



Yuan Zhang

Contato

- ✉ zhang489@hotmail.com
- ☎ (61) 98142-0098
- 📍 Brasília - DF
- 🌐 www.linkedin.com/in/zngyuan/

Formação Acadêmica

- **Pontifícia Universidade Católica de Minas Gerais | Pós-Graduação**
Ciência de Dados
outubro de 2021 - outubro de 2022
- **Universidade de Brasília - UnB**
Master's degree (Interrompido)
2018 - 2022
- **Faculdade Metropolitana | Pós-Graduação**
Big Data Analytics
dezembro de 2020 - novembro de 2021
- **Universidade de Brasília - UnB**
Bachelor's degree, Industrial Chemistry
2012 - 2017

Idiomas

- Português
- Inglês
- Mandarin
- Shanghainese

Sobre mim

Formado pela Universidade de Brasília, e mestrando na Universidade de Brasília (UnB) e Pós-graduado na área de Big Data e Data Science. Tenho experiência com ETL utilizando SQL, PySpark, Python, Pentaho e outras diversas ferramentas, além de atuar no processo de ingestão e validação de dados nos ambientes Azure, Google, Minio, e Dremio Lake, juntamente na criação de indicadores e insights utilizando as bases de dados na áreas da saúde, economia, e educação. Teste e implementação de prompts dentro da esteira de dados.

Experiência Profissional

Programa Justiça 4.0 - Ministro Luís Roberto Barroso - Projeto idealizado por Luís Roberto Barroso ministro do Supremo Tribunal Federal (STF)

➤ **Data Consultant**

Julho de 2024 - Dezembro de 2024

Atividades:

- Uso de CI/CD na integração do desenvolvimento dos códigos do GitHub;
- Levantamento do Terraform - Minio Storage e estruturação do DataLake;
- Extração dos API's com WebScraping com implementação de Bots (Assíncrono ou Paralelo);
- Integração dos servidores com o Lake;
- ETL com PySpark, Python, R, Polars, DuckDB e Scala;
- Orquestração com Apache Airflow, e testes com Dagster;
- Sugestão e implementação do DBT em transformação de dados e na modelagem de tabela fato e dimensional;
- Implementação de teste do Prompt utilizando Llama 3.1;
- Criação de API utilizando FastAPI para a testes de MLOPS com a utilização dos modelos de NLP;
- Utilização do Docker Compose;
- Governança de dados com DBT;

Skills: FastAPI · GCP · Python · SQL · DBT · AirFlow · DuckDB · Git

Certificações

- **Azure Databricks & Spark Core For Data Engineers** (Python/SQL) - IDUC-085bc804-736e-4ab7-9f10-0de45bf3805d
- **Astronomer logo Astronomer Certification DAG Authoring for Apache Airflow**
- **Astronomer Certification for Apache Airflow Fundamentals**
- **ETL - Azure Architecture Center** (Keyrus Academy) - No expiration date
- **Story Telling McSill Story Studio andB!** Issued Nov 2021 - Credential ID7488754
- **Data Analyst Career Track Data** (365 Data Science Credential IDDD-23DA35E347
- **Data Scientist Career Track Data** (365 Data Science Credential IDDD-16126A97CA
- **Git and GitHub** (365 Data Science) Credential IDCC-1BD7181CA1
- **Web Scraping and API Fundamentals in Python Web Scraping and API Fundamentals in Python** (365 Data Science) Credential IDCC-71A945447B
- **Business Analyst Career Track Business Analyst Career Track** (365 Data Science) Credential IDDD-52CB97E66F
- **Huawei Open Class AI** - AI Huawei Open Class Issued Aug 2020 - No expiration date
- **Desenvolvimento ETL/DataStage ID UC-** 1f6ca21d-b9d2-456f-a783-877fc5a80943

Hard Skill

- Python
- Javascript
- SQL
- PySpark
- Pentaho
- Data Factory
- DataBricks
- Apache AirFlow
- Git
- FastAPI
- Flask
- Azure, GCP, AWS, Minio, Dremio
- Data Factory
- Docker
- Luigi
- Power BI
- Unity Catalog
- DBT

Confederação Nacional da Indústria

► Data Engineer | Analytics Engineer | Backend

Outubro de 2021 - Julho de 2024

Atividades:

- Criação de bot para webscraping e realização de ingestão de dados no Blob Storage.
- ETL com PySpark, Python e Pentaho.
- Criação do catálogo de dados utilizando Unity Catalog.
- Criação de pipelines e workflows.
- Orquestração de dados no ambiente Azure com uso de Azure DataFactory e AirFlow.
- Estruturação do banco de dados.
- Criação de indicadores e insights das bases da saúde (absenteísmo), e economia (monitor do emprego).

Skills: Azure DevOps · Azure Data Factory · Azure Databricks · Azure Data Lake · SQL DW · PySpark · Python · SQL · Unity Catalog

Confederação Nacional da Indústria

► Analytics Engineer

Outubro de 2021 - Março de 2022

Atividades:

- Análise de dados e criação de indicadores com pandas, pandera e koala.
- Python e PySpark (DataBricks).
- Modelagem e construção de indicadores com o uso de dados públicos (RAIS, CAGED, CENSO ESCOLA, IBGE).
- Criação de Bot em WebScraping (Bs4, Selenium).
- Análises de dados.
- Criação de BI com o uso de PowerBI.

Skills: Extract, Transform, Load (ETL) · SQL hive · DeltaLake · Koala · Apache Airflow · Azure DevOps · Azure Databricks · Azure Data Lake · PySpark · Python · DBeaver

HartB – Innovation as a Service

► Data Engineer | Analytics Engineer

Outubro de 2020 - Setembro de 2021

Atividades:

- Estruturação de dados IPTU - SP na criação de painel de aluguel.
- Ingestão de dados e criação de Database pelo AWS Crawler.
- Tratamento de dados via AWS Glue.
- Criação de KPI's pelo PySpark Glue e via Python
- Utilização de Dask em processamento.
- Athena, e Redshift para consumo.

Skills: AWS Glue · AWS Lambda · Athena · S3 · GlueSpark · FastAPI · Python · Dask · PySpark · SQL · Insomnia · RedShift · WebScraping

DataDrill

► Analytics Engineer

Agosto de 2019 – Setembro de 2020

Atividades:

- Armazenamento dos dados no MiniO.
- Ingestão de dados e ETL com Python.
- Orquestração com AirFlow.
- Construção do produto.
- Painéis em Power BI.
- Construção do produto por meio das criações de KPI's utilizando-se dos dados do censo escolar, IBGE e RAIS.

Skills: Python · DBeaver · MinIO · PostgreSQL · SQL · AirFlow

Natura & Co

► Data Engineer | Backend

Maio de 2019 – Outubro de 2019

Atividades:

- ETL com Python e Pentaho.
- Implementação dos KPI's.
- Estudo de indicadores dos dados da INEP (CENSO ESCOLAR).
- Armazenamento dos dados no BigQuery.

Skills: Pentaho · FastAPI · Google BigQuery · Python · SQL

Pan American Health Organization (OPAS) | Consultor

► Data Engineer

Novembro de 2018 – Abril de 2019

Atividades:

- Estruturação de dados a partir de MySQL, e RedShift juntamente com dados de DataSus e posteriormente realizando a ingestão dos dados no BigQuery.
- Utilização do Flask para o deploy.
- Utilização do AirFlow em scheduling.
- ETL com Python.
- Processamento dos dados e criação de indicadores.

Skills: AWS Glue · Amazon S3 · Amazon Athena · Amazon Redshift · Flask · Google Cloud Platform (GCP) · Apache Airflow · Google BigQuery · Python · SQL · Cloud Functions

PNUD – ONU | Consultor

► Data Scientist

Maio de 2018 – Outubro de 2018

Atividades:

- WebScraping com Python e Javascript.
- Utilização das libraries: Selenium, BS4, Request, Puppeteer.
- Modelagem com banco de dados utilizando Postgres.
- Levantamento de informações, e criação de relatórios para as tomadas de decisões.

Skills: Selenium · Beautiful Soup · Puppeteer · Scrapy · JavaScript · Python · SQL · Postgres

Shanghai Pudong Development Bank Internship

► Data Analyst

China – Shanghai

Dezembro de 2017 – Abril de 2018

Atividades:

- Análise e desenvolvimento de painéis.
- Levantamento de estudo dos insight.
- Análise de dados pentaho e python.