Ab El-Fake

Big Data Engineer/ Architect
Permanent Res.
☽ **Mobile: (+44) 000000000000**
✉ **Email: wxyz@gmail.com**

Professional Summary

- **5 years of Big Data experience** with hands on Data Engineer and Architect on tech stack using **AWS**, **GCP**, **Spark framework with Python & Scala, Cloudera , ETL/ELT Datawarehousing.**
- **Architect cloud environment following AWS architecture principles, design & build data pipelines, Data modelling and CI-CD.**
- **Technical solutions for business requirements, Estimates for work, Sprint planning.**
- **Experience in real time data ingestion using Spark & Scala and PySpark for ingesting semi-structured and structured data files.**
- Designed scalable architectures for real time data processing using Spark framework, Python and Scala programs and Kafka.
- **Total 15 years of IT experience with expertise in driving and leading end to end Big Data, Data Integration and BI solutions.**
- Proof of Concept (PoC), Request for Proposal (RFP) and Technical Presales for Hadoop Big Data & Oracle DWH & BI.
- **Migrate on premise Datawarehouse to AWS Redshift, GCP BigQuery and cloud agnostic SnowFlake.**
- Designed and implemented data lake integration and ingestion using Hadoop Ecosystem tools like pig, hive, sqoop, flume, Spark.
- Installing, configuring and Maintaining Single & multi node Cluster setup for Hadoop on premise & on Cloud.
- Experience of development in Agile environments, Test Driven Development (TDD), CICD continuous integration continuous development (Jenkins)
- Experienced with SQL databases such as Oracle, Postgress, MySQL, SQLServer and NoSQL databases such as Hbase, MongoDB and Cassandra.
- Data warehousing, Data modelling, Metadata, ETL (Informatica) and ELT (ODI), data cleansing, slicing and dicing of data.
- Experience as solutions architect for an implementation of the DWH & BI solution with public sector clients.
- Performance tuning of BI reports and ETLs. Involved in design of the integration of other reporting systems into Oracle BI platform.
- Consulting in IT industry, handling large projects, On-site Off-shore project coordination.
- Project Management, Release Management, Technical reviews**,** Test Cases, Quality Assurance (QA).
- Software Development Life Cycle (SDLC) using methodology driven environment like Agile, Kanban and Waterfall methodology.
- Experience in supporting the various production issues addressed by users and providing them with required solutions. Customer Support and system maintenance. Scalable
- Versatile communicator dealing with stakeholders and non-technical business users

IT Skills	
Big Data	Apache Spark,Hadoop, MapReduce, YARN, Hive, Pig, Sqoop, Oozie, Zookeeper, Flume, Kafka, Impala. Scala, Python, PySpark, Pandas, NumPy. Cloudera CDH, Hortonworks (HDP) GitHub, Jenkins, Terraform, Ansible, Datadog, ELK HDFS, XML, JSON, Avro, Parquet, sequence file, text file, CSV, Snappy, gzip.
AWS Cloud Technologies	AWS storage S3, EC2, Kinesis, Lambda, boto3, EMR, ECS, ELB, Athena, Glue, AWS RDS, **RedShift**, QuickSight, IAM, VPC Cloud, Direct Connect, CloudWatch, CloudFormation
Google Cloud Platform	**BigQuery**, DataProc, Pub/Sub, Cloud Storage, BigTable, CloudSQL.
Virtualization / Serverless	Docker, AWS Serverless, Vagrant, VirtualBox, VMWare
Programming Languages	Python, Scala, SQL, PL-SQL, Core JAVA, Unix, Pearl.
Version Control and CI-CD	GitHub, Jenkins, CircleCI, MS VSS, PVCS, Siebel Helpdesk, Remedy User.
Database	Snowflake, Redshift, BigQuery, Oracle 11g, Teradata, SQL Server, MySQL, Impala, Postgres, Dynamo DB, Presto, MogoDB, HBase, Cassandra.
ETL/ELT Tools	Matillion, Fivetran, Oracle Data Integrator (ODI), OBI Apps. Informatica ETL, Oracle Datawarehouse Administration Console (DAC)
Reporting	Oracle Business Intelligence Enterprise Edition (OBIEE 10g/11g) OBI Apps for HCM, Financials, Siebel Helpdesk and Siebel Sales. MS Excel Power Pivot 2012, MS SSAS, SSRS
Oracle Tools / Performance Tuning	Oracle Forms-Reports 10G/9i/6i, Oracle Enterprise Manager, SQL*Loader, TOAD Database Performance Tuning, Export/Import, Data Pump, Oracle SQL Developer, Oracle Data Modeler, Oracle Analytics Workspace Manager
Project Management	JIRA, Confluence, MS Visio, Agile methodologies, SCRUM.
Operating Systems	Linux (various versions), Unix, Ubuntu, Centos, Windows Server.

1. Company Name	Inawisdom (www.inawisdom.com)
Period of Employment	Nov-2019 to Apr-2020
Designation	Data Engineer working on AWS tech stack

Real time streaming data project:
Courier company in Dubai wants to see real time transaction entries and updates to parcels and their journey legs till delivery plus schedule time slot for consignee's. Build streaming data pipeline to ingest data from RabbitMQ to push it to AmazonMQ. Then ECS which runs kinesis producer agent to publish data to Kinesis data streams. Now Kinesis Firehose delivery systems consumes data and writes it to Redshift cluster. We used lambda functions in Firehose to transform data. Data Science Algorithms runs on data in Redshift to create courier's delivery slots and updates consignees.
Lambda functions in Kineisis Firehose to cleanse data in-flight then encrypt it with base64 and writes to Redshift in public subnet.

Incremental data pipeline using Glue which pulls data from on-prem SQL server and writes to S3, Use Athena to view raw data in S3.
Orchestrate data pipelines using Step Functions.

Build multiple data pipeline to source data from on-premise SQL Server, S3 using ELT tool Matillion and load into Redshift cluster.
Upgrade Matillion servers and migrate ETL packages to new environment.

2. Company Name	Inmarsat (www.inmarsat.com)
Period of Employment	June-2019 to Sep-2019
Designation	Cloud Data Engineer working on Google Cloud Platform (GCP)

Architect with hands on Data Engineer on Google Cloud Platform.
Design and build data pipeline using GCP's Dataproc , BigQuery. Data Modelling. Metadata creation using labelling.
Extract data from ServiceNow, SalesForce using FiveTran and build datalake.
Schedule deltas using Composer/AirFlow

3. Company Name	Holland and Barrett International (www.hollandandbarrett.com)
Period of Employment	Nov-2018 till Apr-2019
Designation	AWS Cloud Architect & Big Data Engineer.

Design and build Datalake using AWS services like Glue, Athena, EMR, RDS, RedShift, Spark, Python, S3, EC2, Jupyter Notebooks.
Build environments and data pipelines as per the business requirement and proposing the right tool sets to be used as per the requirements.
Extracted data from legacy Oracle DWH into S3 using AWS Glue and EMR. Create Glue ETL jobs to extract & transform the data.
Create source tables using crawler and stored its metadata into data catalog.
Ingest extracted Oracle data from S3 to AWS RDS Aurora.

4. Company Name	Sainsbury's Ltd. (www.sainsburys.co.uk)
Period of Employment	July-2017 till Oct-2018
Designation	Big Data Developer.

Sainsbury's is one of the biggest retail company in UK who wants to excel in big data space with cutting edge technologies.
As a big data developer worked on various projects for real time and batch data ingestion.

Real time Transaction, Customers, Product data ingestion into data lake
Environment: CDH 5.8, HDFS, Spark, Hive, Kafka, HBase, Scala, Python, AWS S3, Lambda, Agile Methodology, JIRA, GitHub, Terraform, Ansible, Jenkins, Docker.
This is to capture all transactions happening in shop across Sainsburys stores real time and ingest data into data lake. Data scientist team will run algorithm to find bestselling products region wise, suggests pricing, promotions.
Responsibilities: Present and persuade the design architecture to the various stakeholders.
Developed complete end to end data processing in Hadoop eco system.
Scala scripts, transformations of XML files to Data frames/SQL and RDD/MapReduce in Spark.
Stack environment is scripted in Terraform. Ansible & Jenkins for automated deployments.

Batch data ingestion to datalake
Environment: CDH 5.8, HDFS, Hive, HBase, Scala, Python, AWS S3, Teradata, Ab-initio, Agile Methodology, JIRA, GitHub, Jenkins.
Data scientist needs more than 3 years data into datalake for analysis as oppose to only 6 plus months of live real time data.
Responsibilities:
Proposed solution and architect tech stack to pull data from existing Teradata data warehouse into data lake.
Writing SQL, shell scripts.

Migrate on-Premise Datawarehouse to SnowFlake DB.
Environment: AWS S3, VPC, Direct Connect, SnowFlake database, Ab Initio, Teradata, Teradata parallel transporter (TPT).
Responsibilities: Sainsburys was the first biggest client of SnowFlake in UK to migrate their legacy on prem Teradata datawarehouse to Snowflake.
As part of this huge migration program we went through professional training from SnowFlake then worked on PoC to prove SnowFlake DB is faster than Teradata. Build ETL pipeline to extract data from on prem enterprise datawarehouse and dump it to S3. Use SnowFlake's snowpipe, COPY, Bulk loading tools.

Smart Product Matching
Environment: AWS, Lambdas, Docker, Postgres, RedShift, Agile Methodology, JIRA, GitHub, Jenkins.
Sainsburys wants to compare their products and prices with other UK retailers. We've ingested data from third party provider and processed it with algorithm written in R and find the barcode matches, ingredients match, near matches, etc.

Responsibilities: Architect, build AWS environment and then build data pipeline.

4. Company Name	Honda Motors Europe Ltd. (www.honda.co.uk)
Period of Employment	July-2015 to May-2017
Designation	Freelance MI Senior Consultant (Hadoop Big Data, ODI, OBIEE)

As a **Big data developer** worked on below projects.

Engine testing data analysis

Environment: CDH 5.0, HDFS, Hive, Scala, Kafka, OBIEE, Agile Methodology.

Honda motors process engines test data via IoT sensors. This data is feed to Kafka cluster and then run transformation logic written in Scala.

Responsibilities: Data modeling and designing in Oracle.

SQL queries to ingest engine test data into data lake.

OBIEE reports to design reports and dashboards.

Data migration and Integration

Environment: CDH 5.0, HDFS, SQOOP, Hive, Oracle data Integrator, OBIEE, Agile Methodology.

This project is to move data from legacy data warehouse to one platform which is Cloudera Hadoop distribution.

Responsibilities: Export data from warehouse into flat csv files and load them using ODI and SQOOP, SQL scripts and build BI reports.

As an **ODI/OBIEE consultant** I was involved in design and development of the integration solution involving the Hyperion products and MS SQL server financial data warehouse. Also involved in integration of OBIEE and Hyperion planning as a source.

- Importing and exporting data into HDFS and Hive using Sqoop.
- Collecting and aggregation of log data using Apache Flume and staging data in HDFS for further analysis.
- Developed multiple MapReduce jobs in Java for data cleaning and preprocessing.
- Creating Hive Internal/ External tables and working on them for data analysis.
- Responsible for Data Models, detail technical design document and architecture solution
- Integration of OBIEE 11g security to AD. Development of reports involving Spatial data linking to Essbase
- Oracle Database 11g, SQL Server 2008, ODI 11g, FDM, EPMA, Hyperion 11.1.2.1

5. Company Name	Price Waterhouse Coopers UK (www.pwc.co.uk)
Period of Employment	April 2014 to Aug 2015
Designation	Manager Oracle Business Intelligence & Big Data

Work as technical manager for Oracle BI & Big Data team. Roadmap to build oracle BI practice at PwC. Build brand new Datawarehouse servers for client demos. Creating custom Datawarehouse and BI reports as per client requirement and demo it to them. Work on proposals and proof of concepts. Big Data Architecture & Consultancy from greenfield, introduced and deployed all technologies. In-depth vendor PoC's, deep technical product analysis, deployments and report writing. Vendor pricing negotiations leveraging vendor-insider knowledge in Big Data, successfully negotiated up to 80% discounts. Built and upgraded many clusters across all major Hadoop distributions, cluster migrations of dozens of TBs of data, loading and processing of multi-terabyte datasets Hive table, multi-structured parallel indexing of 20 billion documents (50TB) to real-time search systems, performance and scalability tuning. Advanced Nagios Plugin Development including Hadoop, MongoDB, HBase, Cassandra, Hortonworks Ambari, and Spark. Advanced Hadoop cluster builds, problem reproductions and debugging to code level in Java. Data Science and distributed Machine Learning with Mahout, Hive with Python UDFs for transformations and data cleansing, Python, R studio & programming.

Major Projects

Client Name : London Stock Exchange Group (www.lseg.com)

Project: Oracle BI Financials & Projects using BIApps 11.1.1.7.1

London stock exchange group is world's renowned stock exchange located in central London.

I've build BI server, load eBS data using Oracle data integrator and then build custom reports on top of out of the box BI reports. Implemented BI Financials & Projects module.

Client Name: City of London (www.cityoflondon.gov.uk)

Project Name: Oracle Business Intelligence, OBIApps 11.1.1.7.1

City of London is municipal council in central London. PwC ERP team upgraded their eBS to 11 and then me & my team implemented Oracle BI from scratch. We've implemented BI Financials, HR, Projects, Procurement & Spend analytics.

City of London implemented Oracle Property Manager(OPN) on eBS and wants to have BI reports. As there is no property manager BI offerings so we've build custom warehouse for property manager. Right from taking business requirements then I convert it to technical offerings then build ETL mappings using oracle data integrator and custom BI reports using OBIEE.

6. Company Name	Mokum Change Management Ltd.(www,mokum.com) Now PWC (www.pwc.co.uk)
Period of Employment	November2012 to March 2014
Designation	Manager

Major Projects

Mokum Change Management Ltd. now merged with Price Waterhouse Coopers (PWC) in March-2014.

Mokum is Oracle gold partner having very good reputation in UK and Europe as specialized Oracle ERP consulting company. At Mokum, I've worked for client John Lewis who is UK's top retail company.

Client Name: John Lewis UK (www.johnlewis.com)   November 2013 – April-2012

Project: Oracle Learning Management (OLM) custom OBI Apps Reporting on eBS R12

John Lewis is using Oracle Learning Management (OLM) application for training and certification of their employees to upscale skills.

Client wants to use OBIEE reporting to analyze their employee trainings and certifications.

My work involves gathering business requirement and converting them into technical specifications.
Then worked on designing data warehouse and created full ETL using Informatica.
OBIEE RPD modelling and creating front end analysis and dashboards.

Client Name: John Lewis UK (www.johnlewis.com)   **June 2013 – October 2013**
Project: HCM OBI Apps Reporting on eBS R12
John Lewis implemented Human Capital Management OBI Reporting on eBS HCM application.
This fully customized ETL and reporting solution.
Worked on end to end implementation where I worked on Informatica ETL to bring data from source eBS,
RPD modelling and building customize analysis and dashboards.
Client Name: John Lewis UK (www.johnlewis.com)   **December 2012 – May-2013**

Project: Spend Classification implementation using OBIApps (www.waitrose.com)
Spend Classification is OBI reporting product from Oracle which gives insight about organizations
Spending patterns. JLP implement this to classify their taxonomy and employee expenses.
Complete installation of Spend Analytics.
Requirement gathering, Documentation, Technical Architect for spend analytics.
Data from Purchase Orders, Requisitions and Invoices of eBS ERP form the sources for Spend
Analytics/Spend Classification. Implement and modify ETL to bring data from eBS to Data warhouse.
Created OBI repository to arrange data in star schema using dimensions and facts
Build Reports and Dashboards. Performance tuning of SQL's in reports.

7. Company Name	TomTomLtd.www.tomtom.com
Period of Employment	February 2011 to September 2012
Designation	Senior Application Engineer- Oracle Technologies

Major Projects

TomTom is the world's leading supplier of location and navigation products and services focused on providing all drivers with the world's best navigation experience. Headquartered in Amsterdam, TomTom has 3,500 employees and sells its products in over 40 countries.

I.   **Reverse Logistics application**
     Tools: Oracle DB 11G, Forms/Reports 9i, Java.
     Description: This application is used to track shipping and returns of tomtom GPS device.
     DBA activities like Design, Create/ Install Database, Performance Tuning, Export/Import of data.
     Developing Forms, Reports in existing system as per business requirements.
     Created Web Services, Java stored procedures.
     Writing SQL queries, PL/SQL packages, performance tuning of database.
     Maintaining existing projects and support.
     Build and release Management, Project Management,
     Test cases, User Acceptance Test (UAT) and Business Acceptance Test (BAT).
II. **Support Oracle eBSR12 (ERP) application:**
     This is TomTom ERP application used for AR, AP and GL.
     Modifying Oracle Forms/Reports as per business requirements.
     Writing Database Packages, Triggers, SQL query, Complex Views.
     Created complex Invoicing reports in OBI (XML) publisher.
III. **OBIEE 10g reporting and ETL on OBI Apps for Oracle eBS, Siebel Sales(CRM):**
     Creating and Modifying Answers/Reports, Dashboards, iBOT's/Alerts. Usage Tracking, Security using LDAP.
     Writing complex SQL queries and Pl-SQL packages for data transformation and loading.
     Creating Dimensions, Measures, Hierarchies. Views and Materialized View.
     Modify and create new mappings for ETL in Informatica which are pre-build and shipped by Oracle corp.
     So done customization in ETL as and when required by business owners.
IV. **OBIEE 10g and ETL in OWB 10g for Nagios Monitoring Tool:**
     TomTom uses Nagios tool for health check of TomTom servers and Network e.g. checking server & network availability, availability and downtime of all enterprise application on hosted servers. This information is stored in MySQL database.

8. Company Name	Motivity Solutions Ltd. Bracknell RG12 1JG, UKwww.motivity.co.uk
Period of Employment	May2005 to Oct-2010
Designation	Senior Oracle Consultant

I.   **Various Projects , Development and Support**

     OBIEE Reporting: **Aug 2009– Oct 2010**
     Organization: Motivity Solution Ltd. Bracknell, UK.
     Description: This is Reporting and Analysis project on OLAP and OLTP oracle database.
     Tools: Oracle Business Intelligence Enterprise Edition (OBIEE), Oracle 10G,
     Discoverer, Oracle Reports 6i/9i, PL-SQL, SQL Loader, Data Pump, Oracle Data Integrator.
     Responsibilities:
     Created complex reports using Oracle BI Publisher, BI Answers, BI Dashboard,
     Excel Analyzer, HTTP feeds, Web Services.
     Experience in BI security model, Administration, Configuration, Data Sources, Delivery options,
     Repository, Runtime Parameters. Creating Repository, Applying Row Level Security, Aggregate Tables.
     Metadata for OLAP using Analytic Workspace Manager.

     Data warehousing: **Dec2008 – Oct2010**
     Tools: Oracle Warehouse Builder 11G
     Dimensional Modeling, Data Modeling, Maintaining and Building data warehouse which involves extraction, transformation and loading of data.
     Building Views, Materialized views, Indexes, Partitioning, Analyze database. SQL/PL-SQL.
     Used oracle export/import, data pump, SQL loader to load/unload data.

     Development, Maintenance and Support: **Jan 2008 – Oct 2010**
     Tools: Oracle 9i/10G, Forms/Reports 6i/9i/11G, Oracle Designer, Oracle JDeveloper, Java,

Oracle Apex, XML.
DBA activities like Design, Create/ Install Database, Performance Tuning, Export/Import of data.
Developing Forms, Reports in existing system as per client requirements.

II. **Vehicle Leasing and Finance Product:**   **May 2005 – June 2008**
Software: Oracle Database 9i/10G, Oracle Application Server, Forms/Reports 6i/9i, XML, Java.

For Client-Server Application developed we used Oracle Database 9i/10G and Front End as
Forms/Reports 6i and Oracle XML Publisher.
For Web based Application we used Oracle Database 10G with RAC architecture on Windows Server.
Used Oracle Forms/Reports 9i front end.Oracle BI Publisher for reporting.
Responsibilities:
Design and Developed complicated Forms including OLE and Active X Controls and Reports.
Design Form/Report Templates, created .pll libraries.
Writes Java Packages for file operations and then uploaded in Oracle.

III. **Reporting Tool:**   **July 2006 – Jan 2007**
Software: Oracle Database 9i/10G, Oracle Application Server, Forms 6i, Java.
This reporting tool is product which has same facility like more of established market reporting Tool. Developed with Oracle Forms, database and java. User can print reports in ASCII, MS Word, Excel, HTML, PDF format. Here user can print report on any tables from database.
Responsibilities:
Involved in Design, User Interface, Development, Testing of application, Client Support.
Creates forms for front end, Database packages, java stored procedures.
Created complex views, Materialized view, Triggers.

IV. **Migrating Forms6i to Forms 9i**   **Jan 2007 – March 2007**
Software: Oracle Database 9i/10G, Oracle Application Server, Forms/Reports 9i, Java.
Responsibilities:
Solely transformed Forms & Reports 6i to 9i.
Replaced OLE & Active X Controls in Forms 6i with Java beans in Forms 9i.
Replaced File IO with Webutil.pll library.

9. Company Name	KPIT Cummins Ltd. INDIA (SEI CMM Level 5 Co.)www.kpitcummins.com
Period of Employment	Sep 2004 to May-2005
Designation	Senior Oracle Consultant

I. **Data Warehouse and Reporting:**   **Oct2004 – May 2005**
Client:CAPONE UK – Credit Card, is global player in credit cards.
Description:Thisapplication takes care for all the credit cards related issues like new card application, Statements checking, Balance Transfer, Retention of customer, Marketing strategy, etc. Build Data Warehouse for reporting.
Software: Oracle, sql, pl-sql, Discoverer, Informatica, Siebel.
Responsibilities:Worked on ETL tool Informatica to extract, transform and load data.
Created many transformations like Filters, Expression, Routers, Lookup, Stored Procedures, etc using Informatica Design Centre.
Created Mapping, Mapplets, Workflow.Developed PL/SQL Packages, Stored Procedures, and Stored Functions.

10. Company Name	ZENSAR Technologies Ltd. Pune India(*SEI CMM Level 5 Co.*)www.zensar.com
Period of Employment	Jan 2004 to Sep 2004
Designation	Oracle Consultant

I. **Shipping and Logistics**   **Jan 2004 – Sep 2004**
Client: P & O Ned Lloyd – Worlds 2nd Largest Cargo Shipping Company.
Software:Oracle*8i (PL-SQL), Oracle Developer*-FORMS*6i*/REPORTS*6i, WIN2000 (O/S)*.
Responsibilities: Developed/Debug PL/SQL Packages, Stored Procedures, and Stored Functions.
Developed/Debug forms using Oracle Forms 6i & Customs Reports in Reports 6i.

11. Company Name	FWL Technologies Ltd. Pune India
Period of Employment	Jun 2003 to Dec 2003
Designation	Software Engineer

I. **LOGISTICS**   **Jun 2003 – Dec 2003**
Client: CMA-CGM, World's Largest Shipping Lines.
Description: Successfully worked for Modules for Shipping Lines and Logistics Project for Clients CMA-CGM. A customer wants to Export/Import goods then he approaches the shipping Company, He makes the booking for that voyage through Agent (of company)

II. **FINANCIAL**   **Feb 2003 – May 2003**
Description: Worked on Modules for Financial Projects - Sales Ledger at company
The project is divided into three Modules as Sales Ledger, Sales Marketing and Bonded Warehouse. Sales Ledger is about keeping the information of sales/purchase to clients/vendors
Software: Oracle*8i (PL-SQL), Oracle Developer*-FORMS*6i*/REPORTS*6i, WIN2000 (O/S)*.

Education

- Bachelor of Mechanical Engineering in from Amravati, India – 2002 (equivalent to UK BSc Degree Level 2.1)
- PG Diploma in C-DAC (Centre for Development of Advanced Computing), Pune, India – 2003, Distinction.

References Available Upon Request.