**OLEKSII SAIUN**

**Email:** [oleksiisaiun@gmail.com](mailto:oleksiisaiun@gmail.com)

**Cell:** +1-832-231-8005

**LinkedIn:** <https://www.linkedin.com/in/oleksii-saiun-37a20b1b0/>

**SENIOR DATA ENGINEER – Summary**

Data Engineer with 10+ years of experience in a design of complex DWH, ETL, Data Lake solutions. Hands on experience with Big Data technologies – Spark, Hive, Sqoop, Impala, Hue, Atomic.

Creating ETL/ ELT dataflows using cloud technologies Azure Data Factory, GCP Data Flow and on-premise ETL tools IBM DataStage, Informatica, MS SSIS, Pentaho, SAP DS.

Extensive experience of working with DBMS Oracle, Teradata, Sybase IQ, MS SQL Serve.

Business domain knowledge covers Oil & Gas, Travel, Retail. Medical and Banking areas.

Master's Degree in Mathematics Science. Certified Oracle Database SQL Expert, Oracle Advanced PL/SQL Developer Certified Professional (OCP) and IBM Certified DataStage, IBM Quality Stage.

**technical skills:**

|  |  |
| --- | --- |
|  |  |
| **Databases:**  **ETL tools:**  **Big Data:**  **NoSQL**  **Reporting:**  **Languages:**  **Frameworks:**  **Others:** | Oracle, Teradata, SQL Server, Sybase IQ, MySQL, PostgreSQL  IBM DataStage, Informatica, MS SSIS, Pentaho, SAP DS  Hive, Sqoop, Impala, HDFS, Kafka, Spark, Oozie, Airflow, Azure DataBricks, Sqoop  Apache Cassandra  Cognos, MicroStrategy, Power BI  Java, PySpark, Python, SQL, PL/SQL, T-SQL, Bash  Spring Boot, Spring Data, Spring MVC, Hibernate  Git, Perforce, SVN, Stash, Source Tree, Jira, Docker, Minikube, Rest API, Azure DevOps |

**WORK HISTORY**

**EPAM Systems**

Dec 2020 – Present

|  |  |
| --- | --- |
| **Client:**  Walmart Labs  **Location:** Houston, TX |  |

**Project Role:** Developer

**Job Position**: Senior Data Engineer

**Project Description:** Data Lake Team

**Responsibilities**:

* Create ETL pipelines to load structural and semi-structural data into Data Lake
* Creation, partitioning and population tables in Hive
* Migration of existing on-premise ETL dataflow into MS Azure Data Factory
* coding APIs in PySpark for Azure DataBricks to transform and process files of Avro and Parquet formats

**Databases:**

* Oracle 12c

**Big Data:** Hadoop, Sqoop, Spark-SQL, Hive, Automic, Hue

Nov 2019 – Nov 2020

|  |  |
| --- | --- |
| **Client:**  Penske  **Location:** Houston, TX |  |

**Project Role:** Developer

**Job Position**: Senior Data Engineer

**Project Description:** Design DWH to analyze data about leasing trucks

**Responsibilities**:

* Batch data load of Parquet files into Cassandra
* Load data into Data Lake using Sqoop, Hive, Kafka
* Data load of events using Kafka into Cassandra
* Scheduling ETL jobs in Oozie
* Create ETL data flow to load dimension and fact tables into DWH using SAP Data Services

**Databases:**

* MS SQL Server 2014, Oracle 11g ,12c

**Big Data:** Cloudera Apache Hadoop, Hive, Ambari, Sqoop, Kafka, Spark 1.6

**ETL tool:** SAP DS

Jun 2016- Present

|  |  |
| --- | --- |
| **Client:**  **Chevron** – MCBU (Mid Continent Business Unit)  **Location:** Houston, TX |  |

**Project Role:** Tech Lead/ Solution Architect

**Job Position**: Lead Software Engineer

**Project Description:** Data integration from system of records to DWH, performing change data capture, reporting in oil and gas industry

**Responsibilities**:

* Design Data Lake to ingest and process structured, unstructured, semi structured data
* Batch processing data in Hive: creating tables, bucketing, partitioning data
* Design Sqoop workflows to load and unload data from the Data Lake
* Preparing and running queries in Impala
* Scheduling jobs in Apache Oozie
* Working with different formats files in Hive: Sequential file, Avro, Parquet
* Design of ETL workflows using IBM DataStage, MS SSIS to load data to DWH
* Migration ETL workflows from IBM DataStage to MS Cloud Azure Data Factory

**Databases:**

* MS SQL Server 2014, Oracle 11g ,12c

**Big Data:** Cloudera Apache Hadoop, Hive, Impala

**Cloud**: MS Azure Data Factory

**Tools, technologies:**

* ETL tool – Sqoop, IBM DataStage 11.5, MS SSIS
* Reporting: Power BI
* Sqoop, Pig, Hue, HDFS, Oozie

Jan 2015- May 2016

**Client:**  **Expedia**

**Location:** Kiev, Ukraine

**Project Role:** Key Developer

**Job Position**: Senior Software Engineer

**Project Description:**

* Design and population Data Lake
* Migration existing DWH platform from DBMS DB2 to Teradata;

**Responsibilities**:

* Design Teradata TPT jobs (Teradata Parallel Process jobs) to move existing data from DB2 into Teradata
* Implementing batch data Big Data pipeline with batch processing
* Load and load data using Sqoop, Hive
* Hadoop development and implementation. Load unstructured and semi-structured data sets using Sqoop. Hive queries

**Databases:**

* DB2, Oracle 11 g, MS SQL Server,DBMS Teradata 15

**Tools, technologies:**

* Teradata utilities: TPump, FastLoad, MultiLoad,FastExport, BTEQ, TPT;
* Apache Hadoop (Hive/Sqoop/Hue)
* Informatica 8.5/9.6
* Stash, Perforce

Dec 2011- Dec 2014

**Client:**  **Chevron** – GOM (Gulf of Mexico Business Unit)

**Location:** Kiev, Ukraine

**Project Role:** Team Leader

**Job Position**: Senior Software Engineer

**Project Description:**

GOM MDM is the ETL project that relates to oil and gas industry. The purpose is pull data from different sources (DBMS Oracle, flat files), integrate and loading data into DWH

**Responsibilities:**

* ETL design using IBM DataStage;
* Development stored procedures, functions, triggers packages using Oracle PL/Sql
* Partitioning tables, indexes, materialized views;
* Tuning SQL and PL/SQL queries (read plan, profiling, tracing);
* Lead a team

**Databases:**

* Oracle, MS SQL Server

**Tools, technologies:**

IBM DataStage and QualityStage, IBM MDM, TFS, Jira

Aug 2008- Dec 2011

**Client:**  **Raiffeisen Bank Aval**

**Location:** Kiev, Ukraine

**Project Role:** Developer

**Job Position**: Software Engineer

**Project Description:**

RBA Client Data Integration Project:

Design and support ETL dataflow to load Data Marts from customer’s clients

Applications. Design Cubes and BI reports

**Duties:**

* Coding stored procedures and triggers in Oracle, Sybase ASE, Sybase ASA, Sybase IQ;
* ETL development using IBM DataStage
* Design OLAP data model;

**EDUCATION**

*National Technical University of Ukraine "Kiev Polytechnic Institute"*

Masters: Mathematics and Computer Science(2009)

**github**

<https://github.com/oleksiisaiun/JAVA/tree/master/ALGORITHM/DATA-STRUCTURES/SOLUTIONS/LEETCODE>

**CERTIFICATES**

* IBM Certified Solution Developer - InfoSphere DataStage 8.7
* IBM Certified Solution Developer - Information Analyzer v8.7
* IBM Certified Solution Developer - Information QualityStage v8.7
* [Oracle Certificate: 1Z0-047 Oracle Database SQL Expert](http://education.oracle.com/pls/web_prod-plq-dad/db_pages.getpage?page_id=5001&get_params=p_exam_id:1Z0-047)
* [Oracle Certificate: 1Z0-144 Oracle Database 11g: Program with PL/SQL](http://education.oracle.com/pls/web_prod-plq-dad/db_pages.getpage?page_id=5001&get_params=p_exam_id:1Z0-144)