

# FIA/P GRADUAÇÃO

# DATA SCIENCE

## DATA GOVERNANCE & DATA SECURITY MANAGEMENT

Prof. Dr. Renê de Ávila Mendes

# Objetivos da disciplina

**DISCIPLINA:** Data Governance & Data Security Management

**OBJETIVOS:** Descubra como funciona um **projeto de banco de dados** dentro de um ambiente corporativo, aplicando **técnicas de levantamento e documentação de requisitos**, aderente aos projetos de bancos de dados e aprenda a representar esses requisitos em arquiteturas de solução tecnológica para Data distribution e Data integration, modelos de estruturas de dados e dicionários de dados buscando **Data quality**. Garanta a qualidade dos dados de uma empresa para prover os melhores subsídios à tomada de decisão de negócio, praticando **Data cleaning** para limpar, harmonizar, complementar e corrigir dados inconsistentes, incompletos ou incorretos. Compreenda como funciona o **ciclo de vida da informação** e as responsabilidades administrativas sobre os dados de negócio, buscando qualidade, segurança e compatibilidade com políticas de administração de informação corporativas auditáveis, aplicando práticas atuais de **Data profiling** e conhecendo os princípios de **Data auditing**, de forma a atender a **Lei Geral de Proteção de Dados (LGPD)**.

# Assuntos – 1º Semestre

- Introdução
- Estratégia Empresarial; Arquitetura Empresarial; TOGAF; Archimate e Archi
- Requisitos para projetos de bancos de dados
- SCRUM
- **BPMN e Bizagi Modeler**
- Qualidade em metadados; Oracle Data Modeler
- Arquiteturas de integração e distribuição física de banco de dados; Estudos de caso
- Master Data Management e Data Hub; Pentaho-DI



**BPMN**

# **BPMN – Business Process Model and Notation**

- Padrão para modelagem de processos de negócio
- Mantida por Object Management Group (OMG)
- Versão atual 2.0.2 (01/2014)
- <https://bpmn.org>

# Vantagens de usar uma notação comum

- **Comunicar** – a modelagem dos processos gera um conhecimento compartilhado entre os *stakeholders*. Para que a comunicação seja comum, recomenda-se adotar uma modelagem simples dos processos.
- **Ensinar** – o diagrama de processos serve para orientar o que cada um deve fazer. Por isso, recomenda-se que o diagrama seja bem documentado.
- **Mostrar as conexões e as dependências entre os processos** – Diagramas do BPMN são uma excelente maneira de detalhar os processos.

# Algumas definições

- **Processo** – uma sequência ou um **fluxo de atividades** em uma organização com o objetivo de realizar um trabalho.
- **Processo de negócio** – um conjunto definido de atividades que representam os **passos** necessários para atingir um objetivo de negócio, incluindo o fluxo e o uso de informação e de recursos.
- **Analista de negócio** – Um especialista que analisa necessidades e problemas de negócios, consulta usuários e *stakeholders* para identificar oportunidades para melhorar o retorno dos negócios por meio da tecnologia da informação e define, gerencia e monitora os requisitos dos processos de negócios.

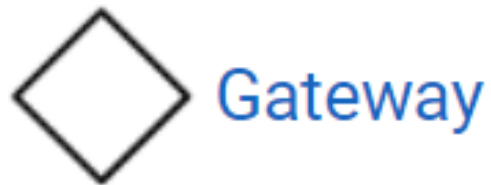


# BPMN – Recursos básicos



Algo que **acontece**. O ícone utilizado denota o tipo do evento: um envelope denota uma mensagem, um relógio denota tempo.

# BPMN – Recursos básicos



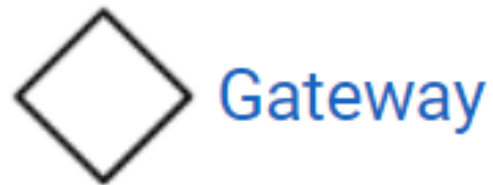
Algo que é **feito**. Trabalho que uma empresa ou organização executa usando processos de negócios. Pode ser atômica ou composta.

# BPMN – Recursos básicos



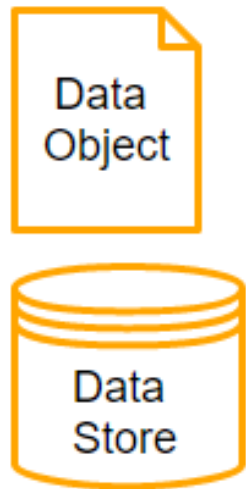
Determina a **bifurcação** e  **fusão** de caminhos, dependendo das condições expressas

# BPMN – Recursos básicos



Representa a **progressão geral** de como um Processo ou segmento de Processo é executado.

# BPMN – Outros recursos – Dados



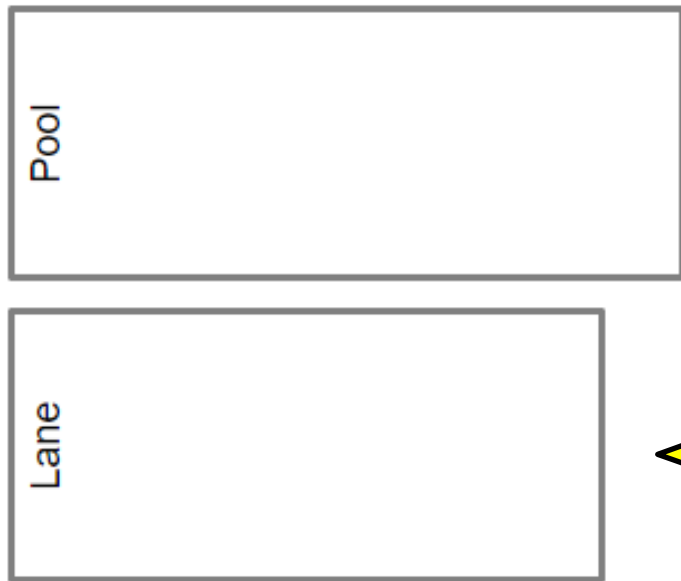
Representa dados circulando através do processo, tais como documentos, emails ou cartas.

# BPMN – Outros recursos – Dados



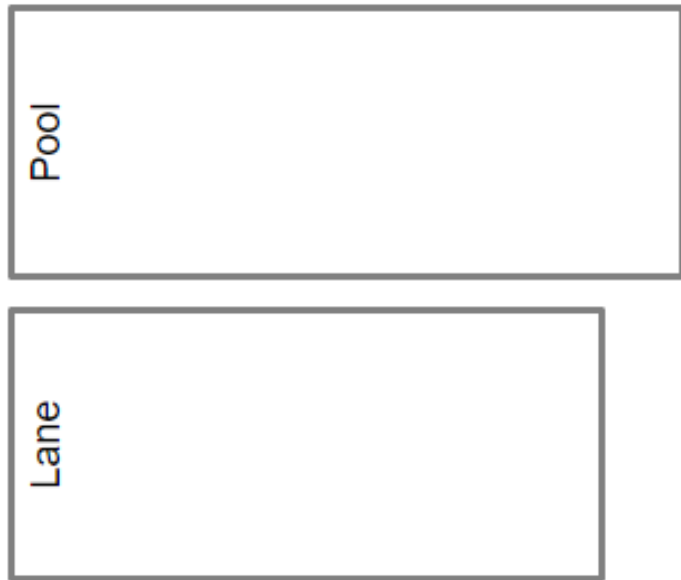
Local onde o processo pode ler ou armazenar dados, como bancos de dados ou arquivos físicos. O dado é persistido além do tempo de vida da instância do processo.

# BPMN – Outros recursos – Participantes



Indica **quem** é responsável por executar  
quais tarefas ou subprocessos.

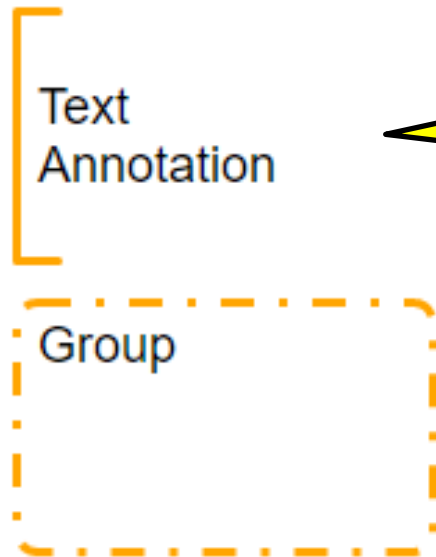
# BPMN – Outros recursos – Participantes



Assume o controle do processo atribuindo as tarefas. Orquestra as tarefas.

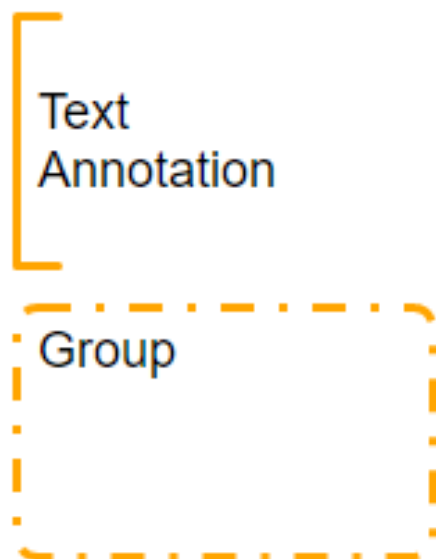


# BPMN – Outros recursos – Artefatos

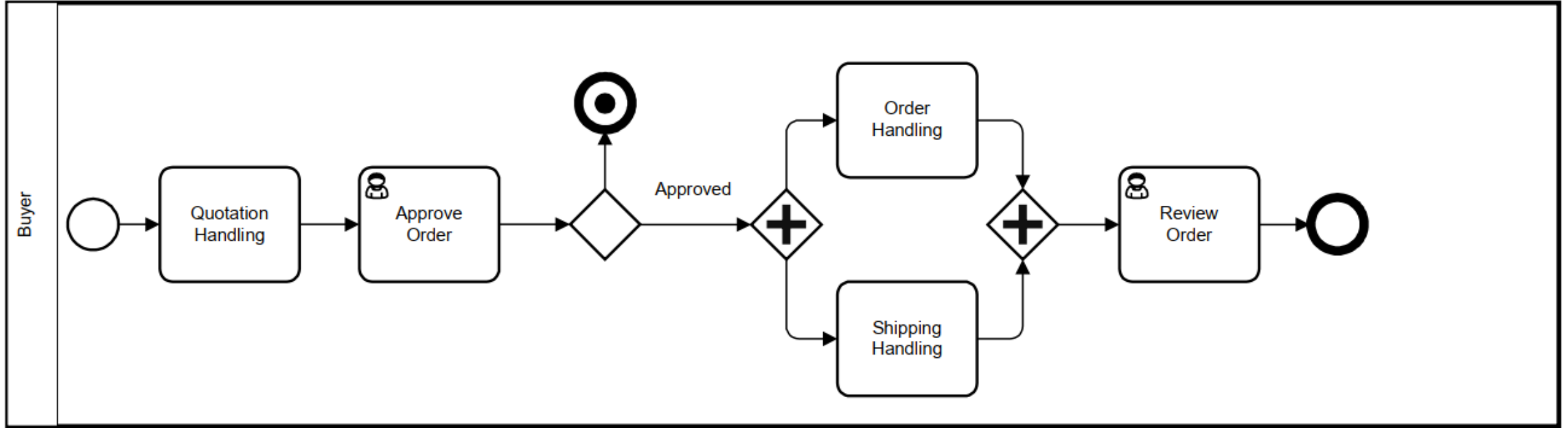


Texto que explica como a tarefa é executada.

# BPMN – Outros recursos – Artefatos



Não tem influência sobre o significado do processo. Pode ser aplicado livremente. Geralmente é utilizado para marcar partes do modelo para as quais convenções específicas são aplicáveis.




# BPMN – Event

- Start Event
- ◉ Intermediate Event
- End Event

Um evento que indica onde um determinado processo começa.

# BPMN – Event

 Start Event

 Intermediate Event

 End Event

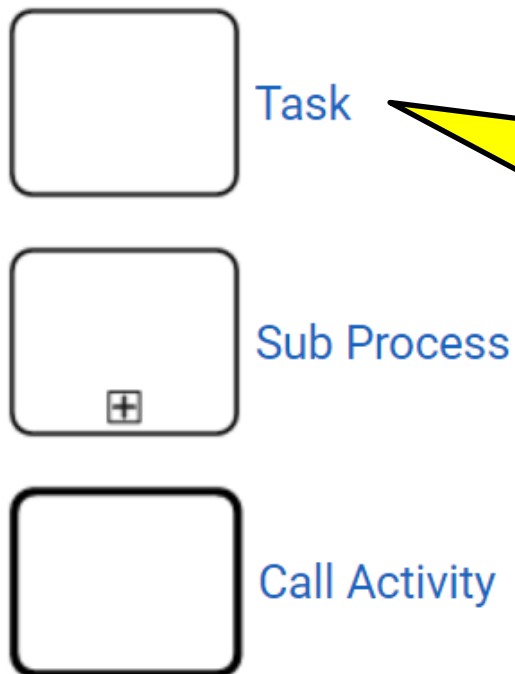
Representa algo que acontece entre os eventos inicial e final. Por exemplo, uma tarefa pode fluir para um evento que lança uma mensagem para outro *pool*, onde um evento subsequente espera para capturar a resposta antes de continuar.

# BPMN – Event

- Start Event
- ◉ Intermediate Event
- End Event

Um evento que indica onde um caminho no processo terminará. Representa o resultado de um processo.

# BPMN – Activity



Uma tarefa representa uma unidade de trabalho que não pode ser dividida em um nível adicional de detalhamento do processo de negócios. **É uma atividade atômica.** Uma tarefa é a atividade de nível mais baixo ilustrada em um diagrama de processo. Um conjunto de tarefas pode representar um procedimento de alto nível.

# BPMN – Activity



Task



Sub Process



Call Activity

Usado para ocultar ou revelar níveis adicionais de detalhes do processo de negócios. **Um subprocesso é uma atividade composta.** Tem seus próprios eventos de início e fim independentes.



# BPMN – Activity



Task










Sub Process



Call Activity


Um ponto no processo em que um processo global ou uma tarefa global é reutilizada.


# BPMN – Gateway

-  Exclusive Gateway - without Marker
-  Exclusive Gateway - with Marker
-  Inclusive Gateway
-  Parallel Gateway
-  Complex Gateway
-  Event-Based Gateway
-  Event-Based Gateway to Start a Process
-  Parallel Event-Based Gateway to Start a Process

Usado para criar fluxos alternativos em um processo. Como apenas um dos caminhos pode ser percorrido, ele é chamado de exclusivo.

# BPMN – Gateway


 Exclusive Gateway - without Marker


 Exclusive Gateway - with Marker


 Inclusive Gateway

 Parallel Gateway

 Complex Gateway


 Event-Based Gateway


 Event-Based Gateway to Start a Process

 Parallel Event-Based Gateway to Start a Process

Usado para criar fluxos alternativos em um processo onde todos os caminhos são considerados.


# BPMN – Gateway


 Exclusive Gateway - without Marker


 Exclusive Gateway - with Marker


 Inclusive Gateway

 Parallel Gateway

 Complex Gateway

 Event-Based Gateway


 Event-Based Gateway to Start a Process

 Parallel Event-Based Gateway to Start a Process

Usado para criar fluxos paralelos sem qualquer avaliação.

# BPMN – Gateway


 Exclusive Gateway - without Marker


 Exclusive Gateway - with Marker


 Inclusive Gateway

 Parallel Gateway

 Complex Gateway


 Event-Based Gateway


 Event-Based Gateway to Start a Process

 Parallel Event-Based Gateway to Start a Process

Usado para modelar um comportamento de sincronização complexo.

# BPMN – Gateway


 Exclusive Gateway - without Marker


 Exclusive Gateway - with Marker


 Inclusive Gateway

 Parallel Gateway

 Complex Gateway

 Event-Based Gateway

 Event-Based Gateway to Start a Process

 Parallel Event-Based Gateway to Start a Process

Dois processos paralelos são iniciados com base em um evento, mas não há avaliação do evento.

# BPMN – Flow

—————→ Sequence Flow

○-----▷ Message Flow

..... Association

.....> Data Association

Mostra em que ordem as atividades são executadas. O fluxo de sequência também pode ter um símbolo no início, um pequeno losango indica um dos vários fluxos condicionais de uma atividade, enquanto uma barra diagonal indica o fluxo padrão de uma decisão ou atividade com fluxos condicionais.

# BPMN – Flow

—————→ Sequence Flow

○-----▷ Message Flow

..... Association

.....> Data Association

Mostra quais mensagens fluem através dos limites organizacionais (ou seja, entre pools). Um fluxo de mensagens nunca pode ser usado para conectar atividades ou eventos no mesmo conjunto.



# BPMN – Flow

—————→ Sequence Flow

○ - - - - - ▷ Message Flow

..... Association

..... ▷ Data Association

É usado para associar um Artefato ou texto a um Objeto de Fluxo .

# BPMN – Flow

—————→ Sequence Flow

○-----▷ Message Flow

..... Association

.....> Data Association

É usado para associar um objeto de dado a um Objeto de Fluxo .

# Ferramentas de modelagem

- Bizagi Modeler
- Draw.io
- Microsoft Visio
- outras

# Mais recursos de aprendizagem

- BPMN Quick Guide <https://www.bpmnquickguide.com/>
- BPMN 2.0 Symbol Reference <https://camunda.com/bpmn/reference/>

# Exercícios

## SQOOP IMPORT

- Primeira fase – o Sqoop lê os metadados da tabela de origem para que sejam convertidos em um tipo de dados Java durante a criação de uma classe que encapsulará um registro dessa tabela
- Segunda fase – nessa fase os dados são lidos da base de origem pelo Sqoop e solicita ao Hadoop a gravação dos dados em HDFS

