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 data tool-chain and accelerate insurer analytics.

CONTACT

Location: Thessaloniki, Greece
Email: kourlitisV@hotmail.com
Tel: +30 6949552002
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

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

- Directed a global, 120-member cross-functional team with Agile methodologies, delivering 24/7 production support and new features for a $\sim$1 M LOC data analysis platform serving more than 3,000 users.
- Achieved 3x reduction in organizational data storage costs by adoption of a lightweight columnar data format ahead of schedule.
- Rolled out streamlined software configuration paradigm, reducing analysts onboarding time by over 95%.
- Planned and supervised early-career developers projects, delivering innovative data engineering products on time and fostering professional growth.

Data Science Researcher
Mar 2023 - Sept 2023

- Awarded €40k grant to modernize legacy C++ ETL tools for big data stream analysis ($\mathcal{O}$(100 TB)) with horizontally scalable array-based solutions, achieving 4x higher throughput and aligning with industry data science standards.

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

- Led a research team of 7 in applying advanced analysis algorithms and deep learning for complex data classification, increasing analysis coverage by 12%.
- Developed an end-to-end CNN computer vision model, with multi-GPU training strategy in `PyTorch` and `MLflow`, to improve simulation software cost-efficiency by up to 20%.
- Established FAIR principles for AI models and standardized methodologies among international research teams, allowing faster review cycles and enhanced transparency across projects.
- Stress-tested Google Cloud Platform's readiness and scalability for big data analysis workflows, 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.