

# Dave Austin, Ph.D.

Charleston, SC | austindi1133@gmail.com | LinkedIn | GitHub

## Summary

---

Data Engineer and Scientific Computing specialist with a Ph.D. in Computational Physics and experience building automated data workflows across HPC and cloud production environments. Skilled in designing reliable pipelines, orchestration systems, and data platforms that transform complex raw inputs into reproducible, analytics-ready datasets. Combines distributed systems experience from large-scale scientific computing with modern data engineering practices including cloud infrastructure, workflow orchestration, and production monitoring.

## Technical Skills

---

**Languages:** Python, SQL, Bash, TypeScript

**Data Platforms:** Snowflake, PostgreSQL, Athena/Presto, Parquet, JSON

**Cloud & Infra:** AWS (S3, EC2, Lambda, CloudWatch), Docker

**Orchestration:** Prefect, SLURM

**Data Engineering:** ETL/ELT, data modeling, batch & event-driven pipelines, observability

**Scientific Computing:** VASP, Quantum ESPRESSO

**Other:** Git, Linux, CI/CD, L<sup>A</sup>T<sub>E</sub>X, MATLAB, Mathematica

## Experience

---

### Data Engineer / Software Engineer

Mar 2025 – Present

*Bear Cognition*

*Charleston, SC*

- Own and architect the company's data platform, designing ingestion, modeling, and transformation workflows supporting analytics, logistics, marketing, and reporting teams.
- Establish data standards, schema design practices, and reliability patterns across batch, scheduled, and event-driven pipelines.
- Serve as primary technical owner of the Snowflake environment, managing performance, governance, and scalable data organization.
- Designed agentic workflows automating ingestion and validation of heterogeneous partner datasets, reducing manual intervention.
- Engineered normalization and validation logic transforming semi-structured partner inputs into production-ready data models.
- Built monitoring utilities summarizing Prefect warnings and failures to improve visibility into missed or delayed data processing.
- Implemented retry logic, scheduling strategies, and secret management for production workflows.
- Partner with cross-functional stakeholders to translate operational needs into data products enabling reporting and decision support.
- Contributed to full-stack development in a Nuxt + TypeScript application including API integrations and backend improvements.

### Postdoctoral Scholar — Scientific Computing & Data Systems

Aug 2024 – Mar 2025

*University of Central Florida*

*Orlando, FL*

- Extended automated computational workflows supporting large-scale electronic structure simulations and reaction modeling.
- Built analysis pipelines converting raw simulation outputs into structured datasets used for downstream modeling and publication.

- Developed validation and monitoring scripts to detect incomplete or non-converged simulations using physics-informed criteria and automated quality checks.
- Maintained reproducible workflows enabling regeneration of analysis artifacts from tracked configuration metadata.
- Led mentorship and coordination across projects, guiding researchers on workflow automation, job scheduling, and reproducible data practices.

**Graduate Research Assistant — Scientific Computing & Data Engineering** Aug 2018 – Aug 2024  
*University of Central Florida* Orlando, FL

- Designed automated HPC orchestration systems managing DFT simulations across distributed supercomputing environments (Stampede2/3, Perlmutter, Anvil).
- Built Python tooling to submit, monitor, and manage batches of 20–30 concurrent jobs with timeout detection, retry logic, and automated restart workflows.
- Eliminated manual monitoring by implementing logging, validation checks, and email notifications, saving days of compute time per project and increasing configuration sampling throughput.
- Developed large-scale ingestion pipelines parsing GB–TB simulation outputs (VASP/Quantum ESPRESSO) into structured datasets (JSON/CSV) for analysis and publication.
- Extracted electronic structure features (DOS, band structures, energies, relaxed geometries, charge density / localization) and built scripts to regenerate analysis artifacts for reproducibility.
- Implemented automated data movement via SSH/SCP from HPC systems to local analysis environments, including selective transfer of large artifacts (e.g., multi-GB WAVECAR outputs).
- Created simulation frameworks generating STM/STS images from electronic structure data; supported multiple peer-reviewed publications through data generation and analysis.
- Mentored graduate and undergraduate researchers on SLURM scheduling, workflow automation, debugging, and reproducible data engineering practices.

**Teaching Assistant** Aug 2019 – May 2020  
*University of Central Florida* Orlando, FL

- Led undergraduate physics laboratories and instructed students in experimental data analysis and quantitative problem solving.
- Developed grading frameworks and provided technical instruction supporting consistent evaluation and student feedback.

## Education

---

**Ph.D. in Physics (Computational Materials Science)** 2024  
University of Central Florida

*Dissertation: First Principles Studies of Nano-Scale Phenomena At Surfaces: From Characteristics of Single Atom Catalysts to Molecular Structure Formation*

**B.S. in Physics** 2018  
College of Charleston

## Research Impact

---

8 peer-reviewed publications in computational physics and materials science demonstrating advanced modeling, large-scale data analysis, and interdisciplinary collaboration.