+0162 74 60 114
Kleyerstr. 39, 60326 Frankfurt,
Germany
marvin.taschenberger@gmail.com

Marvin Taschenberger

Data Engineer / Data Architect

GitHub: taschenbergerm
LinkedIn:
taschenberger-marvin-94531bb2/

Solution oriented data architect with a strong focus on automation, communication and sharing knowledge. Architectures that create and code that i write always puts humans first.

Skills

Tools and Languages	Python, SQL (MySQL, Postgres), NoSQL (MongoDB, Clickhouse, Neo4j), Kafka, Linux, Docker, Cloud (Google), Golang (hobby)
Communication	German (mother tongue), English (business professional)

Technical Experience

Data Engineer / Data Architect — **Dec 2018 — Present**
Solactive — *Frankfurt*

Key Responsibilities

- Design, implementation and enforcement of several ITIL conform standards regarding configuration (CM), test (TM) and release (RM) management. The CM standards were used to point the developers in a direct of a proper micro-service (12-Factor-App) inspired design and promote decoupling, cohesion and SOLID-principles to increase the quality and lower the cost of development. Combining with the new Test-Guidelines to ensure a properly tested software (Unit & Integration Test), the integration into the landscape (system test) and the inclusion of the business user (User Acceptance Test). Increasing both confidence with new features as well as relieve some pressure from the developers. Adding the release management, including a controlled release cycle, communication, preparation and execution reduced the amount of errors significantly.
- Creating several utilities and deployment templates to implement the new Configuration, Test and Release Management. This included building several CI/CD pipelines that allowed for an automated deployment and of quality standards. As well as the creation of several helpers (based on cookiecutter) that create boilerplate code and folder structures to ease the change for developers.
- Designing multiple micro-service based end-to-end data pipelines using kafka-stream to build a data warehouse and inject data points into the life-calculation engine. The origin of the data varied - reaching from open data available through the web up to proprietary data via GUI's. Processes were triggered via events over Kafka-Stream and listened to using pythons Faust Framework. All schedules and configurations were saved in a replicated MongoDB triggered via pythons Advances Scheduler.
- Building a data warehouse for complex-data including data streams and feeds from multiple vendors including end-of-day as well as intraday marketdata. The sources for these reached through various technologies - Restful API's, web scraping as well as ftp/sftps delivery. Moreover as the universe of requested data needed to be flexible - multiple components allowed the interaction of business users as well as automated housekeeping. Transparency was achieved view logging through elastic search and displayed via dashboards build with Dash and Grafana allowing both business users as well as service operation experts to closely monitor the state of the systems. The data as well as the operation procedures were made available due to REST-Interfaces using FastAPI and Celery.
- Development of a cryptocurrency Index Calculation engine for indices and inavs based on CoinMarketCap's API. Schedulers and workers were created using Celery and connected via RabbitMQ deployed within a docker swarm.
- Training of business-user, analysts and developers in python and general programming
- Supervision and training of the departments software-engineering interns and new joiners

Technologies

Python, Kafka, Clickhouse, MySQL, MongoDB, Docker, RabbitMQ, Jenkins, Redis, Airflow, Rundeck, Linux, Grafan, EFK-Stack

Consultant Data Scientist — **Oct 2017 — Nov 2018**
STATWORX — *Frankfurt*

Key Responsibilities

- Support and Manage of major projects by creating architectures, ML models,data pipelines and presentations.
- Holding a biweekly python workshop to train fellow employees.
- Numerous side projects
  - development of a AI based Tool to estimate the the success rate for invites on business portal like Linked-In or Xing.

- support of a development to classify and cost estimation of damages on building based on drone images
- consulting clients by providing concepts of database-systems, data-pipelines and integration of data science into their product and project life-cycle.s

Technologies

R, Python,Docker, PostgreSQL, Linux, Google Cloud

Project: Multinational retailer **Oct. 2017 — Nov. 2018**
STATWORX *Frankfurt*

Outline

Creation of a web based Application to evaluate and forecast customers effect on changes in the pricing strategy. The underlying simulation supported either models either based on statistical elasticity or on a random forest.

Key Responsibilities

- Extending and fine tuning the statistical and machine learning models by enriching the underlying business model and data pipelines.
- Designing the technical architecture to bring the manual POC into production
- Automation of the data preparation via R , python and airflow and deploy it via docker on premises.
- Design and implement the Full-Stack application using R, R-Shiny, JavaScript.
- Deployed the Tech-Stack on multiple platform over its lifetime including on-prem, Azure and Google Cloud.
- Hold trainings, workshops and presentations for the clients to handle the application from the business, technical and management perspective.

Technologies

R, Python, Airflow, Docker, PostgreSQL, Linux, Azure, Google Cloud

Education

M.Sc. Economics — Major in Statistics and Econometrics - Grade: 1.9 , *University of Bonn* 2015 — 2017
B.Sc. Economics 2012 — Major in Macroeconomics - Grade: 2.1, *Christian Albrechts University, Kiel* 2012 — 2015

Interests

- D&D and Boardgames
- Cooking
- 3D-printing and modeling
- offee