

## WOJCIECH BAŻANT

Email: [wojciech.bazant@gmail.com](mailto: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
- Data Modeling
- Workflow Design
- Unix/Linux
- SQL
- Bioinformatics
- Cloud Solutions
- ETL
- Git

## PROFESSIONAL EXPERIENCE

### **Data Engineer, Ethermed** *Jan 2023 — Apr 2024*

- Second tech hire at a healthcare AI startup post-seed funding
- Developed data access code, decision-making modules, and evaluated machine learning models
- Implemented the first client integration and handled deployment and reliability operations until a DevOps hire
- Reviewed code, mentored colleagues, and helped the company meet its key client goals
- Tech Stack: Data Engineer, MLOps, LLM, Databricks, PySpark, Pandas

### **Data Engineer - Freelance, Deep Tech Jobs** *Jul 2022 — Sep 2022*

- Improved processes for a 'deep tech job' search platform using a content management system

- Modeled data boards with Python, integrated web scraping SaaS, and migrated computation to AWS Lambda
- Reduced job search times from over five seconds to under a second by moving data from DynamoDB to an OpenSearch index
- Tech Stack: Data Engineer, Python, Serverless, Elasticsearch, Amazon Lambda, Web scraping, CMS, Pandas

**Programme & Data Analyst, University of Glasgow** *Apr 2022 — Feb 2023*

- Collaborated with the infrastructure team to modernize VEuPathDB's data provision infrastructure
- Developed a web application using Java in the microservices paradigm and a data workflow producing files
- Improved performance and added features, currently live on plasmodb.org and other VEuPathDB sites
- Tech Stack: Data Engineer, Backend Engineer, Python, Java, Spring Boot, SQLite

**Bioinformatics Application Developer, Senior (D), University of Pennsylvania** *Oct 2019 — Jun 2022*

- Responsible for the MicrobiomeDB.org portal.
- Developed solutions for whole genome metagenomics (WGS) data and created a new pipeline using Nextflow.
- Processed about six thousand samples, ensuring open access to publicly funded human microbiome data studies.
- Invented CORRAL, a novel method for identifying eukaryotes in WGS samples  
Published a major journal article as the first author  
(<http://doi.org/10.1186/s40168-023-01505-1>)
- Tech Stack: Data Engineer, Data Scientist, SQL, Unix, Perl, Python, Make, Makefile, Pandas, Nextflow

**Bioinformatician, Wellcome Sanger Institute** *Sep 2017 — Jun 2019*

- Maintained WormBase ParaSite, coordinating large-scale data production processes.
- Developed new pipelines in Perl and supported scientists using the site.
- Tech Stack: Backend Engineer, Database Engineer, Data Engineer, Data Scientist, SQL, Data Modeling, Unix, Perl

**Software Developer, European Bioinformatics Institute** *Mar 2016 — Sep 2017*

- Overhauled the Expression Atlas backend to support more data types and formats
- Improved data ingest code and maintained bioinformatics pipelines
- Ported experiment and gene pages to a single page framework using React Router
- Tech Stack: Full-Stack Engineer, Java, Spring Framework, TDD, Apache Solr, DDD, Scala, R, Bash, Perl, React, JavaScript, NPM, Webpack

**Technology Associate, Morgan Stanley** *Jul 2014 — Mar 2016*

- Core developer for a Java application on top of a Hazelcast distributed cluster
- Developed features and bug fixes for backend applications in Java and Scala
- Tech Stack: Backend Engineer, Java, SQL, Scala

## EDUCATION

**MMath, Mathematics, University of Oxford** 2013 – 2014 First Class Honors

**BA, Mathematics, University of Oxford** 2010 – 2013 Upper Second Class Honors

## QUANTITATIVE SKILLS

**Data Engineering and Bioinformatics:**

- **Data Pipeline Development:** Designed and implemented efficient data pipelines
- **Database Design and Management:** Optimized data ingest processes for integrity and performance
- **Big Data Technologies:** Utilized PySpark, Databricks, and Pandas for large-scale data 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

## 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
- **Client Relations** Builds strong relationships with clients through professionalism and reliability
- **Project Management** Organizes and manages projects effectively using Agile methodologies

## LANGUAGES

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

## INTERESTS

chess, Go, slacklining, hiking, kayaking, gardening