Building a Supercomputing Team

A Case Study of Team Weiming

Yuanhang Sun, Peking University Super Computing Team

About Me

- Education: Peking University, B.S. ('21) & M.S. Candidate ('25)
- Research Interest: MLSys (Machine Learning Systems)
- Member: PKU Supercomputing Team
- ASC Experience:
  - ASC23 Observer, ASC24 Team Captain, ASC25 Student Coach

Agenda

- Introduction
- Competition Preparation
- Team Organization
- Team Building in the AI Era

Team Intro

About the Team

- Founded: 2015
- Members: PKU undergraduates passionate about HPC
- Composition
  - Diverse & Balanced: Distribution across grades, gender, and majors
  - Multidisciplinary: Includes Foreign Languages, Engineering, CS, Electronics, etc.
  - Structure: Strong legacy with a mix of veterans and new members
- Mission
  - Cultivate Awareness: Promote HPC mindset across campus
  - Empower Usage: Enable students from diverse backgrounds to leverage HPC tools

Track Record

2025.11.20
IndySCC大学生超算比赛
线上 第二名

2025.05.14
ASC大学生超算比赛
中国 青海 **亚军**

2024.04.13
ASC大学生超算比赛
中国 上海 **总冠军**

2023.05.10
ASC大学生超算比赛
中国 合肥 **总冠军**

2020.11.09
SC大学生超算比赛
线上 第二名

2019.04.25
ASC大学生超算比赛
中国 大连 一等奖

2016.11.16
SC大学生超算比赛
美国 盐湖城 第六名

2024.11.21
SC大学生超算比赛
美国 亚特兰大 第二名

2024.11.21
SC大学生超算比赛
美国 亚特兰大 第二名

2023.11.15
SC大学生超算比赛
美国 丹佛 第二名

2021.05.12
ASC大学生超算比赛
中国 南科大 一等奖

2019.11.17
SC大学生超算比赛
美国 丹佛

2017.11.11
SC大学生超算比赛
美国 丹佛 第四名

北京大学 PEKING UNIVERSITY

Competition Preparation

Team Structure & Roles

- **Captain x 1**
  - **Team Coordination**: Oversee overall operations.
  - **Management**: Schedule meetings & maintain documentation.
  - **Competition Strategy**: Allocate cluster time & make critical decisions.
  - **Logistics & Support**: Handle logistics and ad-hoc tasks.
- **SysAdmin x 1**
  - Often doubles as an Application Lead
  - **System Design**: Hardware architecture & assembly.
  - **System Tuning**: Software environment & toolchain optimization.
  - **Power Management**: Monitor and control power consumption.
- **For Applications**
  - **Workflow**: 1 Lead + 1-2 Members per application.
  - **Benchmarking**: Conduct baseline tests.
  - **Performance Tuning**: Analyze bottlenecks & optimize code.

Optimization Workflow

- **Step 0: Toolchain Construction**
- **Step 1: Baseline & Comprehension**
  - Run baseline & Understand application logic, I/O formats, and module functions
- **Step 2: Bottleneck Analysis**
  - Identify performance hotspots using profilers (e.g., VTune, Nsight).
- **Step 3: Optimization Strategy**
  - **Software**: Replace libraries, rewrite compute-intensive kernels
  - **Hardware**: Tune hardware configuration for optimal fit

> ***Note:*** *Steps 2 and 3 form an iterative cycle.*

Besides Optimization

- Where to find cluster
  - Taobao: Monthly bare-metal rental with root access.
  - HPC Cluster: Pay-as-you-go (Core-hours / GPU-hours)
  - **Recommendation:** Maintain 1-2 local servers to foster team identity and ownership
- Why Toolchains are important
  - ASC24 Finals: Pre-installed mlnx_ofed drivers failed to support ConnectX-6, many teams failed to run multi-node tasks.
- **Food Matters:**
  - Pizza at meetings significantly boosts attendance and engagement

Team Orgazation

Sustainable Growth: 一顿饱和顿顿饱

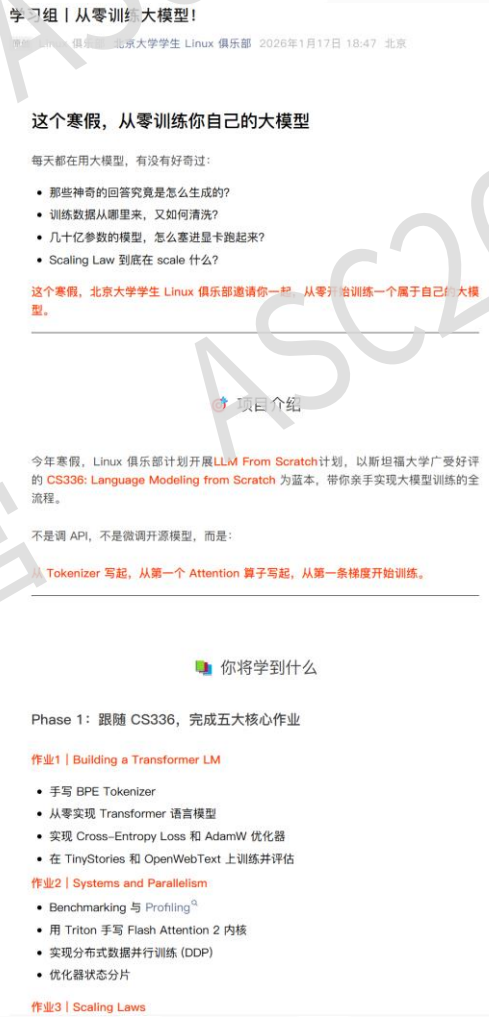
- Old Guides New
  - **Novices:** Participate in actual competitions
  - **Veterans:** Provide experience, strategy, and guidance
  - Member Lifecycle
    - Freshman (Spring): Join & Observe ASC
    - Sophomore (Fall): Compete in SCC
    - Sophomore (Spring): Compete in ASC and then **retire**!
    - Junior: One member remains as Captain.
- Open to all majors and skill levels
  - Passion for HPC is the only prerequisite.
  - Interdisciplinary background.
- Quality over Quantity
  - Better to have fewer, high-quality members

The Competition-Driven Model

- Core Competitions: ASC、SCC and HPCGame
- **Recruitment**
  - Timing: Annual intake in February.
  - Source: Select top Freshmen performers from HPCGame
- **Learning by Doing**
  - Assign actual ASC problems to new members immediately
  - Emphasize self-initiative, hands-on debugging, and peer discussion
- **Advanced Development:** create challenges for HPCGame
  - Tackle unknown optimization tasks
  - Convert solutions into new contest problems

Advanced Training & Workshops

- **Topic Selection**
  - Aligned with academic/industry hotspots (e.g., LLM).
  - Tailored to team members' interests.
- Recent Highlight (Winter 2026)
  - "**LLMs from Scratch**."
  - Co-hosted by the Supercomputing Team & Linux Club.
- Expert Insights
  - Professors & Alumni invited to weekly meetings.
  - Sharing frontier knowledge and industry practices.



学习组丨从零训练大模型！

北京大学学生 Linux 俱乐部 2026年1月17日 18:47 北京

这个寒假，从零训练你自己的大模型

每天都在用大模型，有没有好奇过：

- 那些神奇的回答究竟是怎么生成的？
- 训练数据从哪里来，又如何清洗？
- 几十亿参数的模型，怎么塞进显卡跑起来？
- Scaling Law 到底在 scale 什么？

这个寒假，北京大学学生 Linux 俱乐部邀请你一起，从零开始训练一个属于自己的大模型。

项目介绍

今年寒假，Linux 俱乐部计划开展LLM From Scratch计划，以斯坦福大学广受好评的 CS336: Language Modeling from Scratch 为蓝本，带你亲手实现大模型训练的全流程。

不是调 API，不是微调开源模型，而是：

从 Tokenizer 写起，从第一个 Attention 算子写起，从第一条梯度开始训练。

你将学到什么

Phase 1：跟随 CS336，完成五大核心作业

作业1 | Building a Transformer LM

- 手写 BPE Tokenizer
- 从零实现 Transformer 语言模型
- 实现 Cross-Entropy Loss 和 AdamW 优化器
- 在 TinyStories 和 OpenWebText 上训练并评估

作业2 | Systems and Parallelism

- Benchmarking 与 Profiling
- 用 Triton 手写 Flash Attention 2 内核
- 实现分布式数据并行训练 (DDP)
- 优化器状态分片

作业3 | Scaling Laws

Proposal : HPCWiki

- The Pain Point
  - Extensive repetition in basic documentation across different university teams.
  - Non-Core Focus: Time wasted on generic tutorials rather than unique competitive advantages.
- The Vision
  - Establish a unified, shared HPC knowledge base.
  - Eliminate "reinventing the wheel" and standardize basic training.
- Current Status
  - Website framework deployed at hpcwiki.io.
  - Currently stalled due to manpower shortage.

Proposal: Shared Recruitment Platform

- **Current Challenges**
  - **Timing Gap**: Recruitment (Oct) for ASC Season, but Lack of competitive events
  - **Redundancy**: Duplicated effort in building evaluation platforms across universities
  - **Cost**: High expense for recruitment contest compute resources.
- **Proposed Solutions**
  - **Infrastructure:** Reuse HPCGame infrastructure for judging and hosting.
  - **Joint League:** Inter-university format to share the workload of problem design
  - Sponsorship Model:
    - Platform centralizes sponsor negotiations
    - Sponsors cover Core-hours for **ALL** competitions
    - Platform provides compute to contests.
- **Pilot Success (Sept-Oct 2025)**
  - LZU-NJU-QLU 3-University Joint League.

Paradigm Shift: HPC Teams in the AI Era

HPC Teams in the AI Era

1. **Opportunities: High ROI**
   - **Market Demand:** Exploding need for MLSys talent in industry.
   - **Career Boost:** Higher expected returns for students participating in HPC.
2. **Technical Evolution: "Black-box" to "White-box"**
   - **Code Accessibility:** LLMs help decode massive legacy codebases (e.g., 200k+ lines of Fortran).
   - **Barrier Removal:** Previously "untouchable" code is now open for refactoring.
   - **Competitive Edge:** More dimensions for optimization and rivalry.
3. **The Crisis: Talent & Skills**
   - **Recruitment:** Students prefer pure AI/Algo over Systems Engineering.
   - **Skill Erosion:** Over-reliance on LLMs -> Decline in fundamental coding abilities.
   - **Training Burden:** Harder to cultivate deep understanding in novices.
4. **Strategic Response**
   - **The Pivot:** Transitioning toward AI-integrated workflows is necessary.
   - **New Curriculum:** Shift focus to **AI Literacy**—teaching the *correct* and rigorous use of AI tools.

北京大学
PEKING UNIVERSITY

Thanks!

Best Wishes:
Wishing ASC continued success and growth!