

从分布式追踪 到任务级能耗治理

—— DeepFlow 在数据中心算电协同中应用的思考

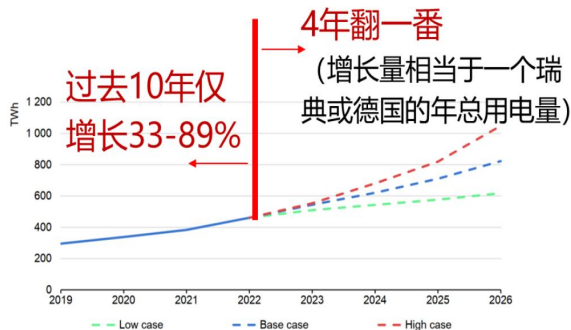
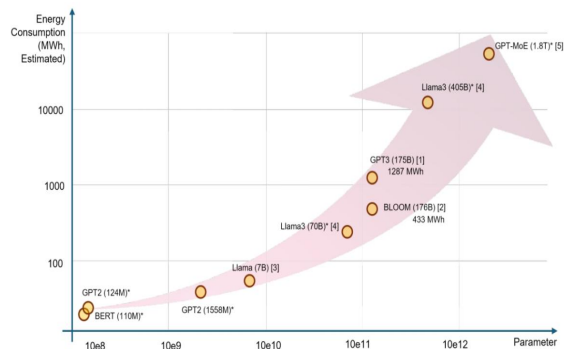
王梓炫 南京邮电大学 博士后

CONTENTS

- 01 数据中心算电协同的背景与意义
- 02 算电协同的场景与挑战
- 03 基于 DeepFlow 的任务追踪与能耗分析
- 04 总结与展望

01

数据中心算电协同的背景和意义



(一) AI驱动算力需求爆发式增长

- 全球AI算力需求年均增长率达**43%**，远超传统IT负载（8%），2030年AI算力用电占比将超**50%**（Citi Research, 2024）。
- 我国智能算力规模占比已达**25.4%**，同比增长**45%**，预计2025年提升至35%（中国信通院, 2024）。

(二) 数据中心能耗呈指数级攀升

- 大型AI模型训练能耗惊人：GPT-3训练耗电1287 MWh，相当于**120个**家庭年用电量（Nature, 2024）。
- 全球数据中心总耗电量4年翻番，增量相当于**瑞典全国年用电量**（IEA, 2024）。
- 我国数据中心年耗电达1500亿 kWh，**占全社会用电量1.6%**，三年复合增长率超30%。

(三) 能耗成为发展的核心瓶颈

- 数据中心PUE居高不下（平均>1.5），电力成本占运营总成本**56.7%**。
- 高功耗密度引发电网风险：单一集群部署10万块H100 GPU（约150MW）**可导致区域电网崩溃**（OpenPipe, 2024）。
- “能源危机”已成为AI行业公认挑战（OpenAI, 2024）

数据中心电力负荷时空灵活性来源(1)		
工作负载时空灵活性(1.1)	IT设备效率管理(1.2)	辅助设备效率管理(1.3)
时间灵活性: • 离线负载 空间灵活性: • 在线负载	目标: • 实现IT设备功率的瞬时改变 方式: • CPU频率调整 • CPU电压调整	目标: • 改变辅助设备功率 方式: • 制冷系统-控时恒温 • 储能系统-灵活充放

算力负荷的特殊性对电力系统构成严峻挑战

- 高功率密度的特点
- 高可靠性与绿色性要求
- 强波动性与周期性



提升需求侧运行灵活性，
缓解新能源消纳

可调节负荷参与电网调度，
获得收益，降低运营成本

算电协同是应对算力能耗危机、实现电力系统与算力产业共赢发展的有效路径。

数据中心与电力系统运行协同(2)		
时间灵活性(2.1)	空间灵活性(2.2)	促进可再生能源消纳(2.3)
超短期灵活性(秒级、分钟级) 调频 短期灵活性(小时级) 调峰 中长期灵活性(日级) 日前市场 长期灵活性(日级以上调节) 电网规划	配电网级空间灵活性 • 解决城市配电网阻塞 输电网级空间灵活性 • 减少电网输电管道投资冗余 • 提高电力系统需求侧灵活性	目标: • 提高可再生能源利用比例，绿色数据中心， • 互联网企业“双碳”目标 方式: • 工作负载时空调度 • 辅助设备灵活控制

协同算力负荷助力新型电力系统建设

- 算力负荷具备独特的时空灵活性
- 协同优化可显著提升系统经济性与绿色性
- 赋能新型电力系统建设



02

算电协同的场景与挑战

01 算电协同的场景

三个方向

方向1

考虑各种内部节能措施以减少能耗，降低能耗成本

实质上仍是在削减负荷，而不是转移负荷

方向2

考虑空间分布，利用电价在空间上的差异转移算力

以转移负荷为主，但该行为可能对电网造成影响

方向3

考虑算力转移对电网的影响，针对性地制定需求响应策略

以转移负荷为主，对数据中心需求响应的引导更为精准

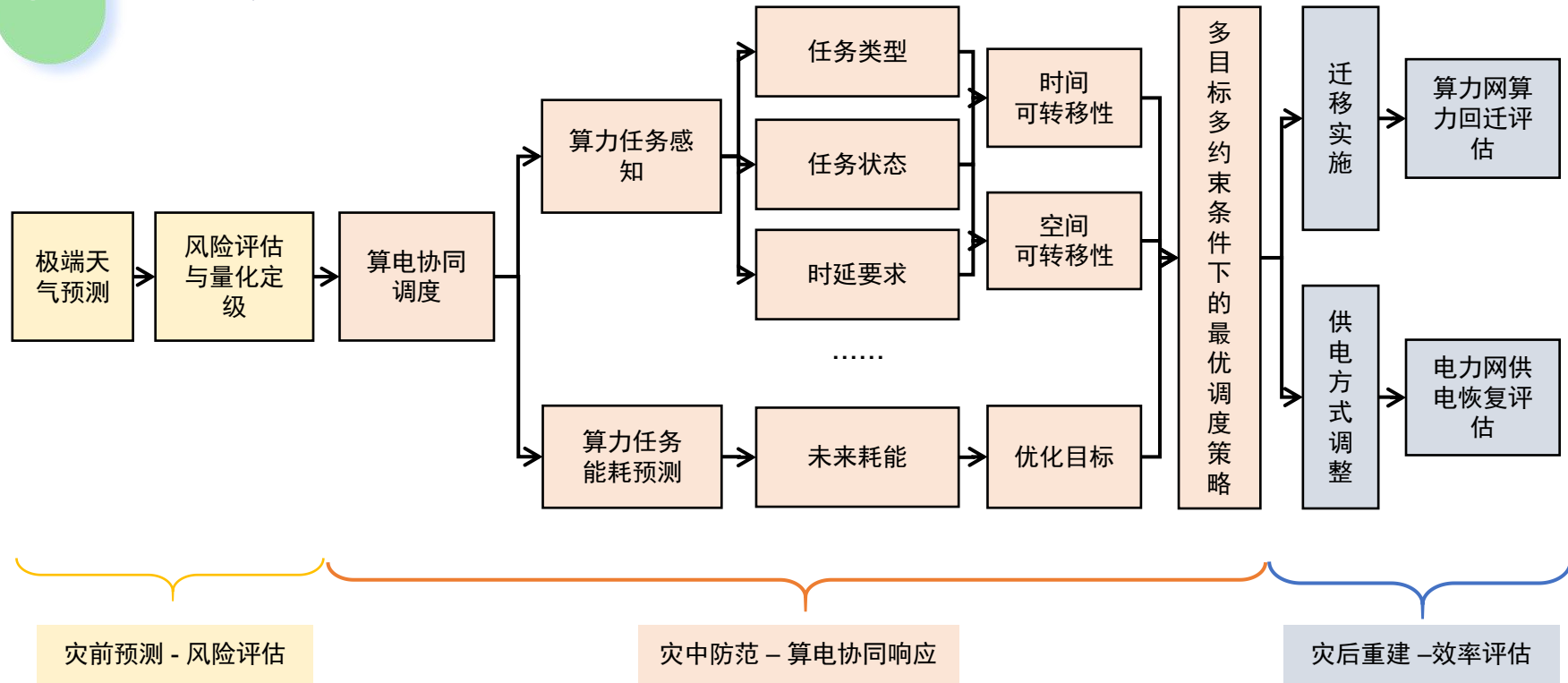
两个挑战

负荷时空迁移能力量化评估建模

多种约束下的联合优化调度框架及方法

02

一个算电协同的例子

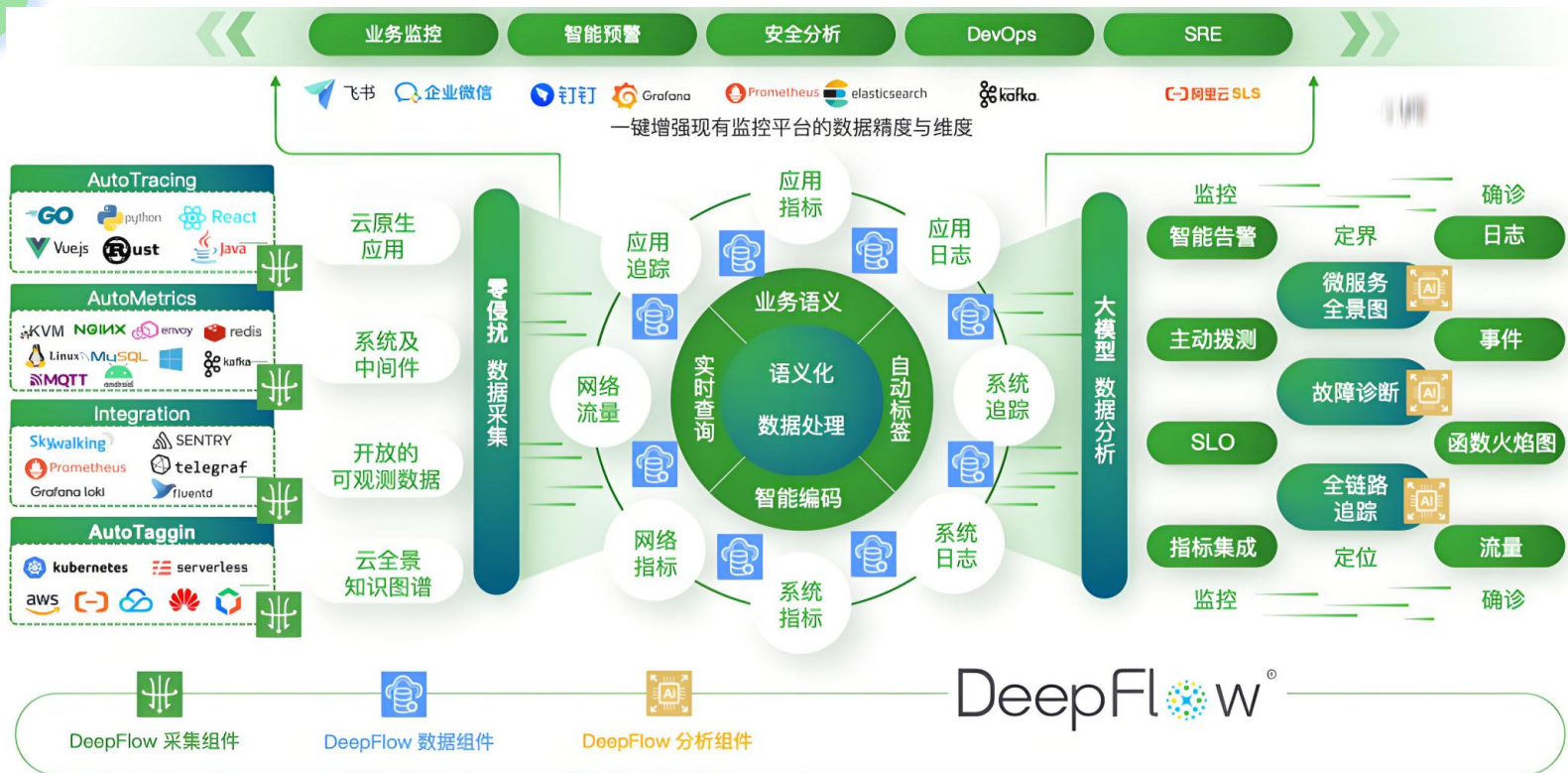




来源:《算力电力协同:思路与探索》白皮书

03

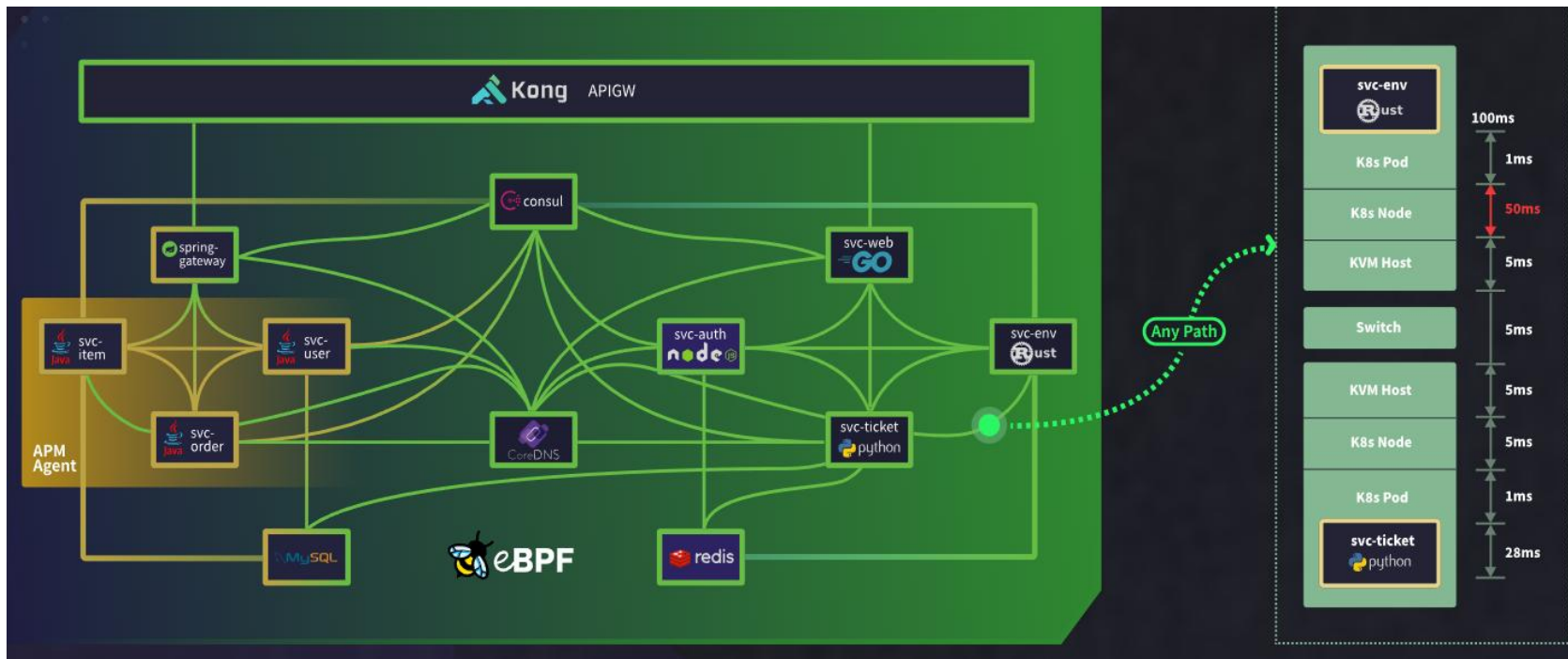
基于DeepFlow的任务追踪
与能耗分析



来源: deepflow官网 <https://deepflow.io/zh/>

02 DeepFlow的优势

任意 Service 的全景图



来源: deepflow官网 <https://deepflow.io/zh/>

03 DeepFlow的优势

任意 Request 的分布式追踪



来源: deepflow官网 <https://deepflow.io/zh/>

04 DeepFlow的优势

任意 Function 的持续性能剖析



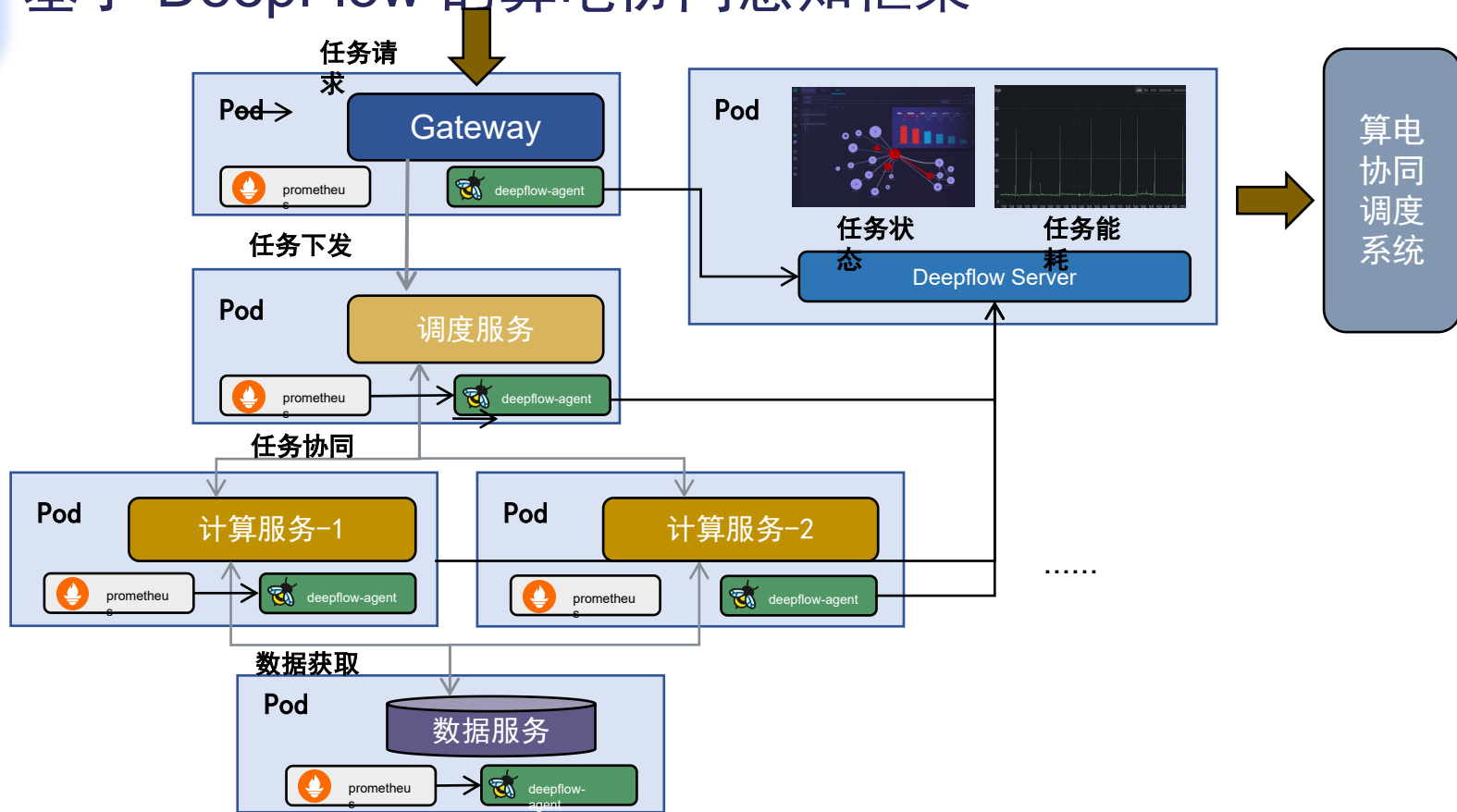
来源: deepflow官网 <https://deepflow.io/zh/>

05 DeepFlow的优势

无缝集成流行的可观测性技术栈

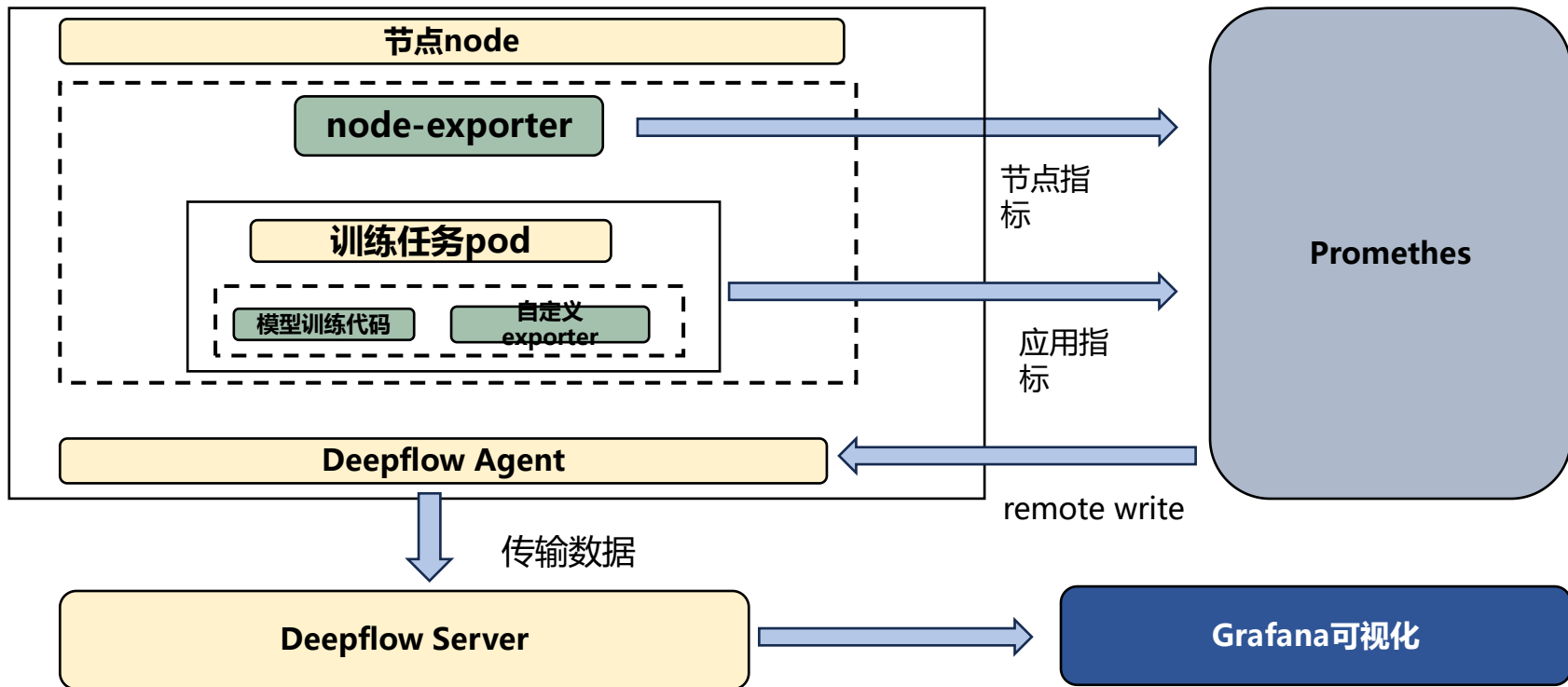


基于 DeepFlow 的算电协同感知框架



07 算电数据采集

Kubernetes集群



remotewrite: 将数据长期写入到远端存储

```
remoteWrite:
- url: http://172.31.15.216:38086/api/v1/prometheus
  replicas: 1
  resources:
    requests:
      memory: 400Mi
  ruleNamespaceSelector: {}
  ruleSelector: {}
  scrapeInterval: 30s
  securityContext:
    fsGroup: 2000
    runAsNonRoot: true
    runAsUser: 1000
  serviceAccountName: prometheus-k8s
  serviceMonitorNamespaceSelector: {}
  serviceMonitorSelector: {}
  version: 3.4.0
```

Deepflow远端写入地址

自动选择所有ServiceMonitor

为自定义exporter配置ServiceMonitor及Service

```
apiVersion: monitoring.coreos.com/v1
kind: ServiceMonitor
metadata:
  name: carbon-exporter-monitor
  namespace: monitoring
spec:
  selector:
    matchLabels:
      app: carbon-exporter
  endpoints:
    - port: metrics
      interval: 30s
      path: /metrics
```

ServiceMonitor的名称

与Service的标签匹配

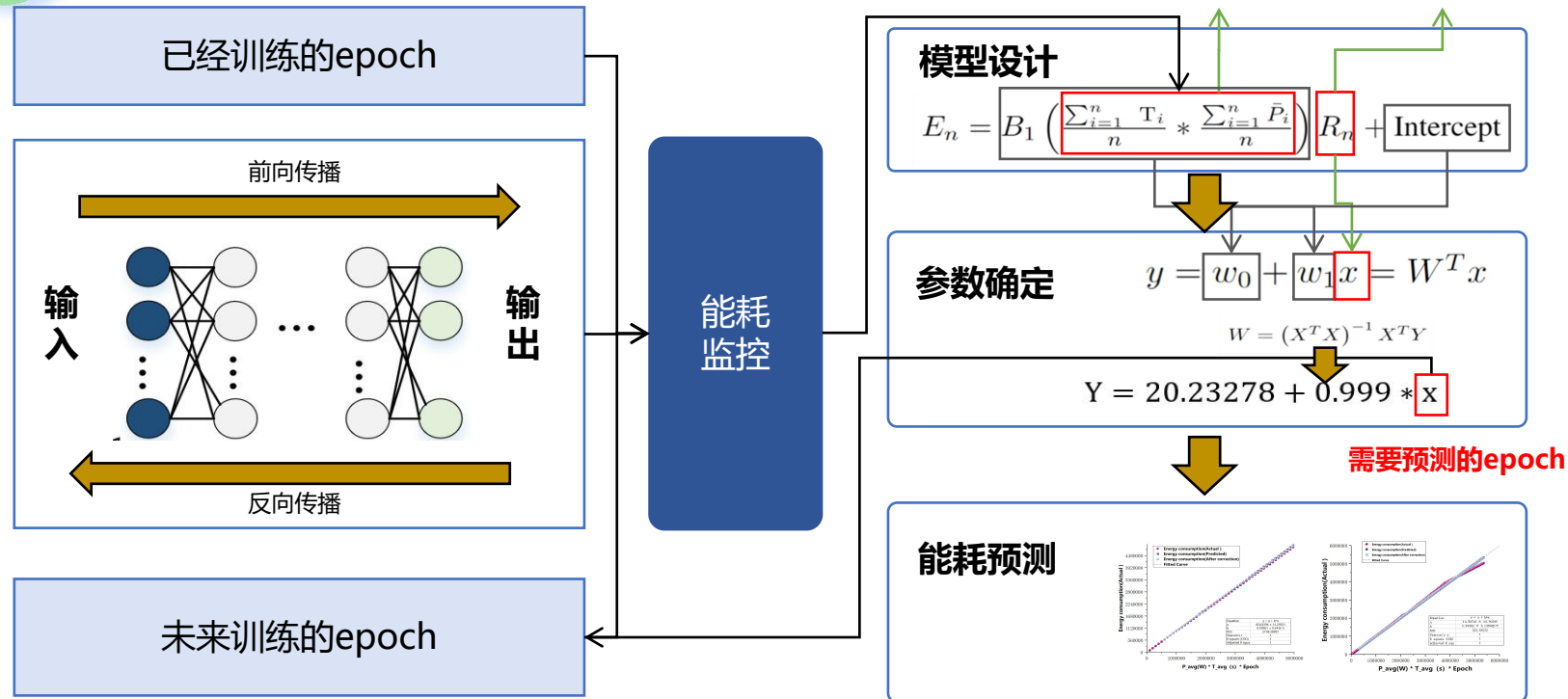
与Service中定义的端口名称匹配

```
apiVersion: v1
kind: Service
metadata:
  name: carbon-exporter-service
  namespace: monitoring
labels:
  app: carbon-exporter
spec:
  selector:
    app: carbon-exporter
  ports:
    - name: metrics
      port: 8000
      targetPort: 8000
```

Service的名称

与ServiceMonitor的selector匹配

与ServiceMonitor的port匹配



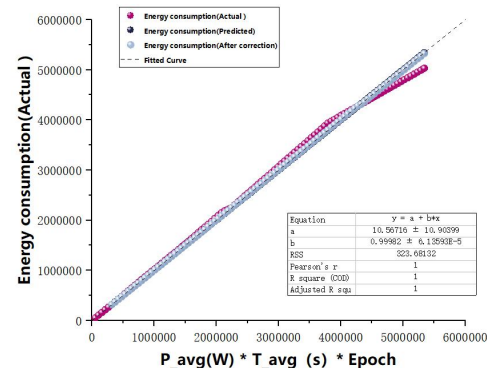
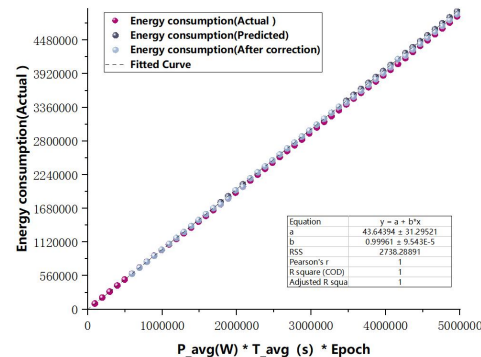
Wang Z X, Lu J R, Wang P, et al. Two-Stage Energy Prediction With Prior Estimation and Dynamic Adaptation for Symbiotic IOV[J]. IEEE Internet of Things Journal, 2025.

训练5轮预测未来50轮的能耗 误差小于0.6%

Epoch	P_avg(W)	E_total (J)	Predict E(J)	Error Rate(%)
1	115.2	96142.2		
2	118.6	195211.1		
3	123.1	294956.4		
4	123.9	396444.1		
5	120.8	495897.0		
6	114.5	595028.0	595018.0	0.00%
7	120.3	694643.5	694180.4	-0.07%
8	116.1	793277.9	793342.8	0.01%
9	117.9	893062.9	892505.2	-0.06%
10	109.8	986133.9	991667.6	0.56%
...	...	...	...	...
40	114.7	3889872.7	3921586.6	0.82%
41	110.0	3985943.3	4019625.2	0.85%
42	114.1	4082085.0	4165038.2	2.03%
43	118.0	4180969.9	4215702.3	0.83%
44	120.4	4282141.1	4313740.9	0.74%
45	122.3	4384080.0	4411779.5	0.63%
46	111.6	4477715.0	4509818.1	0.72%
47	117.4	4576519.2	4607856.7	0.68%
48	117.1	4673955.0	4705895.3	0.68%
49	117.6	4772043.5	4803933.9	0.67%
50	122.8	4873848.5	4901972.5	0.58%

训练5轮预测未来100轮的能耗误差小于6%

Epoch	P_avg (W)	E_total (J)	Predict E(J)	Error Rate(%)
1	89.9	52924.4		
2	90.0	107994.2		
3	91.8	160854.0		
4	92.1	214958.6		
5	90.6	267045.0		
6	93.5	320673.4	320451.9	-0.07%
7	90.5	374701.5	373858.8	-0.22%
8	91.7	428311.9	427265.7	-0.24%
9	89.6	484595.8	480672.6	-0.81%
10	86.9	540640.4	534079.5	-1.21%
...	...	...	...	...
90	103.3	4652520.9	4792619.1	3.01%
91	104.0	4688657.1	4845870.3	3.35%
92	107.2	4726145.1	4899121.5	3.66%
93	103.4	4764768.5	4952372.7	3.94%
94	106.6	4803873.7	5005623.9	4.20%
95	105.0	4840285.0	5058875.1	4.52%
96	108.1	4878188.1	5112126.3	4.80%
97	100.6	4915567.3	5165377.5	5.08%
98	105.5	4953535.4	5218628.7	5.35%
99	105.4	4991947.6	5271879.9	5.61%
100	106.5	5029581.4	5325131.1	5.88%



Wang Z X, Lu J R, Wang P, et al. Two-Stage Energy Prediction With Prior Estimation and Dynamic Adaptation for Symbiotic IOV[J]. IEEE Internet of Things Journal, 2025.

04

总结与展望

01 总结与展望



目前，团队正致力于基于 DeepFlow 构建一套“零侵入的算力-能耗一体化可观测体系”。

未来我们计划从2个方向深化探索：

- 扩展至多云 / 异构算力中心。
- 结合 LLM 进行智能调度与根因分析



THANKYOU

感谢观看

王梓炫 南京邮电大学