WOJCIECH BAŻANT
Email: wojciech.bazant@gmail.com
Code: https://github.com/wbazant

SUMMARY

Highly skilled Data Engineer with a robust background in bioinformatics and software development. Experienced in full-stack development, data pipelines, and backend engineering, with a strong commitment to delivering high-quality, scalable solutions that meet business and research needs. Known for being a proactive team player with excellent problem-solving skills and a dedication to continuous learning and improvement.

CORE SKILLS

- Python, Java, Perl, Javascript, Bash, Scala
- Data Modeling
- Workflow Design
- Unix/Linux
- SQL
- Bioinformatics
- Cloud Solutions
- ETL
- Git

PROFESSIONAL EXPERIENCE

Data Engineer, Ethermed

Remote, Jan 2023 — Apr 2024

- Second tech hire at a healthcare AI startup after seed funding
- Implemented custom decision making modules and their integration with data flow of an early prospective client, leading to the client signing the contract
- Owned data access, helped ML engineers create training features, and evaluated machine learning models they produced
- Created the company's first operation dashboards and reliability procedures
- Provided code reviews and code architecture for a pilot project involving using LLMs for question evaluation in the context of prior authorizations
- Prototyped a requirements-to-code project pairing LLM codegen with abstract syntax tree manipulation
- Wrote a parsing pipeline for insurers' medical guideline documents
- Tech Stack: Data Engineer, Python, MLOps, LLM, Databricks, PySpark, Pandas

Data Engineer - Freelance, Deep Tech Jobs

Remote, *Jul 2022 — Sep 2022*

- Freelancer for a 'deep tech job' search platform
- Designed and developed a replacement for a low-code integration between a web scraping SaaS and a CMS board based on AWS Lambda and Python
- Extended the integration with data cleaning and annotation modules requested by the client
- Tech Stack: Data Engineer, Python, Serverless, Elasticsearch, Amazon Lambda, Web scraping, CMS, Pandas

Programme & Data Analyst, University of Glasgow

United Kingdom, Apr 2021 — Feb 2023

- Bioinformatician on VEuPathDB bioinformatics team with role focusing on genomic sites
- Designed and developed a Java microservice ([code link](#)) that currently powers sequence retrieval across over 10 VEuPathDB sites
- Designed and developed a .bed based interface for sites to query the microservice that facilitated implementing many requested functions and clearing a multi-year backlog
- Deprecated a CGI-based legacy sequence retrieval service using application database storage for sequences, which addressed global performance issues caused by the service
- Tech Stack: Data Engineer, Backend Engineer, Perl, Java, Spring Boot, SQLite

Bioinformatics Application Developer, Senior (D), University of Pennsylvania

United States, Oct 2019 — Jun 2021

- Bioinformatician on VEuPathDB bioinformatics team with role owning MicrobiomeDB.org
- Developed a Nextflow pipeline for whole genome metagenomics data ([code link](#)) which allowed the project to show studies with more accurate results
- Designed and developed an ingestion point for users' own datasets
- Initiated a collaboration with a NIH Nephele, a resource offering free processing of metagenomic sequencing data, that led to over a thousand researchers viewing their data through MicrobiomeDB
- Developed a tool based on novel method for identifying eukaryotes in WGS samples, produced a study that demonstrated its efficacy, and led the writing of a publication accepted by a major journal ([Improved eukaryotic detection compatible with large-scale automated analysis of metagenomes | Microbiome](#))
- Tech Stack: Data Engineer, Data Scientist, SQL, Unix, Perl, Python, Make, Makefile, Pandas, Nextflow

Bioinformatician, Wellcome Sanger Institute

United Kingdom, Sep 2017 — Jun 2019

- Bioinformatician with role split between the Berriman lab at Wellcome Sanger Institute and the Wormbase team at the European Bioinformatics Institute
- Operated and continuously improved data production processes for WormBase ParaSite, resulting in five successful releases (see release notes)
- Conducted automated quality assessment and data loading for the genome of a parasitic nematode *H. contortus*, becoming a named co-author in a publication in one of the best science journals, Nature (publication link)
- Designed and developed an ingestion point paired with an annotation processing pipeline (code link) which resulted in WBPS gene pages being enriched with results of transcriptomic studies and review pages summarising public availability of approximately all transcriptomic data for each organism
- Responsible for supporting users of the site internally in Berriman lab, and externally through help desk queries, helping lab scientists access genomic data relevant to their research
- Made conference posters and operated a booth, representing the resource in three parasitology conferences
- Tech Stack: Bioinformatician, Data Production, Data Modelling, SQL, Unix, Perl

Software Developer, European Bioinformatics Institute

United Kingdom, Mar 2016 — Sep 2017

- Full stack developer in three person development team for the Expression Atlas group
- Worked in the Agile framework on the Java backend (code link) and React frontend (code link)
- Developed an annotation pipeline (code link) which helped the group evaluate Scala as alternative to pipeline development strategy of mixing R, Bash, and Perl
- Tech Stack: Full-Stack Engineer, Java, Spring Framework, TDD, Apache Solr, Scala, R, Bash, Perl, React, JavaScript, NPM, Webpack

Technology Associate, Morgan Stanley

United Kingdom, Jul 2014 — Mar 2016

- Junior developer for an internal product providing P&L and risk assessment calculations for the bank's front office divisions under Interest Rate Derivatives
- Worked in the Agile framework on a Java application, its Scala API, and a Hazelcast distributed cache backend
- As part of on-call rota, diagnosed and triaged urgent production issues
- Tech Stack: Backend Engineer, Java, Hazelcast, Scala

Summer Intern, Infosys

India, 2013

- Summer intern at a research division of a large software development company
- Researched encoding of business compliance rules with model checking

Tutor, Akademeia

Poland, 2010 — 2013

- Math and physics tutor for an elite educational consulting and services company
- Supported over 50 students in preparation for exams and university interviews

EDUCATION

MMath, Mathematics, University of Oxford
United Kingdom, 2013 — 2014
First Class Honors

BA, Mathematics, University of Oxford
United Kingdom, 2010 — 2013
Upper Second Class Honors

NON-PROFIT

Falling Fruit

2024 — present

- Contributor and maintainer for a popular portal for foragers
- Took over stalled project to replace the current aging site and app and lifted it from a prototype-quality alpha to nearing readiness for rollout ([code link](#))
- Organised Hacktoberfest and integrated efforts of twenty participants
- Tech Stack: Frontend Engineer, React, Redux, LLM

United World Colleges

2016 — 2022

- Volunteer developer for the Polish committee
- Built an application platform and a pipeline for automatically anonymising and packaging applications for grading, removing repetitive work and assuring efficiency and fairness
- Operated the platform for eight years, providing tech support for applicants
- Tech Stack: Full Stack Engineer, AngularJS, Python, Google App Engine

QUANTITATIVE SKILLS

Data Engineering and Bioinformatics:

- **Data Pipeline Development:** Designed and implemented efficient data pipelines
- **Database Design and Management:** Optimized data ingest processes and schemas
- **Big Data Technologies:** Utilized PySpark, Databricks, and Pandas for analysis
- **Cloud Solutions:** Implemented AWS and Google App Engine for scalable solutions
- **Microservices and Serverless Architecture:** Developed microservices and serverless architectures
- **ETL Processes:** Ensured efficient data flow and transformation

Mathematical and Statistical Analysis:

- **Statistical Modeling:** Applied statistical methods for genomic data analysis
- **Machine Learning:** Evaluated and implemented machine learning models
- **Data Visualization:** Presented complex data in accessible formats
- **Algorithm Development:** Created algorithms for genomic data processing

Programming and Scripting:

- **Python:** Extensive experience in data engineering and analysis
- **Java and Scala:** Proficient in backend development and data processing
- **Bash and Perl:** Automated data processing tasks efficiently
- **Javascript:** Built modern single page applications

PERSONAL SKILLS

- **Communication** Effective in conveying complex ideas clearly and concisely
- **Teamwork** Collaborates well with cross-functional teams to achieve common goals
- **Problem-Solving** Analytical thinker with a knack for innovative solutions
- **Adaptability** Quickly adapts to new technologies and changing project requirements
- **Time Management** Efficiently prioritizes tasks to meet deadlines and maintain productivity
- **Leadership** Guides and mentors team members to foster a productive environment
- **Attention to Detail** Ensures high-quality output through meticulous review and testing
- **Continuous Learning** Actively seeks new knowledge and skills to stay current in the field
- **Project Management** Organizes and manages projects effectively using Agile methodologies

LANGUAGES

- **Polish:** Native Language
- **English:** Fluent
- **German:** Conversational basics
- **Russian:** Intermediate

HOBBIES

chess, foraging, hiking, kayaking, gardening