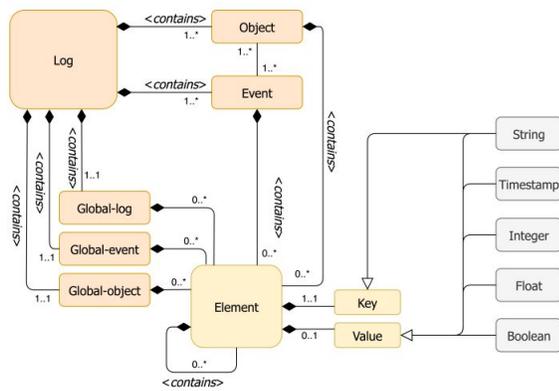


Figura 4: Metamodelo do padrão OCEL (adaptado de [7])

- um modelo de dados conceitual para armazenamento de informações referentes à execução de operações logísticas;
- um projeto de inserção de dados orientado à diferentes situações possíveis na execução das atividades referentes às operações logísticas;
- construção e instanciação de um modelo físico de dados para geração da instância de banco de dados que fornecerá os insumos para criação de logs de eventos.

Uma vez estabelecido esse contexto sintético, os próximos passos para conclusão desse trabalho são:

- modelar logs de eventos seguindo a especificação do padrão XES, a especificação do padrão XOC e a especificação do padrão OCEL;
- construir *scripts* de extração de informação do modelo conceitual e da relação de mudanças a fim de instanciar os logs de eventos;
- evidenciar as diferenças entre os logs de eventos construídos sob à perspectiva de tarefas de descoberta de modelos de processos, checagem de conformidade e execução de melhoria de processos.

## 4 RESULTADOS

A seção 4.1 apresenta o contexto sintético referente ao domínio logístico suportado por um sistema de informação simplificado e o respectivo modelo de dados conceitual para armazenamento de dados associados às atividades executadas no referido domínio. A seção 4.2 descreve variações de instanciação do modelo de dados, simulando algumas situações reais que podem ocorrer no domínio de aplicação modelado. A seção 4.3 mostra um excerto da relação de mudanças que representa a entidade fornecedora de informação para a construção de um log de eventos<sup>3</sup>.

### 4.1 Modelo de dados: operações logísticas

O modelo de dados utilizado foi criado para refletir uma situação relacionada ao contexto logístico, porém de forma simplificada. Seguindo essa ideia, foi definido o modelo ilustrado na figura 5.

O modelo representa uma organização que recebe pedidos de entregas de diferentes itens, possivelmente atrelados a um único pedido no lojista. Porém, na visão simplificada do sistema projetado, os pedidos estão associados ao lojista e têm atributos indicando a origem do pedido, e o destino e destinatário dos itens. Cada item, de forma individual, pode ser colocado em um pacote ou podem ser empacotados em conjunto. Os pacotes são as unidades entregues aos destinatários. Para registrar as diferentes tentativas de entregas, possivelmente realizadas por diferentes entregadores, há uma relação que registra as entregas e o estado dela. Os lojistas enviam itens para entrega, e a organização que utiliza o sistema tem diferentes motivos para desejar ter informação sobre o lojista: desde priorização de lojistas parceiros até redução da redundância no modelo. Os entregadores, que levam os pacotes até seus destinatários, precisam também estar presentes no modelo de dados para que a organização logística possa controlar quais entregas cada um deles realizou.

### 4.2 Variantes dentro do modelo de dados

A instanciação do modelo de dados conceitual considera situações variadas. Por exemplo, como um pacote não está restrito a um item, itens de diferentes pedidos podem ser empacotados juntos caso tenham um mesmo destinatário e itens que pertencem a um mesmo pedido podem ser entregues em pacotes separados. As entregas podem falhar, quando, por exemplo, não há ninguém para receber o pedido e, assim, um mesmo pacote pode constar em diferentes tentativas de entregas até que uma delas seja bem-sucedida. Também há a possibilidade de uma entrega ter vários pacotes, como quando uma primeira entrega de um pacote falha e outro chega para ser entregue ao mesmo destinatário. Para este trabalho, tentamos cobrir algumas dessas variações para identificar como elas influenciam a extração de dados para construção dos logs de eventos.

### 4.3 A relação de mudanças

Definidas as relações do sistema de informação, uma relação adicional foi criada para registrar as interações do sistema com o banco de dados: a relação nomeada “*Change table*”. Essa relação é composta por atributos que registram o tipo de alteração realizado nas relações do banco de dados, a relação afetada pela alteração, os valores alterados e quando essas alterações foram efetuadas. Um exemplo de registro gravado na relação “*Change table*” é mostrado na figura 6. O registro nessa relação é criado a partir de um gatilho que identifica alterações em todas as outras relações do banco de dados. Como ação do gatilho são inseridas informações sobre as mudanças nas relações do sistema de informação. Esse tipo de relação também é conhecida como relação de auditoria. Podemos ver na figura um exemplo da inclusão de uma entrega ocorrendo nas duas primeiras linhas: é criada uma entrega e um pacote é relacionado a ela, e ambas as ações são registradas na “*Change table*”. Em seguida, uma outra entrega recebe uma atualização de seu *status*: ela falha, e em seguida uma outra entrega é criada para entregar os pacotes que a anterior falhou em entregar, além de possíveis novos pacotes que vão para o mesmo destinatário.

## 5 CONSIDERAÇÕES FINAIS

Esse artigo apresentou os progressos realizados em um projeto de engenharia de dados para mineração de processos. Até o momento,

<sup>3</sup>Banco de dados disponível em: <https://github.com/pm-usp/data-engineering>

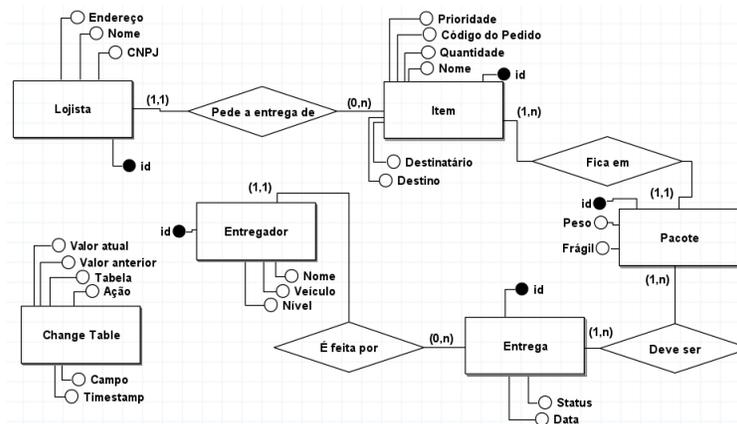


Figura 5: Modelo de dados conceitual - operações logísticas

act	tabela	valor_anterior	valor_atual	campo	timestamp	
	character varying (10)	character varying (40)	character varying (120)	character varying (120)	character varying (40)	timestamp with time zone
28	INSERT	entrega	[null]	('A entregar',2021-11-26,3,3)	all	2021-11-26 09:16:13.631207-03
29	INSERT	pacote_entrega	[null]	(1,3)	all	2021-11-26 09:16:13.631207-03
30	UPDATE	entrega	"A entregar"	"Falha"	status	2021-11-26 09:29:04.979601-03
31	INSERT	entrega	[null]	('A entregar',2021-11-27,2,2)	all	2021-11-26 09:38:49.759988-03
32	INSERT	pacote_entrega	[null]	(2,2)	all	2021-11-26 09:38:49.759988-03
33	INSERT	pacote_entrega	[null]	(3,2)	all	2021-11-26 09:38:49.759988-03

Figura 6: Exemplo de registros na relação de mudanças

as ações referentes ao projeto permitiram fazer um levantamento das principais características de diferentes padrões de logs de eventos usados em mineração de processos e projetar um contexto sintético para suportar a proposição de um *framework* para extração de dados e construção de logs de eventos. Nos próximos passos, diferentes padrões serão usados para construir logs de eventos a partir dos dados presentes na instância do contexto sintético de operações logísticas. Este trabalho está especialmente interessado na discussão da utilidade de logs de eventos centrados em objetos e em fornecer subsídios para o desenvolvimento de mineração de processos aderente ao contexto de operações logísticas. Um exemplo de extração de logs de eventos baseado em sistemas corporativos atuais foi desenvolvido em [1]. Esse exemplo será a base para os próximos passos deste trabalho.

## 6 AGRADECIMENTOS

Este trabalho foi realizado com o apoio da Loggi (www.loggi.com), por meio do Programa de Bolsas de Pesquisa Loggi (PBP-LOGGI).

## REFERÊNCIAS

- [1] Alessandro Berti, Gyunam Park, Majid Rafiei, and Wil van der Aalst. 2021. An Event Data Extraction Approach from SAP ERP for Process Mining. arXiv:2110.03467 [cs.DB]
- [2] A. Berti and W. van der Aalst. 2018. StarStar Models: Process Analysis on top of Databases. In *In CEUR Workshop*.
- [3] B. V. Dongen and W. Aalst. 2005. A Meta Model for Process Mining Data. In *Open Interop Workshop on Enterprise Modelling and Ontologies for Interoperability, Co-located with CAiSE'05 Conference*.
- [4] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. 2013. *Fundamentals of Business Process Management*. Springer Publishing Company, Incorporated.
- [5] Maikel Eck, Natalia Sidorova, and Wil van der Aalst. 2019. Guided Interaction Exploration and Performance Analysis in Artifact-Centric Process Models. *Process Models. Bus. Inf. Syst. Eng.* 61 (2019), 649–663.
- [6] D. Fahland, M. de Leoni, B. F. van Dongen, and W. van der Aalst. 2011. Behavioral Conformance of Artifact-Centric Process Models. In *Lecture Notes in Bus. Inf. Process.* Springer, 37–49.
- [7] A. F. Ghahfarokhi, G. Park, A. Berti, and W. van der Aalst. 2020. Ocel standard.
- [8] Anahita Farhang Ghahfarokhi, Gyunam Park, Alessandro Berti, and Wil M. P. van der Aalst. 2021. OCEL: A Standard for Object-Centric Event Logs. In *New Trends in Database and Information Systems*. Springer, 169–175.
- [9] C. W. Günther. 2009. *Process mining in flexible environments*. Ph.D. Dissertation. Technische Universiteit Eindhoven.
- [10] C. W. Gunther and H. M. W. Verbeek. 2014. *XES - standard definition*. BPMcenter.org.
- [11] R. Hull, E. Damaggio, R. De Masellis, F. Fournier, M. Gupta, F. T. Heath, S. Hobson, M. Linehan, S. Maradugu, A. Nigam, P. N. Sukaviriya, and R. Vaculin. 2011. Business artifacts with guard-stage-milestone lifecycles: Managing artifact interactions with conditions and events. In *Proc. of Intl. Conf. on Distributed Event-Based Syst.* 51–62.
- [12] IEEE. 2016. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams - IEEE Std 1849-2016. , 50 pages.
- [13] X. Lu, M. Nagelkerke, D. van de Wiel, and D. Fahland. 2015. Discovering Interacting Artifacts from ERP Systems. *IEEE Trans. on Serv. Comp.* (2015), 861–873.
- [14] Viara Popova and Marlon Dumas. 2014. Discovering Unbounded Synchronization Conditions in Artifact-Centric Process Models. In *Bus. Process Manage. Workshops. Lecture Notes in Bus. Inf. Process.* Springer.
- [15] W. van der Aalst, P. Barthelmeß, C. A. Ellis, and J. Wainer. 2001. Proclerts: A Framework for Lightweight Interacting Workflow Processes. *Int. J. of Cooperative Inf. Sys.* 10 (2001), 443–481.
- [16] W. van der Aalst and A. Berti. 2020. Discovering Object-Centric Petri Nets. *Fundamenta Informaticae XXI* (2020), 1001–1040.
- [17] W. van der Aalst, G. Li, and M. Montali. 2017. Object-Centric Behavioral Constraints. In *CEUR Workshop Proc. in Description Logics*.
- [18] W. M. P. van der Aalst. 2011. *Process mining - discovery, conformance and enhancement of business processes* (1st ed.). Springer, Berlin, Germany.
- [19] Wil M. P. van der Aalst. 2016. *Process Mining - Data Science in Action*. Springer.
- [20] Wil M. P. van der Aalst. 2019. *Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data*. 3–25. [https://doi.org/10.1007/978-3-030-30446-1\\_1](https://doi.org/10.1007/978-3-030-30446-1_1)
- [21] M. Weske. 2007. *Business Process Management: Concepts, Languages, Architectures* (2nd ed.). Springer.