

# 智慧医研

基础设施面对医疗 AI 的机遇与挑战

Lenovo  
NetApp + Lenovo 联想



CHIMA2021会议资料  
未经授权请勿转载  
[www.chima.org.cn](http://www.chima.org.cn)

# 智慧医研

基础设施面对医疗 AI 的机遇与挑战

Lenovo + NetApp 联想



CHIMA2021会议  
未经授权请勿转载  
[www.chima.org.cn](http://www.chima.org.cn)



# 联想 “新IT”

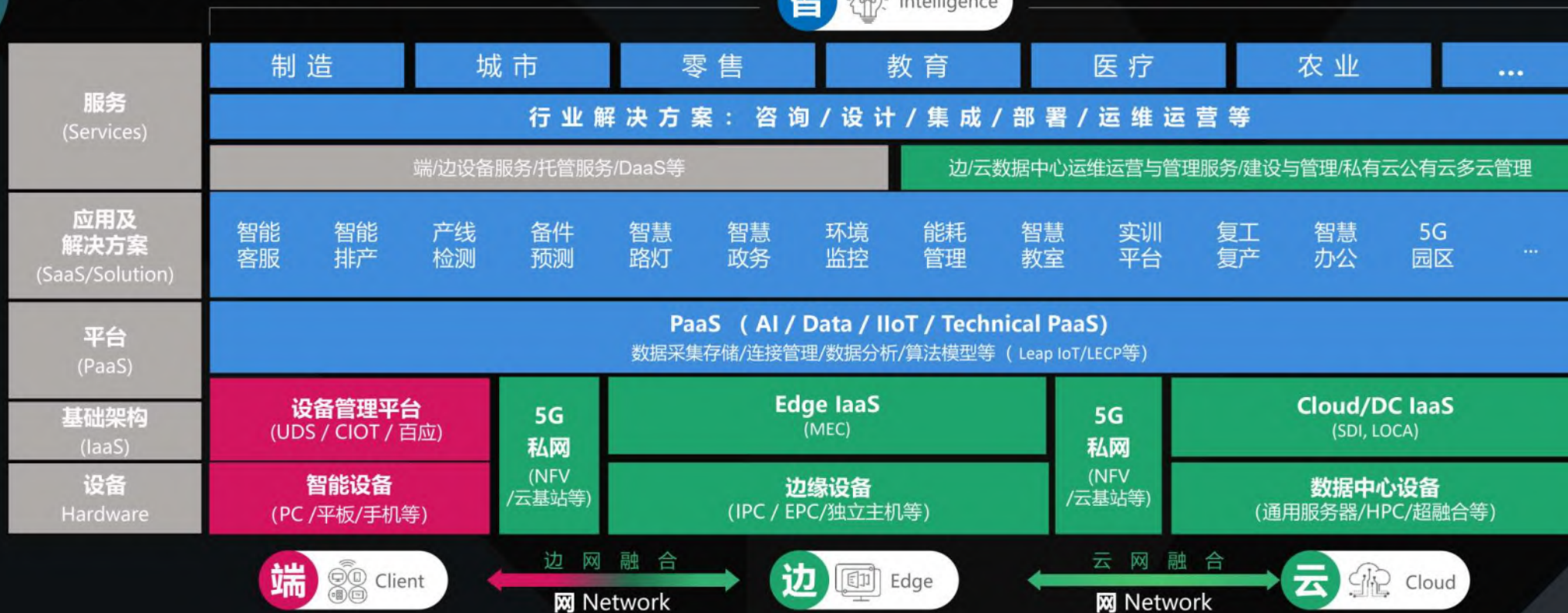
## 技术全景图

■ 智能物联网 ■ 智能基础架构 ■ 行业智能

智



Intelligence

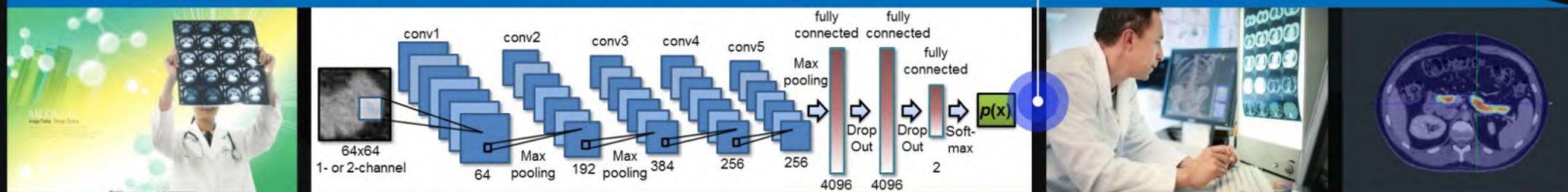


联想在医疗人工智能的实践

E-Health

Powered by  
深度学习一体系统

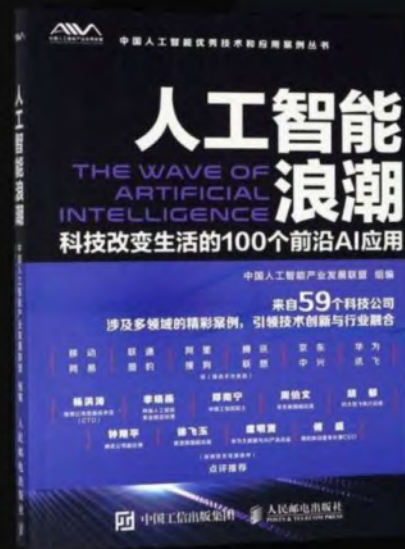
Efficient Assistant



Trainable High Accuracy



# 联想医疗人工智能平台 已 获 奖 项



## BigData Platform

APACHE  
**Spark**

**Flink**

## SaaS



ChatBot



Knowledge Graph



E-Health



Smart Assistant



Surveillance

ASR/NLU

CV

CA

Shared data

Typical models

Pretrained models

### Deep Learning Frameworks



Caffe



theano

mxnet

### Training Platform

Shared Storage

Scheduler

CMD UI

Web portal

### Clustered Resource Pool



Server



Storage



NV GPU



Phi



FPGA

Lenovo AI Appliance

## 医疗AI平台

# 基础设施的挑战

### 标准开源与作业调度

平台兼容性与高效的任务调度

### 融合高效的算力

CPU、GPU、计算芯片

### 高性能的存储力

提取、准备、训练、结果、归档

### 业务反哺

业务与科研数据的安全转变

# 联想提供全栈高效算力引擎

GPU是目前主流计算引擎



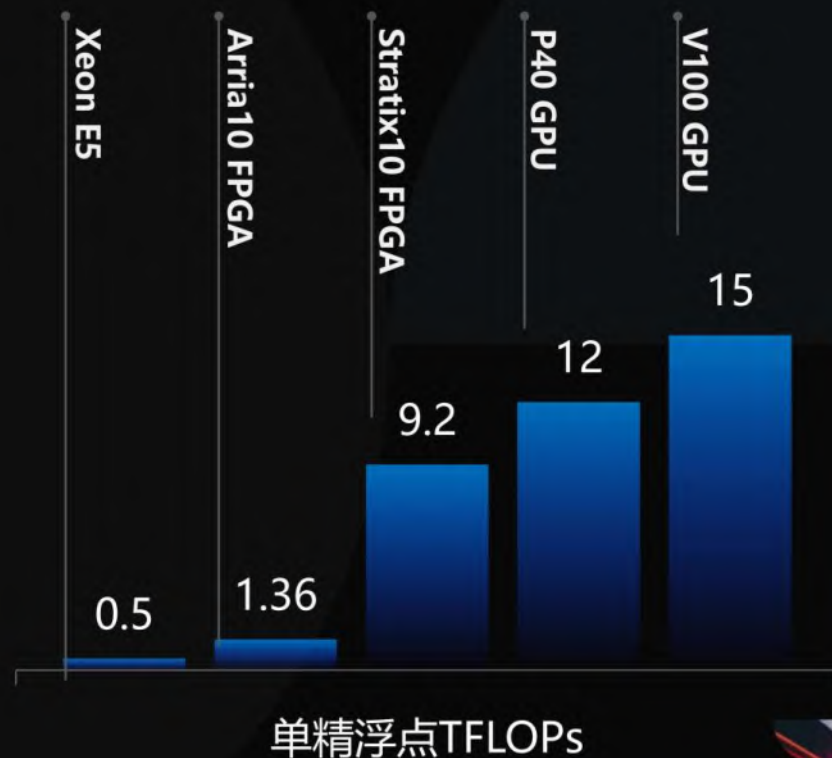
CPU用于控制和同步  
机器学习专用的Phi



FPGA在嵌入式解决方案



专用ASIC芯片  
寒武纪智能处理卡





# 超级算力 联想连续7年全球第一



**水冷机柜后门**

比风冷提高3.5倍的效能

节省的效能可提供4000年LED灯泡的动力

**智能控制软件**

动态调节CPU 浮点计算

最大限度减少能耗

**热转换模块**

比风冷系统提升24%处理能力

每个机架每年节省\$2,100

**水冷节点**

比风冷系统提升45%处理性能

降低 40% 电力成本

**Neptune™**  
液体冷却技术



## INGEST 数据提取

负载类型: 吞吐量优先  
协议类型: NFS、S3、HDFS、SMB  
部署方式: 边缘, 生产中心, 云  
涉及技术: NAS、对象存储、分布式文件系统、  
边缘数据分析

## PREPARATION 数据准备

负载类型: 海量文件 IO, 吞吐量, 海量标签  
小文件 IO  
协议类型: S3、NFS、HDFS、SMB  
部署方式: 生产中心, 云  
涉及技术: NAS、对象存储、数据湖、存储效  
率 (去重、压缩等)

## 模型训练 TRAINING

负载类型: 小文件随机IO、低延时; 大文件  
高吞吐量、高带宽; 元数据处理  
协议类型: NFS  
部署方式: 生产中心, 云  
涉及技术: NAS、SSD 闪存存储介质

## INFERENCE 模型推理

负载类型: 混合读写, 低延时  
协议类型: NFS、SMB  
部署方式: 生产中心, 云  
涉及技术: NAS、SSD 闪存存储介质

## RETAIN 数据保持

负载类型: 顺序读写, 吞吐量  
协议类型: NFS、S3  
部署方式: 生产中心, 云  
涉及技术: NAS、对象存储、公有云

面向医研数据

# 全周期定义数据

基于 NFS 的 NAS 协议是 AI 全流程最具普适性的存储协议

数据生命周期

# 业务数据与医研数据安全转换





# AI 数据性能与存储系统的关系

## 量变引起质变



### 小文件数据集

- 降低数据延时以便更快地向 CPU 或 GPU 提供数据
- 低延时存储介质
  - 内存 < SCM < SSD < 机械硬盘
- 低延时数据传输技术, RDMA over Fabric
- 低延时传输介质, 光纤



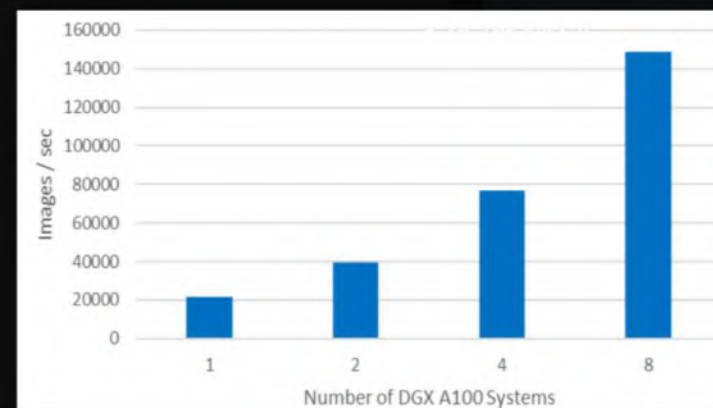
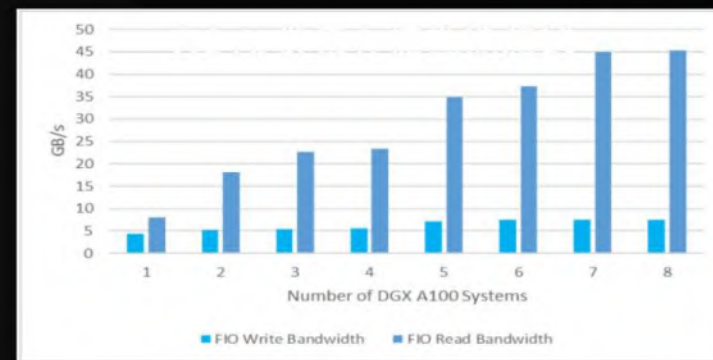
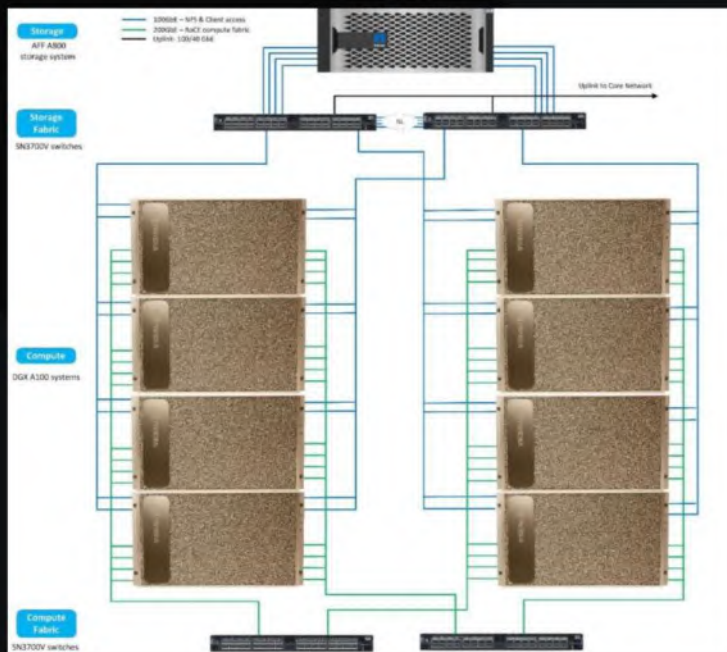
### 大文件数据集

- 增加数据带宽以便在单位时间内向 CPU 或 GPU 提供更多的数据
- 高性能并行文件系统, 提高数据吞吐量性能

提高 AI 系统性能的建议: **避免 CPU 或 GPU 等待数据**

降低 AI 性能成本的建议: **平衡延时和带宽之间的关系**

8 台 NVIDIA DGX A100  
NetApp 1 套2 节点 A800, 配置 48x1.92T NVMe SSD



存储性能实践案例

# 8 节点 GPU群集 联想凌拓存储性能测试





标准科研平台



业界领先算力

# 我们的优势



高效数据管理



安全快速的数据转换

我们期待  
与您的  
进一步  
交流...

展位地点: **B4展厅**

展位号: **BT33**





# thanks.

Lenovo + NetApp 联想



CHIMA2021会议  
未经授权请勿转载  
[www.chima.org.cn](http://www.chima.org.cn)