André Kelpe

SOFTWARE ENGINEER WITH 20 YEARS OF PROFESSIONAL EXPERIENCE

andre@kel.pe
+49 151 70 33 26 72
Berlin, Germany

+ EXPERIENCES

Since April 2021 - Principal Member of Technical Staff at Salesforce (Berlin/remote)

My frist team was the UI platform analytics team, which was developing ML models for predicting page load times. The model is used in the designer and is meant to guide customers to better decisions that influence the load times of their pages. The project was using a novel "cluster less" approach to data processing, indexing, feature extraction and summarizing. The entire platform is event based and runs on AWS. AWS Stepfunctions are the central piece. When I joined the team I added support for more compute environment types. I also moved the ML training code from experimental to deployable/repeatable units. The last project on the team I worked on was event driven cataloging of the data inside AWS Glue Data Catalog.

Since August 2022 I am part of the PKI (Public Key Infrastructure) team in the security organization of Salesforce where I am working with 5 teams on various aspects of the internal PKI infrastructure. Our org is responsible that all data at Salesforce is encrypted while in transit. My role is overseeing, steering and implementing projects from the technical angle. Projects cover a wide variety of PKI releated topics like trust store management, Certificate Transparency Logs monitoring, Internal PKI inventory management, CRL publishing, deployment automation, Root CA tracking and implementing and overseeing review processes. As a PKI lead I also review designs of other teams and provide guidance to other teams across all of Salesforce.

March 2019 – April 2021 - Principal Engineer at Trifacta (Berlin office)

I joined as an Engineer focussing on various platform projects. The first project was an open-tracing/Zipkin inspired lightweight tracing framework to forward request- and session-ids across all of Trifactas services. Services are written in JS, Java, Scala and Python. The projects goal was to make the general supportability of the software stack at Trifacta easier. This feature was well received by on-prem customers, support and also by the on-call engineers supporting Trifactas cloud products. The next big project was an integration with Databricks Delta-Lake on Databricks for Azure. This project allows customers of Trifacta to read and write from/to Databricks tables as well as Delta Tables.

After the above I became tech lead for https://clouddataprep.com the Google cloud integrated offering by Trifacta. I led and implemented various projects (google sheets integration, adoption of google cloud CDN, enhanced security features with google IAM). However the main focus for 2020 and beyond was to integrate with the google cloud marketplace to replace a custom one-of-a-kind integration with googles backends we had. The marketplace platform we integrated with was relatively new and unstable, which made the project complex from technical as well a project management point of view. The project required in-house and external coordination with Googles engineering teams.

I was also responsible for all prodution pushes of https://clouddataprep.com (organizing, follow-up with other teams, processes as well as executing the pushes) and incident responses in European time. Part of this work was to shorten the release cycle from 3 times a year to monthly, which I successfully transitioned to. The deployment is deeply integrated with GCP, running on top of GKE/Kubernetes and Helm.

November 2016 – January 2019 - Principal Software Engineer at SnapLogic (San Mateo/Berlin remote)

Senior member of the BigData team. First assignment was BigData platform stability. Made on-site deployments for customers stable and predictable. Improved general stability of an early stage product leading to the biggest deal of SnapLogic at that time. The product is build on top of Apache Spark with a graphical designer that allows non-programmers to express BigData processes.

Desinged and implemented first iteration of the managed BigData integration platform for SnapLogic ("eXtreme"), which enables customers to use BigData technologies in the cloud without having to manage a cluster or know how to write code. The integration is mainly based on EMR, S3, IAM with micro-services orchestrating everything.

June 2013 – September 2016 - Senior Software Engineer at Driven Inc. (San Francisco/Berlin remote)

Major contributor to Cascading core as part of the open source team. Build and maintained integrations for JDBC (MySQL, Oracle, Postgres, Redshift), HBase, memcached, elasticsearch for Cascading. Maintained Cascading SDK and worked with Hadoop vendors to run Cascading compatibility suite. Worked with and contributed to upstream (hadoop) and downstream (scalding, cascalog, elasticsearch, flink) projects. Built integration for Cascading and Hive.

Presented Cascading at international meetups and conferences (Devoxx Belgium, HUG France in Paris, Apache BigData Budapest, TU Berlin, Big Data Beers Berlin).

Build JVM level agent using byte code manipulation to capture telemetry of Hive queries to be monitored with driven, the commercial monitoring product of Concurrent.

Worked on stability, performance and reliability of the product. Build new indexing strategy for the product on top of elasticsearch.

Presented Driven at the company booth at Hadoop Summit Brussels, Strata London & Hadoop Summit San Jose.

September - December 2012 - Quality Assurance Engineer at Awingu (Lochristi, Belgium)

Responsible for quality and release management of the IAAS cloud solution we were building for the biggest Belgian telco (Belgacom). Part of the job involved managing remote engineers in another time zone.

October 2010 - August 2012 - Expert Software Engineer at TomTom Belgium

Design and implementation of new map product creation and delivery platform including migration of legacy processes to the new platform. The stack is based on hadoop avro, crunch, zookeeper, giraph. The platform had to integrate data in all shapes and forms like RDBMS, XML, images, ESRI shapefile among others while presenting them in a uniform way through a special hadoop input format. It also had to be able to stream this data over to non hadoop consumers, which was done with a custom RPC protocol based on avro and orchestrated by zookeeper.

Implementation of business contingency/disaster recovery project on AWS EC2.

Implementation of an e-delivery system based on AWS (S3/cloudfront).

July 2009 - September 2010 - Senior Software Engineer at TeleAtlas Data Gent, Belgium

Implementation of versioning and release management of software and non software artifacts with subversion, hudson and maven.

Maintenance of the cash-cow processes within the TomTom maps unit.

2005-2009 - Senior Software Engineer at Tele Atlas Deutschland GmbH, Harsum, Germany

Implementation of a software system to integrate user generated map updates in the mother database of Tele Atlas. The system is based on java, oracle and the clustering technology terracotta and interfacing with other older internal technologies.

Implementation of new product line geared towards navigation customer for Tele Atlas, this included everything from data access layers to geomtry generalization, format conversions and large scale processing of geospatial data.

+ EDUCATION

2001 - 2005	**Bachelor Informationsmanagement & Informationstechnology (IM/IT)**
Stiftung Universität Hildesheim	Universität Hildesheim Universitätsplatz 1 31141 Hildesheim, Germany

+ LANGUAGES

German Native **English** Level C2 **Dutch** Level B1

+ IT PROFESSIONAL SKILLS

Languages/Technologies	**Golang, X.509, JVM (Java), Python, Spark, Hadoop, Hive, Elasticsearch, gradle, git, maven, AWS, EMR, S3, Linux, parquet, PostgreSQL, IAM, EC2, AWS Lambda, AWS Stepfunctions, GCP, GCS, Dataflow/Beam, Spring-Boot, express.js, Kubernetes, GKE, Helm, bash, sqlite**