

Tech World '18

普“会”的人工智能产品

AI for Everyone

联想

Lenovo

罗军 | 2018.9.26

DeepNEX

深度学习平台

联想

Lenovo

深度学习的使用

- 搭建深度学习系统的过程：
 - ✓ 购买服务器：GPU选购、CPU及内存配置、磁盘阵列选购
 - ✓ 网络连接：网卡及交换机选择、网络拓扑设计
 - ✓ 配置操作系统：Linux OS, GPU驱动, CUDA及cuDNN安装
 - ✓ 深度学习框架安装：TensorFlow, MXNet, Caffe, Theano, PyTorch, Keras, ...
 - ✓ 应用开发、研究
- 多人使用的痛点：
 - ✓ 众多深度学习框架、不同版本不同依赖
 - ✓ 如何共享GPU等硬件资源？
 - ✓ 如何提供分布式数据存储与共享？
 - ✓ 加快速度：如何进行分布式训练？
 - ✓ 如何从繁杂的诸如运行环境运维等工作中解脱出来？

过程繁琐



谁在抢我的计算资源？

谁删了我的训练数据？

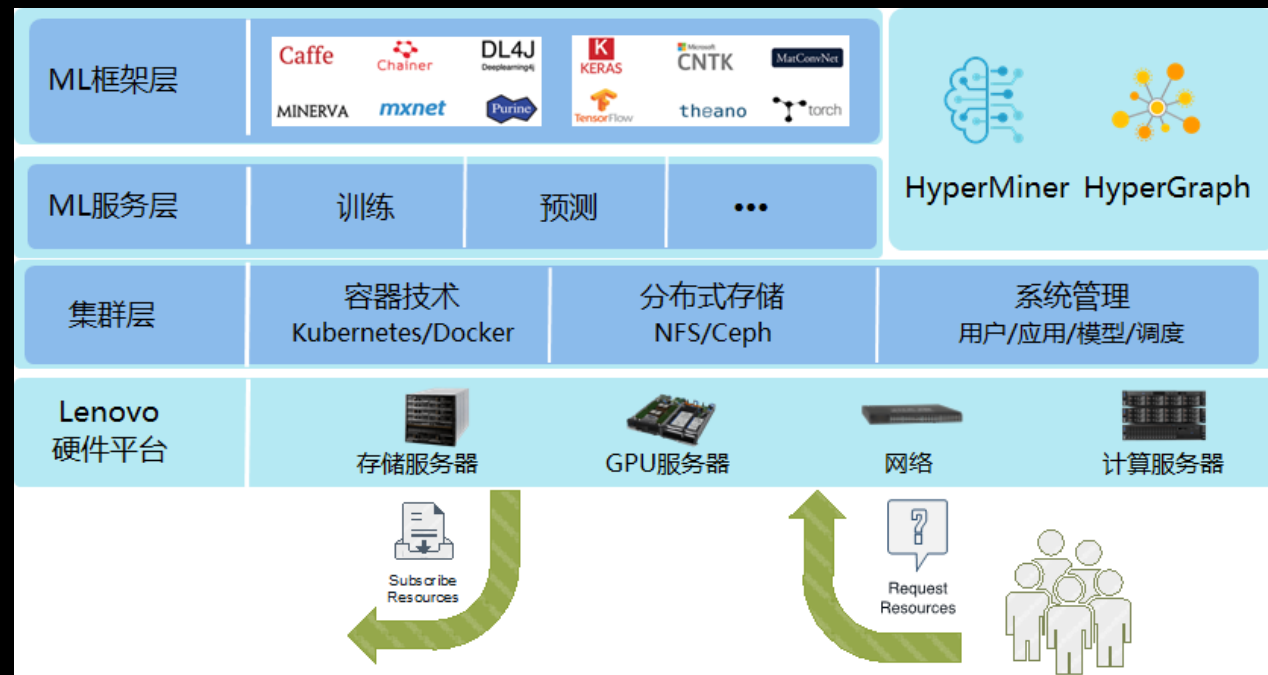
分布式训练怎么分配节点？

Tensorflow 依赖CUDA哪个版本？

重启之后Nvidia驱动不见了？！

DeepNEX

- DeepNEX平台特性
 - 支持异构服务器
 - 容器化平台：多租户、灵活资源分配
 - 支持分布式机器学习训练及预测
 - 支持海量数据的机器学习
 - 融合多个流行的深度学习框架
 - GPU资源调度：一卡多用、多机多卡等
 - 计算资源管理：支持资源预先分配和动态申请（排队、优先级）
 - 高容错、高扩展、高可用
 - 支持镜像定制、集群分发及删除
 - 提供平台监控功能



DeepNEX

**快速、低成本
构建
属于你自己的
深度学习应用平台**

DeepNEX Demo

HyperMiner

自动化机器学习平台

联想

Lenovo

定位：自动化机器学习平台

自动化机器学习平台 HyperMiner

一个多租户的机器学习云平台，支持多种机器学习模型和架构。

用户不需要太多机器学习相关的背景知识即可非常简单的在本平台上进行机器学习相关的应用开发。

流程自动化

自动特征工程，自动建模，自动调优



模型集成

支持常见机器学习模型



简单易用

简单的界面，丰富的解决方案

产品功能 - 行业解决方案库

The screenshot displays the 'Commodity Recommendation' workflow in the Lenovo AI solution library. The interface includes a top navigation bar with 'Lenovo', 'Profile', 'Design', 'Datasource', and a user profile 'Ryan Johnson'. The workflow consists of six steps: 1. Industry Template, 2. Upload data, 3. Check data type, 4. Choose ID & Target, 5. Running, and 6. Result & Evaluation. Below the workflow, users are prompted to select an industry template from five options: Finance & Insurance, Retail & Marketing, Public Security, Health & Medical, and Education. A detailed view of the 'Finance & Insurance' template is shown, describing a claim category prediction task. Below this, four specific solution cards are presented, each with a 'Select' button. A blue circular icon with a brain and a pencil is overlaid on the interface, pointing to the 'Automatic Model Optimization' feature description on the right.

Lenovo Profile Design Datasource Ryan Johnson RJ

CommodityRecommendation

1 Industry Template 2 Upload data 3 Check data type 4 Choose ID & Target 5 Running 6 Result & Evaluation

Please select the most appropriate Industry Template below:

Finance & Insurance Retail & Marketing Public Security Health & Medical Education

Predict the category of a claim based on features available early in the process, helping accelerate claims process. [learn more >](#)

Customer Segmentation of Wholesale Customers
Classification
In this experiment, we perform customer segmentation of wholesale customers.
Select

Demand Forecasting and Price Optimization
Regression
Utilizing historical transaction data to train a demand forecasting model. Pricing of products in a competing group is also incorporated to predict.
Select

Clustering: Find similar companies
Classification
This experiment clusters similar companies into same group given their Wikipedia articles and can be used to assign cluster to new .
Select

Demand Forecasting and Price Optimization
Regression
Utilizing historical transaction data to train a demand forecasting model. Pricing of products in a competing group is also incorporated to predict.
Select



解决方案优化

针对不同行业特点，对行业模板定制优化，同时也支持从通用方案模板创建项目



自动模型优化

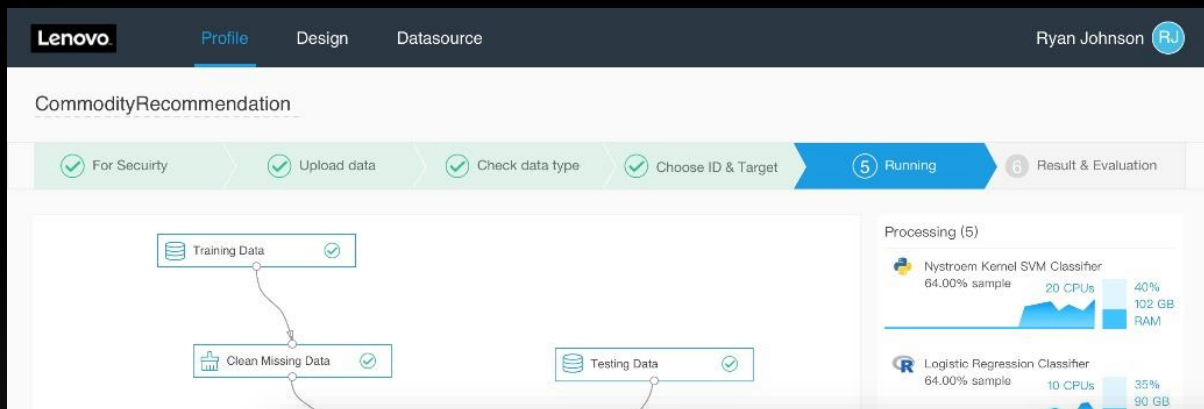
针对具体数据，自动进行模型选择，和模型参数调优



自动特征工程

针对不同行业特点，定制化的自动特征工程

产品功能 - 直观的操作流程



特征探查

特征数据预览, 特征类型查看, 统计特性展示, 数据分布可视化



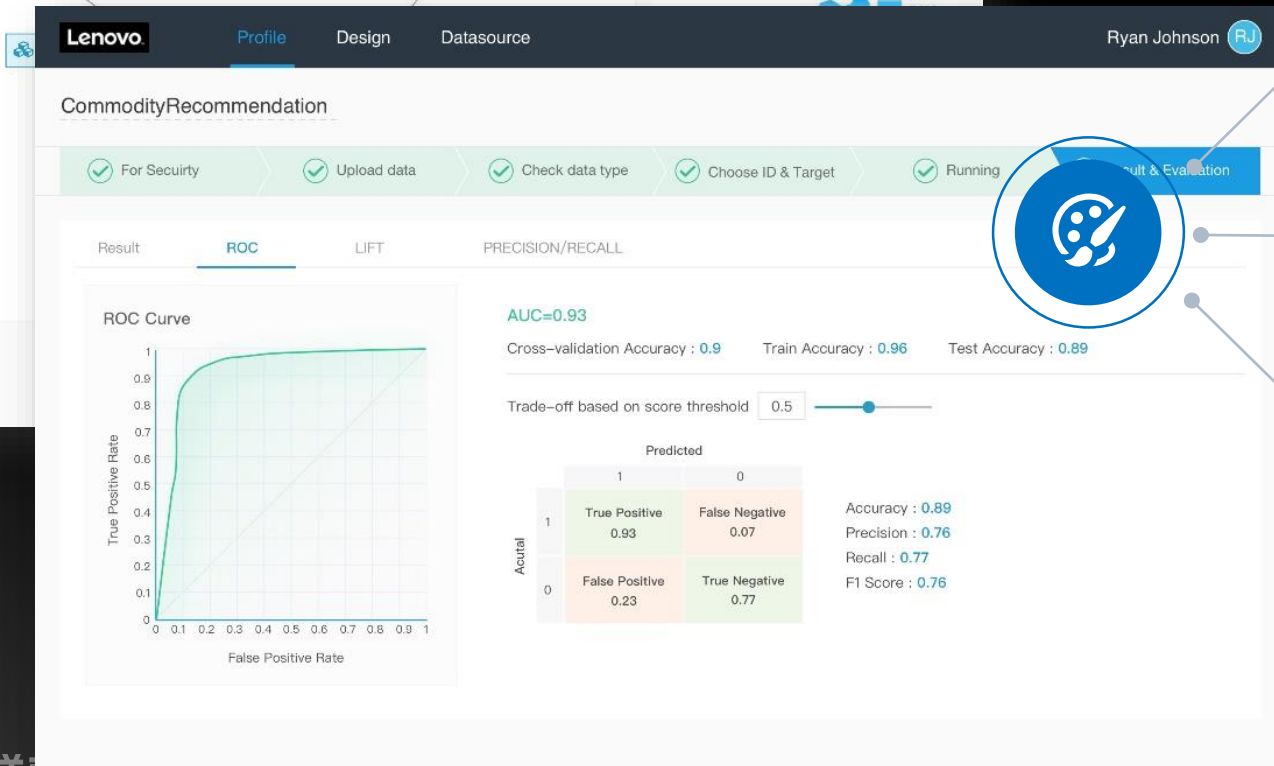
设置简单

可以采用模板默认参数, 也支持自定义参数, 向模型引入专家知识



可视化模型评估

模型评估结果可视化, 提供多种指标衡量模型效果

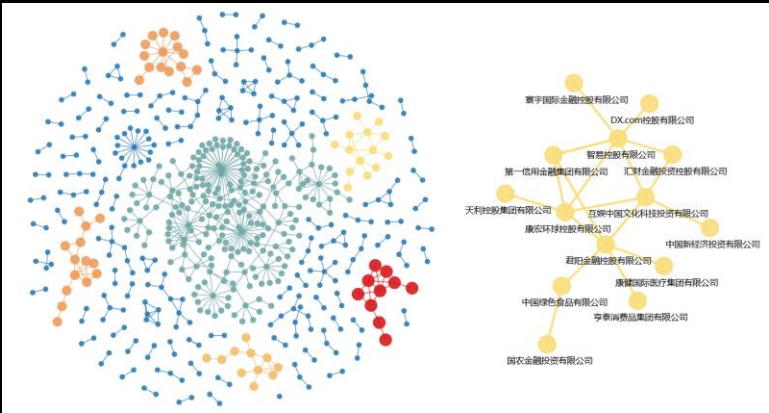
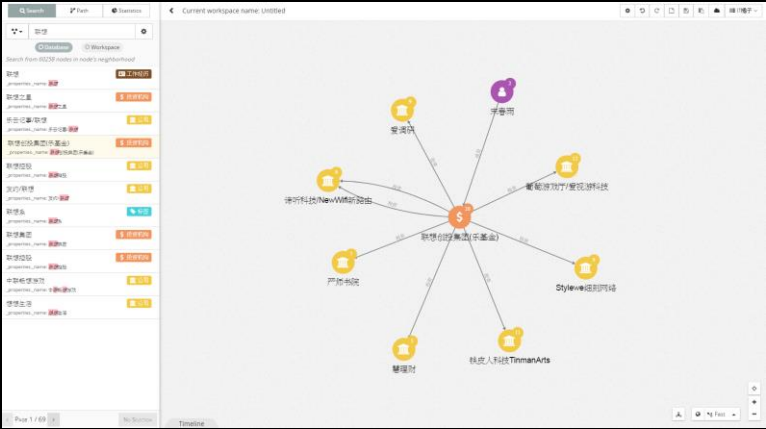
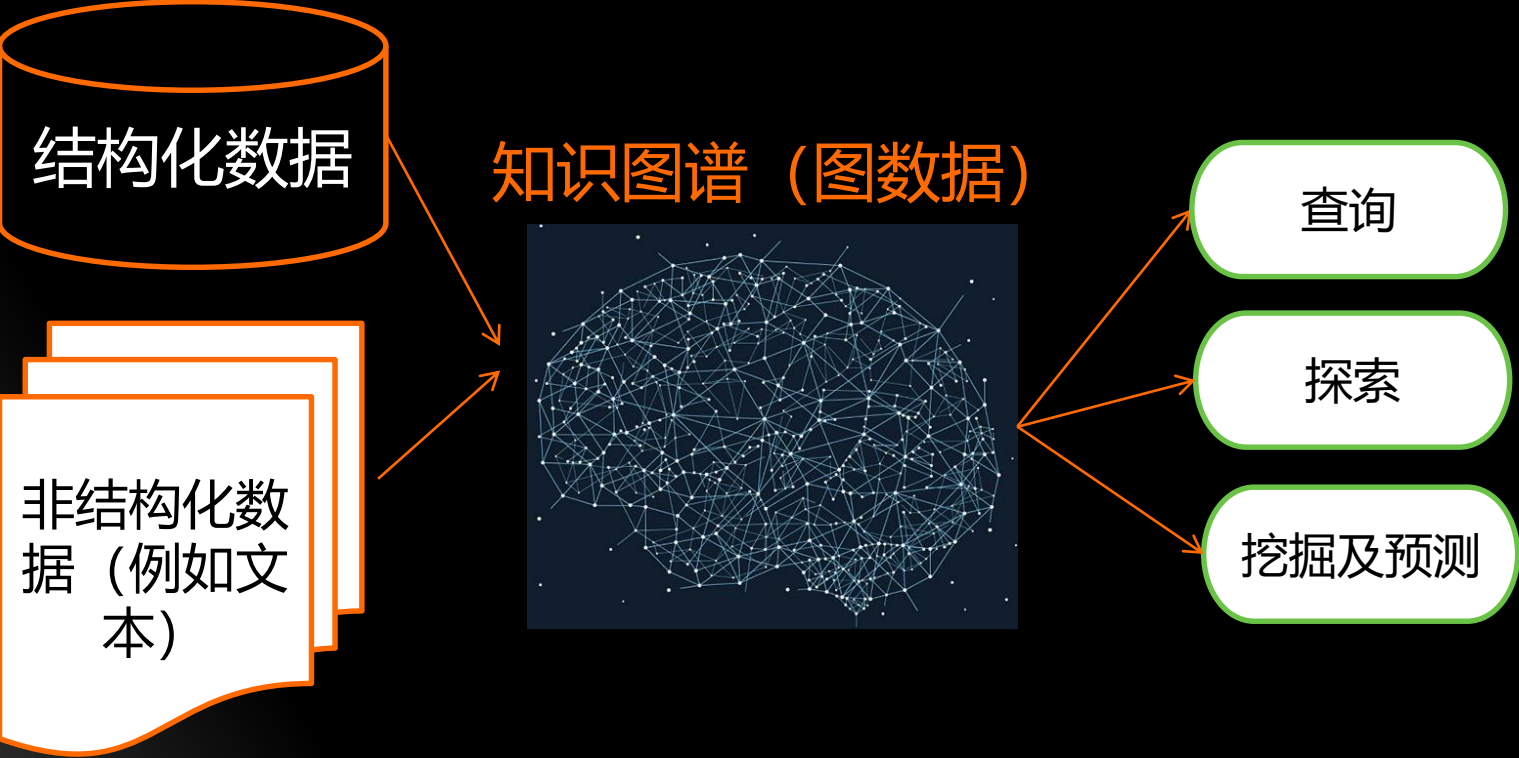


HyperMiner Demo

HyperGraph

大规模图数据平台

知识图谱工具



- 结构化、非结构化数据的关联表示
- 交互式数据查询及探索分析
- 基于图数据的数据挖掘及预测

目标一：大规模图数据交互式查询及探索分析

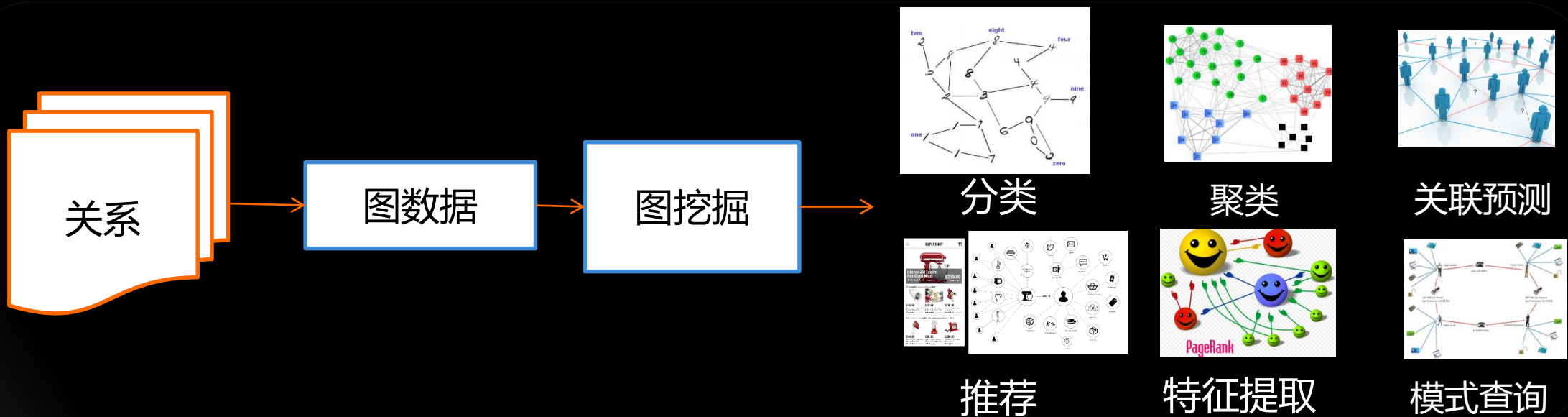
- 在**秒级别**对**百亿千亿规模**图数据进行分析的大数据平台，为企业用户提供一个可以不断拓展数据容量和维度、打通多源数据、自由探索和挖掘关联数据的工具

- **更易的操作性**：系统提供了包括数据导入、查询交互在内的一站式服务；
- **最佳的扩展性**：系统支持从单机到多机集群的扩展与部署，支持高性能秒级延迟的计算；
- **更快的交互性**：系统可以快速响应复杂数据可视化渲染需求，提供用户友好的交互方式；
- **更好的接口性**：系统提供多种API接口，用户可以根据自己的需求进行二次开发。



目标二：大规模图数据挖掘及预测

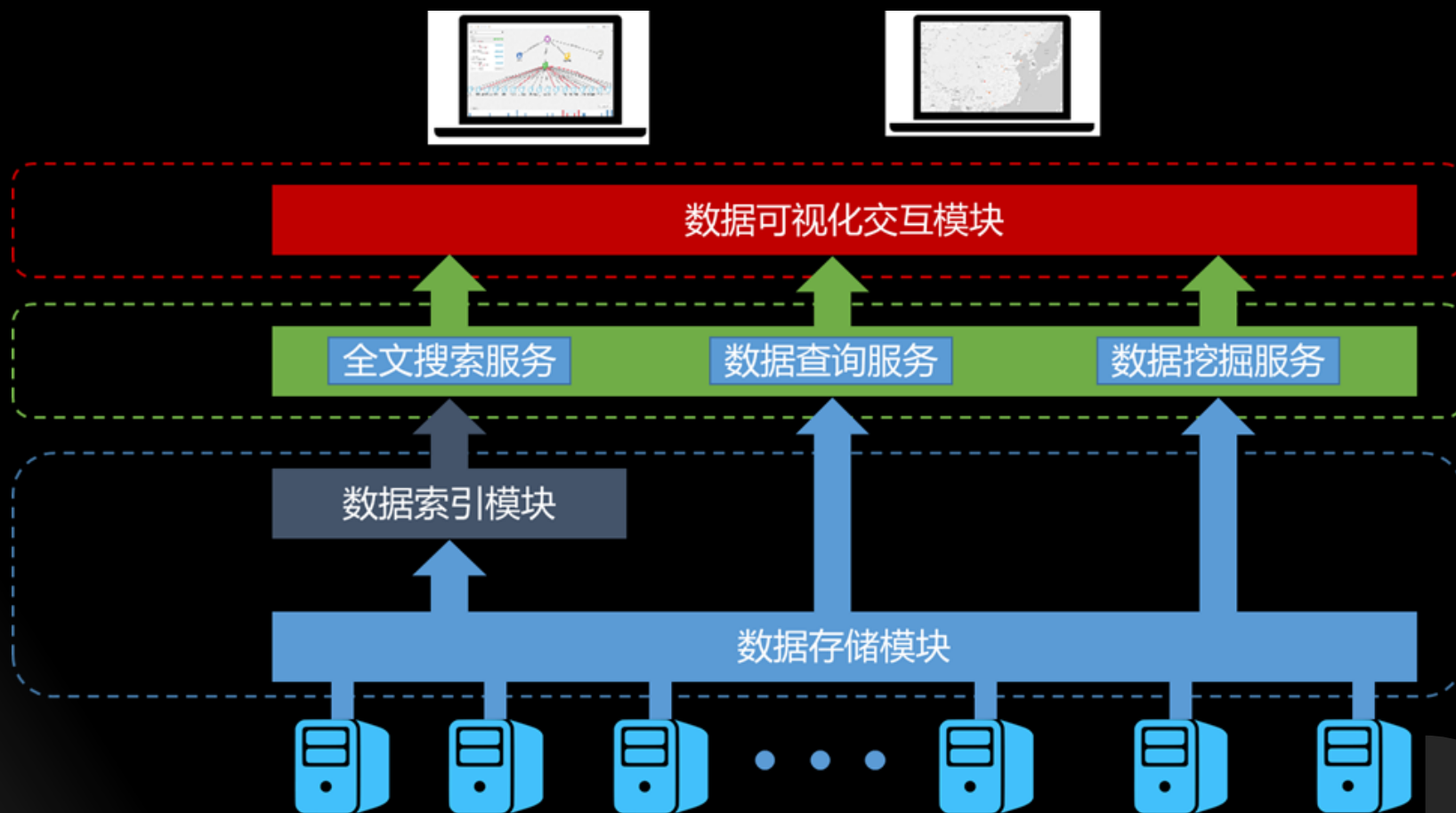
图挖掘



- 基于图数据的各类算法集API
- 基于图数据的数据挖掘及预测模块
- 基于图数据分析的行业解决方案案例

技术框架

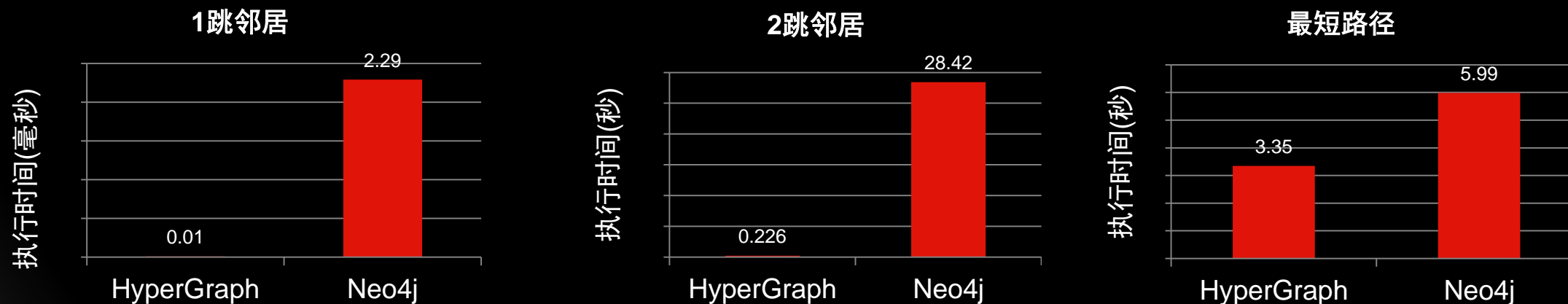
自主研发的图数据分析平台架构



同行对比-查询分析

■ 单机性能实验

- 数据集：DBPedia（百科数据）8 千万个结点，2.6 亿条边



• 分布式性能实验

- 数据集：HyperLink Graph（互联网网页链接图），35 亿个结点，1280 亿条边
- 硬件：5 台机器
- 对比方案：
 - Neo4j：数据集需要 10T 左右的存储空间，对单机硬件要求高，不适合普通的机器硬件
 - Titan(DSE)：耗时比较长，对于复杂的最短路径查询，基本不能得到结果
 - HyperGraph：可以在 10s 以内得到最短路径查询结果

HyperGraph Demo

Tech World '18

普“会”的人工智能产品

AI for Everyone

联想

Lenovo

罗军 jluo1@lenovo.com

thanks.

Lenovo 联想

Tech
World '18

