MARCO ANTONIO PEREIRA
DATA ARCHITECT

City/Country: Viseu/Portugal
Mobile: +351 912 616 357
Email: marco77ap@gmail.com
Linkedin: https://www.linkedin.com/in/marcoap/
Github: https://github.com/MarcoAP

Skill	*Demonstrated in CV?*	Where
Problem Solving	Yes	Recognized for the ability to learn new technologies and implement best practices in various projects.
AWS	Yes	Experienced in AWS services and migrating Big Data environments from on-premises to AWS.
Azure	Yes	Experienced in Azure services and migrating Big Data environments from on-premises to Azure.
Google Cloud	Yes	Participated in projects utilizing Google Cloud services for data analytics and storage solutions.
Databricks	Yes	Utilized Databricks for developing and optimizing data analytics workflows and pipelines.
Continuous Integration	Yes	Experienced with CI tools such as Git, Jenkins, AWS CodeCommit, and AWS CodePipeline.
NoSQL Database	Yes	Worked with HBase (Cloudera stack), DynamoDB (AWS), Redis, MongoDB (On-Premises), CosmosDB (Azure).
RDBMS	Yes	Worked with MySQL, Oracle, MS SQL Server, both on-premises and cloud.
Hadoop	Yes	Vast experience with Hadoop, including design, implementation, and maintenance.
Python	Yes	Developed data processing pipelines and ETL processes using Python.
Apache Spark	Yes	Deep experience developing data pipelines with PySpark and Scala, and Spark.

Profile Summary

Professional Architect and Data Engineer with over 19 years of industry experience, specializing in big data technologies for nearly six years. Proficient in automation tools, Hadoop/Spark development, and possessing a comprehensive understanding of the end-to-end software design lifecycle.

Demonstrates proficiency in cloud technologies, particularly Azure and AWS, including expertise in migrating from on-premises solutions to cloud environments.

Skilled across all phases of the Software Development Life Cycle (SDLC) with a keen emphasis on design, development, maintenance, analysis, programming, testing, and implementation across diverse applications.

Recognized for exceptional interpersonal skills, fostering strong rapport with colleagues and stakeholders. Proven track record of initiating, promoting, and maintaining robust interpersonal relations with an ability to navigate various circumstances courteously, professionally, and tactfully.

Certifications

Brand/Vendor	Certification	License
databricks	Databricks Certified Data Engineer Professional	Link
	Databricks Certified Associate Developer for Apache Spark 3.0	Link
	Databricks Certified Hadoop Migration Architect	Link
aws	SAA-C03 AWS Certified Solutions Architect – Associate	F63C6P81VM4E1SSM / Link
	CLF-C01 AWS Certified Cloud Practitioner	B1ZVRHYB3J4E1P9T / Link
Azure	AZ-305 Microsoft Certified: Azure Solutions Architect Expert	Link
	AZ-400 + AZ-104 Microsoft Certified: DevOps Engineer Expert	Link
	DP-203 Microsoft Certified: Azure Data Engineer Associate	Link
	AZ-104 Microsoft Certified: Azure Administrator Associate	Link
	AZ-900 Microsoft Certified: Azure Fundamentals	Link
Google Cloud	Google Cloud Certified - Professional Cloud Architect	Link
	Google Cloud Certified - Professional Data Engineer Certification	Link
CCA cloudera SPARK & HADOOP DEVELOPER	CCA-175 Cloudera Certified - CCA Spark and Hadoop Developer	100-023-557 / Link
Scrum.org	PSM II - Professional Scrum Master	Link
	PSM I - Professional Scrum Master	
	PSPO I - Professional Scrum Product Owner	
ITIL® Foundation V3	ITIL® v3 Foundation in IT Service Management	5136198.20309974
EXIN IT Service Management ISO/IEC 20000 FOUNDATION	EXIN IT Service Management Foundation based on ISO/IEC 20000	5136198.20327260
EXIN Information Security Management ISO/IEC 27001 FOUNDATION	EXIN Information Security Foundation based on ISO/IEC 27001	5136198.20332538

Technologies - Knowledge Base

Brand/Vendor	Tool/Service
databricks	• *Cloud Provider: AWS, Azure and GCP* • *Lake: Delta Lake* • *Tables: Delta Tables, Delta Live Tables* • *Type engine: Spark SQL and Spark Streaming* • *Governance: Unity Catalog*
aws	• Storage: AWS S3 • RDBMS: AWS RDS • NoSQL: AWS DynamoDB • Compute: AWS Lambda, AWS EC2 • Hadoop: AWS EMR • ETL: AWS Glue • DW: AWS RedShift • Container and K8s: AWS ECS, AWS EKS • Networking: VPC, Subnet, Security Groups, ACLs • Management and Security: AWS IAM, AWS KMS • IaaC: CloudFormation, Terraform • CI/CD: AWS CodeCommit, AWS CodePipeline, AWS CodeBuild • Logging: AWS CloudWatch • Streaming: AWS KMS
Azure	• *Storage: Azure Blob Storage, Azure Data Lake Storage gen2* • Data Pipeline: Azure Data Factory • *Compute: Azure Functions* • *Data Engineering: Azure Databricks* • *NoSQL: CosmosDB* • *DW: Azure Synapse Analytics* • *Data Cataloging: Azure Purview* • IaaC: Terraform • Management and Security: Azure EntraID • Streaming: Azure Event Hub
Cloud Architect · Google Cloud Certified · Professional	• *Storage: Buckets* • *Compute: VM Instances* • *Serverless: Google Cloud Functions* • IaaC: Terraform
Apache	• *Apache Airflow* • *Apache Camel* • *Apache JMeter* • *Apache Spark* • *Apache Kafka* • *Apache Pig* • *Apache HBase* • *Apache Flume* • *Apache Nifi* • *Apache Hadoop*
Dev Ops	• *Git* • *Docker* • *Kubernetes*

Professional Experience

TEKEVER ***Tekever***	***From:*** *November 2024* ***To:*** *Present* ***Portugal - Hybrid***	***Role:*** *Data Architect* ***Permanent Employee***	
Project	Data Platform	**Sector**	Surveillance-as-a-service solution
Tasks	At Tekever, I am actively involved in the design and evolution of the company's Data Platform, built on Kubernetes with Helm as the deployment standard. The platform integrates and manages a wide range of services, including: ● **Apache Kafka (Bitnami)** as the messaging backbone ● **Apache Spark Operator (Kubeflow)** with Apache Spark 3.5 for large-scale data processing ● **Traefik** as the ingress controller ● **Keycloak** for authentication and authorization ● **WireGuard** providing secure VPN services ● **Apache Airflow** and **Apache NiFi** for orchestration and DataFlow management ● **Neo4j** as the Graph Database ● **Apache Druid** as the Time Series Database ● **ArgoCD** for Helm-based environment management ● **Rancher** for multi-region cluster orchestration ● **MLflow** for machine learning model lifecycle management The platform is designed with a **Cloud-Agnostic mindset**, ensuring flexibility, scalability, and independence from vendor lock-in.		
	Key Technologies: Kubernetes, GKE (Google Kubernetes Engine), Google Cloud, Apache Spark, Apache Kafka, Terraform, Bitbucket, Neo4J, Apache Nifi		

ULTRA TENDENCY pioneering the future ***Ultra Tendency***	***From:*** *October 2022* ***To:*** *October 2024* ***100% Remote from Portugal to Germany***	***Role:*** *Big Data Architect* ***Permanent Employee***	
Project	Drei Österreich (Hutchison Drei Austria) – Austria	**Sector**	Telecom

Tasks	● **Data Migration from On-Premises to Cloud** o I evaluated the on-premises infrastructure and defined the migration requirements and objectives. o I compared cloud providers (Azure, GCP, AWS) based on cost-effectiveness, performance, security, and tools. o I developed proof-of-concepts on Azure, GCP, AWS, Databricks, and Snowflake to assess performance and cost. ● **Defining KPIs (Key Performance Indicators)** o I identified and defined relevant KPIs, such as daily data ingestion rate (10 TB/day) and total volume stored (500 TB). o I monitored resource utilization (CPU 70%, memory 60%, disk 50%) and job success rate (95%). o I measured data processing time (2 hours on average), data availability (99.5%), and access latency (100 ms). ● **Review of Technologies Used in the Cluster** o I used Apache Hive and HDFS for data storage and Apache Spark (PySpark) for data processing. o I monitored with Grafana and InfluxDB, managed with Cloudera Manager, Apache Ranger, and Kerberos. o I tested Apache Airflow for workflow scheduling and used Tableau for result visualization.		
	Key Technologies: PoCs with AWS, Azure, GCP, Databricks and Snowflake		
Project	ECB (European Central Bank) - Germany	**Sector**	Banking
Tasks	● Design the architectural project for the Proof of Concept and Final Project. ● Estimate hours and create a Sprint plan to inform senior management. ● Validate architecture with the client's architecture team. ● Create documentation with High-Level Design and Low-Level Design. ● Understand technical restrictions (IP ranges, VPC, etc.). ● Present estimated cost plan for using the architecture based on Cloud AWS. ● Support DevOps team on architectural issues. ● Progress with the "lift-and-shift" migration (Cloudera On-Premises to Cloud AWS). **Key Technologies:** AWS Stack (MSK, Lambda, EKS, EMR, VPC, IAM, Terraform).		
Project	Internal - MLOps - Germany	**Sector**	Sales
Tasks	● Develop internal initiative (asset) based on MLOps to enable the sales team to sell services based on pre-programming or pre-defined machine learning algorithms or enable customization from development to deployment. **Key Technologies:** Azure Stack (focused on Data and Analytics), AWS Stack (focused on Data and Analytics), Databricks, Kubernetes, Docker, Terraform, Apache Airflow.		
Project	Cerascreen - Germany	**Sector**	Health/ Medicine

Tasks	• Manage the client's internal IT team. • Draw up a plan based on goals and priorities and present it to senior management. • Seek continuous improvement on the AWS architecture and implement it. • Maintain direct contact with AWS when necessary. • Promote internal training on AWS technologies/services. **Key Technologies:** AWS Stack (focused on Data and Analytics).

NTT DATA ***NTT Data***	***From:*** *February 2018* ***To:*** *September 2022*	***Role:*** *Chief Architect* ***Permanent Employee***	
Project	TT Club (Thomas Miller) - UK	**Sector**	Insurance
Tasks	• Migrate the current Data Warehouse On-Premises to Cloud (Microsoft Azure). • Present architecture design based on Microsoft Azure focused on Data and Analytics. • Create PoC (Proof of Concept). • Ensure all components (Datasets, integration runtime, pipelines, workflows, ETL, triggers) are working as expected. • Create monitoring. • Support and train the team. • Team Lead. **Key Technologies:** MS SQL Server On-Premises, SyBase, Azure SQL, Azure Data Factory, Databricks.		
Project	Santander Portugal - Portugal	**Sector**	Banking
Tasks	• Migrate Cloudera workflows to Azure Databricks and Kafka Confluent. • Design target architecture for migration and data-source mapping. • Develop MVP for use cases. • Lead team development. **Key Technologies:** Hadoop, Hive, Spark, Apache Kafka, Confluent, Control M, Java 8, Spring Boot, Python 3, Azure Databricks, Azure Data Factory, CosmosDB, OpenShift.		
Project	Santander UK - UK	**Sector**	Banking
Tasks	• Migrate data lake to AWS Cloud. • Design target architecture and map existing technologies to AWS solutions. • Migrate Hive-based transformation jobs to AWS Glue workflows using Spark. • Lead team development. **Key Technologies:** Hadoop, Hive, Spark, Yarn, Python, Terraform, AWS Glue (PySpark), VPC/Subnet, IAM, RDS/Oracle, AWS Athena.		

Project	Itau - Brazil	**Sector**	Banking
Tasks	• Migrate on-premises Cloudera Data Lake to AWS services. • Design target architecture and technology mapping. • Migrate hive-based transformation jobs to EMR and AWS Glue. • Research, analyze, recommend, and select technical approaches. • Develop templates for best practices. • Implement advanced search engine for bank statements. • Design architecture based on AWS services. • Develop data processing pipelines consuming Kafka topics with customer transactions. • Integrate with Lucidworks Fusion and Apache Solr. **Key Technologies:** Nifi, HBase, Lucidworks, EKS, CloudFormation, AWS Keyspaces (Cassandra of AWS), ECS, EC2, Glue (PySpark), Storage Gateway, S3, CloudWatch, CodeCommit/CodePipeline, Scala with Spark, AWS Athena, Hadoop, Hive, Spark - PySpark, Yarn, Python, EMR (PySpark), VPC/Subnet, IAM, RDS/Oracle.		
Project	Enel - Chile	**Sector**	Energy
Tasks	• Deploy and implement big data solutions for Enel based on Spark SQL/PySpark. • Analyze issues, design target architecture, deploy new cluster, implement best practices. • Train customer teams. • Manage Cloudera CDH cluster / Cloudera Administrator **Key Technologies:** Cloudera, PySpark.		
Project	Banco de Bogotá - Colombia	**Sector**	Banking
Tasks	• Deploy data processing architecture based on Scala and Spark. • Design and build data processing and ETL pipelines. • Implement and configure big data technologies. • Lead team of big data engineers. **Key Technologies:** Hortonworks, Scala, Spark.		
Project	Enel - Brazil	**Sector**	Energy
Tasks	• Develop and transform data for the information área based on Spark SQL. • Implement data quality processes. **Key Technologies:** PySpark.		
Project	GPA Grupo Pão de Açúcar - Brazil	**Sector**	Retail
Tasks	• Develop Tacaca framework for big data developments. • Design and develop Maven archetypes, CI/CD pipelines, and train data engineering team. **Key Technologies:** Scala, Java, Spark, Kudu, Hive, Jenkins, Apache Kylin.		

Project	Telefónica Vivo - Brazil	**Sector**	Telecom
Tasks	• Big data technical reference for Brazilian and Spanish teams. • Mentor junior team members. **Key Technologies:** PySpark, Hive.		

Sodexo	***From:*** *September 2017* ***To:*** *February 2018*	***Role:*** *Big Data Developer / Data Engineer* ***Contractor***	
Project	SPI DataRoadMap	**Sector**	Employee Benefits and Rewards Services
Tasks	• Implement proof of concept for Hadoop cluster. • Migrate Data Lake On-Premises to AWS. • Load charges on data lake Hive and Impala. • Implement data processing with Pig and Spark. • Train IT support team. **Key Technologies:** Hadoop, Hive, Impala, Scala, Spark, Flume, Kafka.		

IBM	***From:*** *August 2016* ***To:*** *August 2017*	***Role:*** *Big Data Developer / Data Engineer* ***Permanent Employee***	
Project	Santander Brazil - Brazil	**Sector**	Banking
Tasks	• Implement lambda architecture for real-time geolocation recommendations. • Develop EndPoints REST using Apache Camel. **Key Technologies:** Architectures (Lambda and Kappa) Cloudera CDH, Java 1.7, Apache Pig, Apache Spark/PySpark, Hbase, Hive, Impala.		

EAI Brasil	***From:*** *December 2013* ***To:*** *August 2016*	***Role:*** *Business Intelligence Developer* ***Contractor***	
Project	América Móvil - Brazil	**Sector**	Telecom
Tasks	• Development of Dimensions and DW foundations. • Analysis, Development and Integration with native applications such as Portlet Indicators of BI, developer Dashboards for Tableau, Obiee (Oracle Business Intelligence Enterprise Edition 11g) and Qlikview. • Development of PHP applications for quick-shots, using Symfony and Laravel frameworks, Mysql and Oracle 11g about database and ETL charge files.		

	Key Technologies: Oracle 9i, PHP.		
Project	GRU Airport - International Airport of Sao Paulo - Brazil	**Sector**	Airport
Tasks	● Developer new windows in billing system about departures and arrives with tax billing local based on weight, size, passengers numbers. ● Developer all new levels hierarchy menus of billing system, because in the old model all people it was using admin user (unique user; losing rastreability). ● Integrate, with Web Services, the SAP erp with the billing system. **Key Technologies:** C# and SQL Server database.		

Telefónica | Vivo	***From:*** *May 2011* ***To:*** *November 2013*	***Role:*** *Senior Software Engineer and Team Leader* ***Contractor***	
Project	Point Program System - Brazil	**Sector**	Telecom
Tasks	● System based on SMS "interactive" ie the client sends SMS to the Large Account / Short-code set in Middleware as a short message with the reserved word identified by Regular Expression intercepted by Erb and forwarded to the Central Via SMS Services which is managed by the Weblogic version 10MP2; Optin customers perform activation of the Points Program. **Key Technologies:** Java 1.6, Struts, JSP, Hibernate, EJB 3.0, Spring, Oracle 9i Database, JBoss and Weblogic Server.		
Project	PTT (PUSH TO TALK - Radio) - Brazil	**Sector**	Telecom
Tasks	● Based on Java, Oracle 10g, Weblogic 10MP2, configuration Large Account / Short-code set in the Middleware. Acted as Technical Leader, I managed scope, time, risk, and realized the development and implementation of the Reset Password PTT functionality via SMS where customers of Telefonica operator | Vivo could perform the reset password via SMS by service via SOA bus if the list of keywords sent in the SMS were identified. **Key Technologies:** Java 1.6, Struts, JSP, Hibernate, EJB 3.0, Spring, Oracle 9i Database, JBoss and Weblogic Server.		
Project	System Telefonica Vivo Cinemark - Brazil	**Sector**	Telecom
Tasks			
	● Based on Java, Oracle 10g, Weblogic 10MP2, configuration Large Account / Short-code set in the Middleware and mailing received via CTRL-M,		

	customers elect to receive promotional SMS from Brazilian Cinemas network, Cinemark, offering 50% discount, based on recharge statistics. **Key Technologies:** Java 1.6, Struts, JSP, Hibernate, EJB 3.0, Spring, Oracle 9i Database, JBoss and Weblogic Server.

Education

BSC Education – English Language Exchange Program

- Short-term English immersion program - London, United Kingdom

Master of Business Administration

- Fundação Getúlio Vargas - Brazil

Bachelor of Information System (Technology)

- Universidade Presbiteriana Mackenzie - Brazil

Languages

English	Advanced English (C1)
Portuguese	Native