José Armenteros

Sr. Principal Data Engineer

Contact

Address
McLean, VA, 22101

Phone
(305) 962-4701

E-mail
jose.r.armenteros@gmail.com

Skills

ETL Pipelines
Excellent

Data Management
Excellent

People Leadership
Excellent

Machine Learning & AI
Very Good

Software

Python
Excellent

SQL
Excellent

Spark
Excellent

Cloud Services
Very Good

Docker & Kubernetes
Good

Data engineer experienced with end-to-end data pipeline development. I am skilled at setting up data systems and other frameworks, managing infrastructure, and aligning procedures with business needs. I thrive in cooperative environments and enjoy mentoring staff.

Relevant Work History

2022.08 – Present

Senior Delivery Principal

Slalom, Tysons Corner, VA
- Developed a pipeline for bulk geocoding of hundreds of millions of data records using Smarty and PySpark
- Led migration of on-prem enterprise data platforms towards COTS SaaS offering
- Redesigned data platform environment models that reduced compute costs by 50% and ~1 petabyte of object storage
- Led a D&A community of practice
- Developed protocols that led to 30% reduction of ETL deployment lifecycle
- Leads a nine-person data engineering team
- Manages and mentors 3 mid-to-senior level developers
- Works with leaders in other markets to launch and support companywide business development and learning initiatives
- Mentors multiple individuals across PM, D&A, and SE specialties

2021.10 – 2022.08

Principal

Slalom, Tysons Corner, VA
- Terraformed Databricks platforms using a custom-made python helper applications
- Designed a comprehensive EIM system that utilizes all top market technologies in data ingestion, ETL, machine learning, and data governance
- Led a five-person data engineering team
- Managed and mentored four direct reports
- Developed and trained junior staff on data engineer, machine learning, and cloud systems architecture topics
- Participated in response generation and oral presentations for federal government bids/contracts

2021.08 – 2021.10

Principal Data Scientist

Oracle, Reston, VA
- Analyzed unstructured text to extract key terms and themes relevant to current business problems through Large Language Model (LLM) analysis
- Updated and expanded functionality for an internal support chat bot using python's rasa software packages
- Provided technical guidance and mentorship to 6 mid-level/senior developers on team

2021.03 –
2021.07

Senior Data Scientist

Oracle, Reston, VA
- Engineered machine-learning classification and clustering models to improve OCI operations and change management
- Generated topic clustering algorithms that allowed for classification of root cause problems in major operational programs into generic topics for consumption by teams

2018.03 –
2021.02

Senior Data Engineer

Oracle, Seattle, WA
- Designed and implemented ETLs to transform highly nested, semi-structured JSON blobs into relational database models
- Migrated ETL pipelines to Spark and cloud-native services, improving refresh cadences tenfold and increasing column coverage from ~30% to 98% (1500 fields).

2016.01 –
2018.02

Data Analyst

Decisive Data, Redmond, WA
- Generated a system that predicted Microsoft BizSpark usage data using internal datasets and social media activity.
- Developed an application that expedited conversion of thousands of Costco's old IDB2 queries to their new Oracle DB infrastructure by a factor of 10.

Education

2007.09 –
2011.06

Bachelor of Science: Biological Engineering

Massachusetts Institute of Technology - Cambridge, MA