

# Khasan Al-Ammori

## Technical Team Lead

Khasan is a Technical Team Lead in Capgemini. He has 12 years of combined experience serving e-commerce, hospitality, social media and health industries. His industry tenure has afforded him involvement in three big data projects, all of which included leading international teams of engineers. He also has entrepreneurial experience from several startup involvement either from the technical side or being a founder of one. His technical toolset includes Java, Python, Spark, Hive, Impala, Hadoop, HBase, Airflow, Lucene, Elastic Search, PostgreSQL, GreenPlum and Oracle.

Khasan's industry experience includes data warehousing and analytics solutions implementation and integration. He also has experience developing and contributing to existing big data technologies. Furthermore, Khasan has managerial skills and is an experienced team lead.

## Tech skills

### Main skills

- Data engineering
- Backend development

### Programming languages

- Proficient in: Java
- Familiar with: Python, Scala, Ruby, Kotlin, R

### Technologies

- BigData: Spark, Hive, Trino (Presto), Impala, Airflow, Superset
- Databases: PostgreSQL, GreenPlum, Oracle
- Frameworks: Spring, Hibernate, Lucene, Elastic, Selenium
- Servers: Linux, FreeBSD
- Limited knowledge: NiFi, HBase, Cascading, DL4J, WEKA, ClickHouse, Cassandra

### Languages

- Ukrainian: native
- Russian: native

- English: advanced
- Arabic: elementary
- German: elementary

## Professional Experience

### Capgemini - Technical Team Lead

October 2021 - Present [Intracompany transfer]

- Main tools: Spark, Hive, Trino, Cascading, Airflow, Python, Scala, Java
- Leading team of engineers: interviewing; conducting onboarding sessions and support; gathering and grooming requirements from stakeholders on the client side; task formation and delegation; technical support for the team and the client; conducting ongoing technical trainings; process formation and implementation;
- Implementing data-processing pipelines: gathering and documenting domain knowledge; business goals and non-goals determination; implementing technical solutions using technologies like Spark, Hive, Trino, Airflow, FlowHub, QueryBook, etc; integration with existing data pipelines; data validation; ongoing performance tuning;
- Data job optimization: identifying long running and/or expensive jobs; identifying pipelines missing SLA; analyzing existing implementations in search for bottlenecks; assessing possibility of improvement and identifying the highest-yield approach; rewriting existing jobs to implement select solutions; continuous data validation and monitoring;
- Incidents and other technical issue resolution: conducting post-mortem analysis; identifying flaws of failed solutions; timely workaround implementations; designing and implementing long-term solutions;
- Data warehousing cost optimization: assessing data growth trends; identifying stale data; working with data owners to clean up unused data; implementing data retention policies;

### Altran - Team Lead

February 2020 - September 2021 [Intracompany transfer]

- Main tools: Spark, Hive, Trino, Airflow, Livy, Kerberos, Python, Scala, Java
- Leading team of engineers: interviewing; conducting onboarding sessions and support; gathering and grooming requirements from stakeholders on the client side; task formation and delegation; technical support for the team and the client; conducting ongoing technical trainings; process formation and implementation;

- Implementing data-processing pipelines: gathering and documenting domain knowledge; business goals and non-goals determination; implementing technical solutions using technologies like Spark, Hive, Trino, Airflow, FlowHub, QueryBook, etc; integration with existing data pipelines; data validation; ongoing performance tuning;
- Spark SQL migration from Hive: rewriting queries to support new standart; performance tuning; data validation;
- Spark SQL migration automation: designing and implementing an automatic solution to scale migration process to tens of thousands data jobs;
- Data warehousing cost optimization: feasibility study on compression algorithms, file formats and data ordering storage size implications;
- L&D: writing quick start guides for new data engineers; writing performance optimization guides; improving existing documentation; conducting data edu trainings;
- Data matchup framework: designing and implementing a solution to validate data job changes, estimate and locate mismatches between datasets;
- Securing Hadoop cluster with Kerberos: patching Livy client to allow for security hardening connections with Kerberos; configuring kerberized Livy client to work with Envoy proxy; integrating with other service in the cluster; installation, configuration and distribution automation;

## Lohika - Team Lead Engineer

September 2018 - February 2020

- Tools: Hive, Presto, Airflow, Superset, Java, Python, Scala, Spark
- Leading team of engineers: interviewing; conducting onboarding sessions and support; gathering and grooming requirements from stakeholders on the client side; task formation and delegation; technical support for the team and the client; conducting ongoing technical trainings; process formation and implementation;
- Creating a framework for data pipeline scheduling and execution: designing and implementing an ad-hoc data pipeline execution framework, that is designed for tight-SLA workflows;
- Working on ETL architecture development: redesigning some of the existing ETL pipelines in order to improve performance and stability of tier-1 data sources;
- Cross pipeline error detection: gathering business expectations on dependent data pipelines; performing data validation; identifying sources of issues; working cross-team to design and implement solutions that would eliminate existing data discrepancies;
- Error detection automation: designing and implementing a solution to validate data between versions of the same pipeline; alert and monitoring solutions creation;
- Implementing data-processing pipelines: gathering and documenting domain knowledge; business goals and non-goals determination; implementing technical solutions using technologies like Spark, Hive, Presto, Airflow, Python, Scala; integration with existing data pipelines; data validation; ongoing performance tuning;

- Spark SQL migration from Hive: rewriting queries to support new standard; performance tuning; data validation;

## Global Logic - Big Data Developer

September 2017 - September 2018

- Tools: Hive, Impala, MapR, Cloudera, HBase, PostgreSQL, GreenPlum, Redis, Java 8, Ruby, Python, Ansible
- Development of encryption utility library used in distributed environment: designing and implementing a solution to encrypt and share configuration between services in a secure manner;
- Data migration from Oracle and PostgreSQL to Hadoop: rewriting old functionalities to support Hive and Impala instead of RDBMS; data validation; performance tuning;
- Security hardening for Redis: patching Redis connector library to use Kerberos as a part of GDPR compliance;
- Software installation and deployment automation: designing and implementing a tool to migrate existing installation scripts from in-house proprietary solution to Ansible;
- Feasibility study on data warehouse solutions: assessing different datawarehouse tools to determine optimal solution for storing and analyzing data; analysis included comparison of file formats, compression techniques, memory footprint, processing cost implications and more;

## TeaLao - Developer and Co-Founder

January 2017 - February 2020

- Tools: Java 8, Java 11, Spring, Spring MVC, Hibernate, PostgreSQL, Lucene, Elasticsearch, Logstash
- Created ecommerce project TeaLao from zero to fully developed functional product, with hundreds of users reading, learning about tea and purchasing it every day;
- Application design: creating technical specification and architecture for data storage, DB model, media storage, payment processing, user interface integration;
- Technical solutions design and implementation, includes, but not limited to: media storage, data analytics, currency conversion service, internationalization layer, full text search service, SEO module;
- Managing a technical team, consisting of two engineers and one designer;
- Managing a creative team, consisting of one marketing specialist and four content creators;
- Logistics and supply chain management: working with overseas partners to select, package and ship produce;

## Lubkin Consulting LLC (Medex) - Team Lead, later CTO

August 2014 - January 2017

- Tools: Java 7, Java 8, Spring, Hibernate, Android, iOS, PostgreSQL, Lucene, JS, HL7
- Designing application architecture, includes, but not limited to: client portal website, doctor portal website, client mobile application, notification center, encrypted storage module, CRM;
- Leading a team of three engineers and one designer to implement technical solution for a client;
- Working with a large variety of specialists, including marketing team, help desc, doctors and others, in order to gather domain-specific information, provide ongoing support, gather feedback, create product pivot strategy;
- Technical presentation for investors: market analysis, strategy, financial planning, operation costs breakdown, competition analysis, business plan, investment offers;

## Chasopys - Developer

June 2012 - January 2013

- Tools: Java 7, Swing, Spring, PostgreSQL, JavaFX, RFID, NFC
- Development of a time-based ticketing system. The system includes: permit system, ticket and pass issuing module, CRM tool, financial reporting, marketing tools, RFID and NFC modules;

## Kolo Global - Automation Engineer

September 2010 - August 2012

- Tools: Java 6/7, Swing, AWT, Hibernate, Spring, HtmlUnit, Ruby
- Creating a large tool set for marketing department, automating their tasks

## Education and Professional Affiliations

Kyiv National Transport University - PhD Student 2017 - ongoing

Kyiv National University of Technologies and Design - Master's in Computer Science 2013

Kyiv National University of Technologies and Design - Bachelor's in Computer Science 2012