

Cenários de Uso · Sistema Cartorial · R5

Autor: Leonardo Menezes **Cliente:** 27º Tabelionato de Notas de São Paulo **Escopo:** MVP back-office puro · arquitetura R5 **Versão:** R5 · 2026-05-12 **Status:** versão final · base para UAT e aceite contratual

1. Visão geral

O Sistema Cartorial (Stoix) é uma plataforma de back-office que auxilia o Tabelião e seus escreventes em três frentes: redação de minutas de atos a partir de modelos curados, validação cruzada com a base normativa do próprio cartório e cálculo do emolumento devido conforme a tabela oficial vigente. A plataforma **não substitui o ato notarial**: a lavratura permanece sendo feita em livro físico, com assinatura presencial das partes e do Tabelião. O software é instrumento de apoio à decisão — quem decide é sempre o profissional.

1.1 Os quatro módulos curados pelo Tabelião

MÓDULO	CONTEÚDO	QUEM ALIMENTA	QUEM USA EM PRODUÇÃO
Knowledge Base (KB)	Provimentos CGJ-SP, NSCGJ-SP, jurisprudência selecionada, doutrina	Tabelião sobe PDFs oficiais, IA extrai, Tabelião aprova	Pipeline de geração de minuta (RAG) e busca manual
Tabela de Emolumentos	Tabela CGJ-SP anual, com FUNDREJUS e ISS como linhas próprias	Tabelião sobe XLSX oficial, IA estrutura, Tabelião ativa versão	Cálculo de emolumento no fim do Fluxo B
Catálogo de Modelos	Modelos de ato (Escritura V&C, Procuração, Ata, etc.)	Tabelião monta no wizard de 4 passos do Fluxo A	Escrevente seleciona modelo no início do Fluxo B
Procedimentos	Wiki interna em markdown, runbooks e checklists	Tabelião ou escrevente sênior, com histórico das 10 últimas versões	Consulta operacional do dia a dia

1.2 Os dois fluxos operacionais

Fluxo A — Tabelião cria ou atualiza modelo de ato. Wizard de quatro passos: upload do modelo (DOCX ou edição inline), revisão das sugestões de IA, parametrização de regras e

variáveis, aprovação final com desafio MFA TOTP. O resultado é uma `act_template_version` imutável, com `act_template_version_hash` calculado sobre o markdown normalizado.

Fluxo B — Escrevente lavra um ato. Seleção do modelo do catálogo, upload dos documentos das partes (RG, CPF, certidões, matrícula), OCR via Textract, preenchimento das variáveis pela LLM (Bedrock Sonnet 4.5 primário, Haiku 4.5 fallback), validações locais e alertas baseados na KB, geração da minuta final e cálculo do emolumento. O escrevente imprime a minuta e o ato é lavrado em livro.

1.3 As quatro personas

- **Lucas C. — Tabelião (notary).** Único responsável pela curadoria normativa, aprovação de modelos, ativação de novas tabelas de emolumentos, decisões de cobrança em caso de divergência. Toda ação sensível pede MFA TOTP.
- **Maria S. e João P. — Escreventes (clerks).** Executam o Fluxo B, atendem o público, geram minutas. Acesso reduzido — apenas cinco itens no menu lateral (Dashboard, Novo ato, Catálogo de modelos, Procedimentos, Sair).
- **DPO — Encarregado pela LGPD.** No 27º Tabelionato SP o DPO é o próprio Tabelião por força do Provimento CNJ 134/2022, com canal técnico operado por escrevente designado. Atende `lgpd_request`, exporta dados de titulares, autoriza anonimização e acompanha auditoria.
- **Stoix Ops.** Equipe Stoix. Provisionamento de tenant, suporte, atualizações de software, monitoração de SLA. Não acessa dados do tenant em operação normal.

1.4 Princípios operacionais inegociáveis

- **Tabelião é o curador.** A Stoix não decide qual provimento é o vigente, qual modelo é admissível ou qual emolumento se aplica. A Stoix entrega a plataforma e garante fidelidade técnica.
- **Auditoria imutável.** Cada evento sensível é gravado em `audit_event` com hash chain HMAC e dump diário cruzado em bucket master Stoix. O livro perpétuo de minutas finais fica em S3 Object Lock por 30 anos.
- **Billing reproduzível.** Cada `billing_event` carrega três snapshots imutáveis: a versão do modelo, a versão da tabela de emolumentos e o snapshot das consultas RAG que apoiaram a redação. Em fiscalização, a cobrança é reproduzível mesmo anos depois.
- **MFA em decisões sensíveis.** Toda aprovação que muda a base regulatória ou econômica do cartório (KB, fee schedule, modelo) exige TOTP recente (≤ 5 minutos).
- **PII fora da LLM.** Antes de qualquer chamada Bedrock, o `LlmClient` aplica `PIIScrubber.strip()` no prompt — CPF, RG, e-mail, telefone, endereço.

2. Cenários · Onboarding

2.1 Provisionamento de tenant para o 27º Tabelionato SP

Persona: Stoix Ops **Pré-condição:** contrato e DPA assinados; slug `notary-27sp` aprovado pelo cliente; chave KMS dedicada já provisionada na conta AWS Stoix. **Gatilho:** ticket de provisionamento aberto após assinatura, com payload mínimo: razão social, CNPJ, comarca, e-mail do Tabelião, e-mail do DPO.

Passos: 1. Operador Stoix executa o stack Pulumi `stoix-cartorio-tenant` com a variável `tenant_slug=27sp`. 2. Pulumi cria os recursos isolados do tenant: RDS `notary-27sp-prd` (db.t4g.micro multi-AZ, pgvector e pgcrypto habilitados), buckets S3 `notary-27sp-uploads`, `notary-27sp-kb`, `notary-27sp-fee-schedule` e `notary-27sp-livro-perpetuo` (este último com Object Lock Compliance 30y), filas SQS Standard (`ocr-queue`, `llm-queue`, `kb-extract-queue`, `fee-extract-queue`) e tópicos SNS de alerta. 3. Pulumi cria o realm `notary-27sp` no Keycloak shared, com política de senha forte e MFA TOTP obrigatório. 4. Pulumi roda as migrações Flyway no RDS dedicado, criando o schema completo (tabelas mutáveis e tabelas append-only com `audit_event`). 5. Pulumi emite o evento `tenant.provisioned` no log master Stoix. 6. Operador cria manualmente o usuário inicial do Tabelião no realm `notary-27sp` com perfil `notary`, senha temporária e flag `must_change_password=true`. E-mail de boas-vindas é disparado pela SES com link para a tela `01-login`.

Resultado esperado: tenant operacional, sem dados, com Tabelião capaz de fazer o primeiro login. `audit_event` do tenant tem 0 linhas e cadeia HMAC inicializada com `prev_hash = NULL`.

Eventos emitidos: `tenant.provisioned`, `auth.realm_admin_action` (criação do usuário inicial).

Métricas operacionais: - Tempo esperado: 25 minutos do `pulumi up` à entrega do link de acesso. - Indicadores de sucesso: stack Pulumi sem drift, MFA challenge alcançável, primeira query no RDS dedicado retornando schema completo.

Variações: - Se o Pulumi falhar em qualquer recurso, o stack é destruído integralmente (`pulumi destroy`) e o operador reabre o ticket — não se entrega tenant parcialmente provisionado. - Se o slug `notary-27sp` já existir em outro tenant, o pipeline aborta com erro de unicidade.

2.2 Tabelião acessa pela primeira vez e configura MFA

Persona: Lucas C. (Tabelião) **Pré-condição:** tenant provisionado (cenário 2.1); usuário criado com senha temporária; e-mail de boas-vindas recebido. **Gatilho:** Tabelião clica no link do e-mail.

Passos: 1. Tabelião abre a tela `01-login` e insere e-mail e senha temporária. 2. Sistema valida credenciais no Keycloak realm `notary-27sp` e redireciona para a tela de troca obrigatória de senha. 3. Tabelião define nova senha (mínimo 12 caracteres, mistura de classes, contra dicionário de senhas vazadas). 4. Sistema apresenta o enrollment de MFA TOTP: QR code para Google Authenticator, Authy ou similar, com fallback em código manual. 5. Tabelião escaneia, digita o código TOTP de seis dígitos para confirmar e o sistema marca `mfa_enrolled=true`. 6. Tabelião é redirecionado para a tela `02-dashboard-tabeliao`, que exibe os quatro módulos com indicadores zerados (0 modelos, 0 itens na KB, 0 tabelas de emolumento, 0 procedimentos).

Resultado esperado: sessão ativa, MFA TOTP habilitado, senha definitiva. Próximas operações sensíveis pedem TOTP novo.

Eventos emitidos: `auth.user_logged_in` (com `mfa_used=false` no primeiro acesso), `auth.mfa_enrolled`, `auth.user_logged_in` (com `mfa_used=true` após enrollment).

Métricas operacionais: - Tempo esperado: 4 minutos do clique no e-mail ao dashboard. - Indicadores de sucesso: enrollment TOTP confirmado, dashboard renderizado sem erro.

Variações: - Se o Tabelião digitar TOTP errado três vezes no enrollment, sistema reinicia o processo do passo 4 sem bloquear conta. - Se o link do e-mail expirar (TTL 72 horas), Stoix Ops reemite com auditoria do motivo.

2.3 Tabelião sobe o primeiro normativo na Knowledge Base

Persona: Lucas C. (Tabelião) **Pré-condição:** Tabelião logado; nenhuma `kb_category` cadastrada ainda. **Gatilho:** Tabelião decide subir o Provimento CGJ-SP nº 58/2024 (atos notariais).

Passos: 1. No dashboard `02-dashboard-tabeliao`, Tabelião clica em “Knowledge Base”. 2. Sistema abre a tela `23-kb-dashboard`, sem itens. Há um banner orientando criar primeiro uma categoria. 3. Tabelião clica em “Gerenciar categorias” e abre `22-kb-admin-categorias`. 4. Tabelião cria três categorias iniciais: “Provimentos CGJ-SP” (peso 0,9), “Jurisprudência CSM/CGJ” (peso 0,7), “Doutrina e referências” (peso 0,5). O peso pondera o rerank final do RAG sobre o score de similaridade. 5. Tabelião volta para `23-kb-dashboard` e clica em “Subir documento”. 6. Tela `24-kb-upload-passo-1` pede categoria (seleciona “Provimentos CGJ-SP”), arquivo (PDF oficial do Provimento 58/2024), vigência (`effective_from = 2024-09-01`, `effective_until = 2099-12-31` para vigência indeterminada). Tabelião confirma. 7. Sistema cria `kb_document`, salva PDF em `notary-27sp-kb/`, enfileira mensagem em `kb-extract-queue` e mostra `25-kb-upload-passo-2` com status “Em extração”. 8. `kb-extract-worker` consome a mensagem, executa Textract, segmenta o texto em chunks hierárquicos (Tier 1 por Art./§/Inciso), gera embeddings com `amazon.titan-embed-text-v2` em `sa-east-1`, persiste `kb_chunk` no pgvector HNSW (`m=16`, `ef_construction=64`). 9. Quando termina, sistema notifica o Tabelião via SES. Ao reabrir, tela `26-kb-extracao` mostra preview lado a lado: PDF original à esquerda, chunks estruturados à direita, com diff visual destacando títulos, parágrafos e referências. 10. Tabelião revisa, ajusta dois chunks (correção de quebra de seção), clica “Aprovar versão”. 11. Sistema pede desafio MFA TOTP. Tabelião digita o código, sistema aprova. 12. `kb.version_approved` é registrado no `audit_event` com hash chain. Índice pgvector marca a versão como `status=active`. Tela `28-kb-detalle-versao` mostra a versão vigente.

Resultado esperado: uma `kb_category` “Provimentos CGJ-SP” peso 0,9 com uma `kb_version` ativa contendo o Provimento 58/2024, indexada e disponível para o RAG do Fluxo B.

Eventos emitidos: `kb.document_uploaded`, `kb.extraction_started`, `kb.extraction_completed`, `auth.mfa_challenge_passed`, `kb.version_approved`.

Métricas operacionais: - Tempo esperado: 8 a 12 minutos para PDF de até 40 páginas (Textract p95 ≤ 6s/pág). - Indicadores de sucesso: revisão diff visual sem ajustes massivos (>30% de chunks editados sinaliza falha de extração).

Variações: - Se Textract retornar `confidence_avg < 0,85`, sistema emite `kb.extraction_failed`, manda alerta SES ao Tabelião e mantém o documento como rascunho até reprocessamento manual. -

Se o Tabelião rejeitar a extração, registra `kb.version_rejected` com `reason` livre e o documento volta para a fila de extração com novo tentativa.

2.4 Tabelião sobe a primeira tabela de emolumentos (XLSX CGJ-SP)

Persona: Lucas C. (Tabelião) **Pré-condição:** Tabelião logado; nenhuma `fee_schedule` ativa. **Gatilho:** Tabelião precisa habilitar cálculo de emolumento para iniciar atendimento ao público.

Passos: 1. No dashboard, Tabelião clica em “Tabela de Emolumentos”, abrindo `20-tabela-emolumentos`. Tela mostra “Nenhuma tabela vigente — atos econômicos bloqueados”. 2. Tabelião clica em “Subir nova versão”, abrindo `21-nova-versao-emolumentos`. 3. Tabelião seleciona XLSX oficial CGJ-SP 2026, indica `effective_from = 2026-01-01` e `effective_until = 2026-12-31`, descreve a fonte (“Provimento CG nº 38/2025 — Anexo I”). 4. Sistema salva o arquivo em `notary-27sp-fee-schedule/`, cria `fee_schedule` e enfileira mensagem em `fee-extract-queue`. 5. `fee-schedule-extract-worker` parseia o XLSX, normaliza linhas com `(act_code, value_range)` no padrão `[min, max)` (fechado em lower, aberto em upper), separa FUNDREJUS e ISS como linhas próprias e popula `fee_schedule_line`. Emite `fee_schedule.extracted` com `version_id_draft`. 6. Sistema mostra preview tabular ao Tabelião: cada linha com `act_code`, `value_range`, valor base, FUNDREJUS, ISS. Tabelião revisa amostras críticas (faixas de maior volume — Escritura V&C, Procuração, Reconhecimento de firma). 7. Tabelião clica “Ativar versão”. Sistema pede MFA TOTP. Tabelião digita código. 8. `fee_schedule.version_activated` é registrado com hash chain. Versão fica vigente; cobrança no Fluxo B liberada.

Resultado esperado: `fee_schedule_version` ativa com hash sha256 do conteúdo normalizado, `effective_from = 2026-01-01`, lookup exato `(act_code, value_range)` disponível para a `BillingService`.

Eventos emitidos: `fee_schedule.uploaded`, `fee_schedule.extracted`, `auth.mfa_challenge_passed`, `fee_schedule.version_activated`.

Métricas operacionais: - Tempo esperado: 6 a 10 minutos do upload à ativação. - Indicadores de sucesso: 100% das linhas com `act_code` reconhecido pela CGJ-SP e `value_range` consistente sem sobreposições.

Variações: - Se a extração detectar `value_range` sobreposto (mesma `act_code` com faixas que se sobrepõem), sistema sinaliza erro estrutural e bloqueia ativação até correção. - Se houver linha com `act_code` desconhecido pela master list CGJ-SP, sistema sinaliza como `warning` mas permite ativação após confirmação explícita.

2.5 Tabelião cria o primeiro modelo de ato (Escritura de V&C)

Persona: Lucas C. (Tabelião) **Pré-condição:** KB com ao menos um provimento aprovado (cenário 2.3); `fee schedule` ativa (cenário 2.4). **Gatilho:** Tabelião quer ter o modelo de Escritura de Venda e Compra pronto antes do primeiro atendimento.

Passos: 1. No dashboard, Tabelião clica em “Catálogo de Modelos”, abrindo `08-catalogo-atos`. Tela vazia, com botão “Novo modelo”. 2. Tabelião clica e inicia o Fluxo A, passo 1, na tela `04-fluxo-a-01-novo-ato-upload-modelo`. Escolhe entre upload DOCX ou edição inline. Faz upload de um DOCX base

de Escritura V&C que já usava no cartório. 3. Sistema converte DOCX para markdown normalizado (Unicode NFC, LF, trim trailing whitespace), identifica variáveis candidatas com colchetes ou marcadores conhecidos. 4. Tela `05-fluxo-a-02-revisao-sugestoes-ia` apresenta sugestões da IA: variáveis detectadas (`[nome_vendedor]`, `[nome_comprador]`, `[matricula]`, `[valor]`, etc.), cláusulas opcionais (resolutiva, alienação fiduciária), referências a provimentos relevantes (vindas da KB via RAG hermenêutico — K=20, N=5). Tabelião aceita 14 variáveis e rejeita 2 (que não fazem sentido na prática do 27º). 5. Tela `06-fluxo-a-03-parametros-regras` permite configurar regras de validação por variável: `valor` é numérico positivo, `cpf_vendedor` é CPF válido (DV), `matricula` é alfanumérico não vazio, e `data_nascimento_*` deve estar no passado. 6. Tabelião marca o `act_code` correspondente (“Escritura de Venda e Compra de Imóvel”), seleciona a faixa-padrão de valor para fins de billing. 7. Tela `07-fluxo-a-04-aprovacao-final-publicacao` mostra preview final + diff entre o DOCX original e o modelo parametrizado. Tabelião revisa e clica “Publicar versão”. 8. Sistema pede MFA TOTP. Tabelião digita código. 9. Sistema calcula `act_template_version_hash` (sha256 do markdown normalizado), persiste `act_template_version` imutável, emite `act_template_version.published` com hash chain.

Resultado esperado: `act_template` “Escritura de Venda e Compra de Imóvel” com `act_template_version` v1 publicada, parametrizada, pronta para o Fluxo B.

Eventos emitidos: `act_template.created`, `act_template_version.drafted` (a cada passo do wizard), `auth.mfa_challenge_passed`, `act_template_version.published`.

Métricas operacionais: - Tempo esperado: 25 a 40 minutos para o primeiro modelo (curva de aprendizado). - Indicadores de sucesso: hash gerado e persistido; modelo aparece no catálogo `08-catalogo-atos`.

Variações: - Se o Tabelião abandonar o wizard antes do passo 4, o rascunho fica salvo em `act_template_version.drafted` e pode ser retomado por 30 dias. - Se houver conflito de `act_code` (já existe modelo ativo com o mesmo código), sistema oferece deprecar a versão anterior ou criar nova versão dentro do mesmo `act_template`.

3. Cenários · Operação diária do Tabelião

3.1 Aprovar versão de modelo no Fluxo A com MFA TOTP

Persona: Lucas C. (Tabelião) **Pré-condição:** modelo “Procuração Pública” em rascunho no wizard, passo 4. **Gatilho:** Tabelião terminou o ajuste e quer publicar.

Passos: 1. Na tela `07-fluxo-a-04-aprovacao-final-publicacao`, Tabelião revê diff e clica “Publicar versão”. 2. Modal de MFA TOTP abre, pedindo código de 6 dígitos. A sessão MFA tem janela máxima de 5 minutos a partir da última `auth.mfa_challenge_passed`. 3. Tabelião digita o código. Sistema valida no Keycloak. 4. Se válido, sistema persiste a versão, calcula `act_template_version_hash`, emite `act_template_version.published` com hash chain HMAC. Modal fecha, tela mostra “Versão v2 publicada”.

Resultado esperado: versão imutável publicada; aparece no catálogo. Versão v1 (se existir) permanece histórica, não é apagada.

Eventos emitidos: `auth.mfa_challenge_passed` , `act_template_version.published` .

Métricas operacionais: - Tempo esperado: menos de 30 segundos do clique à confirmação. - Indicadores de sucesso: hash persistido; cadeia HMAC válida no self-check noturno.

Variações: - TOTP errado: sistema registra `auth.mfa_challenge_failed` , exibe mensagem genérica (“código inválido”), permite até 3 tentativas antes de pedir novo login. - Sessão MFA expirada: sistema reabre modal de login MFA e, após sucesso, retoma o passo 4.

3.2 Subir nova versão de modelo já publicado

Persona: Lucas C. (Tabelião) **Pré-condição:** modelo “Escritura V&C” v1 publicado. **Gatilho:** Tabelião quer adicionar cláusula resolutive opcional ao modelo.

Passos: 1. No catálogo `08-catalogo-atos` , Tabelião clica no modelo “Escritura V&C”, abrindo detalhe com histórico de versões. 2. Tabelião clica “Nova versão”. Sistema clona o conteúdo da v1 como rascunho v2 e inicia o wizard no passo 2 (revisão de sugestões IA). 3. Tabelião adiciona a cláusula resolutive, cria variável `flag_clausula_resolutiva` (booleana, default `false`) e variáveis condicionais `condicao_resolutiva_texto` e `prazo_resolutiva_dias` . 4. Passa pelos passos 3 e 4, aprova com MFA. 5. Sistema persiste v2; v1 vira histórica mas continua acessível para reproduzir billings antigos.

Resultado esperado: duas versões coexistem em `act_template_version` ; somente v2 é selecionável para novos atos no Fluxo B; `billing_event` antigos referenciam v1.

Eventos emitidos: `act_template_version.drafted` , `auth.mfa_challenge_passed` , `act_template_version.published` .

Métricas operacionais: - Tempo esperado: 10 a 15 minutos para diff incremental sobre versão existente. - Indicadores de sucesso: v1 permanece consultável; novos atos pegam v2.

Variações: - Se a v2 contém regra de validação incompatível com atos em andamento (improvável no MVP, pois Fluxo B usa snapshot da versão no momento do `act.started`), sistema apenas registra a publicação — atos abertos não são forçados a migrar.

3.3 Atualizar a KB com novo provimento CGJ-SP

Persona: Lucas C. (Tabelião) **Pré-condição:** KB já com `kb_category` “Provimentos CGJ-SP”. **Gatilho:** CGJ-SP publica Provimento 12/2026 modificando emolumentos de averbação.

Passos: 1. Tabelião abre `23-kb-dashboard` , clica “Subir documento”, seleciona categoria “Provimentos CGJ-SP”, carrega o PDF e define `effective_from = 2026-04-01` , `effective_until = 2099-12-31` . 2. Pipeline de extração roda igual ao cenário 2.3. 3. Ao aprovar a nova versão, sistema emite `kb.version_approved` para o novo documento. Como o Provimento 12/2026 não substitui um anterior, não há `kb.version_superseded` automático — coexistem. 4. Se o Tabelião quiser explicitamente marcar que um provimento anterior foi revogado, ele edita a versão antiga e ajusta `effective_until` para a data da revogação, emitindo `kb.version_superseded` .

Resultado esperado: novo provimento indexado e disponível ao RAG. Em consultas RAG futuras, o reranker pondera com base no `kb_category.weight` (0,9 para provimentos).

Eventos emitidos: `kb.document_uploaded`, `kb.extraction_completed`, `kb.version_approved`, opcionalmente `kb.version_superseded`.

Métricas operacionais: - Tempo esperado: igual a 2.3 (8 a 12 min). - Indicadores de sucesso: chunks aparecem em consulta manual `23-kb-dashboard` com filtros de busca.

Variações: - Se o PDF estiver protegido por senha ou OCR-incompatível, Textract retorna `confidence_avg` muito baixo. Tabelião precisa reabrir o PDF, fazer OCR localmente e resubmeter.

3.4 Adicionar nova categoria de KB com peso

Persona: Lucas C. (Tabelião) **Pré-condição:** Tabelião logado. **Gatilho:** Tabelião quer separar súmulas do CSM como categoria própria.

Passos: 1. Tabelião abre `22-kb-admin-categorias`. 2. Clica “Nova categoria”, informa nome “Súmulas CSM-SP” e peso 0,8. 3. Sistema persiste `kb_category` com `is_active=true`, emite `kb.category_weight_updated` (com `old_weight=NULL`) sob hash chain. 4. Categoria fica disponível para receber documentos no próximo upload.

Resultado esperado: categoria criada, peso aplicável imediatamente em queries futuras de RAG.

Eventos emitidos: `kb.category_weight_updated`.

Métricas operacionais: - Tempo esperado: menos de 1 minuto. - Indicadores de sucesso: categoria visível em `24-kb-upload-passo-1`.

Variações: - Se o Tabelião editar o peso de categoria existente, evento `kb.category_weight_updated` registra `old_weight` e `new_weight`. Mudança afeta apenas consultas RAG **futuras** — `kb_consultations_snapshot` em billings passados continua representando a realidade do momento do ato.

3.5 Consultar auditoria de billing do mês

Persona: Lucas C. (Tabelião) **Pré-condição:** mês com pelo menos um `billing_event` consolidado. **Gatilho:** virada de mês — Tabelião quer revisar fatura interna antes de pagar emolumentos ao Estado.

Passos: 1. Tabelião abre `15-relatorios-billing`. Filtra por mês competência (abril/2026). 2. Sistema lista todos os `billing_event` do período: data, `act_code`, modelo usado (com `act_template_version_hash` truncado), versão da tabela de emolumentos, valor base, FUNDREJUS, ISS, total. 3. Tabelião pode clicar em qualquer linha para ver o `kb_consultations_snapshot` que apoiou o ato — lista das consultas RAG (oversample K=20, com flag `in_top_n` para os que entraram no top-5 final) e hashes das versões de KB usadas. 4. Tabelião exporta CSV para conferência contábil.

Resultado esperado: Tabelião valida que cada cobrança é reproduzível e que os snapshots batem com o que esperava.

Eventos emitidos: nenhum evento de domínio — leitura apenas. Acesso é registrado em log de aplicação separado (não na hash chain).

Métricas operacionais: - Tempo esperado: relatório p95 < 2s para até 500 atos/mês. - Indicadores de sucesso: total mensal bate com soma das linhas; nenhum `billing_event` com `kb_consultations_snapshot` vazio (gate aplicativo impede a consolidação se RAG não foi consultado).

Variações: - Se houver divergência entre o relatório e a contabilidade do Tabelaio, ele abre ticket de suporte (cenário 6.3).

4. Cenários · Operação diária do Escrevente

4.1 Lavrar Escritura de V&C com Cláusula Resolutiva (cenário completo)

Persona: Maria S. (Escrevente) **Pré-condição:** modelo “Escritura V&C” v2 publicado (cenário 3.2); KB com Provimento CGJ-SP 58/2024 e Provimento 12/2026; fee schedule 2026 ativa; partes presentes no balcão com documentos físicos. **Gatilho:** atendimento ao balcão — vendedor José da Silva e comprador Ana Pereira chegam para lavrar Escritura de Venda e Compra de imóvel matrícula 12.345 do 18° RI-SP, valor R\$ 850.000, com cláusula resolutiva pelo prazo de 60 dias condicionada à quitação de hipoteca anterior.

Passos: 1. Maria abre `03-dashboard-escrevente`. Menu lateral mostra os cinco itens: Dashboard, Novo ato, Catálogo de modelos, Procedimentos, Sair. 2. Clica “Novo ato”, abrindo `09-fluxo-b-01-novo-documento-selecao-ato`. 3. Tela apresenta busca e filtro por `act_code`. Maria digita “venda” e o catálogo filtra para “Escritura de Venda e Compra de Imóvel”. Seleciona o modelo, sistema mostra `act_template_version_hash` (v2) e descrição resumida. 4. Maria clica “Iniciar ato”. Sistema cria `act_instance` com `act_template_version_id` referenciando o snapshot da v2, emite `act.started`, abre `10-fluxo-b-02-upload-docs-ocr`. 5. Tela `10` mostra grid de upload de documentos: RG vendedor, CPF vendedor, comprovante endereço vendedor, RG comprador, CPF comprador, comprovante endereço comprador, matrícula atualizada do imóvel, certidão negativa do imóvel. Maria escaneia os documentos físicos com o scanner de mesa. 6. Cada upload cria um `document` no sistema, salva em `notary-27sp-uploads/`, enfileira mensagem em `ocr-queue`. Sistema mostra status individual (“Em OCR”, “Concluído”, barra de progresso). 7. `ocr-worker` consome mensagens da fila, chama Textract em `sa-east-1`, retorna texto + `confidence_avg` por documento. Para os 8 documentos: 7 voltam com `confidence_avg ≥ 0,93`; a matrícula volta com `0,88` (manuscrito em averbação manual). Sistema marca a matrícula com tag “Confiança média — revisar”. 8. Maria clica em “Avançar”. Tela `11-fluxo-b-03-preenchimento-llm-revisao` abre. Sistema empacota o texto OCR (sem PII — `PIIScrubber.strip()` remove CPF, RG, telefone, endereço antes do prompt), envia ao Bedrock Sonnet 4.5 em `sa-east-1` com o template parametrizado, recebe preenchimento das variáveis. Tempo decorrido ≤ 4s (p95). 9. Tela mostra o markdown final com as variáveis preenchidas em destaque visual. Maria revisa: nome do vendedor (correto), CPF vendedor (correto após decrypt na camada de serviço), nome do comprador (correto), CPF comprador (correto), matrícula 12.345 (correto), área do imóvel 87 m² (correto), valor R\$ 850.000 (correto). 10. Maria ativa o toggle `flag_clausula_resolutiva = true`. Tela revela as variáveis condicionais. Maria digita: `condicao_resolutiva_texto = "Quitação integral do saldo devedor da hipoteca registrada na matrícula 12.345, R-7"`, `prazo_resolutiva_dias = 60`. 11. Sistema reabre a chamada LLM

(novo `act.llm_fill_started` com `revision_n = 2`) para incorporar as novas variáveis ao texto, mantendo as anteriores. Volta em 3s. 12. Tela `12-fluxo-b-04-alertas-validacoes` é acionada automaticamente conforme a revisão progride. Sistema executa: - Validações locais: CPF com DV correto (vendedor e comprador), datas coerentes (vendedor não pode ter idade < 18), prazo da cláusula resolutive > 0. - RAG hermenêutico (K=15, N=3 para Fluxo B preenchimento): consulta KB sobre cláusula resolutive em escrituras de V&C. Recupera trechos do Provimento 58/2024 art. 17 e súmula CSM-SP nº 9. Sistema avalia: cláusula está conforme. - Alerta visual: “Cláusula resolutive com prazo > 30 dias — verifique necessidade de registro complementar conforme art. 17 §2º do Prov. 58/2024” (alerta, não bloqueio). 13. Maria lê o alerta, confirma com o Tabelião verbalmente (decisão notarial — não é decisão da máquina), marca “Ciente” no alerta. 14. Maria clica “Gerar minuta final”. Sistema vai para a tela `14-fluxo-b-06-minuta-final-emolumento`. 15. Sistema executa, em transação ACID: - Marca `act.human_reviewed` (com hash chain) — gate de consolidação cumprido. - Renderiza a minuta final em PDF, salva em `notary-27sp-livro-perpetuo/` com S3 Object Lock Compliance 30y, calcula `content_hash`. - Emite `act.consolidated` com `final_minuta_s3_key`, `content_hash`, `act_template_version_id` (snapshot da v2 do momento do `act.started`). - `BillingService` inline (sem fila separada — decisão R5.1) executa o lookup exato no `fee_schedule_line` por `(act_code, value_range)` para R\$ 850.000: encontra a linha aplicável, calcula valor base, FUNDREJUS, ISS (cada um como linha separada da tabela), totaliza. - Fecha o `kb_consultations_snapshot`: lista de todas as consultas RAG do ato (oversample K=15 ou K=20 dependendo do tipo), com `{kb_chunk_id, kb_version_hash, score, in_top_n: bool}`. Tamanho típico ~50KB/ato. - Persiste `billing_event` com os três snapshots imutáveis (`act_template_version_id + hash`, `fee_schedule_version_id + hash`, `kb_consultations_snapshot`). Emite `billing_event.computed` com hash chain. - Tudo na mesma TX SQL: ou ambos os eventos comitam, ou nenhum. Decisão R5 #3. 16. Tela `14` mostra: minuta final em preview (botão “Imprimir”), valor total do emolumento detalhado (base + FUNDREJUS + ISS), botões “Imprimir minuta”, “Gerar guia de recolhimento”, “Marcar como lavrado”. 17. Maria imprime a minuta em duas vias (livro e arquivo do cartório), entrega ao Tabelião para conferência presencial, colhe assinaturas físicas das partes e do Tabelião. 18. Maria volta à tela `14`, clica “Marcar como lavrado”. Sistema apenas atualiza um campo de UI; o livro perpétuo digital já foi consolidado no passo 15 e é imutável.

Resultado esperado: ato lavrado em livro físico; minuta consolidada e imutável em S3 Object Lock 30y; `billing_event` registrado com snapshots reproduzíveis; relatório do mês incorpora a cobrança automaticamente.

Eventos emitidos: `act.started`, `document.uploaded` (×8), `document.ocr_completed` (×8), `act.llm_fill_started` (×2, revisões 1 e 2), `act.llm_fill_completed` (×2), `act.party_added` (×2 — vendedor e comprador), `act.alert_raised` (cláusula resolutive > 30 dias), `act.human_reviewed`, `act.consolidated`, `billing_event.computed`, e `kb.consulted` em cada query RAG (telemétrico, sem hash chain).

Métricas operacionais: - Tempo esperado end-to-end ≤ 90s (p95) para 3 documentos; ~120-150s para 8 documentos com revisão humana incluída. - Indicadores de sucesso: minuta consolidada com `act_template_version_hash` correto; `billing_event` com `kb_consultations_snapshot` não-nulo; OCR avg `confidence` ≥ 0,90 em pelo menos 6 dos 8 docs.

Variações: - Se o RAG retornar um alerta marcado como `severity=blocking` (caso raro — só ocorre se a KB contém norma que invalida o ato proposto), sistema **bloqueia** a geração da minuta com

mensagem clara da norma violada. Cenário coberto separadamente em 4.4 e 4.5. - Se o `BillingService` falhar no lookup (faixa de valor não coberta pela tabela), sistema reverte a TX inteira — não consolida nem cobra. Maria recebe erro pedindo revisão do valor declarado. Esse é o gate que evita ato lavrado sem cobrança correspondente (cenário 6.3 detalha o tratamento). - Se a sessão MFA expirar entre passos longos, sistema pede reautenticação leve (senha apenas, sem TOTP — TOTP só é exigido em decisões do Tabelião). Maria não tem decisão sensível neste fluxo.

4.2 Reconhecimento de firma e autenticação de cópia

Persona: João P. (Escrevente) **Pré-condição:** modelos “Reconhecimento de firma” e “Autenticação de cópia” publicados; KB ativa. **Gatilho:** cliente apresenta documento com firma a reconhecer.

Passos: 1. João abre `03-dashboard-escrevente`, clica “Novo ato”. 2. Em `09-fluxo-b-01-novo-documento-selecao-ato`, busca “reconhecimento”, seleciona “Reconhecimento de firma por semelhança”, inicia. 3. Sistema cria `act_instance`, emite `act.started`. Abre `10-fluxo-b-02-upload-docs-ocr`. 4. João escaneia o documento físico apresentado pelo cliente e o RG/CPF do titular da firma. Apenas 2 documentos — fluxo é curto. 5. OCR é executado, LLM preenche variáveis mínimas (nome do titular, número do RG, CPF, data do documento). Sem cláusulas condicionais. 6. Tela `12-fluxo-b-04-alertas-validacoes` valida CPF DV; sem alertas RAG (ato simples). 7. João gera minuta, sistema consolida, `BillingService` cobra o emolumento de “Reconhecimento de firma” da tabela vigente. 8. João imprime o selo/etiqueta, cola no documento físico, assina em conjunto com o Tabelião (que conferiu visualmente a firma).

Resultado esperado: ato lavrado; selo apostado no documento físico; `billing_event` gerado.

Eventos emitidos: `act.started`, `document.uploaded` (x2), `document.ocr_completed` (x2), `act.llm_fill_completed`, `act.human_reviewed`, `act.consolidated`, `billing_event.computed`.

Métricas operacionais: - Tempo esperado: 30 a 60 segundos end-to-end. - Indicadores de sucesso: cobrança gerada com valor compatível com a tabela vigente.

Variações: - Para autenticação de cópia, mesmo fluxo com modelo “Autenticação de cópia”. OCR opcional — escrevente pode pular upload se o documento físico é autoexplicativo. - Volume alto: múltiplos reconhecimentos para o mesmo cliente — cada um é `act_instance` separado; o sistema não agrupa.

4.3 Procuração pública

Persona: Maria S. (Escrevente) **Pré-condição:** modelo “Procuração Pública” v1 publicado. **Gatilho:** outorgante presente para conferir poderes a procurador.

Passos: 1. Maria seleciona modelo “Procuração Pública” e inicia ato. 2. Upload de documentos: RG e CPF do outorgante; identificação do outorgado (RG/CPF se presente, ou dados informados pelo outorgante). 3. OCR e LLM preenchem dados das partes. 4. Tela `11-fluxo-b-03-preenchimento-llm-revisao` apresenta blocos de poderes: gerais, específicos (alienação, hipoteca, representação judicial, recebimento de valores). Maria marca os poderes solicitados pelo outorgante. 5. Sistema valida: poderes específicos exigem texto livre que descreva o objeto e o limite (sistema exige preenchimento, não permite continuar com campo vazio). 6. RAG hermenêutico consulta a KB sobre poderes vedados em procuração pública. Se algum poder atípico for selecionado (ex.: “doação a herdeiro necessário sem

autorização judicial”), sistema emite alerta `act.alert_raised`. 7. Maria conclui, gera minuta, consolida ato.

Resultado esperado: ato lavrado; cobrança conforme tabela; trilha de poderes registrada em snapshot.

Eventos emitidos: `act.started`, `document.uploaded` (x2 a 4), `act.llm_fill_completed`, `act.alert_raised` (se atípico), `act.human_reviewed`, `act.consolidated`, `billing_event.computed`.

Métricas operacionais: - Tempo esperado: 5 a 10 minutos com revisão de poderes. - Indicadores de sucesso: poderes específicos com texto descritivo não-nulo; minuta gerada sem alertas bloqueantes.

Variações: - Procuração com substabelecimento: variável adicional, IA preenche se identificada nos documentos.

4.4 Edge — OCR com baixa confiança em matrícula

Persona: Maria S. (Escrevente) **Pré-condição:** ato em andamento com upload de matrícula imobiliária antiga, manuscrita. **Gatilho:** Textract retorna `confidence_avg = 0,78` para a matrícula (abaixo do threshold de aceitação automática 0,85).

Passos: 1. Após o upload, na tela `10-fluxo-b-02-upload-docs-ocr`, a matrícula aparece com tag amarela “Confiança baixa — revisar manualmente”. 2. Maria clica no documento, sistema mostra o PDF da matrícula lado a lado com o texto OCR. Campos com confiança individual baixa estão destacados (números de averbação, nomes próprios em caligrafia irregular). 3. Maria edita os campos críticos no painel direito: corrige nome do proprietário anterior, número de registro de hipoteca, área. Cada edição é registrada como `act.party_added` ou `act.validation_failed` corrigida. 4. Maria salva. Sistema marca o documento como “Revisado manualmente” e prossegue para o passo LLM com o texto corrigido.

Resultado esperado: dados OCR enriquecidos pela revisão humana; LLM recebe versão corrigida; ato consolida normalmente.

Eventos emitidos: `document.ocr_completed` (com confidence baixa), eventos de revisão manual via API BFF (registrados em `audit_event`).

Métricas operacionais: - Tempo esperado: +2 a 5 minutos adicionais ao fluxo padrão. - Indicadores de sucesso: após revisão, o ato segue sem mais alertas de OCR.

Variações: - Se Maria decidir que o documento não é utilizável (página rasgada, ilegível), ela exclui o upload e pede ao cliente nova certidão. `act_instance` permanece em rascunho.

4.5 Edge — CPF inválido detectado pela validação local

Persona: Maria S. (Escrevente) **Pré-condição:** ato em andamento, LLM preenche `cpf_vendedor` com base no OCR do CPF físico. **Gatilho:** validação local com algoritmo de DV detecta CPF inválido (DV não confere com os 9 primeiros dígitos).

Passos: 1. Tela `12-fluxo-b-04-alertas-validacoes` mostra alerta vermelho de validação: “CPF do vendedor inválido — DV não confere”. 2. Sistema **bloqueia** a geração da minuta enquanto o erro persistir. Botão “Gerar minuta” fica desabilitado. 3. Maria verifica o documento físico: o cliente forneceu

CPF errado. Confere com o cliente. 4. Maria corrige o número no painel de revisão. Sistema revalida em tempo real, DV confere, alerta some. 5. Botão “Gerar minuta” reabilita. Fluxo continua.

Resultado esperado: ato só consolida com CPFs com DV válido. Sistema não permite cobrança baseada em dado claramente inconsistente.

Eventos emitidos: `act.validation_failed` (registro do erro), correção via API BFF, eventual revalidação registrada como `act.party_added` atualizado.

Métricas operacionais: - Tempo esperado: +1 a 3 minutos para resolver a inconsistência. - Indicadores de sucesso: validação volta a passar; ato consolida com CPFs válidos.

Variações: - Se o cliente realmente não tiver CPF válido (situação cadastral irregular RFB), sistema não consolida — Maria orienta o cliente a regularizar antes do ato. Não há override para gerar minuta com CPF inválido.

5. Cenários · Manutenção sazonal

5.1 Nova tabela de emolumentos anual (CGJ-SP publica em dezembro)

Persona: Lucas C. (Tabelião) **Pré-condição:** fee schedule 2026 ativa com `effective_until = 2026-12-31`. **Gatilho:** CGJ-SP publica Provimento CG nº 42/2026 com tabela 2027 em 10/dezembro/2026.

Passos: 1. Em 01/dez/2026, sistema dispara `kb-vigencia-worker` (escopo expandido para fee schedule). Detecta `effective_until = 2026-12-31` da tabela ativa; gera alerta proativo (30 dias antes). E-mail SES para o Tabelião e tarefa ClickUp para Stoix Ops. 2. Em 15/dez/2026, novo alerta (15 dias antes). 3. Tabelião baixa o XLSX oficial CGJ-SP 2027. Executa o cenário 2.4 com `effective_from = 2027-01-01`, `effective_until = 2027-12-31`. Ativa com MFA. 4. Em 26/dez/2026, alerta de 5 dias é suprimido porque já existe versão ativa para o período subsequente. 5. Em 01/jan/2027, sistema automaticamente faz transição de versão ativa: `BillingService` passa a usar a versão 2027. Versão 2026 fica histórica, preservada para reproduzir billings de 2026.

Resultado esperado: transição transparente; nenhum ato de jan/2027 cobrado com tabela 2026; nenhum ato de 2026 reprocessado.

Eventos emitidos: `kb.expiry_warning_emitted` (em 30, 15 dias), `fee_schedule.uploaded`, `fee_schedule.extracted`, `fee_schedule.version_activated`, e `fee_schedule.expired` em 01/jan/2027 para a versão antiga.

Métricas operacionais: - Tempo esperado: 15 a 30 minutos de trabalho do Tabelião na virada do ano. - Indicadores de sucesso: dia 01/jan/2027 transcorre sem atos bloqueados por falta de tabela vigente.

Variações: - Se o Tabelião não subir a tabela 2027 até 31/dez/2026, em 01/jan/2027 o sistema **bloqueia** atos econômicos (`BillingService` recusa lookup em tabela expirada). Atos não-econômicos (reconhecimento sem valor declarado) seguem. Tela `30-kb-bloqueio-fluxob-vencida` é exibida com explicação.

5.2 KB com normativo expirando (alerta de vigência)

Persona: Lucas C. (Tabelião) **Pré-condição:** Provimento com `effective_until` definido próximo (caso atípico — normalmente provimentos têm vigência aberta). **Gatilho:** `kb-vigencia-worker` cron diário detecta T-30/T-15/T-5 dias do vencimento.

Passos: 1. Sistema emite `kb.expiry_warning_emitted` em T-30, manda e-mail SES para o Tabelião com link para a tela `29-kb-modal-vigencia-expirando`. 2. Tabelião abre o modal, vê a versão, decide: prorrogar (editar `effective_until`), substituir por nova versão, ou deixar expirar. 3. Se prorrogar, é uma operação textual sobre a versão existente — não muda hash de conteúdo, apenas metadado de vigência. Registra `kb.category_weight_updated` -like (operação registrada com hash chain). 4. Se substituir, executa cenário 3.3. 5. Se deixar expirar, em `effective_until` sistema emite `kb.version_expired`, RAG passa a filtrar essa versão fora dos resultados (`status=active` AND data no range).

Resultado esperado: vigência sob controle do Tabelião; nenhuma surpresa em fiscalização.

Eventos emitidos: `kb.expiry_warning_emitted` (×3 thresholds), eventualmente `kb.version_expired`.

Métricas operacionais: - Tempo esperado: 2 a 5 minutos por revisão. - Indicadores de sucesso: nenhuma versão vencida ainda servindo RAG (`status=active` invalidada por filtro de vigência).

Variações: - Múltiplas versões expirando no mesmo dia — alertas consolidados em um único e-mail diário.

5.3 Errata pontual de modelo (mid-version)

Persona: Lucas C. (Tabelião) **Pré-condição:** modelo “Procuração Pública” v1 publicado e em uso ativo. **Gatilho:** Tabelião identifica erro de digitação em cláusula padrão.

Passos: 1. Tabelião abre `08-catalogo-atos`, seleciona “Procuração Pública”, clica “Nova versão” (não há edição in-place de versão publicada — imutabilidade R5). 2. Sistema clona v1 → rascunho v2. Tabelião corrige a digitação, sem mudar lógica. 3. Passa pelos passos 2-4 do Fluxo A. Aprova com MFA. 4. v2 publicada. Atos novos passam a usar v2; atos consolidados em v1 permanecem com snapshot da v1 (defesa material em fiscalização).

Resultado esperado: correção propagada para o futuro, sem alterar o passado.

Eventos emitidos: `act_template_version.published` (v2), `act_template_version.deprecated` (v1, opcional — Tabelião decide se a v1 fica disponível como histórica não-selecionável).

Métricas operacionais: - Tempo esperado: 5 a 15 minutos para errata simples. - Indicadores de sucesso: hash da v2 difere do hash da v1; novos atos referenciam v2.

Variações: - Errata em modelo muito complexo (ex.: Escritura V&C com 18 variáveis): Tabelião deve fazer revisão integral via wizard; é caso de versão maior, não errata.

5.4 Sucessão de Tabelaio (mudança de responsável legal)

Persona: Stoix Ops + Lucas C. (Tabelaio saindo) + novo Tabelaio entrando **Pré-condição:** CGJ-SP comunica formalmente a sucessão da delegação; documentação legal em ordem. **Gatilho:** notificação oficial do novo titular.

Passos: 1. Stoix Ops abre ticket de sucessão. Confirma identidade do sucessor com a CGJ-SP. 2. Stoix executa rotina `tenant_export : pg_dump` do RDS `notary-27sp-prd`, dump dos buckets S3 (incluindo `notary-27sp-livro-perpetuo` com objetos cifrados + chave de envelope KMS portada via BYOK auditado), export do realm Keycloak em JSON. Pacote completo entregue ao sucessor em 60 dias máximo (R4 REQ-24). 3. Stoix Ops cria usuário do novo Tabelaio no realm `notary-27sp` com perfil `notary` e flag `must_change_password=true`. E-mail de boas-vindas enviado. 4. Antigo Tabelaio tem acesso revogado (`auth.realm_admin_action`). 5. Novo Tabelaio faz primeiro login (cenário 2.2), revisa modelos e KB ativos, decide manter, deprecuar ou substituir cada item. Cada decisão é auditada com hash chain. 6. Livro perpétuo permanece intocável — atos lavrados na gestão anterior continuam acessíveis ao novo Tabelaio, com `act_template_version_hash` da época preservado.

Resultado esperado: continuidade do serviço público (Lei 8.935 art. 4º); chave KMS dedicada transferida via escrow auditado; novo Tabelaio opera com base intacta.

Eventos emitidos: `auth.realm_admin_action` (revogação + criação), `tenant.provisioned` -like na master log Stoix com sucessão.

Métricas operacionais: - Tempo esperado: 30 a 60 dias para sucessão completa, incluindo entrega do pacote e validação. - Indicadores de sucesso: novo Tabelaio acessa minutas antigas; hash chain de `audit_event` permanece válido na verificação noturna.

Variações: - Sucessão em situação litigiosa (perda de delegação por decisão judicial): orientação da CGJ-SP define quem pode acessar; Stoix coopera.

6. Cenários · Incidentes e suporte

6.1 Indisponibilidade parcial (RDS failover de AZ)

Persona: Stoix Ops + escreventes em operação **Pré-condição:** RDS multi-AZ provisionado normalmente. **Gatilho:** falha de hardware na AZ primária; RDS executa failover automático para AZ secundária.

Passos: 1. CloudWatch detecta falha de health-check no RDS primário; alarme dispara. 2. RDS executa failover Multi-AZ (RTO típico 60-120s). 3. Durante o failover, escreventes em operação no Fluxo B vêem erro transitório (“Sistema indisponível, tente novamente em segundos”) na próxima ação que toca banco. Fila SQS continua aceitando mensagens (workers Lambda retentam em DLQ se necessário). 4. SNS → Lambda router → ClickUp abre tarefa “P1 RDS failover notary-27sp” automaticamente. SES envia alerta para o e-mail de plantão Stoix. 5. Conexões de aplicação reestabelecem em < 2 min. Escreventes retomam de onde estavam (rascunho de ato é persistente — Maria não perde o trabalho). 6. Stoix Ops valida no dashboard CloudWatch que tudo voltou ao normal; comunica o Tabelaio por e-mail informando duração da indisponibilidade.

Resultado esperado: interrupção de 1 a 3 minutos; nenhum dado perdido; nenhum ato consolidado em estado inconsistente.

Eventos emitidos: alarmes CW, eventos de auditoria em padrão normal.

Métricas operacionais: - Tempo esperado de recuperação: ≤ 3 min. - Indicadores de sucesso: nenhum `billing_event` órfão (consolidação sem cobrança ou cobrança sem consolidação); cadeia HMAC íntegra.

Variações: - Failover demorando > 5 min: escalation para AWS Support; comunicação imediata ao Tabelião via SES. - Falha de AZ + falha de region (improvável): cenário fora do MVP (DR cross-region é roadmap futuro).

6.2 Bug de cálculo reportado pelo Tabelião

Persona: Lucas C. (Tabelião) + Stoix Ops **Pré-condição:** ato lavrado; billing reportado pelo Tabelião como incorreto. **Gatilho:** Tabelião abre ticket no canal de suporte com `billing_event_id`.

Passos: 1. Tabelião acessa `15-relatorios-billing`, copia o `billing_event_id` da linha divergente. 2. Abre canal de suporte (e-mail `dpo@stoix` ou portal Stoix). Inclui printscreen e descrição. 3. Stoix Ops abre tarefa ClickUp com label `bug-billing`. SLA P2: resposta em 4h úteis. 4. Engenharia Stoix reproduz o cálculo usando os três snapshots imutáveis do `billing_event` (template hash, fee schedule hash, kb_consultations_snapshot). Confirma se o cálculo do sistema bateu ou não com a expectativa do Tabelião. 5. Se o sistema estiver certo (Tabelião confundiu faixa ou ato): explicação detalhada com referência ao snapshot. Caso encerrado. 6. Se o sistema estiver errado (bug genuíno): patch é desenvolvido. Para o ato específico, Stoix gera `billing_adjustment` (registro de ajuste contábil, append-only, com referência ao `billing_event` original — não altera o evento original, que permanece imutável). Tabelião e Stoix concordam no valor corrigido. Patch é deployado.

Resultado esperado: transparência absoluta sobre como cada cobrança foi calculada; correções não destroem histórico — somam ajustes.

Eventos emitidos: `billing_event.computed` permanece; `billing_adjustment.created` (ajuste contábil append-only).

Métricas operacionais: - Tempo esperado: 4h para resposta inicial; 1 a 5 dias para patch se for bug. - Indicadores de sucesso: ajuste contábil documentado; auditoria reproduzível.

Variações: - Se múltiplos atos foram impactados pelo mesmo bug, Stoix Ops fornece lista completa de `billing_event_id` afetados e gera ajustes em lote, com aprovação do Tabelião.

6.3 Falha de billing — ato lavrado mas billing_event não gerado

Persona: Stoix Ops (alerta automático) **Pré-condição:** transação ACID falha em condição rara (ex.: integridade quebrada pelo lookup `fee_schedule_line` em faixa não coberta). **Gatilho:** decisão R5 #3 garante que `act.consolidated` e `billing_event.computed` ocorrem na mesma TX — então, em condições normais, ou ambos comitam ou nenhum. Este cenário cobre o “ou nenhum”.

Passos: 1. Fluxo B chega ao passo de consolidação. `BillingService` inline executa o lookup. Falha (ex.: valor declarado fora de qualquer faixa coberta pela tabela). 2. TX SQL é revertida integralmente. `act.consolidated` não é emitido; `act_instance` permanece em rascunho com flag de erro. 3.

Escrevente vê mensagem: “Não foi possível calcular o emolumento — valor declarado (R\$ X) fora das faixas da tabela vigente. Confirme o valor com o cliente antes de prosseguir.” 4. Se erro estrutural (a tabela vigente realmente não cobre o caso), escrevente escala ao Tabelião. Tabelião pode (a) revisar o valor com o cliente, (b) abrir ticket Stoix para investigar gap na tabela, (c) optar por não gerar o ato no sistema (lavatura em livro físico sem suporte da plataforma — modo offline emergencial). 5. Logs CW capturam a falha de lookup; alerta P2 abre tarefa ClickUp.

Resultado esperado: nunca há ato consolidado sem billing correspondente. A garantia é estrutural (TX ACID), não procedural.

Eventos emitidos: nenhum evento de domínio para o ato (rollback completo); CW Logs com stack trace; alerta ClickUp.

Métricas operacionais: - Tempo esperado de detecção: imediato (na própria UI do escrevente). - Indicadores de sucesso: zero atos consolidados sem `billing_event` correspondente no relatório mensal.

Variações: - Se a falha for transitória (ex.: deadlock raro), o retry pelo escrevente resolve. SQS retentativas para workers são automáticas.

6.4 Canal Stoix de suporte e SLA

Persona: todas as personas (Tabelião, escreventes, DPO) **Pré-condição:** contrato ativo. **Gatilho:** qualquer dúvida, incidente ou pedido.

Passos: 1. Canal único: e-mail `suporte@stoix.dev` (P2 e P3) ou alarme automático CW → ClickUp (P0 e P1). 2. Classificação automática por triagem na recepção: - **P0** — sistema indisponível, atos não podem ser lavrados. SLA: resposta em 30 min, mitigação em 4h. - **P1** — funcionalidade crítica degradada (RAG fora, OCR fora, billing fora). SLA: resposta em 2h, mitigação em 24h. - **P2** — bug não-bloqueante, divergência em cálculo. SLA: resposta em 4h úteis, mitigação em 5 dias úteis. - **P3** — dúvida operacional, sugestão de melhoria. SLA: resposta em 2 dias úteis. 3. Cada ticket gera tarefa ClickUp visível ao Tabelião num portal `15-relatorios-sla` (read-only).

Resultado esperado: cliente sabe a qualquer momento o status de seus tickets e os SLAs aplicáveis.

Eventos emitidos: alertas CW conforme aplicável.

Métricas operacionais: - Tempo esperado: SLAs acima. - Indicadores de sucesso: ≥ 95% dos tickets resolvidos dentro do SLA do trimestre.

Variações: - Fora do horário comercial (P0/P1 apenas): plantão acionado via SNS → SES e ClickUp; resposta no SLA. P2/P3 aguardam dia útil.

7. Cenários · Auditoria e LGPD

7.1 Atendimento a direito de acesso LGPD (titular pede dados)

Persona: DPO (escrevente designado pelo Tabelião) + Lucas C. (Tabelião como Encarregado) **Pré-condição:** portal `19-area-dpo-lgpd` operacional. **Gatilho:** titular José da Silva envia e-mail pedindo lista de atos em que figura como parte.

Passos: 1. DPO recebe o pedido por e-mail e abre `19-area-dpo-lgpd`. Clica “Nova solicitação LGPD”. 2. Preenche: tipo (`access`), e-mail do requerente, identificação (CPF, validado por DV e cruzado contra base local), descrição livre. 3. Sistema cria `lgpd_request`, emite `lgpd_request.submitted` com hash chain. Calcula prazo legal: 15 dias corridos a partir do recebimento (LGPD art. 19), com alertas internos em T-5/T-2/T-0 para o DPO (escalation de 5 dias úteis é interno — REQ-09 R5). 4. DPO confirma recebimento (`lgpd_request.acknowledged`). 5. DPO usa busca interna por `tax_id` (decrypt na service layer do API BFF, decisão R5 #7 — nunca SQL function pública). Lista todos os `act_party` em que o CPF aparece. 6. DPO gera o relatório de acesso em PDF (lista de atos, datas, papel da parte — vendedor, comprador, outorgante, etc.), envia ao titular por canal seguro. 7. DPO marca `lgpd_request.fulfilled` com `fulfillment_kind=report` e `evidence_s3_key` apontando ao PDF gerado. Hash chained.

Resultado esperado: titular atendido em até 15 dias; trilha de cumprimento auditável.

Eventos emitidos: `lgpd_request.submitted`, `lgpd_request.acknowledged`, `lgpd_request.fulfilled`.

Métricas operacionais: - Tempo esperado: 1 a 5 dias úteis na maioria dos casos. - Indicadores de sucesso: zero pedidos `lgpd_request` ultrapassando 15 dias corridos.

Variações: - Se o CPF não retornar nenhum ato, DPO responde negativamente com fundamento. `lgpd_request.fulfilled` registra `fulfillment_kind=no_data_found`. - Se a identidade do requerente não for confirmada, DPO solicita documentação adicional antes de atender. Prazo de 15 dias é interrompido até confirmação.

7.2 Anonimização preservando audit_event

Persona: DPO **Pré-condição:** titular solicita anonimização de dados não vinculados a livro perpétuo. **Gatilho:** pedido recebido após 5 anos da última interação não-notarial (sessões, logs operacionais).

Passos: 1. DPO abre `19-area-dpo-lgpd`, cria `lgpd_request` tipo `anonymization`. 2. Sistema executa busca: identifica registros em tabelas operacionais (sessões antigas, telemetria de UI) e separa do livro perpétuo (`act_party` em atos consolidados — protegido por Lei 8.935 art. 41). 3. DPO revisa o escopo, aprova com MFA TOTP. Sistema executa script de anonimização: substitui campos PII por placeholders (`<anonymized>` em texto livre; nulificação de FKs onde permitido). 4. **audit_event permanece intocável** — eventos passados referenciam `act_party_id` por FK; o conteúdo cifrado do `tax_id` permanece, mas o registro principal pode ser anonimizado conforme o escopo definido. Decisão R5: `audit_event` é append-only; nenhuma anonimização toca essa tabela. 5. DPO comunica o titular: o que foi anonimizado, o que permanece (livro perpétuo) e o fundamento legal. 6. `lgpd_request.fulfilled` com hash chained.

Resultado esperado: dados não-livro anonimizados; livro intacto; trilha auditável preservada.

Eventos emitidos: `lgpd_request.submitted` , `auth.mfa_challenge_passed` , `lgpd_request.fulfilled` .

Métricas operacionais: - Tempo esperado: 5 a 10 dias úteis, com revisão de escopo. - Indicadores de sucesso: anonimização não quebra hash chain de `audit_event` ; consultas RAG futuras não retornam dados anonimizados.

Variações: - Pedido de exclusão completa: bloqueado para livro com mensagem fundada no art. 16, I da LGPD c/c Lei 8.935 art. 41. `lgpd_request.rejected` com `legal_basis` preenchido.

7.3 Exportação de portabilidade

Persona: DPO **Pré-condição:** titular solicita portabilidade dos seus dados. **Gatilho:** pedido LGPD art. 18, V.

Passos: 1. DPO cria `lgpd_request` tipo `portability` em `19-area-dpo-lgpd` . 2. Sistema gera pacote: JSON estruturado com todos os `act_party` do CPF, papéis em cada ato, datas, `act_template_version_hash` , sem expor PII de outras partes (separação por papel). Inclui PDF das minutas em que o titular figura (em formato aberto — PDF/A). 3. DPO valida, aprova com MFA TOTP, envia ao titular por canal seguro com hash de integridade do pacote. 4. `lgpd_request.fulfilled` com `evidence_s3_key` .

Resultado esperado: titular recebe seus dados em formato aberto; outros titulares co-partícipes dos atos não são expostos.

Eventos emitidos: `lgpd_request.submitted` , `auth.mfa_challenge_passed` , `lgpd_request.fulfilled` .

Métricas operacionais: - Tempo esperado: 5 a 15 dias. - Indicadores de sucesso: pacote aberto e legível em ferramenta padrão (Adobe Reader, navegador, jq).

Variações: - Se o titular pedir somente os dados não-notariais (sessões, logs), pacote é menor e exclui referências a atos.

7.4 Defesa material em fiscalização CGJ-SP (provar integridade do livro)

Persona: Lucas C. (Tabelião) **Pré-condição:** notificação de fiscalização CGJ-SP recebida; auditor pede demonstração de integridade do livro digital. **Gatilho:** chegada do auditor ao cartório.

Passos: 1. Tabelião abre `19-area-dpo-lgpd` (ou módulo de auditoria equivalente). Apresenta: - Lista de atos do período fiscalizado, com `act_template_version_hash` , `fee_schedule_version_hash` e `kb_consultations_snapshot` de cada um. - Verificação da cadeia HMAC do `audit_event` : o sistema executa em tempo real um `audit.hash_chain_verified` ad-hoc e mostra “Cadeia íntegra desde primeiro evento até o último”. - Dump diário do chain no bucket master Stoix (cross-tenant, decisão R5 #2) como evidência externa: se algum evento do tenant foi adulterado, o dump master discordaria. - Objetos no `notary-27sp-livro-perpetuo` com Object Lock Compliance 30y — auditor pode verificar via console AWS (ou via Stoix Ops) que os hashes batem e que os objetos não foram modificáveis desde a consolidação. 2. Para qualquer ato específico, Tabelião reproduz o cálculo do emolumento

usando os três snapshots (sistema mostra a TX ACID que produziu o `billing_event`). 3. Tabelião exporta o relatório completo do período em PDF assinado pela Stoix (assinatura digital do operador, não A3 ICP-Brasil — escopo MVP).

Resultado esperado: auditor convencido da integridade; relatório arquivado pela CGJ-SP.

Eventos emitidos: `audit.hash_chain_verified` (manual, sob demanda).

Métricas operacionais: - Tempo esperado: 1 a 3 horas de demonstração presencial. - Indicadores de sucesso: cadeia HMAC sem quebras; nenhum objeto livro perpétuo modificado fora da consolidação inicial.

Variações: - Se a fiscalização demandar análise forense independente, Tabelião usa o `tenant_export` (rotina de portabilidade total — R4 REQ-24) para entregar pacote auditável ao perito designado. - Se for detectada quebra de hash chain (cenário improvável — protegido por Object Lock + dump master), incidente P0 é aberto imediatamente, Stoix comunica o Tabelião em $\leq 24h$ e juntos comunicam a CGJ-SP e a ANPD conforme aplicável.

8. Aceite e UAT

8.1 Critérios de aceite para assinatura do contrato

A assinatura do contrato está condicionada à validação prévia do Tabelião sobre os módulos e fluxos abaixo, conduzida em UAT presencial ou remoto assistido.

BLOCO	CRITÉRIO DE ACEITE	CENÁRIOS DE REFERÊNCIA
Onboarding	Tabelião consegue acessar, configurar MFA, subir 1 normativo na KB, 1 tabela de emolumentos, 1 modelo de ato sem assistência	2.1 a 2.5
Fluxo A	Tabelião conclui criação, edição e nova versão de modelo, com MFA funcional em cada aprovação	3.1, 3.2
Fluxo B	Escrevente conclui Escritura V&C com cláusula resolutive em $\leq 90s$ (p95) com 3 documentos	4.1
KB	Tabelião aprova ou rejeita extração com diff visual; vigências dispararam alertas; bloqueio com KB vencida funciona	2.3, 5.2
Fee Schedule	Tabela 2026 ativa cobra corretamente Escritura V&C valor R\$ 850.000 com FUNDREJUS e ISS conforme planilha CGJ-SP	2.4, 4.1
Billing	<code>billing_event</code> inclui os 3 snapshots; relatório mensal totaliza corretamente; export CSV abre em Excel	3.5, 4.1
Auditoria	Cadeia HMAC verificável; Object Lock 30y no livro perpétuo confirmado em console AWS; dump diário master Stoix presente	7.4
LGPD	DPO consegue criar <code>lgpd_request</code> , atender e marcar fulfilled em todos os tipos previstos	7.1 a 7.3

8.2 Como o Tabelião valida cada módulo

- KB:** subir 3 PDFs representativos (um provimento, uma súmula, uma orientação), aprovar todos com MFA, conferir busca manual em `23-kb-dashboard` retornando os 3.
- Fee Schedule:** subir XLSX 2026 oficial, conferir as faixas mais usadas (V&C, Procuração, Reconhecimento), ativar com MFA.
- Catálogo de Modelos:** criar Escritura V&C v1, fazer V&C com cláusula resolutive (4.1) end-to-end, conferir cobrança correta.
- Procedimentos:** criar um procedimento markdown (“Atendimento ao público — escritura de imóvel”), editar 3 vezes, conferir que o histórico jsonb mantém as 10 últimas revisões.

8.3 Sign-off por bloco

UAT é conduzido em uma sessão de 4 horas com o Tabelião e ao menos um escrevente sênior. Cada bloco é validado individualmente. Sign-off por checklist com assinatura do Tabelião e do representante Stoix. Pendências P0/P1 bloqueiam assinatura; P2/P3 viram backlog pós-go-live com prazos contratuais.

9. Métricas operacionais visíveis ao cliente

MÉTRICA	VALOR-ALVO	ONDE O CLIENTE VÊ
Disponibilidade do serviço (uptime mensal)	≥ 99,5%	16-relatorios-sla , e-mail mensal de status
Fluxo B end-to-end (p95, 3 docs)	≤ 90s	16-relatorios-sla
RAG p95	≤ 800ms	16-relatorios-sla
LLM p95 (Bedrock Sonnet 4.5)	≤ 4s	16-relatorios-sla
Textract p95 por página	≤ 6s	16-relatorios-sla
Acurácia OCR (campos extraídos sem revisão)	≥ 90% nos campos não-manuscritos	relatório trimestral
Acurácia preenchimento LLM (campos aceitos sem edição)	≥ 85%	relatório trimestral
Taxa de alertas RAG vs bloqueios	informativa: 95% alertas, 5% bloqueios é o esperado	16-relatorios-sla
Tempo médio de aprovação de modelo (Fluxo A end-to-end)	< 30 min	relatório trimestral
Total de atos lavrados/mês com billing reconciliado	bater 100% com contabilidade do Tabelião	15-relatorios-billing
LGPD requests dentro do prazo legal	100% em ≤ 15 dias corridos	19-area-dpo-lgpd

10. Governança operacional

10.1 Matriz de aprovação – quem aprova o quê

AÇÃO	TABELIÃO	ESCREVENTE	DPO	STOIX OPS
Provisionar tenant	informa	—	—	executa
Criar usuário escrevente	aprova + MFA	—	—	executa (sob pedido)
Subir KB e aprovar versão	aprova + MFA	—	—	—
Definir peso de <code>kb_category</code>	aprova + MFA	—	—	—
Subir fee schedule e ativar	aprova + MFA	—	—	—
Publicar <code>act_template_version</code>	aprova + MFA	—	—	—
Lavrar ato (Fluxo B)	confere presencial	executa	—	—
Atender <code>lgpd_request</code>	(Encarregado)	—	executa	apoia técnico
Anonimização parcial	aprova + MFA	—	propõe	—
Sucessão de Tabelião	(sai/entra)	—	—	executa export
Patch de software	informado	—	—	executa
Resposta a incidente P0/P1	informado em ≤ 24h	—	informado se LGPD	executa

10.2 Cadência de revisão

- **Diária:** verificação automatizada da cadeia HMAC do `audit_event` e do dump master Stoix.
- **Semanal:** revisão de tickets ClickUp por Stoix Ops com Tabelião (15 min).
- **Mensal:** fechamento do `billing_event` consolidado; revisão da reconciliação contábil pelo Tabelião; relatório SLA.
- **Trimestral:** revisão da KB (categorias e pesos); revisão do catálogo de modelos (deprecated vs ativos); acurácia OCR/LLM.
- **Anual:** nova tabela de emolumentos CGJ-SP; pen-test focado em isolamento físico de tenant; revisão integrada Stoix + Tabelião + DPO; renovação contratual.

10.3 Canal Stoix de suporte

- **E-mail:** `suporte@stoix.dev` (P2/P3) e `dpo@stoix.dev` (questões LGPD).

- **Alarmes automáticos:** CW → SNS → Lambda router → ClickUp + SES.
 - **Portal interno:** [16-relatorios-sla](#) para o Tabelião acompanhar tickets do tenant.
 - **SLA:** P0 30min/4h, P1 2h/24h, P2 4h úteis/5 dias úteis, P3 2 dias úteis. Detalhamento em 6.4.
 - **Plantão:** P0/P1 cobertos 24x7 via SNS → SES; P2/P3 dias úteis 08h-18h SP.
-

Leonardo Menezes — Stoix leonardo@stoix.dev · 2026-05-12