

# Fredy Reyes Sánchez Data Engineer

Ciudad del Carmen, Campeche, México | <https://www.linkedin.com/in/fredy-reyes-sanchez/> | <https://github.com/fredyliniux> | +13235575050 [fredyreysa@gmail.com](mailto:fredyreysa@gmail.com)

## Career Summary

Data Engineer with over 10 years of extensive experience in Big Data Development, including a deep understanding of LLM (Large Language Models), AI (Artificial Intelligence), ML (Machine Learning), and predictive modeling. I specialize in Azure, AWS, and GCP platforms, utilizing tools such as Hadoop, Databricks, Spark, and PySpark. With a strong background in Azure Data Factory, I've successfully designed and implemented solutions leveraging Azure Data Analytics services, including Azure Storage, Azure SQL Data Warehouse, Azure Data Lake, Azure Cosmos DB, and Azure Data Factory. Additionally, my expertise extends to data modeling, working with star and snowflake schemas, and I've contributed to various data warehouse projects using Snowflake and Databricks. My proficiency in predictive modeling and machine learning enhances data-driven decision-making processes.

## Summary of Qualifications

Hadoop/Big Data	Kafka, HDFS, YARN, Hive, Apache Spark, Presto DB, Horton Works, Cloudera, MapReduce
NoSQL Database	HBase, Cassandra, MongoDB
Build and Deployment	Docker, Maven, Sbt, Git, SVN, Jenkins, Terraform
Scripting	Python, Scala, Pyspark, SQL, Shell Scripting, Pig Latin, HiveQL, R, Bash
Databases	PostgreSQL, MariaDB, Teradata, BigQuery, Oracle, MySQL, MS SQL Server, SQLite
Analytics	Tableau, Power BI, Microsoft SSIS, SSAS and SSRS, Looker, QlikView
Web Dev. Technologies	Streamlit, Flask, HTML, XML, JSON, CSS
ETL	AWS Glue, Informatica PowerCenter, Coalesce
Operating Systems	Linux, Unix, macOS, Windows 8, Windows 7, Windows Server 2008/2003
AWS	EC2, EMR, S3, Redshift, EMR, Lambda, Athena, RDS, Step Functions
Azure	Azure DevOps, Azure Data Factory, Azure SQL Data Warehouse, Cosmos DB, Azure Functions
Google GCP	Composer, Cloud Storage, DataFlow, Cloud Run, Cloud Pub/Sub
Python	NumPy, Pandas, PySpark, Geopandas, Scikit Learn, Matplotlib, Seaborn, Tensorflow, Keras
Cloud-Based Data Processing	Databricks, Snowflake, Azure Synapse, Redshift, Big Query, MongoDB Atlas

## Languages

- Spanish (native speaker)
- English (advanced)

## Career History

**Ascendion , Remote**  
**Senior Engineer**

**Nov, 2023 – Today**

- Research, evaluate, and recommend generative AI frameworks, libraries, and tools suitable for specific use cases, considering factors such as model complexity, scalability, and integration capabilities with Snowflake.
- Collaborate with data scientists and machine learning engineers to design and implement proof of concept projects for generative AI applications, leveraging Snowflake's data infrastructure and machine learning capabilities.
- Develop and deploy machine learning models for natural language processing (NLP), text generation, image synthesis, or other generative AI tasks using frameworks like TensorFlow, PyTorch, or Hugging Face Transformers.
- Conduct performance testing and benchmarking of generative AI models on Snowflake, analyzing factors like inference speed, resource utilization, and model accuracy to optimize model deployment and usage.

- Integrate Apache Spark-based data processing workflows with Snowflake for end-to-end data pipelines, leveraging Spark's distributed computing capabilities for preprocessing, feature engineering, and model training.
- Design and implement automation processes using Apache Spark and Snowflake for data integration, transformation, and loading tasks, streamlining data pipelines and reducing manual effort.
- Monitor and manage Spark jobs and Snowflake tasks using monitoring tools and dashboards, ensuring the reliability, scalability, and performance of data processing workflows.
- Collaborate with cross-functional teams to identify opportunities for process automation and optimization, leveraging Snowflake's data platform and Spark's computing capabilities to drive operational efficiency and agility.
- Stay updated on the latest trends and advancements in generative AI, machine learning, and data processing technologies, actively participating in industry forums, conferences, and training programs to enhance expertise and knowledge.
- Continuously iterate and improve data engineering and AI solutions based on feedback, performance metrics, and emerging business requirements, ensuring alignment with organizational goals and objectives.

***Walmart, Mexico City***

**Apr, 2022 – Dec, 2023**

***Senior Big Data Engineer***

- Designing and implementing highly performant data ingestion pipelines from multiple sources using Apache Spark and/or Azure Databricks.
- Solution designing using the Azure Data Analytics platform Microsoft Azure services.
- Development and expert knowledge of Azure Data Lake storage and ingestion using Azure Data Factory.
- Proficient in developing and implementing highly performant data ingestion pipelines using Apache Spark and PySpark for processing large volumes of data from diverse sources.
- Experienced in designing and deploying structured Spark streaming pipelines, leveraging distributed big data technologies and modern architectural approaches.
- Experienced in optimizing the performance of Spark jobs by fine-tuning parameters, optimizing resource utilization, and enhancing overall job efficiency.
- Proficient in designing, building, and unit testing applications on the Spark framework using Python, ensuring robustness and reliability of Spark-based solutions.
- Development and expert knowledge of Azure Databricks for transformations and processing data.
- Develop modern data architectural approaches for business intelligence reporting and analytics, including that for machine learning models and data science, ensuring effectiveness, scalability, and reliability.
- Develop real-time data pipelines, streaming analytics, distributed big data, and machine learning infrastructure.
- Design, build and unit test applications on Spark framework on Python.

***eCivis Inc Pasadena, California (USA)***

**Sep, 2021 – Apr, 2022**

***Senior Data Engineer***

- Implement database schema for ETL and BI reporting solutions in multiple databases including Microsoft SQL Server, Amazon Relational Database Services for SQL Server, AWS Redshift/Postgres, and MySQL.
- Design, develop, implement, and optimize ETL processes that merge data from disparate sources into Microsoft SQL Server and AWS Redshift for consumption by enterprise SaaS applications.
- Mining administrative data for security and behavioral compliance using AI/ML.
- Conduct impact analysis and provide policy recommendations.
- Analyze large volumes of structured and unstructured data.
- Explore and enhance data models (text, numerical).

- Implement security best practices and other compliance requirements.

**Scrumlink, Santa Clara (USA)**

**Feb, 2021 – Sep, 2021**

*Senior Data Engineer / Scientist*

- Design, build and unit test applications on Spark framework on Python.
- Building Data Pipelines using Databricks and scheduling Databricks jobs.
- Airflow scheduling Data Pipelines of Databricks.
- Building Databricks notebooks in Scala, PySpark and SQL.
- Used Broadcast Join in SPARK for making smaller datasets to large datasets without shuffling of data across nodes.
- Designed and implemented Sqoop for the incremental job to read data from DB2 and load to Hive tables and connected to Tableau for generating interactive reports using Hive server2.
- Develop and deploy the outcome using spark and Scala code in Hadoop cluster running on GCP.

**MC1 Win the Market, Mexico City**

**Mar, 2019 – Feb, 2021**

*Data Engineer / Scientist*

- ETL Developer, with extensive knowledge of coding languages and warehouse architecture as SQData.
- Integration with cloud applications, real-time business intelligence and legacy data migrations with SQData.
- Data replication and Integration process with the most important tools in the marketplace as: Hadoop, SQData, Spark, Kafka.
- ETL processes to design and fill a bespoke data warehousing environment for WTM's application.
- Developed new storage systems and deploy new cloud instances and testing of new data environments.
- Guidance to data scientists and full-stack engineers to build sustainable and high-quality software solutions using the latest advances in data science, data engineer, and cloud-engineering

**Insituto Nacional Electoral (INE), Mexico City**

**Ene, 2018 – Feb, 2019**

*Data Engineer*

- Data migration Oracle databases to AWS services.
- Migration and replication of multiple databases to AWS cloud infrastructure.
- Systems Administrator of Red Hat, Solaris, Windows, and Linux servers.
- Member of technical architecture team for migration to AWS cloud services to ensure high availability of PREP website.
- Scripted and coded design in Python, Java and Power Shell.
- Implementation of technical security configuration guidance to enterprise information systems teams.
- Vulnerability scanning, penetration testing, and security assessments in support of risk management and audit activities. Establish configuration of security policies under standard ISO 27001.

## **Education**

**Metropolitan Autonomous University (UAM)**

**2015**

Electronic Engineering

**Monterrey Institute of Technology (ITESM)**

**2023**

Master's in Applied Artificial Intelligence.

## **Courses and Certifications**

- Web Development, Data Science, Angular, Java, Monitoring Software, AWS, Network, Operative Systems, Programming, Project Management, Anaconda, Python, Scrum, SQL, Sales and marketing.