

Demystify

the data realm



+33 6 74 20 08 54
mehdi@tazimehdi.com

MEHDI TAZI
DATA SPECIALIST

EDUCATION AND CERTIFICATIONS

- 2024** - Book Author: The Definitive Guide to Data Integration (Editor Packt)
- 2019** - AWS Big Data Specialty & Architect Associate
- 2016** - PhD at Paris-EST in the Internet of Things (IoT) Virtualization field (In Progress)
- Hadoop Hortonworks administrator certification
- Fundamentals of Data Science
- 2014-2015** - JAVA « OCAJ » Certification & Spring CORE Certification
- Apache Cassandra for Java developers Certification
- 2013** - Scrum Master Official Certification - Scrum Alliance
- MongoDB for DBAs and MongoDB for Java developers Certifications
- 2011** - Master's degree MIAGE (Business Informatics) in Engineering of Distributed Information Systems at (ESIAG [Paris - France](#)) (honors)
- 2010** - Engineering degree in Software Engineering [Fez – Morocco](#) (honors)
- Cisco CCNA Certification [Rabat – Morocco](#) (honors)



EXPERTISE

Technical & Architectural skills

- Defining a Data Strategy roadmap in alignment with business needs and top-level management strategy.
- Consulting and Designing Data Platforms, LakeHouses, Datalakes, Datalabs and Modern data stack solutions: Organization, Architecture, Modeling, Governance, Security and Monitoring and Data Observability.
- Participating in the development and maintenance of the IT architectural strategy and vision of enterprise architecture by continually evaluating trends in technology and business needs as well as selecting, designing and supporting implementation of new solutions for business transformation.
- Designing and integrating existing systems, BI, analytical systems, data science systems with the (big)data platform.
- Collecting business requirements, setting-up functional & technical solutions, then implementing monitoring & audits
- Designing ingestion, exchange, logging/archiving, modeling and exposition big data layers/zones.
- Setting-up BRP & BCP and Designing highly available, reactive, real-time, secure and governed solutions
- Coding, Consulting and designing architectures distributed/Parallel scalable BigData/NoSQL solutions.
- Setting-up comprehensive FinOps governance and optimization strategies, implementing standards and best practices for cost management, infrastructure optimization, and technical efficiencies.
- Monitoring & forecasting cost usage, contributing to purchasing optimizations and community engagement through webinars and workshops to foster a culture of FinOps excellence.
- Data Engineering using Java/Scala/Python : AirFlow, Spark, DBT, AWS(EMR,GLUE,Athena,Lambda), Databricks...
- Infrastructures & projects industrialization/Automation: DevOps, Infrastructure as code, MLOPS & Engineering stacks
- Pre-sales and Writing technical specifications.
- Technical Keynote speaker and trainer.

Storage

DB: Cassandra, MongoDB, DynamoDB, CosmosDB, HBase, Redis & RDBMS(Postgres, RDS ...)
FILESYSTEMS/ObjectStorage: HDFS, S3/Glacier/DeepArchive, ADLS2,...
FORMAT : DELTA, PARQUET, CSV, XML, JSON, YAML, AVRO...
STREAM : KAFKA, KINESIS...

Processing

Spark, HADOOP Technologies (Hive, SQOOP...), AWS (Glue, EMR, Databrew, DMS, Lambda, ECS, EKS ...), DBT, Talend for big data, Dataiku, AZURE ADF...

Consumption & Exposition

Athena, Impala, Trino/Presto, GraphQL/REST API, PowerBI, QuickSight SageMaker, ...

Automation

Docker, Airflow, Ansible, CloudFormation, Azure DevOPS, Jenkins, Terraform (in progress) ...

Environments

Infrastructure : OnPremise, Cloud, Multi-Cloud, Hybride architectures
Operation Systems: Windows, Unix, MacOS
Distributions/Providers: AWS, ~~Azure~~, Cloudera, Hortonworks, Confluent/Aiven

Tools

Intellij IDEA, Visual Studio, Jupyter, Git, Maven/Sbt, Sonar, JIRA, QC, SNow, UML, Merise...

Prog languages

Python, Scala, Java/JEE, C#, HTML5/CSS3, SQL, CQL, JS

- **Project management skills**

- Accompany teams setuping & onboarding their use cases
- Coach, Lead, Estimate, Plan, Organize and monitor projects.
- Animate meetings and motivate teams, Facilitate relations between and within teams.
- Communicate on the goals & visions and write reports to the top management.
- Prepare the releases, iterations and update the user stories and the backlog with the product owner.
- Split the product into Use cases and prepare the releases, then Split Use cases into tasks and prepares the iterations.
- Establishment of resources availability and of the production capacity tables.
- Generating indicators like (BurnUp/Down charts, Team velocity, Tasks monitoring, Code Quality,Scope).
- Identifications and risks management.

- **FinOPS skills**

- Design and implementation of FinOps governance, standards & best practices
- Implementation of FinOps solutions to eliminate orphans (auto-tagging, CMDB, etc.)
- Technical optimization for cost optimizations using Engineering patterns like bucketing, partitioning, zordering, etc.. and by proposing light stacks and proposing decisions tree to choose the right technology.
- Infrastructure optimization by choosing the right size of machines and clusters
- Participation in architecture committees to accompany projects from the design phase to the most optimized choices in terms of costs and architecture
- Purchasing optimizations of cloud services
- Cost usage, spend monitoring and regular reporting with KPIs
- Forecasting & financial estimations
- FinOps community management: evangelization of FinOps activities, culture and tools via webinars, workshops, FinOps event
- Automation of FinOps activities via in-house tools like dashboarding (tableau,powerbi) and CMDB
- Storage optimization of storage tiering and data lifecycle management

- **Tierce fields knowledge**

- Stock Exchange, Real Estate, Travel, Web accessibility.
- Business fields: Energy, Stock market, Retail banking, Business Travel, E-Commerce...



- **Principal Data Architect**
- **Employer:** Brainlyze – <https://brainlyze.com/>
- **Clients:** Retail Client
- **Job:** Core Data Architect, Datalake Architect
- **Mission:**

4 Years

- **Design of the Data Platform Architecture:** Definition of the target architecture based on Lakehouse and Data Mesh paradigms, integrating observability, governance, and compliance with PII/sensitive data requirements.
- **IT Roadmap Establishment:** Design and maintenance of a strategic roadmap for the evolution of the Data Platform, from current state to target operating model.
- **Core Features Implementation Support:** Contribution to the setup and implementation of essential platform features to enable efficient data management and analytics.
- **Data Solutions Architecture:** Design of both high-level and low-level data solution architectures to meet identified business and technical needs.
- **Requirements Gathering and Qualification:** Collection of teams' needs, qualification of requirements, and proposal of appropriate technical solutions.
- **Use Case Enablement:** Accompaniment and guidance of business and technical teams in implementing their use cases on the platform.
- **Access Automation:** Implementation of automated processes for granting and managing access to the Data Platform.
- **FinOps Activities:** Cost optimization and efficiency improvements through architecture reviews and technical enhancements.
- **Sensitive Data Management:** Implementation of platform features enabling the management and protection of PII and sensitive data.
- **Multi-Cloud Initiatives:** Participation in initiatives aiming at extending the Data Platform to multi-cloud environments.
- **Databricks Migration Program:** Transition from native AWS Data Platform to a Databricks-oriented platform, including the implementation of Unity Catalog and compute services.
- **Catalog Strategy and Migration:** Progressive migration strategy from AWS Glue Catalog to a hybrid approach (Glue + Unity) in the short term, then full adoption of Unity Catalog.
- **Third-Party Integrations:** Integration of Databricks with external tools such as Collibra, AWS services, and complementary technologies.

Technologies: AWS, Databricks + Unity, DBT, EMR, Glue, Athena, Lambda, EKS/K8s, etc.



4 Years

- **ENGIE : Principal Lead (Big) Data/Cloud Architect (application & infrastructure)**

- **Employer:** IT INSIGHT – www.itinsight.fr
- **Clients:** Engie GEM, Engie IT **Paris – France** – www.engie.fr
- **Job:** I'm the lead big data architect at Engie GEM, I've done many projects and studies as Big Data architect
- **Missions :**
 - Responsible for Big Data architectures in a hybrid & multi-cloud context (OnPrem-Azure-AWS)
 - Support & Consulting to promote the Big Data / Cloud approaches and patterns to develop high-performance use cases.
 - Onboarding and support for teams in the implementation of their business use cases.
 - Collaborate with the Business and IT to provide the overall guidance and structure for creating and promoting the Data approach and use cases
 - Provide recommendations to the business in order to develop use cases using Data
 - Collaboration with application architects, infrastructures, APIs, Security, DevOPS to guarantee the consistency of the data offer.
 - Collaboration with the Engie Group for the evolution & improvement of data products (DataCatalog, Data services, organization ...).
 - Implementation of a set of tools accelerating teams onboarding, code development, easing learning curve and promoting the big data / cloud culture: Code samples, Wiki / How to, Frameworks, Workshops, documents good practices and overall guidance.
 - Collaboration for the continuous setup of the data strategy & vision through the evaluation and implementation of new technologies and approaches/patterns.
 - Automation of standard data platform processes: FineOps & billing management, Monitoring, users & groups management...
 - Study and implementation of several advanced subjects: API on Datalake, Governance, Security, Optimization, Automation & DevOPS ...
 - Integration of the platform with third-party solutions: Collibra, Dataiku, PowereBI, Authentication, Splunk ...
 - Industrialization of solutions thanks to DevOps and Infra as code
 - Industrialization of log analysis for the automation of predictive maintenance & monitoring as well as the proposal of architectures meeting business requirements.
 - Promoting & defining data guidance and structure for creating and promoting the Data projects and use cases.

Participating in the development and maintenance of the IT architectural strategy and vision of enterprise architecture by continually evaluating trends in technology and business needs as well as selecting, designing and supporting implementation of new solutions for business transformation

PROFESIONAL EXPERIENCES

- Collaborating with Engie Group in evolving and enhancing the data products and data catalog for all Engie entities.
- Global data products offer (Data, SDK, Catalog, Billing)
- Cost and performance optimization of big data applications
- Management of development teams
- Monitoring and optimzing Billings
- Implementing an automatic solution for tracking un-optimized big data / cloud solutions (cloud principes, queries respecting the partitioning ...)

Tools: Databricks, AWS, Spark, Glue services, S3, ATHENA, Kinesis / SNS, DMS, SageMaker, DynamoDB, Lambda, Docker, EC2, API GATEWAY, CloudFormation, CloudWatch, CloudTrails, AWS BATCH, EKS, Azure (DataFactory, ADLS, U-SQL, HDINSIGHT, EventHub), Scala /Python, Hadoop...



2 Years

- Big Data – NoSQL Architect (Application & Platform)

- **Employer:** IT INSIGHT – www.itinsight.fr

- **Clients:** AXA WORLD – TECH & AXA – DATA INNOVATION LAB Paris – France – www.axa.fr

- **Job:** I have done many projects and studies as Big Data architect and maintained the AXA Cluster

- Missions :

- Setting up diagrams of: technical architecture, high and low-level design, classes, sequence, component, de deployment and infrastructure & flows. Setup & Conducts projects and PoCs
- Studying big data architectures, solutions and products (cloudera navigator, Kafka Navigator integration ...)
- Main lead on cloudera navigator which is used for data governance and access monitoring
- Main Designer & Architect of "SelfService" project which is a solution for automating operations on the big data platform starting from the ticketing app to the ansible scripts.
- Administering the clusters and providing level three support: CDH5, HDFS, HIVE, SPARK, OOOZIE...
- Coding and integrating tools for the big data platform with third party systems (RDMS, logging systems, ftp servers, LDAP, Kerberos ...) : CDH5, Kerberos, OpenLDAP, CentOS, postgresql ...
- Developing automation tools and scripts for deployment and operational procedures: Ansible, python ...
- ❖ **Tools:** Hadoop Cloudera, Ansible, LDAP, Kerberos, HDFS, HIVE, KAFKA, SPARK, OOOZIE, FLUME, PYTHON, JAVA, SILVA, HAPROXY, UNIX, PostgreSQL, Docker...



7 Years

- Chief Technical Officer (CTO) & Principal Data Architect

- **Employer:** BI-NEWVISION – <https://bi-newvision.com/>

- **Clients:** Société Generale, Decathlon, ONDA, etc

- **Job:** Chief Technical Officer & Lead Data Architect. Responsible for the Data offering, I lead all Data projects, ensure their successful delivery, and provide expertise in both Data and technology.



9months

- ONDA : Designing and building the ONDA Platform and its use cases to transform the airports of Casablanca and Rabat by leveraging Data and IA.

- **Project Scoping and Requirements Gathering :** Identify business and technical expectations (business use cases, regulatory needs, pain points and limitations, etc.),
- **Analysis of the Organization and Existing Data Foundation:** Understand the current infrastructure and its dependencies, limitations, existing applications, and the roles of each team.
- **Technical Architecture Document:** Drafting the Technical Architecture Document and detailing the technical architecture that address the identified needs.
- **Installation of Data Platform :** Setup and Implementation of the Data Platform "Lakehouse" based on Cloudera Private Cloud base. The Platform was setup for big data management and analytics, enabling the centralization and real-time processing of data from both airports.
- **Use cases Implementations:** Leading teams technically to efficiently implement the solutions on top of Cloudera Private Cloud base, by providing expertise and guidelines around Data Management and Big Data Technologies. The use cases includes:
 - Development of a dynamic dashboard for real-time monitoring of key performance indicators (KPIs).
 - Use of artificial intelligence and machine learning to analyze passenger flows, anticipate peak traffic, and adjust operational priorities accordingly.

- SGMA : Design of SGMA Data Platform "DataLake" to centralize data from the group and its subsidiaries.

- **Business Requirements Gathering:** Collecting business needs, regulatory needs, etc.
- **Pain Points & Limitations Assessment:** Evaluate the challenges & needs of the group and its subsidiaries.
- **Analysis of the Organization and Existing Data Foundation:** Understand current infrastructures, applications, and organization.
- **Target Architecture Design:** Proposal of the Functional and Technical architectures that address the identified needs. The proposal was addressing both the long-term architecture and a short-term architectures.
- **Implementation Scenarios Proposal:** Present implementation scenarios with varying complexity levels and pricing models. Assistance in decision-making.
- **Definition of the Implementation Roadmap:** Establish the deployment trajectory, guiding principles and best practices.
- **Governance Programs Identification:** Define requirements for data catalog, lineage, common dictionary, processes, etc.



1 year 3 months

PROFESIONAL EXPERIENCES

OCTO
Technology
2 years

- Big Data – NoSQL Architect / Agile Coach (Application)

- **Employer:** OCTO Technology Paris – France – www.octo.com

- **Clients:** BNP PARIBAS, LA POSTE, COLISSIMO, EDF-DOAAT, ERDF, AL BARID BANK, SMCP, OCTO ACADEMY

- **Job:** I have done several projects mainly as a Big Data architect, and occasionally as an agile coach

- Missions:

- Identification of needs, use cases and Big Data perimeter; establishing roadmaps.
- MVPs and POCs Identification; Scooping, management and implementation of projects.
- BigData Consulting and accompaniment: Administration, Architecture, Governance and Security...
- Designing Big Data technical, conceptual, functional, organizational and methodological architectures.
- Designing Datalakes & Datalabs: Organization, Architecture, Modeling and Governance.
- Designing and integrating existing, BI and analytical systems with the big data.
- Designing ingestion, exchange, logging/archiving, modeling and exposition big data zones.
- Setting-up BRP & BCP and Designing highly available, reactive, real-time and secure solutions
- Design, Management and development of Big Data solutions.
- Scripts, infrastructures, deployments and configurations automation and industrialization.
- Pre-sales and Big Data training; Agile coaching
- ❖ **Tools:** Java/JEE, Scala, Hadoop, Spark, Kafka, Ansible, Docker, Cassandra, HBase, Hive, Sqoop, Parquet, Avro, HDFS, Ambari, Zookeeper, Confluent.io, Jupyter, Git, Jenkins, JHipster, Scrum...

REFERENCES :

Thierry GRIMA	GROUP CHIEF ANALYTICS OFFICER	Engie & AXA	https://www.linkedin.com/in/thierrygrima/
Raphael MANSUY	DATA ARCHITECT	DECATHLON	https://www.linkedin.com/in/raphaelmansuy/

LANGUAGES

English: Fluent

French: Bilingual

Arabic: Mother tongue

PUBLICATIONS AND PUBLIC INTERVENTIONS

- Book Author : Data Integration Definitive Guide (November 2023)
- Co-Organizer of Casablanca **MongoDB User Group**
- Internet of Things Virtualization Poster for the "NIST Cloud Computing Forum & Workshop IX" at the National Institute of Standards and Technology – USA
- PhD paper about the Internet of things(IoT) virtualization focuses on « A Micro-Clouds sensors virtualization and composition approach for the Internet of Things »
- « **Games Programming** » Workshop animator & speaker for the JNJD at INPT Morocco
- Keynote speaker of technical and scrum management fields.
- Creating and participating to opensource projects: Chimera2D, MoovinParis, Itefacio.com
- Book Reviewer «**MongoDB for Java Developers**» edited by packtpub: goo.gl/YxskzV