

Estrutura Integrada de Dados do TJCE Geridos por Data Lake Local Contendo Dados Judiciais e Administrativos: um Relatório Técnico

Sérgio Maia Raulino

Tribunal de Justiça do Estado do Ceará (TJCE)

sergio.raulino@tjce.jus.br

Márcio Bezerra de Menezes Serpa Filho

Tribunal de Justiça do Estado do Ceará

marcio.filho@tjce.jus.br

Junior Regis Batista Cysne

Tribunal de Justiça do Estado do Ceará (TJCE)

junior.cysne@tjce.jus.br

Gestão de dados administrativos e judiciais visando economicidade e ações estratégicas.

RESUMO

A transformação digital no Poder Judiciário brasileiro tem se intensificado diante dos desafios impostos pela Sociedade da Informação, caracterizada pela crescente produção, diversidade e circulação de dados em múltiplos formatos e fontes, como: planilhas, arquivos em formato PDF, bancos de dados, interfaces programáticas com aplicações diversas etc. Nesse contexto, o projeto **DataJust – Estrutura Integrada de Dados do TJCE**, conduzido pela Gerência da Plataforma de Dados do Tribunal de Justiça do Estado do Ceará (TJCE), surge como uma resposta inovadora e estratégica às demandas por eficiência, transparência, inteligência e governança na gestão de dados judiciais e administrativos. A iniciativa busca consolidar informações até então dispersas, despadronizadas e de difícil acesso, em um repositório centralizado, performático e alta disponibilidade, promovendo análises avançadas, decisões mais qualificadas e ações estratégicas, fundamentadas em evidências.

O conceito de *data lake* e/ou *house lake*, fundamental para o DataJust, consiste em um repositório capaz de armazenar grandes volumes de dados em seu formato bruto, permitindo flexibilidade no tratamento e análises conforme diferentes necessidades analíticas e operacionais.

1

| | | | | |
|---|--|---|--|---|
|  Programa de Pós-graduação em Administração UFPB |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  Universidade de Brasília |  PPGD |  Universidade Potiguar |
|  IESB Centro Universitário |  1290 FACULDADE DE DIREITO UNIVERSIDADE DE COIMBRA |  DIREÇÃO-GERAL DA POLÍTICA DE JUSTIÇA |  Iluris Instituto de Investigação Interdisciplinar |  AJUS Administração da Justiça |
|  Grupo de Pesquisa em Administração, Governo e Políticas Públicas do Poder Judiciário |  GEJUD Grupo de Pesquisa Gestão, Desempenho e Efetividade do Judiciário |  InfoJus Núcleo de Pesquisa em Informação, Direito e Sociedade |  LIOrg Linguagem, Instituições e Organizações | |

A implantação do DataJust está sendo realizada em infraestrutura local (on-premise), aproveitando e otimizando os recursos computacionais já existentes no datacenter do TJCE, o que vem garantindo a contenção dos gastos com soluções em nuvem e reforçou o compromisso com a sustentabilidade fiscal e a autonomia tecnológica institucional. A arquitetura foi construída com base em soluções open-source amplamente reconhecidas no ecossistema de Big Data.

Palavras-chave: gestão pública inteligente; economicidade; inovação no judiciário; governança e centralização de dados; *data lake*.

Introdução

A inovação introduzida com a utilização de uma plataforma de dados integrada, tendo como elementos centrais os conceitos de *data lake* e /ou de *house lake*, que implanta as melhores práticas e ferramentas do ecossistema de Big Data, vem melhorando de forma significativa a gestão dos dados administrativos e judiciais do Tribunal de Justiça do Estado do Ceará.

Para Khine e Wang (2018), “*Data Lake consiste em um repositório em que todos os dados de uma organização, ou seja, dados estruturados, semiestruturados e não estruturados, são armazenados juntos, independentemente dos tipos, formato ou estrutura*” (p. 280). Essa arquitetura, contendo diversas camadas, onde realizam-se transformações e melhorias de qualidade incremental dos dados, vem permitindo superar as limitações dos tradicionais *data warehouses (DWs)*, conceito até então adotado pelas equipes responsáveis pela gestão dos dados para coleta de informações voltadas para às área de negócio, que exigem modelagens rígidas e processos de cargas não-escaláveis, dificultando ajustes dinâmicos e gerando interferências nos sistemas transacionais. Como destacado por Dixon (2010), os *data lakes* “facilitam análises avançadas, aprendizado de máquina e inteligência artificial, fornecendo uma visão abrangente e unificada dos dados”. Isso é particularmente relevante no âmbito do TJCE, onde existe uma grande quantidade de bases de dados e fontes de dados diversas e sem seguir padronizações.

2

| | | | | |
|---|---|---|--|---|
|  |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  |  |  |
|  |  |  |  |  |
|  |  |  |  | |

A implantação do DataJust está sendo realizada em infraestrutura local (on-premise), aproveitando e otimizando os recursos computacionais já existentes no datacenter do TJCE, o que vem garantindo a contenção dos gastos com soluções em nuvem e reforçou o compromisso com a sustentabilidade fiscal e a autonomia tecnológica institucional. A arquitetura foi construída com base em soluções open-source amplamente reconhecidas no ecossistema de Big Data, como o **Apache Spark** (processamento distribuído de dados em larga escala), o **Trino** (motor de consulta SQL distribuído), o **MinIO** (armazenamento de objetos compatível com S3) e o **Hive Metastore** (catálogo de dados e metadados). Adota-se o **Apache Airflow** como componente central para orquestração das cargas dos dados. Essa composição tecnológica garante alta performance, escalabilidade horizontal, interoperabilidade entre sistemas e independência de fornecedores, mantendo a robustez e segurança da solução.

A integração de dados de sistemas judiciais como PJe, SAJ e SEEU, bem como de sistemas administrativos relacionados a pessoal e custos operacionais, além de dados de acessos à rede e ativos do TJCE, vem permitindo uma visão holística das operações do TJCE. Recentemente, ampliamos o valor do DataJust quando efetivamos a extração e disponibilização segura de dados inéditos por meio da coleta de dados críticos de órgãos parceiros, como: **Secretaria de Administração Penitenciária** (SAP/CE) e **Banco Nacional de Medidas Penais** (BNMP/CNJ). Essa consolidação de informações está apoiando, eficientemente, a construção de painéis com ferramentas BI, com dados estatísticos e predição de padrões, a detecção de gargalos e a implementação de melhorias nos processos judiciais e administrativos.

Conforme Singh (2019), os data lakes oferecem "*flexibilidade para armazenamento, processamento e análise de dados*" (p. 245), características essenciais para instituições que buscam aprimorar sua eficiência operacional. A plataforma conta com mecanismos de segurança robustos, incluindo criptografia dos dados em trânsito, controle de acesso baseado em perfis de negócio e em conformidade com o Princípio do Privilégio Mínimo (PoLP). A organização e centralização dos dados administrativos promovidas pelo DataJust vêm trazendo benefícios concretos, como sua participação na arquitetura do projeto estratégico do “DataMart Administrativo” do TJCE, promovendo o aprimoramento do monitoramento de **custos com pessoal** (servidores, contratados e terceirizados), **custos estruturais** (água, energia elétrica, telefonia, internet, bens patrimoniais etc.), **despesas com passagens, aluguéis de carros, combustíveis e diárias**. A consolidação dessas informações em uma plataforma única facilita a transparência ativa, a prestação de contas à sociedade e a identificação de oportunidades de racionalização e eficiência.

Desde sua implantação em março de 2025, a arquitetura do DataJust vem sendo expandida com velocidade, tanto no número de fontes integradas quanto na capacidade de processamento, com

3



previsão de clusterização dos serviços para ampliar a resiliência e a escalabilidade da solução. O modelo permite a replicação da prática em outras instituições do Judiciário, respeitadas suas particularidades técnicas e operacionais, podendo, assim, ser considerado um modelo de boas práticas e inovação tecnológica na Justiça brasileira.

A implementação do DataJust evidencia como o Judiciário pode responder de forma estratégica aos desafios da Sociedade da Informação, utilizando tecnologias referência no contexto de Big Data, contudo sem geração de custos de aquisição de produtos e serviços proprietários. Vale ressaltar, que o projeto vem sendo desenvolvido com força de trabalho de servidores da casa, em sua grande maioria, visando manter o rico conhecimento adquirido no TJCE a longo prazo, e prover mecanismos de disseminá-lo para outros tribunais brasileiros.

Caracterização da boa prática, mudança e/ou inovação

O DataJust é uma plataforma integrada composta por diversas ferramentas referências no contexto de Big Data global. A intensão é prover formas de ingerir dados de diversas fontes com autonomia de fornecedores, garantir performance, mas também fornecer **alternativas flexíveis de democratização dos dados**.

Nesse contexto, nomeamos cada componente de acordo com sua função:

DataJust Core: Elemento central contendo a representação do Data Lake dos dados ingeridos das diversas origens. Contém: Armazenamento em repositório (MinIO com S3), Catálogo dos dados (Hive Metastore) e ferramenta de consulta/exposição dos dados (Trino).

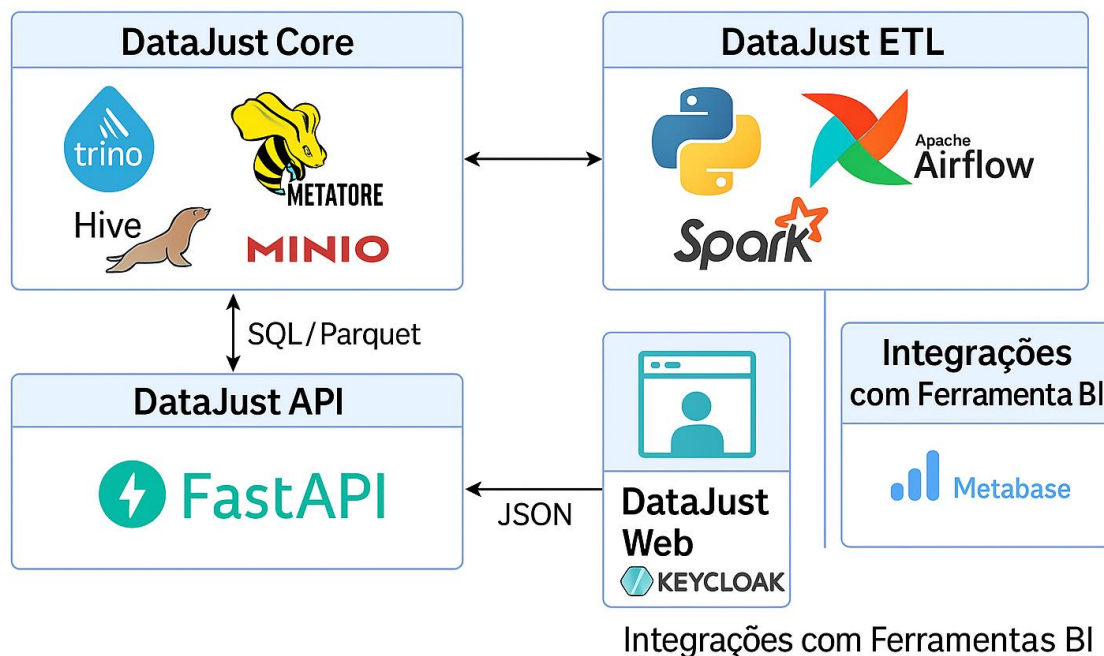
DataJust ETL (Spark/Airflow): Componente responsável pela automação e controle das cargas / ingestões, com foco no histórico e auditoria com logs completos e exportáveis.

DataJust API: Componente criado com a moderna biblioteca python Fast API, que provê interfaces com qualquer dado/entidade existente no DataJust Core. Seu diferencial é permitir acesso por meio de filtros dinâmicos e com segmentação de acessos por perfis.

DataJust Web / Ferramentas BI: Soluções para os usuários de negócio, com painéis pré-definidos e alternativas permitindo democratização dos dados, por meio de extrações personalizadas (Visão atual: Integra-se ao Power BI / Visão futura: validações das integrações com criação de solução Web interna e Metabase, também open-source)

Imagem 1 – Arquitetura geral por macro-componente





Ampliando o componente **DataJust Core**, temos o conceito das diversas camadas, que se diferenciam pelo objetivo dentro do contexto de enriquecimento dos dados até a efetiva entrega para as equipes de entregas de painéis BI. Dessa forma, temos as cinco camadas, conforme quadro abaixo:

Quadro 1 – Camadas existentes no DataJust e suas finalidades

| Camada | Finalidade |
|----------------|---|
| Sandbox | espaço de experimentação, permitindo que equipes parceiras testem hipóteses e explorem novas combinações de dados já ofertados nas diversas camadas do DataJust. Essa camada surgiu em resposta às solicitações das equipes internas, que tinha familiaridade com a linguagem tradicional de banco de dados: SQL. |

5



| | |
|------------------------|---|
| Landing | armazenamento inicial, apenas otimizando o formato original dos dados para o padrão “Parquet” promovendo acessos com performance otimizada e rastreabilidade; |
| Raw (Bronze) | camada bruta, em que os dados são padronizados e organizados, contudo já incorporando elementos de controle histórico com armazenamento no formato Iceberg; |
| Trusted (Prata) | dados validados e enriquecidos, prontos para uso analítico; |
| Refined (Ouro) | bases estruturadas para relatórios e <i>dashboards</i> estratégicos; |

Fontes: Autores (2025)

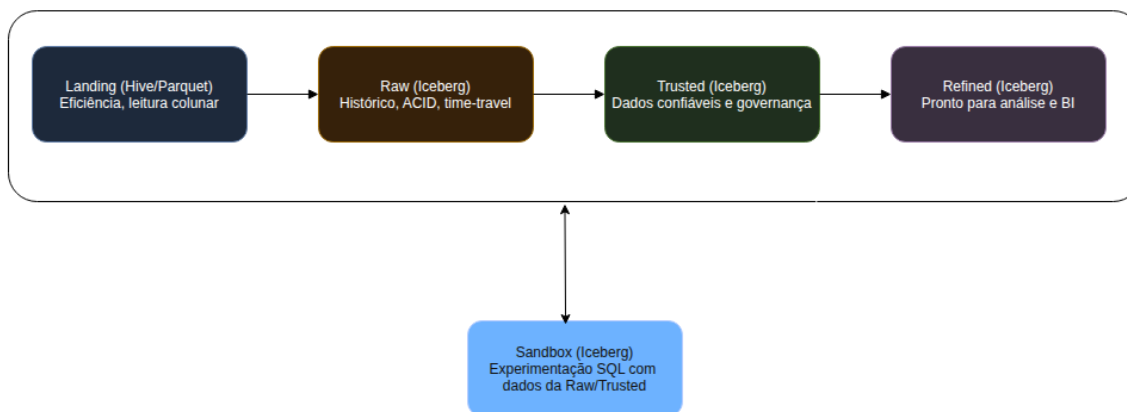
O processo inicia-se na camada Sandbox, onde surge a necessidade de negócio expressa pelos demandantes e é traduzida em experimentações de dados que validam hipóteses iniciais. A partir daí, o fluxo segue de forma estruturada pelas camadas do DataJust: na Landing, ocorre a ingestão bruta dos dados oriundos das fontes, preservando-os em seu formato original; na Raw, esses dados são organizados e padronizados, garantindo integridade e rastreabilidade; na Trusted, aplicam-se regras de negócio, enriquecimento e consistência, consolidando informações confiáveis; e, por fim, na Refined, os dados são modelados de forma analítica para responder diretamente às demandas estratégicas. Esse percurso possibilita que os insights de negócio entregues aos usuários finais sejam fundamentados, auditáveis e relevantes para a tomada de decisão. Na imagem 2, esse processo é explicitado de forma visual a fim de torná-lo mais compreensível:

Imagem 2 – Processo de criação e enriquecimento dos dados

6



Evolução de Camadas no DataJust



Fontes: Autores (2025)

Em todas as camadas, incluindo a *landing*, os dados oriundos das mais diversas fontes são armazenados no padrão Apache Parquet, que oferece benefícios expressivos em termos de desempenho e governança. Por ser um formato colunar, o Parquet possibilita alta eficiência de compressão e economia de espaço, além de permitir que ferramentas analíticas leiam apenas as colunas necessárias, reduzindo significativamente o tempo de consulta e processamento. Esse padrão aberto, amplamente adotado no ecossistema Big Data (Spark, Trino, Hive, Impala, Athena, entre outros), garante interoperabilidade, rastreabilidade e evolução de esquema sem comprometer a consistência histórica dos dados. Como enfatiza Tom White em *Hadoop: The Definitive Guide* (O'Reilly, 2015), os formatos colunares como o Parquet são fundamentais para análises eficientes em larga escala, pois reduzem a E/S de disco e otimizam o uso de recursos de processamento, tornando-se, portanto, a escolha estratégica para o armazenamento inicial do DataJust.

Um dos grandes diferenciais do DataJust em todas as camadas após a *landing* (pouso dos dados) é o uso do Apache Iceberg, um formato de tabela aberto projetado para grandes volumes de dados em ambientes distribuídos. O Iceberg possibilita:

- Controle de versões e dados históricos (*time-travel queries*), permitindo consultas em estados anteriores da tabela;
- Gerenciamento eficiente de *metadados*, garantindo desempenho mesmo com bilhões de registros e milhões de arquivos;

7

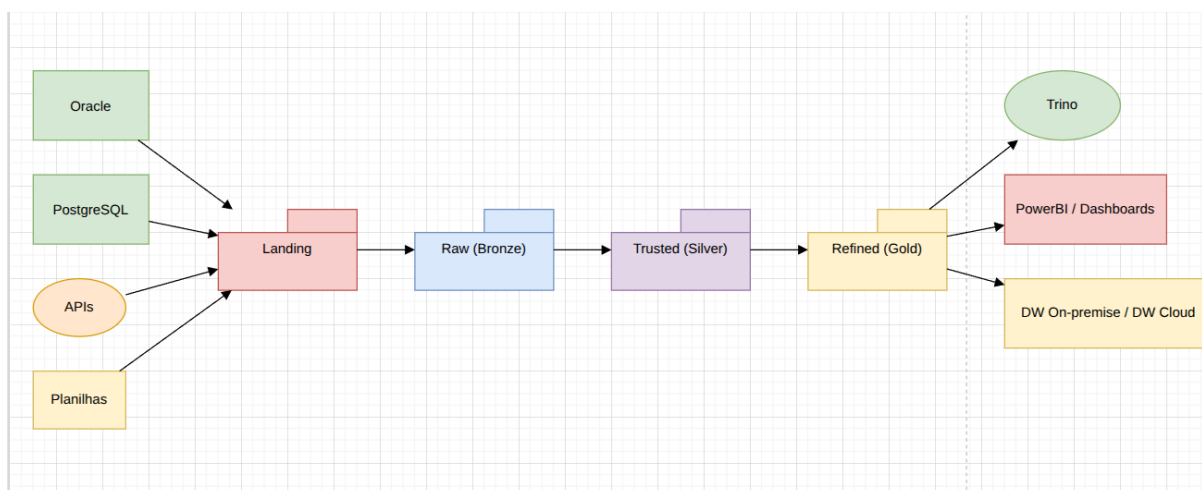


- Evolução de esquema e partições sem necessidade de recriações complexas;
- Integração nativa com motores analíticos como Spark, Trino e Hive (todos elementos utilizamos pela plataforma de dados);
- Consistência ACID, assegurando integridade transacional em operações de escrita/leitura concorrentes.

De acordo com Ryan Blue (criador do Apache Iceberg, Netflix/Tabular), o Iceberg foi desenvolvido para superar limitações de formatos anteriores como *Hive Tables*, trazendo escalabilidade e confiabilidade.

Por fim, ainda no contexto da organização dos dados no DataJust Core, segue uma visão das diversas origens que são de onde os dados são ingeridos e as diversas formas de entregar os dados enriquecidos para as equipes interessadas.

Imagem 3 – Processo de extração das bases de origem entrega dos dados enriquecidos



Fontes: Autores (2025)

O segundo componente arquitetural e elemento-chave para o sucesso da solução é o **DataJust ETL**. Ele tem a responsabilidade de realizar a ingestão dos dados de forma assertiva e garantindo ciclo de vida dos dados. O DataJust ETL foi, recentemente, modernizado para operar em um ambiente clusterizado e escalável, unindo o Apache Airflow, que garante

8



controle e monitoramento das rotinas, ao Apache Spark, responsável pelo processamento distribuído das cargas de dados. Essa arquitetura assegura maior confiabilidade e eficiência na execução das ingestões.

Com essa evolução, o DataJust oferece mais desempenho, disponibilidade e flexibilidade, reduzindo tempos de processamento e ampliando a capacidade de resposta às demandas. Assim, a plataforma passa a sustentar de forma ágil e segura as necessidades estratégicas de dados do TJCE, fortalecendo a tomada de decisão baseada em informações consistentes.

Continuando a explanação, chegamos ao componente **DataJust API**, que foi desenvolvido com a biblioteca moderna FastAPI, reconhecida pela alta performance e simplicidade na criação de serviços em Python. Ele atua como facilitador para obtenção de qualquer dado ou entidade presente no DataJust Core, disponibilizando interfaces padronizadas e seguras. Sua estrutura permite atender diferentes demandas de consumo de dados, desde aplicações internas até sistemas externos que necessitam de integração em tempo real.

O grande diferencial desse componente é a possibilidade de aplicar filtros dinâmicos, facilitando a obtenção dos dados desejados, por qualquer campo das entidades contidas no DataJust Core. Além disso, por meio da ferramenta institucional de gestão de identidade e controle de acessos, o Keycloak, permite-se estabelecer segmentação de acesso por perfis. Por meio desse controle, garante-se que cada usuário visualize apenas as informações que realmente lhe dizem respeito. Esse modelo promove maior segurança, governança e controle, ao mesmo tempo em que oferece flexibilidade para consultas específicas, atendendo tanto necessidades operacionais quanto análises estratégicas.

O componente arquitetural **DataJust Web / Ferramentas BI** foi concebido para aproximar os usuários de negócio do universo de dados, entretanto a ferramenta adquirida pelo TJCE para tal papel está sendo, via de regra, o **Power BI da Microsoft**. Na visão atual, a ferramenta citada entrega está consolidada com a responsabilidade de gerar os relatórios e *dashboards* interativos. Para o futuro, o planejamento prevê validações de novas integrações e a criação de uma solução Web interna, além da incorporação do **Metabase**, uma ferramenta open-source que reforçará a independência tecnológica e ampliará as possibilidades de análise. Com isso, o DataJust avança em direção a um ecossistema cada vez mais acessível, transparente e sustentável.

Para os componentes DataJust API e DataJust Web, realizamos uma PoC (*Proof of Concept*) para validar a viabilidade da solução proposta, utilizando-se a entidade de usuários e grupos do sistema de controle de acesso à rede do TJCE: LDAP. Contudo, no momento, a alta gestão vem priorizando evoluções nos componentes centrais: DataJust Core e DataJust ETL.

9

| | | | | |
|---|--|---|---|--|
|  |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  |  |  Universidade Potiguar |
|  |  |  DIREÇÃO-GERAL DA POLÍTICA DE JUSTIÇA |  Instituto de Investigação Interdisciplinar |  Administração da Justiça |
|  |  Grupo de Pesquisa Gestão, Desempenho e Efetividade do Judiciário |  Núcleo de Pesquisa em Informação, Direito e Sociedade |  LINGUAGEM, INSTITUIÇÕES E ORGANIZAÇÕES | |

Planejamento, desenvolvimento e implementação

Iniciamos a jornada de inovação do DataJust em paralelo a um trabalho conduzido por uma consultoria externa, cujo foco estava tanto na criação e modernização de novos papéis e cargos quanto na implantação de uma plataforma moderna e robusta para a gestão de dados do TJCE. Como essa frente possuía alta complexidade e envolvia a migração dos dados para a nuvem, possivelmente para o ecossistema da AWS, nosso time interno assumiu o desafio de desenvolver uma solução própria on-premise, estabelecendo as bases de um data lake local capaz de atender às necessidades imediatas da instituição.

Ao longo de aproximadamente sete meses de trabalho intenso, de março até setembro de 2025, uma equipe reduzida (dois servidores e um contratado), mas altamente dedicada, adquiriu conhecimentos técnicos em diferentes frentes — desde arquiteturas distribuídas, ferramentas de orquestração e processamento, até práticas de governança de dados. Esse esforço coletivo possibilitou a construção de uma plataforma que, embora em constante evolução, já entrega resultados concretos e perceptíveis. Os ganhos alcançam desde o aumento da confiabilidade e transparência dos dados até a geração de valor para a tomada de decisão, beneficiando diretamente magistrados, gestores e integrantes da alta administração do TJCE.

Como marco desse processo, além do ambiente de desenvolvimento, foram estruturados e entregues também os ambientes de homologação e produção, assegurando ciclos completos de testes, validação e disponibilização segura das soluções. Dessa forma, o DataJust se consolida como uma plataforma estratégica para a modernização da gestão de dados do Tribunal, reforçando a visão de futuro e a autonomia tecnológica da instituição.

Um dos primeiros desafios enfrentados foi a confiabilidade das cargas de dados. Inicialmente, variações nos sistemas de origem ou falhas de rede podiam comprometer a atualização de bases críticas. Com a adoção do **Apache Airflow** como orquestrador, as cargas passaram a ser monitoradas, registradas e reexecutadas automaticamente em caso de falhas. Essa mudança reduziu o risco de interrupções e aumentou a confiança da gestão nos relatórios gerados.

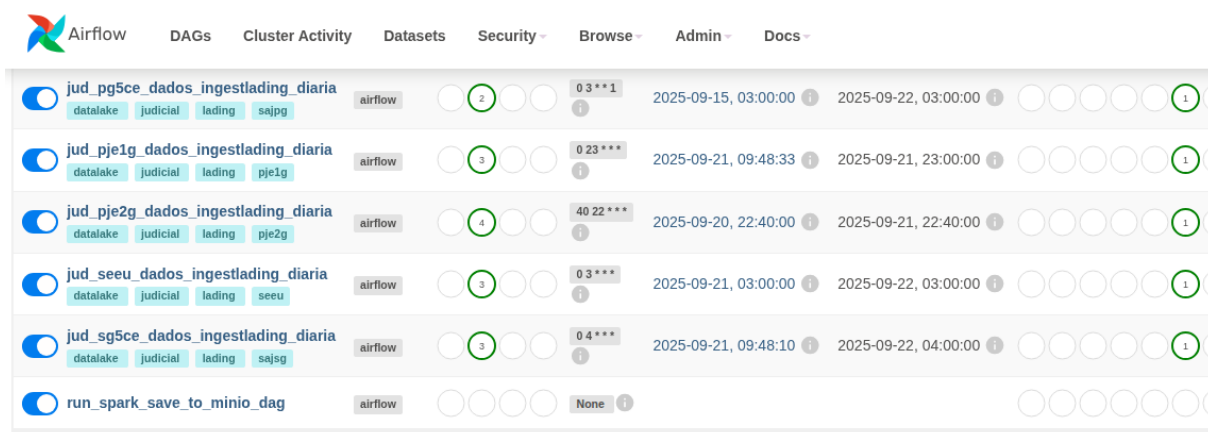
Hoje, os gestores podem ter certeza de que as informações apresentadas em painéis e dashboards estão atualizadas, fortalecendo a tomada de decisão baseada em evidências.

10



Segue tela do Apache Airflow de produção, hospedado em servidor interno da infraestrutura local do TJCE, com a gestão da orquestração das cargas de extrações das bases de origens e transformações de dados administrativos e judiciais:

Imagem 4 – Lista de cargas ETLs em ambiente produtivo



| DAG Name | Category | Next Run | Last Run | Status |
|-------------------------------------|----------|----------------------|----------------------|---------|
| jud_pg5ce_dados_ingestlading_diaria | airflow | 2025-09-15, 03:00:00 | 2025-09-22, 03:00:00 | Success |
| jud_pje1g_dados_ingestlading_diaria | airflow | 2025-09-21, 09:48:33 | 2025-09-21, 23:00:00 | Success |
| jud_pje2g_dados_ingestlading_diaria | airflow | 2025-09-20, 22:40:00 | 2025-09-21, 22:40:00 | Success |
| jud_seeu_dados_ingestlading_diaria | airflow | 2025-09-21, 03:00:00 | 2025-09-22, 03:00:00 | Success |
| jud_sg5ce_dados_ingestlading_diaria | airflow | 2025-09-21, 09:48:10 | 2025-09-22, 04:00:00 | Success |
| run_spark_save_to_minio_dag | airflow | None | | Failed |

Fonte: Autores (2025)

Outro avanço significativo foi a criação de um cluster Spark robusto e adaptável a demanda das cargas e tamanho das ingestões e transformações dos dados. A solução criada possui um nó mestre e cinco nós trabalhadores, responsáveis pela ingestão propriamente dita. Essa nova solução foi recentemente remodelada e ampliou consideravelmente a capacidade de processamento paralelo. Essa expansão reduziu drasticamente o tempo necessário para ingestão de grandes volumes de dados. Cargas que antes levavam horas passaram a ser concluídas em minutos, liberando recursos e possibilitando análises mais rápidas.

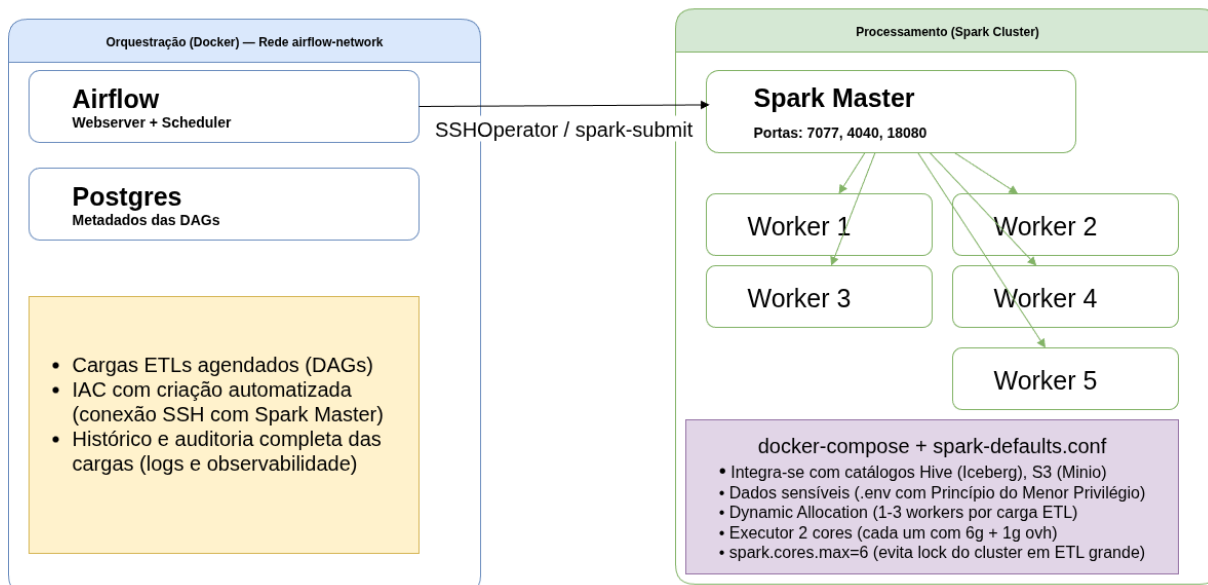
Esse ganho em eficiência não é apenas técnico: ele impacta diretamente a agilidade administrativa e a celeridade judicial, permitindo respostas mais rápidas a demandas internas e externas.

Abaixo, segue um desenho da arquitetura distribuída e com tolerância a falhas, criada para o componente DataJust ETL:

Imagem 5 – Arquitetura detalhada do DataJust ETL



DataJust ETL — Orquestração (Airflow) e Processamento Distribuído (Spark)



Fonte: Autores (2025)

Resultados alcançados

Com a consolidação de bases internas e externas, o TJCE iniciou o desenvolvimento de painéis de apoio à decisão inéditos, capazes de cruzar dados judiciais com informações penitenciárias e de medidas penais. Esses painéis estão sendo desenhados para apoiar magistrados, gestores e órgãos parceiros na identificação de padrões, gargalos e oportunidades de melhoria na gestão da Justiça criminal.

Esse movimento reforça o papel do DataJust como instrumento estratégico, capaz de orientar não apenas a administração do tribunal, mas também a formulação de políticas públicas em cooperação com outras instituições.

Dentre os principais painéis construídos estão:

Quadro 2 – Painéis BI que utilizam dados oriundos do DataJust

| Painel | Objetivo dos painéis para as áreas fim |
|--------|--|
|--------|--|

12



| | |
|---|---|
| Gestão de Acessos à rede TJCE | Apresenta dados dos acessos à rede do TJCE, gerando informações valiosas para gestores de projetos e à área de Segurança da Informação para tomada de decisões; |
| Painel Pena Justa | Com dados fornecidos pelo órgão SAP/CE (Secretaria da Administração Penitenciária) e construído pela SEGOV, o painel exibe dados dos custodiados das penitenciárias do Estado como: situação, periculosidade, faixa etária, totalizadores por unidade carcerária, etc. Com essas informações, um magistrado pode tomar decisões mais assertivas e justas, garantindo proporcionalidade nas medidas aplicadas, melhorando a gestão das unidades prisionais e promovendo maior equilíbrio entre a segurança pública e os direitos fundamentais dos custodiados no Estado do Ceará. |
| Painel de Acompanhamento de Despesas Administrativas | Em processo de expansão na obtenção de dados, via DataJust, esse painel visa oferecer transparência e controle aprimorado sobre gastos com energia, água, telefonia e demais contratos administrativos, permitindo identificar padrões de consumo, otimizar recursos, apoiar auditorias e subsidiar gestores na tomada de decisões baseadas em evidências. |
| BNMP (Banco Nacional de Medidas Penais) | Os dados foram ingeridos recentemente foram entregues, via acesso controlado ao DataJust Core para a SEGOV construir os painéis. Esses dados são fornecidos pelo CNJ e relacionam-se a medidas restritivas de liberdade em escala nacional, possibilitando cruzamentos inéditos com dados processuais locais (TJCE). |

Fonte: autores (2025)

Além dos benefícios encontrados na entrega dos dados otimizados e preparados para os painéis de ferramentas BI, onde os times que atuam diretamente com o Power Bi são os principais beneficiários, consegue-se identificar diversos outros, como ganhos imediatos e estruturantes no modo como os dados são coletados, armazenados e disponibilizados para as diversas áreas da instituição. Um dos principais benefícios é a redução do impacto nas bases de origem, uma vez que cada sistema precisa ser acessado apenas uma vez no momento da ingestão. A partir daí, todas as demais soluções – relatórios, dashboards, APIs e análises avançadas – passam a consumir os dados diretamente do data lake, preservando o desempenho dos sistemas operacionais e garantindo maior estabilidade no ambiente de produção.

Outro benefício importante está no controle centralizado de acessos, viabilizado pela camada de segurança e governança do DataJust. Esse modelo elimina a necessidade de abrir acessos

13



individuais a cada base de origem para novos usuários, reduzindo a pressão sobre os times responsáveis por sistemas específicos. Além disso, o DataJust centraliza a gestão de dados oriundos de planilhas e fontes diversas, conferindo maior consistência, rastreabilidade e confiabilidade às informações que antes ficavam dispersas e sob responsabilidades isoladas.

No horizonte futuro, espera-se que a plataforma viabilize a organização de dados até então sem semântica clara ou dicionário estruturado, por meio de soluções de inteligência artificial generativa aplicadas ao mapeamento e estruturação das informações. Esse esforço trará ganhos significativos em padronização, interoperabilidade e entendimento institucional sobre os dados. Também está prevista uma redução no volume de chamados para criação de relatórios e dashboards customizados, pois as ferramentas de self-service analytics – aliadas a usuários capacitados e conhecedores da semântica dos dados – passarão a atender a maior parte das necessidades.

Outros benefícios percebidos já começam a se consolidar: o aumento da agilidade na entrega de insights para a alta gestão, a democratização do acesso aos dados, que deixa de ser um privilégio restrito a poucos setores técnicos, e o fortalecimento da governança da informação, com trilhas de auditoria e controle sobre todas as camadas do fluxo de dados. Dessa forma, o DataJust se afirma não apenas como um repositório de informações, mas como a base estruturante de uma cultura organizacional orientada a dados no TJCE.

Entendemos que essa solução pode ser replicada em outros órgãos, tanto na esfera do Poder Judiciário, como dos demais poderes, visto que a plataforma trabalha com excelência na gestão de dados, sejam eles relacionados a qualquer esfera de conhecimento.

Conclusões e recomendações

Atualmente, a plataforma de dados integradas do TJCE, o DataJust, está gerando valor real e superando as expectativas de diversas áreas do TJCE, especialmente considerando que sua construção se deu a partir de um contexto inicial muito próximo ao zero, em termos de conhecimento dos integrantes do time e gama de ferramentas disponíveis. Essa conquista reflete a capacidade institucional de desenvolver soluções próprias. Assim, recomenda-se valorizar os servidores da casa, pois são o maior ativo de uma organização, quer seja ela pública ou privada. Recomenda-se capacitar incrementalmente o corpo de servidores dos órgãos públicos e, ponderar racionalmente sobre a contratação de profissionais e consultores externos para a criação das soluções internas.

14

| | | | | |
|---|--|---|--|---|
|  |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  |  |  |
|  |  |  DGPJ DIREÇÃO-GERAL DA POLÍTICA DE JUSTIÇA |  Iluris Instituto de Investigação Interdisciplinar |  |
|  |  GEJUD Grupo de Pesquisa Gestão, Desempenho e Efetividade do Judiciário |  |  | |

A implantação do DataJust no TJCE marca um avanço significativo na forma como o Tribunal passa a gerir seus dados, consolidando um ecossistema on-premise moderno e integrado que já oferece benefícios concretos para unidades judiciais e administrativas. Essa iniciativa rompeu com práticas fragmentadas, estabelecendo uma visão holística de governança e gestão de dados, capaz de sustentar decisões mais rápidas, fundamentadas e alinhadas às demandas institucionais.

Para garantir a plena institucionalização dessa inovação, recomenda-se a formalização de um plano consistente de evolução tecnológica, incluindo o mapeamento detalhado da jornada as-is / to-be, de modo a alinhar processos atuais com as melhores práticas de Big Data e Analytics. Além disso, torna-se essencial o patrocínio da alta gestão no enfrentamento das barreiras culturais e organizacionais, assegurando investimentos contínuos em capacitação, disseminação de boas práticas e incentivos à adoção do novo modelo.

No horizonte futuro, o DataJust deve expandir suas capacidades com a incorporação de recursos de inteligência artificial generativa para estruturação semântica e automação de metadados, bem como a ampliação de ferramentas de self-service analytics que empoderem os usuários finais. A consolidação de ambientes de desenvolvimento, homologação e produção garante a base necessária para essa evolução, mas será a adoção institucional e a governança orientada à inovação que sustentarão o ciclo de modernização contínua.

Entretanto, não podem ser descartadas as soluções em nuvem disponibilizadas por grandes players globais, que oferecem escalabilidade quase ilimitada, serviços de governança avançados e integração nativa com ferramentas de Inteligência Artificial. Entre as principais referências de mercado destacam-se:

- Amazon Web Services (AWS) – com o AWS Lake Formation, Glue, Athena e integração ao Amazon SageMaker para machine learning.
- Microsoft Fabric – unificando data lake, data warehouse e analytics, com nativa integração ao Power BI e modelos de IA no Azure OpenAI Service.
- Google Cloud Platform (GCP) – com BigQuery, Dataplex e integração ao Vertex AI, que potencializa uso de modelos generativos.
- Snowflake – plataforma de dados em nuvem que combina data lakehouse com forte governança, escalabilidade e integrações com IA.
- Databricks – referência global em Lakehouse Architecture, com suporte a Delta Lake e recursos nativos de IA generativa.

15

| | | | | |
|---|--|---|---|---|
|  |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  |  |  |
|  |  |  DIREÇÃO-GERAL DA POLÍTICA DE JUSTIÇA |  Instituto de Investigação Interdisciplinar |  |
|  |  Grupo de Pesquisa Gestão, Desempenho e Efetividade do Judiciário |  Núcleo de Pesquisa em Informação, Direito e Sociedade |  LINGUAGEM, INSTITUIÇÕES E ORGANIZAÇÕES | |

Esse ecossistema global mostra o caminho para uma evolução híbrida: manter a autonomia conquistada com a solução on-premise, enquanto se avalia, de forma criteriosa, a adoção de serviços em nuvem para ampliar capacidade analítica e acelerar a integração com soluções de IA que fortalecem a tomada de decisão estratégica.

Por fim, expressa-se o sentimento de missão cumprida ao ver um projeto de tamanha complexidade ser entregue com excelência, fruto de dedicação, resiliência e orgulho pelo potencial transformador de uma solução de TI concebida passo a passo. Cada configuração realizada e cada serviço disponibilizado representam conquistas que fortaleceram não apenas a plataforma, mas também a confiança no trabalho colaborativo e inovador de um time composto, em sua maioria, por servidores da casa.

Reconhecemos que os desafios futuros permanecem numerosos e complexos, mas reafirmamos a importância de experimentar o novo e enfrentar os desafios com coragem. Pois acredita-se que quando atuamos com paixão e convicção, os resultados extrapolam fronteiras organizacionais e alcançam dimensões pessoais, consolidando o DataJust não apenas como uma solução tecnológica, mas como um marco de inovação e inspiração para todos os que se arriscam em inovar.

Referências

Listar todas as obras citadas no texto, de acordo com o padrão APA (American Psychological Association).

- Blue, R. (2018). *Introducing Iceberg: Tables in data lakes at scale*. Retrieved from <https://iceberg.apache.org>
- Dixon, J. (2010). *Pentaho, Hadoop, and data lakes*. [Blog post]. Retrieved from <https://jamesdixon.wordpress.com>
- Gorelik, A. (2023). *The enterprise big data lake: Delivering the promise of big data and data science*. Novatec.
- Khine, P. P., & Wang, Z. S. (2018). Data Lake: A new ideology in big data era. *ITM Web of Conferences*, 17, 03025. EDP Sciences. <https://doi.org/10.1051/itmconf/20181703025>

16



- Reis, J., & Housley, M. (2019). *Fundamentos de engenharia de dados: Projete e construa sistemas de dados robustos*. O'Reilly Media.
- ResearchGate. (2025). *Data lake: Suas funcionalidades e aplicações*. Retrieved from https://www.researchgate.net/publication/388486864_DATA_LAKE_suas_funcionalidades_e_aplicacoes
- Singh, R. (2019). *Big data analytics: Concepts, technologies, and applications*. Springer.
- White, T. (2015). *Hadoop: The definitive guide*. O'Reilly Media.

17

| | | | | |
|---|--|---|--|---|
|  |  INSTITUTO BRASILEIRO DE ESTUDOS E PESQUISAS SOCIAIS |  Universidade de Brasília |  |  Universidade Potiguar |
|  |  |  DGP DIREÇÃO-GERAL DA POLÍTICA DE JUSTIÇA |  Iluris Instituto de Investigação Interdisciplinar |  AJUS Administração da Justiça |
|  |  GEJUD Grupo de Pesquisa Gestão, Desempenho e Efetividade do Judiciário |  InfoJus Núcleo de Pesquisa em Informação, Direito e Sociedade |  LIOrg LINGUAGEM, INSTITUIÇÕES E ORGANIZAÇÕES | |