VISHAL SOOD

(+41) 79 942 0489 · vishal.chandra.sood@protonmail.com · Bussigny, CH-1030
Swiss Permit C · github.com/visood · linkedin.com/in/physicistvisha

SENIOR SOFTWARE DEVELOPER

Designing platforms that abstract computational complexity while preserving analytical power

A Systems Architect and Senior Engineer with a proven track record of designing and building robust, scalable platforms for data-intensive applications. Combines deep, first-principles expertise in statistical modeling and algorithms from a PhD in Physics with hands-on experience engineering high-performance backends (C++, Python) and complex workflow engines for distributed systems. Eager to leverage deep systems programming experience to quickly master Go and contribute to building next-generation AI infrastructure.

CORE EXPERTISE

Analytical Thinking | Research Design | Scientific Writing | Problem Solving | Project Management | Data Interpretation | Process Optimization | Critical Reasoning | Cross-Disciplinary Collaboration | Experimental Design | Knowledge Management | Technical Documentation | Workflow Development | Strategic Planning | Quality Assurance

TECHNICAL SKILLS

- Programming: Python (Expert), C++ (11/14/17/20) (Expert), SQL, Bash, Julia, Haskell, Scala, Object Oriented and Functional, C++/Template and Python Metaprogramming
- Scientific Computing: Algorithm Development, HPC (SLURM), Data Pipelines, NumPy, Pandas, HDF5, Dask, Performance Optimization, Statistical Modeling
- *Architecture & Design*: Systems Architecture, Distributed Systems, API Design, Microservices, Workflow Orchestration, Plugin-based & Extensible Frameworks
- Software Architecture: Scientific Workflow Automation, API Design, Plugin-based Frameworks, Distributed Systems, Declarative &Compositional APIs
- Data Engineering & DevOps: Data Modelling, CI/CD (Jenkins, GitHub Actions), Docker, Build Systems (Waf, CMake), Database Systems (MariaDB, PostgreSQL, PostGIS), Data Provenance, FAIR
- Research & Validation: Principled Validation Frameworks, Reproducibility by Design, Statistical Analysis, Hypothesis Testing, Monte Carlo Methods.

PROFESSIONAL EXPERIENCE

SENIOR SCIENTIFIC SOFTWARE DEVELOPER 2017 – 2024

Blue Brain Project, EPFL

Architected and built the core computational frameworks that enabled large-scale analysis and validation of the world's most detailed brain circuit models, directly empowering a large team of neuroscientists.

- Translated complex neuroscientific inquiry into robust, automated computational workflows; architected configuration-driven systems (Python, HPC/Slurm) that allowed scientists to execute and manage petabyte-scale analyses reproducibly.
- Led the design and development of a novel, plugin-based scientific validation framework (Python), creating an extensible ecosystem for rigorously comparing computational models against diverse experimental data.
- Engineered user-centric Python APIs that provided an intuitive, high-level interface to complex, terabyte-scale hierarchical data stores (HDF5), significantly lowering the barrier for exploratory data analysis.
- Drove significant improvements in operational efficiency by automating complex pipeline orchestration and dependency management, accelerating the research-to-results cycle.

- Designed and implemented massively parallel data processing workflows (**Python**, Dask) on a 100+ node distributed cluster (**SLURM**), mastering principles of distributed resource management, I/O optimization, and parallelization directly applicable to cloud infrastructure like GCP.

SENIOR PRODUCT DEVELOPER 2015 – 2017

Saphetor SA

Applied software engineering rigour to build a high-performance clinical genomics platform from the ground up, rapidly mastering a new scientific domain to deliver production-critical systems.

- Engineered the high-performance C++ backend for a real-time genomic variant annotation service, translating complex clinical and bioinformatics requirements into a scalable, low-latency system.
- Developed the end-to-end data processing pipeline for integrating and querying multiple large-scale genomic databases, creating the core knowledge base for variant interpretation.
- Designed and implemented a robust Python REST API to serve complex clinical queries, bridging the gap between the high-performance backend and user-facing applications.
- Authored a high-performance computational parser in C++ for a complex, domain-specific data format, demonstrating expertise in efficiently processing structured, high-throughput data streams.
- Led the migration and modernization of the core C++ build system (Make to Waf), significantly improving dependency management and developer productivity.

DATA SCIENTIST 2015- 2015

Citiviz Sárl, EPFL Innovation Park, Switzerland

MAÎTRE ASSISTANT 2011- 2014

Lab of Statistical Biophysics, EPFL, Switzerland

POST-DOCTORAL FELLOW 2008- 2010

Niels Bohr Institute, Copenhagen, Denmark

RESEARCH FELLOW 2006- 2008

Physics Dept., University of Calgary, AB, Canada

ACADEMIC FOUNDATION

Doctor of Philosophy in Physics, Boston University, Boston, MA, USA 2006 Thesis: Interacting Particle Systems on Graphs. Developed theoretical and computational models for stochastic processes, providing a deep, first-principles foundation for modelling complex systems and developing data-driven algorithms.

Bachelor of Technology in Engr. Physics, Indian Institute of Technology Bombay, India 2000

SPECIALIZED EXPERTISE

- **Principled System Design for AI Platforms**
  - Leverages a PhD in Statistical Physics to model complex systems from first principles—the same mathematical foundation underlying modern AI algorithms like diffusion models, Bayesian inference, and large-scale simulations.
  - Implements numerical methods, including Monte Carlo and optimization algorithms, in C++ and Python to solve complex simulation and parameter estimation problems.
- **High-Performance Systems Engineering**
  - Engineers and optimizes high-performance C++ and Python applications for real-time data processing, sharing core patterns with data ingestors and model inference services in an AI stack.
  - Masters modern C++ and Python performance tuning, memory management, and I/O optimization to build robust and scalable data-intensive systems.
- **Robust Validation & MLOps Architecture**
  - Architects formal validation frameworks for the rigorous, systematic benchmarking of competing predictive models against empirical data—a key pillar of MLOps.
  - Enforces reproducibility and auditability by design, creating systems where every result is traceable to its inputs and assumptions—a foundational principle for trustworthy AI.

- **Scalable Infrastructure for AI / ML**
  - Designs automated, configuration-driven pipelines to orchestrate large-scale model training, hyperparameter sweeps, and data preprocessing jobs on distributed systems.
  - Engineers integrated data systems that transform terabyte-scale raw data into queryable, interactive assets for model development and analysis.

KEY ARCHITECTURAL EXPERIENCE

My work has focused on building domain-agnostic frameworks that operationalize quantitative research and data processing at scale.

1. *High-Performance Backend for Real-Time Data Processing*

   Engineered the core C++ backend and Python REST API for a high-throughput, low-latency data annotation service (**Saphetor**).

   - **System Design:** Built a backend service capable of processing thousands of complex data transactions per second, handling structured data streams analogous to financial or event-driven protocols.
   - **Data Pipelines:** Architected robust, end-to-end pipelines managing the full data lifecycle from raw input to a queryable SQL knowledge base.

2. *Framework for Scalable & Reproducible Computation*

   Architected a workflow engine (**connsense-TAP**) to manage petabyte-scale analyses on distributed HPC systems.

   - **Core Principle:** Decoupled a user's declarative intent from the complexity of parallel execution and resource management—a direct parallel to streamlining structured processes on an AI platform.
   - **Auditable by Design:** Elevated the configuration file to be a complete, version-controlled, and executable record of a process, ensuring perfect reproducibility.

3. *Framework for Principled Model Validation*

   Conceived and built a novel software ecosystem (**DMT**) for rigorous, automated model-data comparison.

   - **Interoperable Core:** Used the Adapter Pattern and formal API contracts to allow any predictive model to be tested against any dataset, a key pattern for MLOps.
   - **Automated Reporting:** Engineered a system to programmatically generate rich validation reports, a critical component for any production AI system.

4. *Architecting a Knowledge Portal for a Digital Brain Reconstruction*

   Architected the end-to-end data and analysis pipelines that power the SSCx Portal, a platform making one of the world's most detailed brain reconstructions accessible to the global neuroscience community.

   - **Data Processing Engine**: Designed the core engine (connsense-TAP) to transform raw, petabyte-scale simulation outputs from HPC into a structured, queryable HDF5 knowledge base.
   - **Scientific Content Generation**: Engineered the analysis framework (factology) to programmatically generate the portal's content hierarchical "factsheets" with integrated data and figures.