

DÁVID SZAKÁLLAS

Software & Data Engineer | K8s, Big Data

 szakallas.eu

 david.szakallas@gmail.com

 [/in/dszakallas](https://in.linkedin.com/in/dszakallas)

 github.com/dszakallas

SUMMARY

I am a software engineer experienced in developing/operating data driven applications running on public clouds. I currently work at Pure Storage. In the past I led migrating a science platform to AKS and led establishing containerization and DevOps practices in a data team at Turbine.ai. Before that, I worked on data pipelines running in k8s at Hiya. Previously, I developed the main social database ingestion pipeline, and the credit fraud detection ML training pipeline at Ekata. I hold an MSc in Computer Science.

SKILLS

prog Go, Python, Scala, TypeScript, Rust, C++
datalake Spark, EMR, Databricks, Airflow, PostgreSQL, S3
k8s k8s, AKS, EKS
devops GitOps, Terraform, Docker, Ansible
os Linux, macOS, emacs
locale en, hu, de, cs

EXPERIENCE

Pure Storage

11/2022 – now

Software Engineer (MTS4)

Prague

- Developing Portworx Data Services' control plane written in Go. K8s operations, security & hardening.
- Developing client SDKs for the Fusion product.

Go /K8s /Python /AKS

- Developing the control plane of an undisclosed product written in Go and Python.
- Organizing engineering practices meetings.

Turbine.ai

05/2021 – 06/2022

Software Engineer (Staff)

medtech | startup | Budapest

- Developing the scientific data processing pipeline using Azure Databricks & ADLSv2. Performance tuning, optimization, containerization.
- Training data scientists and junior developers on distributed data processing. Enhancing software development practices and DevOps. Introducing unit testing.

Azure Databricks /AKS /K8s /Azure Functions /Python /Apache Airflow /Earthly /Flux

- Prototyping containerized development environments in Databricks for biologists.
- Introducing the cloud native stack, Kubernetes and GitOps in the company.
- Migrating the science platform orchestrator from Azure VMs to AKS.

Hiya

06/2020 – 05/2021

Data Engineer (Senior)

fraud detection | startup | Budapest

- Developing our internal data platform based on AWS Glue, Apache Spark, Apache Airflow and Kubernetes.
- Integrating Prometheus metrics collection for

Scala /Python /K8s /AWS S3 /AWS Glue /Apache Airflow /Apache Spark /Prometheus

- Apache Spark in K8S.
- Developing and maintaining internal data platform libraries in Scala.

Ekata

02/2017 – 04/2021

Data Engineer (Medior)

fraud detection | startup | Budapest

- Developing the ingestion pipeline powering the social network search engine. Cutting down ingestion costs by 80%.

Scala /Python /Clojure /AWS S3 /Databricks /Apache Airflow /Apache Spark /XGBoost

- Developing our fraud detection ML training pipeline. Developing an ML feature engineering application for data scientists.

EDUCATION

BME

2015–2018

MSc Computer Science

Thesis: Evaluation of openCypher graph queries with a local search-based algorithm in Clojure
Major: Critical systems

BME

2011–2015

BSc Computer Science

PROJECTS

hobby

current

mixxx-launch

github.com/dszakallas/mixxx-launchpad

HID controller mappings for an open-source DJ software.

hobby

current

szakallas.eu

szakallas.eu

My personal programming blog updated occasionally.

LDBC

past

SNB Datagen

github.com/ldbc/ldbc_snb_datagen_spark

Benchmark data generator for graph databases written in Spark running in EMR.