ALEKSEI DIAZ

SUMMARY

Aleksei is an accomplished Data Engineer with more than 5 years of experience in designing and developing end-to-end ETL pipelines across AWS, Azure, and GCP. Proficient in Python, PySpark, ETL processes, and SQL, he has a proven track record of delivering effective ad-hoc data solutions.

SKILLS

- SQL (5 years)
- Python (4 years)
- ETL (4 years)
- PowerBI (3 years)
- AWS (2 years)
- APIs (2 years)
- Azure Databricks
- Azure DevOps
- Azure Synapse
- Google Cloud Platform
- Athena
- SharePoint
- Microsoft Power Query
- Microsoft Power Apps
- Git; GitHub Actions
- DAX

EXPERIENCE

Globant

July 2024 - November 2024

Globant is a tech services firm specializing in digital products, bridging businesses and consumers through innovative technology and AI expertise.

- Develops a Lakehouse architecture from the ground up based on the Medallion framework, structuring the data pipeline into three distinct stages.
- To improve performance and scalability, design and optimize table structures, focusing on partitioning strategies and data type selection.
- Establishes development standards, including notebook structure, coding conventions, and naming guidelines for objects (tables, jobs, views), organizing ETL processes for consistency and maintainability.
- Creates ETL notebooks for data ingestion, cleansing, and transformation, ensuring efficient and reliable data processing workflows.
- Automates deployment workflows using GitHub Actions, ensuring reliable CI/CD across all environments.

Role: Senior Data Engineer

Technologies: PySpark, SQL, Git, GitHub Actions, ETL, Azure Databricks

Indra

July 2023 - July 2024

Indra is a leading software company renowned for its innovative solutions and cutting-edge technology, specializing in enterprise software, cloud computing, and artificial intelligence.

- Worked on a lift & shift migration project to Databricks.
- Migrated 100% of bronze-layer on-prem procedures to their equivalent in Databricks Notebooks (Python & SQL).
- Provided support and troubleshooting for deployed workflows, identifying and resolving bugs to maintain seamless data operations.
- Conducted programmatic validation to ensure that migrated processes produced consistent results with their on-premises equivalents, maintaining data integrity and quality.

Role: Data Engineer

Technologies: Python, PySpark, SQL, Azure Databricks, Azure Data Factory, Jenkins, Git

Ernst & Young (EY)

May 2022 - July 2023

Ernst & Young (EY) is a global leader in assurance, tax, transaction, and advisory services. With a presence in over 150 countries, EY helps businesses navigate complex challenges, capitalize on opportunities, and achieve sustainable growth.

- Worked as a Data Engineer in a team to build a Lakehouse from scratch.
- Developed and optimized data transformation and cleaning processes using serverless PySpark.
- Designed and implemented data pipelines using AWS Step Functions, orchestrating complex workflows with scalability and fault tolerance.
- Utilized AWS Lambda functions for executing short, lightweight tasks, enhancing data automation through webhooks and workflow integrations.
- Leveraged S3 with Iceberg to build a data lake.
- Worked as a Data Engineer in a team to migrate 100% of processes to Azure Synapse in a separate project.
- Migrated 100% of procedures from a cloud vendor to Azure Synapse, enabling better integration with native Azure services such as Power BI.
- Enhanced maintenance by scripting stored procedures, reducing reliance on Python for database operations.
- Documented the new architecture to clearly communicate its status and details to all stakeholders.

Role: Data Engineer

Technologies: PySpark, SQL, AWS (AWS Glue, S3, Athena), Azure DevOps, Azure Data Factory, Power BI, Git

Neo Consulting

June 2021 - May 2022

Neo Consulting is a digital consulting firm specializing in Strategy, Innovation, and Digital Transformation, with a focus on data activation and Machine Learning.

- Served as a key stakeholder for all cloud solutions, deciding on architecture, table schemas, and services based on the final client's requirements.
- Implemented cloud-native pipelines in GCP and AWS to provide ready-to-use data for building dashboards and ML models.
- Automated data ingestion through APIs, eliminating manual processes.
- Defined database and data warehouse architecture, considering storage, normalization, and final usage to create a BI ecosystem.

Role: Cloud Engineer

Technologies: PySpark, SQL, AWS (S3, Lambda), Azure Databricks, Google Cloud Platform, BigQuery, Power BI

Ransa

January 2021 - June 2021

Ransa is a leading end-to-end logistics operator in Latin America, with 85 years of experience in the market.

- Assessed customer needs to design and build dashboards that effectively displayed data according to their requirements.
- Collaborated with internal customers to define descriptive KPIs, maximizing the value and usability of data.
- Automated data wrangling with Power Query, eliminating the need for manual data cleaning.
- Designed and implemented multiple highly customized Power BI dashboards across the company to support data-driven decision-making.

Role: Data Scientist

Technologies: Power Query, Power BI, SQL, DAX

Nexa Resources

April 2020 - January 2021

Nexa Resources is a global mining company that is part of Votorantim and plays a significant role in the daily life of society.

- Took charge of the preparation and automation of reports, achieving a 90% reduction in time.
- Designed the information relationship model for the cost data of five mining units located in two countries.
- Proposed and implemented the standardization of databases and their use in shared spaces.

Role: Business Intelligence Developer

Technologies: Power BI, DAX, SharePoint, Microsoft Power Query

APM Terminals

April 2019 - April 2020

APM Terminals is lifting global trade and the standards of proactivity, connectivity, reliability, responsibility, and efficiency.

- Prepared weekly reports on the status of initiatives and maintained records of SOPs.
- Designed, managed, and developed process documentation.
- Analyzed, improved, and automated processes.
- Mapped existing processes (as-is).
- Developed interactive dashboards with remote accessibility for the entire company.
- Successfully migrated data from Excel databases to SQL databases.
- Provided training on accurate record-keeping.
- Conducted training on the new use of mobile applications (Power BI – Power Apps).

Role: Business Intelligence Developer

Technologies: SQL, Power BI, Azure, DAX, Microsoft Power Apps, Microsoft Power Query

EDUCATION

- **Industrial Engineering,** Pontifical Catholic University of Peru - Peru (2013 - 2020)
- **Mechanical Engineering,** University of Texas - United States (2018)

LANGUAGES

- **Spanish:** Native
- **English:** Proficient (B2)