Elemem加速卡安装指南

文档信息

- **版本**：2.0.0
- **更新日期**：2025年7月
- **适用范围**：Elemem向量数据加速卡

目录

安装前准备

系统要求

推荐系统

- Ubuntu 24.04 LTS

推荐硬件配置

- CPU：AMD EPYC Rome 7H12（2.60G/256M/64C/128T/280W）
- 操作系统：ubuntu 24.04 kernel 6.1.8
- 内存：内存 256GB DDR4 ECC RDIMM * 4
- 主板：支持 PCIe 3.0 x4
- 电源：至少 650W（推荐 750W 80+ 金牌认证）
- 硬盘：SSD 960GB U.2 2.5 Nvme 读取密集型 * 2 + HDD 6T SATA 企业级 3.5 7200 * 3

工具准备

- 螺丝刀
- 防静电手套或防静电腕带

硬件安装

第一步：断电准备

1. 完全关闭计算机并断开电源线
2. 按住电源键10秒以释放残余电荷
3. 佩戴防静电手套或防静电腕带

第二步：打开服务器盖板

1. 拆除机箱面板螺丝
2. 小心滑开或取下面板
3. 找到主板上的PCIe插槽位置

第三步：选择合适的PCIe插槽

1. 优先选择最靠近CPU的PCIe x16插槽
2. 确保插槽周围有足够空间散热
3. 检查插槽是否支持PCIe 3.0或4.0规格

第四步：安装加速卡

1. 拆除机箱后挡板对应位置的挡片
2. 取出Elemem加速卡，避免触碰金手指
3. 将加速卡垂直插入PCIe插槽
4. 确保卡牢固插入，听到"咔哒"声
5. 使用螺丝将加速卡固定在机箱挡板上

第五步：检查安装

1. 确认加速卡已正确安装
2. 检查所有连接线是否牢固
3. 合上服务器

驱动程序安装

获取安装程序（FAE提供）

```
# FAE提供的安装包，名称规格如下
elemem-vector-engine.[2.0.8].tar.gz

# 解压
tar xvf elemem-vector-engine.[2.0.8].tar.gz

# 目录结构
release
├── elemem-driver-2.0.7.202507161739.run          // 驱动
├── elemem-firmware-2.0.2.8.bin                   // 固件
├── elemem_sdk_2.0.1.202507151532_ubuntu24.04.tar // 软件
```

Ubuntu系统

```
# 更新包管理器
sudo apt update

# 安装驱动
sudo bash elemem-driver-2.0.7.202507161739.run

# 检查驱动加载
lsmod | grep elem

# 输出下面内容说明安装成功
elem                   110592  5

# 查看设备状态
elem-smi -L
+-------------------------+---------------------------------+------------------
-----+-------------------------+-------------------------+
|        overview         |                                 |
|                         |                         |
+-------------------------+---------------------------------+------------------
-----+-------------------------+-------------------------+
|       card_index        |            group_num            |       chip_num
|        bank_size        |          alive          |
+-------------------------+---------------------------------+------------------
-----+-------------------------+-------------------------+
|           0             |                26               |          78
|          16             |            1            |
+-------------------------+---------------------------------+------------------
-----+-------------------------+-------------------------+
|      soft_version       |          driver_version         |      smi_version
|    firmware_version     |      fpga_version       |
+-------------------------+---------------------------------+------------------
```

```
-----+------------------------+------------------------+
|          2.0.0           |        2.0.2.202507171648        |
2.0.13.202508111142  |        25070801          |        25080610          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|          card            |             card_idx             |           0
|                          |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|          name            |               dna                |       power.cap
|        power.use         |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|        ELEM-CH21         | 00000000400200000171426 32D40C245 |           0
|          3870            |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|          temp            |             cycle                |         index
|    utilization.rram      |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|           59℃            |                0                 |           0
|            0             |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|          ddr             |             card_idx             |           0
|                          |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|      memory.total        |        memory.used(Mbyte)        |
memory.h2c_buffer    |    memory.c2h_buffer     |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|          8GB             |               1251               |          0%
|          0%              |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|         driver           |             card_idx             |           0
|                          |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|     driver_version       |        2.0.2.202507171648        |       pci_speed
| Speed 8.0GT/s, Width x4  |                          |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
| driver.h2c.speed (MB/s) |          driver.h2c.qps          |
driver.h2c.hugepacket |    driver.h2c.packet    | driver.h2c.packing.rate |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
|            0             |                0                 |           0
|            0             |            0             |
+------------------------+--------------------------------+-----------------
-----+------------------------+------------------------+
```

```
| driver.c2h.speed (MB/s) |          driver.c2h.qps          |
driver.c2h.hugepacket |    driver.c2h.packet    | driver.c2h.packing.rate |
+-------------------------+----------------------------------+-----------------
-----+-------------------------+-------------------------+
|            0            |                0                 |           0
|            0            |            0            |
+-------------------------+----------------------------------+-----------------
-----+-------------------------+-------------------------+
|                         |                                  |
|                         |                        |
+-------------------------+----------------------------------+-----------------
-----+-------------------------+-------------------------+

# 温度监控
elem-smi -q -d temp -i 0 -l 1
59℃
```

elem-smi管理工具更多使用方法见 4-系统管理文档

固件程序安装

```
# 安装固件
sudo elem-update 0 elemem-firmware-[2.0.2.8].bin # 0代表第一张卡，1代表第二张卡

+-------------+--------------+-----------------------------+------------------
-----------+
|   Magic Key | PRODUCT MODEL |              Packet Version |
打包时间 |
+-------------+--------------+-----------------------------+------------------
-----------+
|    VD2-1000 |      BIN@ELEM |                  0x25022101 |         2025-08-11
17:52:04 |
+-------------+--------------+-----------------------------+------------------
-----------+
| FPGA CUSTOM | FPGA REVISION |             Firmware CUSTOM |
Firmware REVISION |
+-------------+--------------+-----------------------------+------------------
-----------+
|  0x25080601 |    0x25080601 |                  0x43483131 |
0x0 |
+-------------+--------------+-----------------------------+------------------
-----------+
|             |              | Normal Main Bin File Length | Normal Main Flash
Start Addr |
+-------------+--------------+-----------------------------+------------------
-----------+
|             |              |                    40516796 |
67108864 |
+-------------+--------------+-----------------------------+------------------
```

```
-----------+

check head success input 'go' to continue

# 版本确认无误后，输入go，然后回车，开始升级固件，升级完成之后，需要掉电重启机器

# 重启系统
sudo poweroff

# 查看固件版本
elem-smi -L
+-------------------------+---------------------------------+-----------------
-----+-------------------------+-------------------------+
|        overview         |                                 |
|                         |                         |
+-------------------------+---------------------------------+-----------------
-----+-------------------------+-------------------------+
|       card_index        |            group_num            |       chip_num
|        bank_size        |          alive          |
+-------------------------+---------------------------------+-----------------
-----+-------------------------+-------------------------+
|           0             |                26               |          78
|          16             |            1            |
+-------------------------+---------------------------------+-----------------
-----+-------------------------+-------------------------+
|      soft_version       |          driver_version         |      smi_version
|    firmware_version     |      fpga_version       |
+-------------------------+---------------------------------+-----------------
-----+-------------------------+-------------------------+
|   2.0.0.202507171131    |        2.0.1.202507031525       |
2.0.6.202507141135   |        25070801         |        25071401         |
```

软件安装

方案1：docker安装

```
# 进入工程目录
cd docker

# 目录中包含如下文件
- docker-compose.yml

# 启动加速卡引擎
sudo docker compose up -d server # -d是为了让容器在后台运行，不使用此参数会直接在当前运行，并直接打印日志到当前窗口。docker-compose.yml 中配置了本地端口8000映射到容器内端口8000
可以通过sudo docker logs elemem_server查看容器启动的日志，容器内的服务是通过supervisor控制的。
关于compose使用的一些说明：
```

```
在旧版中，可能需要使用sudo docker-compose up -d server。在旧版docker时，docker-compose是一个独立的命令，属于Compose v1(2023年标记为deprecated)，新版docker(≥20.10.13)，compse v2(2020年推出的)可以作为一个插件安装，安装后compse是docker的一个子命令，建议使用最新版。

# 运行 C++ Demo
sudo docker compose run --rm client # --rm 代表退出后就删除本次创建的容器，请根据自己需要修改运行参数
or
sudo docker compose up -d client
sudo docker exec -it elemem_client /bin/bash

# 127.0.0.1 可更换为docker宿主机的ip
# --hdf5 后可配置为本地数据文件的路径
# compose文件中可以看到挂载了/mnt/到容器内的/mnt/
# entrypoint.sh 详细写了如何运行各个demo，可以按需修改
cd /root/hilbert
bash entrypoint.sh --server 127.0.0.1:7000  --hdf5 ./c++/SIFT_1M.hdf5

# 运行bench_test
cd c++
./bazel-bin/test_qps_recall config.ini

# 查看运行状态
sudo docker compose ps -a

# 其他未尽docker相关命令，请参考docker文档，在此不一一赘述
```

方案2：主机安装

安装软件运行时依赖

```
#!/bin/bash

set -ue

apt update
apt install -y build-essential
apt install -y libgoogle-perftools-dev
apt install -y libhdf5-dev
apt install -y libhiredis-dev
apt install -y libopenblas-dev
apt install -y wget
apt install -y cmake
apt install -y redis-server redis-tools

WORK_DIR=$(mktemp -d "./elem_XXXXXX")
FAISS_DIR="$WORK_DIR/faiss"
mkdir -p "$FAISS_DIR"
cd "$FAISS_DIR"
```

```
FAISS_PACKAGE="faiss.tar.gz"
wget https://github.com/facebookresearch/faiss/archive/refs/tags/v1.11.0.tar.gz
-O ${FAISS_PACKAGE}
tar --strip-components=1 -xzf ${FAISS_PACKAGE}
cmake -B build . -DFAISS_ENABLE_GPU=OFF -DFAISS_ENABLE_PYTHON=OFF -
DBUILD_TESTING=OFF -DBUILD_SHARED_LIBS=ON
make -C build -j faiss
make -C build install
echo 'export LD_LIBRARY_PATH=/usr/local/lib:${LD_LIBRARY_PATH:-}' >>
/etc/profile
source /etc/profile

echo "deploy.sh completed successfully."
```

安装软件

```
# 解压
tar xvf elemem_sdk_[2.0.1.202507151532_ubuntu24.04].tar

# 启动服务
sudo bash start.sh

# 执行启动后输出内容如下
Waiting up to 1200s for port 6378 to be listened...
Port 6378 is now listening (after 1s).
Waiting up to 1200s for port 8000 to be listened...
Port 8000 is now listening (after 2s).
Waiting up to 1200s for port 7000 to be listened...
Port 7000 is now listening (after 1s).
所有服务已后台启动，日志请查看各自目录下的 *.log 文件。

# 检查启动是否成功
curl http://localhost:8000/health # 返回ok说明安装成功
curl http://localhost:7000/health # 返回ok说明安装成功

# 停止服务
sudo bash stop.sh
```