Cosmin Catalin Sanda

Phone: +45 27 30 60 35
Email: cosmincatalin@gmail.com
Web: https://cosminsanda.com
LinkedIn: https://linkedin.com/in/cosmincatalinsanda

Professional Experience

Data Engineer at **LEGO** (started in *May 2023*)

Part of the Enterprise Transparency team focusing on collecting and manipulating data for use by upper management.

- Developed the GraphQL solution powering our data delivery to stakeholders.
- Expanded and modernized the use of message-buses for real-time consumption and delivery of events.
- Created a bespoke version of the Apollo GraphQL Router with authentication geared towards LEGO needs.
- Introduced and matured elements of code quality such as code coverage gates and static analysis.
- Consolidated the IaC and DevOps best practices across environments and reduced the security attack surface area.
- Mentoring.
- Pioneered the use of Rust for CLI tools and performance sensitive applications.
- Technologies: (C#, Python, Rust), (Azure, AWS, Databricks), (Terraform, GraphQL, Pulsar, ArgoCD, K8)

Data Engineer at **Saxo Bank** (*Dec 2020 - Apr 2023*)

Member of the client modelling team that is responsible for using data to analyze customer behaviour, churn and lifetime value.

- Lead the data transformation projects that our analysis and data science efforts rely on.
- Introduced new elements of DataOps in the data lifecycle (Data Quality checks, Monitoring).
- Containerized and deployed ML based APIs to Kubernetes.
- Pioneered the CI/CD pipelines for our department's projects.
- Mentoring.
- Out-of-band contributions.
- Open source activity: [trades-importer][finance-library]
- Technologies: (Spark, Kubernetes, Kafka, Airflow, MLflow), (Python, C#), (scikit-learn), (Azure, Databricks, Delta Lake)

Data Engineer at **AudienceProject** (*Jan 2015 - Nov 2020*)

Member of the data science team that researches, evaluates and implements technical solutions for Big Data / ML initiatives.

- Designed and implemented the distributed pipeline system for processing 1.5 billion events daily using a mix of batch and streaming elements [blog post].
- Established patterns for processing of data following principles of idempotence, resiliency, automatic recovery, functional approach, monitoring and eventual consistency.
- Bootstrapped the open-source mindset that allowed the release and contribution to external projects such as [MLflow][kinesis-writer][simple-arguments].
- Created and maintained the CI/CD/Artifacts infrastructure and pipelines.
- Maintained a cost-conscious mindset and created a service to allow for easy usage of spot-priced resources complete with failover to on-demand.
- Created the first available customer-facing API [ref].
- Introduced Python and MXNet in the company's stack [blog post].
- Recognized as AWS ML Hero for contributions and resources related to ML on AWS.
- Technologies: (Spark), (Scala, Python, Java, C#), (MXNet, ML.NET), (AWS, GCP, Databricks)

Reference: Henrik Lauritzen (Founder, CEO, CTO)

Data Science Engineer at **Tradeshift** (*May 2014 - Dec 2014*)

- Built a workflow that reconstructs events' lineage using graph queries and distributed data processing.
- Created live dashboards and reports of performance metrics of the platform.
- Technologies: (PostgreSQL, OrientDB, Hive, Talend), (Java, Python), (AWS)

Reference: Anders Nickelsen (Data Scientist)

Senior Software Engineer at **UserReport** (*Sep 2013 - Apr 2014*)

- Established ways of working and best practices with a DevOps mindset.
- Automated builds, testing and releases using Jenkins.
- Further developed the existing data pipeline architecture and implementation.
- Expanded the capabilities of the platform with a focus on backend work.
- Mentoring.
- Technologies: (Hadoop, Hive, Pig), (Java, PHP, NodeJS), (AWS, GCP)

Senior Software Engineer at **Secunia** (*Jul 2011 - Aug 2013*)

- Implemented the Active Directory integration for automating detection of machines.
- Created a replication mechanism using binary logs for providing our vulnerability database to hosted clients.
- Built an embedded database synchronisation system between frontend and backend.
- Created the first version of our scanning software as an Android application [ref].
- Technologies: (Java, PHP, JavaScript), (SQL Server, MySQL), (Android)

Reference: Martin Leopold (Director of Development)

Software Systems Engineer at **2Checkout** (*Apr 2008 - Jun 2011*)
- Built the backend and frontend of the reporting engine used by clients to slice and dice their exact financial activity.
- Conceived the initial version of the email newsletter engine.
- Technologies: (PHP, JavaScript), (MySQL)

Software Systems Engineer at **Stefanini** (*Jun 2006 - Apr 2008*)
- Development of the CRM application used by call center agents
- Practiced Release & Change Management as well as Agile methodologies
- Technologies: (Perl, JavaScript), (Oracle)

Programmer at **NetEngine** (*Jan 2004 - May 2006*)
- Developed the in-house CRM and ERP
- Technologies: (PHP, Javascript), (MySQL)

Education

- **Copenhagen IT University** - **MSc.** - Database Systems (*Feb 2012 - Sep 2015*)

- **Bucharest Polytechnics University** - **B.A. & MSc.** - Applied Informatics (*Sep 2002 - Jul 2007*)

Blog Articles

- https://www.audienceproject.com/blog/tech/pegasus-deep-learning-inference-at-massive-scale/

- https://www.audienceproject.com/blog/tech/aqueduct-real-time-data-processing-scale/

- Experiment tracking in ML pipelines with MLflow

- MXNet to ONNX to ML.NET with Amazon SageMaker, ECS and ECR

- Voice Recognition with MXNet and Amazon SageMaker

- Making a DSL for Generating Images

Languages

- English - advanced

- Romanian - native

- Danish - basic