Vangelis Kourlitis PhD

Data Engineering | ETL | ex-CERN

PROFILE

Data engineer and software developer with 10 years experience in technical projects and leadership positions at CERN. Specialized in developing scalable pipelines in big data analytics and machine learning domains, with a proven track record of optimizing workflows and delivering impactful solutions. Now pursuing to master Chubb's cloud-native data tool-chain and accelerate insurer analytics.

CONTACT

Location: Thessaloniki, Greece
Email: kourlitisV@hotmail.com
LinkedIn: vangelis-kourlitis
GitHub: ekourlit

SKILLS

Programming
Python • SQL • C++ • Bash

Data Pipelines
Apache Kafka

Cloud & DevOps
GCP • AWS • Docker • Git (CI/CD)

Distributed Compute
Dask • Apache Spark • Ray

Workflow Schedulers
HTCondor • Slurm

EDUCATION

UNIVERSITY OF SHEFFIELD
PhD in Physics
2015 - 2019 | Sheffield, UK

ARISTOTLE UNIVERSITY OF THESSALONIKI
B.Sc. in Physics
2009 - 2015 | Thessaloniki, GR

EXPERIENCE

TECHNICAL UNIVERSITY OF MUNICH & CERN | Geneva, CH
Analysis Model Group Coordinator
Oct 2023 - Dec 2024

- Directed a global, 120-member cross-functional team providing 24/7 production support for a $\sim$1 M LOC data analysis stack and formats; used Agile methodologies to deliver robust new features aligned with organizational goals.
- Achieved 3x reduction in data storage costs by adoption of a lightweight columnar data format ahead of schedule.
- Rolled out streamlined software configuration paradigm, reducing onboarding time by over 95%.
- Supervised early-career developers projects, delivering innovative data engineering products on time and fostering professional growth.

Data Science Researcher
Mar 2023 - Sept 2023

- Secured €40k funding to modernize legacy ETL tools for big data stream handling ($\mathcal{O}$(100TB)), achieved 4x increased throughput by horizontally scalable array-based solution.

ARGONNE NATIONAL LABORATORY | Chicago, US
Data Science Researcher
Nov 2019 - Feb 2023

- Led a team of 7 researchers in applying advanced analysis algorithms and AI for complex data classification, increasing experimental reach by 12%.
- Developed computer vision model with multi-GPU training strategy in `PyTorch` and `MLflow` to restore sensor images and accelerate physics simulation workflows by up to 20%.
- Facilitated collaboration among 10 international teams by establishing FAIR principles for AI models and documenting standardized methodologies for robust and explainable AI-driven results.
- Stress-tested Google Cloud Platform's readiness and scalability for large-scale data analysis workflows O(100 TB), minimizing TCO by optimizing egress.

UNIVERSITY OF SHEFFIELD | Sheffield, UK
Doctoral Researcher
Nov 2015 - Oct 2019

- Developed and maintained C++ data-analysis software across the full development lifecycle, enabling high-throughput distributed computations on a 500 k core computing grid.
- Monitored and troubleshoot real-time Kafka-like data streams assessing data quality during daily experiment operation.
- Authored the first technical report at CERN to open source a complete statistical model, setting a precedent for reinterpretation and transparency in the field.