

Arquitetura Agnóstica para Transcrição, Armazenamento e Análise de Áudios Operacionais no Setor Elétrico Utilizando Serviços de Nuvem e Elasticsearch

Patrícia C. Wang¹, Maurício S. Magalhães¹, Hugo D. Rebelo¹,
Leonardo T. Vianna¹, Hugo R. V. Portuita¹, Tássio Simioni¹, Rodrigo M. Rocha²

¹Development and Innovation Dept (RD&I),
Radix Engineering and Software, Rio de Janeiro – RJ – Brazil

²Operations Center, CPFL, Campinas – SP – Brazil

{pdi, patricia.wang}@radixeng.com

Abstract. This article presents a cloud-based solution for the transcription, structured storage, and analysis of audio data from power system operations. The developed tool leverages Elasticsearch for efficient storage and indexing of transcribed data and integrates with data science pipelines for semi-automatic labeling, operator evaluation, and visualization through Business Intelligence (BI). The proposed architecture is cloud-agnostic and has been validated in AWS, GCP, and Azure environments. Experiments demonstrate an average reduction of 83.3% in the time required to search for operational events, highlighting the solution's potential for auditing, informational efficiency, and reliability in the power sector.

Resumo. Este artigo apresenta uma solução baseada em computação em nuvem para a transcrição, armazenamento estruturado e análise de áudios provenientes da operação de sistemas elétricos. A ferramenta desenvolvida utiliza o Elasticsearch para armazenamento e indexação eficiente dos dados transcritos, e integra-se a pipelines de ciência de dados para rotulagem semiautomática, avaliação de operadores e visualização via Business Intelligence (BI). A arquitetura proposta é agnóstica ao provedor de nuvem e foi validada em ambientes AWS, GCP e Azure. Experimentos demonstram uma redução média de 83,3% no tempo de busca por eventos operacionais, evidenciando o potencial da solução para auditoria, eficiência informacional e confiabilidade do setor elétrico.

1. Introdução

As operações do setor elétrico exigem registros precisos de comunicação para garantir a rastreabilidade e a eficiência nos processos decisórios. Essa demanda se estende por toda as etapas da rede de geração, transmissão e distribuição elétrica, onde as interações verbais desempenham um papel central na coordenação de ações operacionais em tempo real. Com o aumento do volume de dados não estruturados nesse contexto, especialmente aqueles provenientes de gravações de chamadas telefônicas, torna-se cada vez mais necessário o uso de mecanismos automatizados de transcrição e indexação que possibilitem a exploração eficiente dessas informações [Jurafsky and Martin 2023].

Esse desafio é particularmente relevante em centros de operação, onde decisões críticas são tomadas com base em interações verbais [Toledo 2017]. Essas comunicações, muitas vezes não registradas formalmente em sistemas estruturados, carregam informações operacionais essenciais, como instruções de manobra, alertas de falha ou coordenações de desligamento de equipamentos. Automatizar a transcrição desses registros permite transformar esses dados não estruturados em documentos textuais pesquisáveis e auditáveis, o que é fundamental para a transparência e conformidade regulatória, como exigido por entidades reguladoras e operacionais do setor elétrico brasileiro, como o Operador Nacional do Sistema Elétrico (ONS) [Operador Nacional do Sistema Elétrico (ONS) 2023].

Além da conformidade regulatória, a transcrição automatizada das comunicações verbais oferece benefícios operacionais significativos. Ao converter gravações de voz em texto estruturado, torna-se possível correlacionar informações contidas nas conversas com eventos registrados nos sistemas supervisórios e de gestão da rede elétrica [Floriano 2021]. Essa integração facilita a reconstrução de ocorrências e alimenta os processos de pós-operação, como a análise de falhas, identificação de não conformidades e emissão de relatórios técnicos. Dessa forma, reduz-se substancialmente o tempo necessário para localizar comunicações relevantes contidas nos áudios, aumentando-se a produtividade e otimizando-se o emprego do capital humano. Além disso, as transcrições dos registros de áudios também possibilitam a automatização da avaliação e monitoramento da qualidade da comunicação dos operadores.

Diante desse cenário, este trabalho propõe uma arquitetura de dados voltada ao armazenamento, transcrição e análise de áudios operacionais no setor elétrico. A solução, baseada em técnicas de processamento de linguagem natural (NLP) e serviços em nuvem, foi concebida para ser empregada no desenvolvimento de uma ferramenta para uma companhia do setor elétrico brasileiro. Seu objetivo é atender às demandas crescentes por auditoria, análise de eventos e avaliação do desempenho comunicacional em centros de operação, contribuindo para a conformidade regulatória e eficiência operacional.

2. Trabalhos Relacionados

Ferramentas de transcrição automática e análise de áudio têm ganhado destaque em setores críticos, como o de saúde [Gonçalves et al. 2024] e jurídico [Carrari 2024]. No entanto, aplicações voltadas ao setor elétrico, integrando essas tecnologias em pipelines robustos com foco em confiabilidade operacional, ainda são escassas.

No contexto de bancos de dados, o *Elasticsearch* é utilizado para otimização de buscas em grandes volumes de texto [Gormley and Tong 2015], tendo avanços nos mecanismos de indexação e recuperação eficiente de dados textuais. Por exemplo, [Cavalcante et al. 2020] utiliza o *Elasticsearch* como motor de busca em seu trabalho. [Soares et al. 2020] propõe a utilização do *Elasticsearch* para indexar impressões digitais convertidas em representações textuais, visando melhorar a eficiência na busca e recuperação dessas informações. Enquanto [de Andrade et al. 2023] propõe uma metodologia para o reconhecimento de entidades nomeadas em textos não estruturados, utilizando modelos de aprendizado de máquina para identificar e classificar informações relevantes em grandes volumes de dados textuais. Ainda que esses trabalhos avancem na modelagem e acesso eficiente a dados textuais, não abordam o desafio de ingestão e

processamento de áudios operacionais em setores como o elétrico.

Do ponto de vista de arquiteturas de dados em nuvem, [Silva 2022] propuseram *pipelines* baseados em *Apache Airflow* para integração de dados meteorológicos em tempo real, destacando a flexibilidade e escalabilidade da orquestração em ambientes distribuídos. Nosso trabalho diferencia-se ao adaptar essa abordagem para dados de áudio e texto provenientes de gravações telefônicas, com múltiplos pontos de entrada, validação de integridade e correlação com dados operacionais estruturados.

Observa-se uma crescente adoção de técnicas de NLP, em especial aquelas fundamentadas em modelos baseados em BERT [Devlin et al. 2019], para a identificação de padrões semânticos em distintos domínios. Tais abordagens têm sido empregadas em contextos que vão desde a análise automatizada de documentos governamentais até a extração de informações relevantes em sistemas de apoio à mobilidade urbana.

No entanto, essas abordagens geralmente se concentram em dados já textuais e estruturados, ao passo que nosso trabalho aborda toda a cadeia, desde a transcrição automática de áudio até a indexação, rotulagem e análise orientada à operação. Contudo, no contexto do setor elétrico, a combinação de *embeddings* semânticos com dados operacionais, como planos de manobra e identificadores de chave, ainda é inexplorada.

Portanto, este trabalho contribui ao estado da arte ao propor uma solução completa e aplicada, que automatiza o processo de transcrição de áudios operacionais com foco em conformidade regulatória, utiliza *Elasticsearch* para indexação e consulta eficiente dos dados transcritos, integra pipelines distribuídos com *Apache Airflow* em ambientes de nuvem híbrida (AWS, Azure, GCP) e oferece módulos específicos para rotulagem semântica, avaliação de desempenho humano e visualização de indicadores via BI.

3. Arquitetura da Solução

A solução é composta por uma arquitetura híbrida, com componentes locais (*on-premise*) e na nuvem. A parte local é responsável pela ingestão e tratamento inicial dos áudios, enquanto a parte em nuvem executa os processos de transcrição, análise e armazenamento. A Figura 1 apresenta em resumo o diagrama da arquitetura.

No ambiente local, as gravações são recebidas via *Secure File Transfer Protocol* (SFTP), passando por um processamento inicial que inclui a remoção de silêncio e separação de canais. Em seguida, os arquivos são comprimidos e enviados para o *Blob Storage*. Também ocorre a extração de dados de sistemas legados, comunicação via texto (CWSi), *Advanced Distribution Management Systems* (ADMS) e sistema de execuções de ordens de serviço (*InService*), por meio de consultas SQL específicas. Esses dados são compactados e enviados para a nuvem.

Na infraestrutura em nuvem, os áudios são reconstruídos e enviados para transcrição. Os textos e metadados gerados são carregados no *Elasticsearch*. Posteriormente, ocorre a correlação desses dados com as informações extraídas das fontes operacionais (como ADMS e CWSi). A solução ainda contempla integração para permitir a visualização dos dados por meio do *Power BI*.

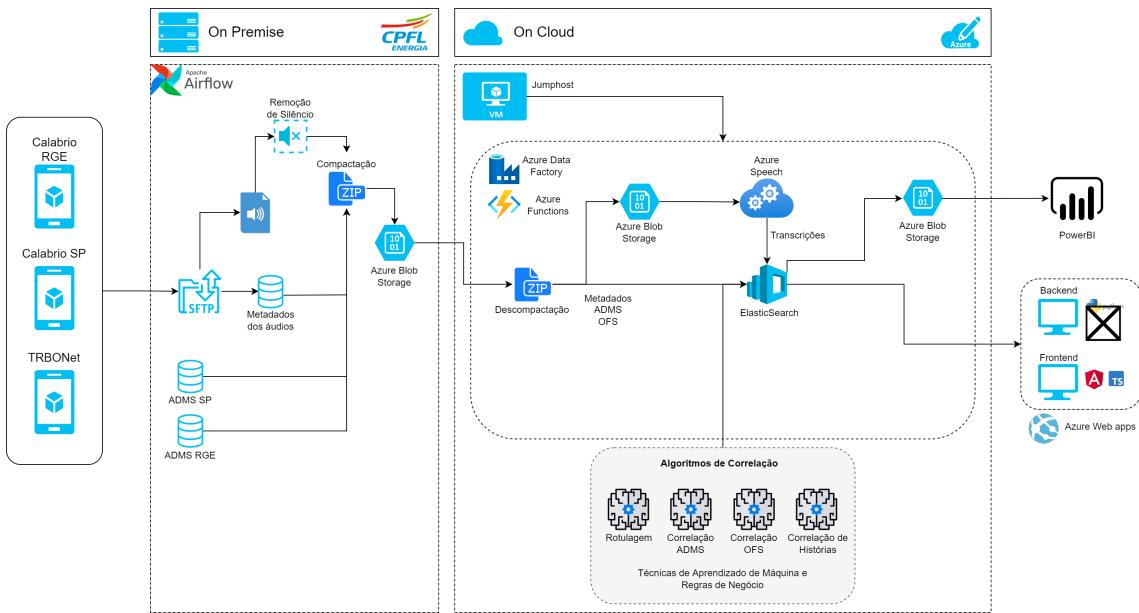


Figura 1. Diagrama de arquitetura da solução.

3.1. Ingestão, ETL e Armazenamento

O processo de ingestão utiliza o *Apache Airflow* para orquestrar as tarefas em dois níveis distintos. No ambiente local (*Airflow on-premise*), o sistema recebe arquivos via SFTP, realiza o pré-processamento dos áudios, compacta e envia os dados para o *Blob Storage*, além de extrair e compactar dados oriundos dos sistemas internos. Já na nuvem (*Airflow na nuvem*), são executadas as tarefas de transcrição dos áudios, reconstrução a partir de fragmentos, carregamento dos dados no *Elasticsearch*, e aplicação de correlações com dados operacionais relevantes.

3.2. Visualização e Autenticação

O *front-end* acessa os dados via *API REST*, com autenticação via *Active Directory Federation Services (ADFS)*. A visualização dos dados é realizada via *PowerBI*, diretamente conectado ao *Elasticsearch*, permitindo geração de *dashboards* e relatórios operacionais.

4. Módulos Funcionais

Três módulos principais compõem o núcleo funcional da solução: o Módulo de Rotulagem, o Módulo de Avaliação do Operador e o Módulo de BI. Cada um desses módulos desempenha um papel específico na transformação dos dados de áudio em informação estruturada e útil para o setor elétrico.

O Módulo de Rotulagem tem como objetivo agrupar automaticamente transcrições que abordam temas similares, facilitando a identificação de padrões de eventos e a busca por casos correlatos. Para isso, são utilizadas técnicas de NLP, baseadas em *embeddings* e medidas de similaridade semântica, como aquelas derivadas do modelo BERT. Isso permite que eventos operacionais com descrições distintas, mas semanticamente equivalentes, sejam agrupados de forma eficiente, otimizando a análise por parte dos especialistas.

Já o Módulo de Avaliação do Operador se dedica à análise da comunicação dos operadores, medindo atributos como clareza, completude, redundância e tempo de res-

posta. A aderência aos protocolos operacionais é também considerada, permitindo o cálculo de indicadores que refletem a confiabilidade da atuação humana durante os eventos registrados. Essa avaliação contribui diretamente para ações de capacitação, *feedback* e melhoria contínua da equipe operacional.

Por fim, o Módulo de BI provê uma interface analítica que possibilita a consulta ágil a diversas informações, como o número de eventos operacionais registrados, as notas atribuídas a cada operador, a qualidade média das transcrições por lote de dados e as tendências operacionais observadas ao longo do tempo. Com isso, a solução oferece suporte direto à tomada de decisão, promovendo maior transparência e rastreabilidade.

5. Resultados e Avaliação Experimental

Para estimar o desempenho da ferramenta desenvolvida na identificação de gravações de áudio relevantes, foi conduzido um estudo comparativo entre a nova aplicação e o sistema PCS, até então utilizado pela CPFL para esse fim. O experimento baseou-se em uma demanda real da equipe de pós-operação do Centro de Operações da Subtransmissão (COS), relacionada à localização de áudios associados a uma ocorrência operacional registrada em 06/02/2022, entre 08h e 09h30. A ocorrência em questão envolveu manobras na rede de 34,5 kV responsável pelo atendimento à subestação particular Jaguar Plásticos, sob responsabilidade da CPFL Santa Cruz. Durante esse intervalo, foram identificados 87 áudios totalizando 3 horas e 31 minutos de gravação.

A consulta realizada no PCS utilizou como critério de filtro apenas o intervalo de tempo da ocorrência. O sistema retornou 87 gravações de áudio referentes às operações das distribuidoras CPFL Paulista, CPFL Piratininga e CPFL Santa Cruz. Para identificar os registros específicos da ocorrência, foi necessário escutar integralmente ou parcialmente todas as gravações. O tempo total despendido por um operador experiente para localizar os áudios relevantes foi de aproximadamente 2 horas.

Com a ferramenta desenvolvida, foram realizadas três abordagens distintas de pesquisa para localização dos áudios da mesma ocorrência.

Na primeira abordagem, aplicou-se o mesmo filtro temporal utilizado no PCS, acrescido do filtro de unidade operacional, restringindo os resultados à CPFL Santa Cruz. Essa operação retornou 53 transcrições, representando uma redução de 39,1% na quantidade de áudios a serem analisados.

Na segunda abordagem, foi adicionado ao filtro temporal o termo de pesquisa “Jaguar”, em referência à subestação envolvida na ocorrência. O sistema retornou todos os áudios/transcrições do período em que houve menção à subestação Jaguar Plásticos, facilitando significativamente a identificação do contexto relevante.

Na terceira abordagem, foi utilizado o filtro de data combinado ao número operativo da chave “978007”, conforme descrito no plano de manobras. O resultado foi a recuperação precisa do áudio/transcrição que mencionava a referida chave. Essa abordagem obteve o menor tempo necessário para localização dos áudios pertinentes, aproximadamente 20 minutos.

A comparação entre as duas ferramentas evidencia ganhos substanciais em eficiência e usabilidade. A nova aplicação possibilitou a aplicação de filtros mais refinados, como o agrupamento por empresa e a busca textual por termos específicos (lo-

calidades, equipamentos ou códigos operacionais). Isso resultou em uma redução significativa do tempo de análise, de 2 horas para 20 minutos — uma economia de aproximadamente 83,3%. Essa otimização se traduziu em uma busca até 6 vezes mais rápida por informações específicas, acelerando o diagnóstico de ocorrências críticas. Adicionalmente, a ferramenta oferece maior precisão na recuperação de informações e contribui para a melhoria da rastreabilidade dos eventos operacionais, aspectos essenciais para auditoria, resposta a incidentes e aprendizado organizacional.

A arquitetura agnóstica, testada com sucesso em três grandes provedores (AWS, GCP e Azure), provou-se escalável, segura e eficiente. A adoção de ferramentas como *Elasticsearch* e *Airflow* permitiu flexibilidade na ingestão e transformação de dados multimodais (áudio, metadados e históricos).

O uso de NLP para rotulagem e avaliação de operadores representa um avanço significativo no uso de ciência de dados em aplicações críticas do setor elétrico, alinhando-se às demandas por transparência, confiabilidade e automação.

6. Conclusões e Trabalhos Futuros

Este artigo apresentou uma ferramenta para transcrição, armazenamento e análise de dados operacionais de áudio no setor elétrico. A construção dessa solução enfrentou desafios importantes, especialmente no que diz respeito à organização e padronização dos dados em um ambiente marcado por fontes heterogêneas, formatos distintos de gravação e ausência de estrutura nos registros de voz. A complexidade aumenta quando se considera a necessidade de integrar esses dados com sistemas legados, exigindo *pipelines* robustos de ETL e mecanismos avançados de correlação semântica.

Apesar desses obstáculos, os ganhos de eficiência e auditabilidade demonstram o valor da aplicação nesse contexto crítico. A solução proposta foi projetada para operar de forma agnóstica em relação ao ambiente tecnológico das concessionárias, podendo ser adaptada a diferentes infraestruturas e realidades operacionais. Isso amplia seu potencial de replicação e escalabilidade em outros cenários.

Para trabalhos futuros o foco é avaliar o desempenho de novos modelos fundamentais de *Automatic Speech Recognition* (ASR) com suporte multilíngue e sensibilidade a ataques regionais, particularmente aqueles focados no português brasileiro. Além disso, incorporar métricas de robustez e escalabilidade em ambientes operacionais contínuos, incluindo avaliações em cenários de edge computing.

A análise das gravações de áudio em forma escrita possibilitou a utilização de filtros estruturados para buscar gravações que contenham palavras ou siglas, permitindo a redução do tempo gasto pelos responsáveis na análise das gravações e, por conseguinte, maior eficiência operacional. Dessa forma, a arquitetura proposta e a ferramenta desenvolvida tem grande potencial para otimizar as atividades e viabilizar análises quantitativas e qualitativas a respeito da operação e das ações tomadas pelos operadores.

Agradecimentos

Agradecer à CPFL Energia pelo apoio técnico e financeiro, bem como ao Programa de Pesquisa e Desenvolvimento da ANEEL, no qual este trabalho foi desenvolvido e financiado pelo subsídio PD00063-3071/2020.

Referências

- Carrari, G. M. (2024). Desenvolvimento de modelo de Inteligência Artificial para degravação de depoimentos judiciais. Dissertação de Mestrado, Instituto Federal do Espírito Santo.
- Caçalante, L., Lima, U., Barbosa, L., Gomes, A. L., Santana, E., and Martins, T. (2020). Improving search quality with automatic ranking evaluation and tuning. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 157–162. SBC.
- de Andrade, C. M., França, C., Belém, F., Jallais, G., Ganem, M. A., Texeira, G., Landerer, A. H., and Gonçalves, M. A. (2023). Promptner: Uma abordagem para reconhecimento de entidades nomeadas em dados sensíveis a partir de instâncias rotuladas automaticamente. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 269–281. SBC.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.
- Floriano, A. C. (2021). Transcrições Automatizadas de Áudios para Centros de Operação do Setor Elétrico. Dissertação de mestrado, Instituto Federal de Santa Catarina (IFSC), Florianópolis, SC. Acesso em 14/05/2025.
- Gonçalves, Y. T., Alves, J. V. B., Sá, B. A. D., da Silva, L. N., de Macedo, J. A. F., and da Silva, T. L. C. (2024). Speech recognition models in assisting medical history. In *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados (SBBD 2024)*, pages 485–497, Florianópolis, SC. Sociedade Brasileira de Computação.
- Gormley, C. and Tong, Z. (2015). *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. O'Reilly Media, Inc.
- Jurafsky, D. and Martin, J. H. (2023). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd)[draft].
- Operador Nacional do Sistema Elétrico (ONS) (2023). Procedimentos de Rede – Regras e Critérios para a Operação do Sistema Interligado Nacional. <https://www.ons.org.br/paginas/sobre-o-sin/rede-basica/procedimentos-de-rede>. Acesso em 13/05/2025.
- Silva, R. F. C. (2022). Arquitectura para Processamento Integrado no Repositório de Dados e Metadados de Observação Terrestre. Dissertação de Mestrado, Universidade NOVA de Lisboa (Portugal).
- Soares, J. M. S., Barbosa, L., Rego, P. A. L., Magalhães, R. P., and de Macêdo, J. A. F. (2020). Indexando impressões digitais utilizando índice invertido: Uma investigação inicial. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 181–186. SBC.
- Toledo, L. H. S. d. (2017). Estudo de caso de um centro de controle de sistema elétrico de potência: Uma abordagem tridimensional. Dissertação de Mestrado, Universidade Federal de Minas Gerais. Acesso em: 14/05/2025.