



新一代云数据仓库OushuDB的演进之路

用科技让人类只为兴趣而工作 www.oushu.io

CONTENTS

目录

1

偶数简介

2

AI和数据版图

3

OushuDB和LittleBoy

4

客户案例

偶数简介

1. 打造新一代AI和数据仓库平台, 让企业更智能
2. Apache HAWQ创始人及核心研发团队成员创立
3. 研发人员大多为Apache Committer & PMC成员, 来自EMC, Oracle, IBM, Teradata, Google等云计算和大数据公司
4. 团队成员大多毕业于北大、清华、CMU等国内外顶级学府, 以及多个ACM程序大赛奖牌获得者
5. 团队研究成果发布在SIGMOD等国际顶级学术会议上, 并拥有多项国际专利
6. 获得国际顶级投资机构红杉中国(A轮)和红点中国投资(天使轮)
7. 微软加速器.北京11期校友企业
8. 美国三大商业杂志Fast Company《中国最佳创新公司50》
9. 工信部第十六届开源 软件推进论坛《中日韩开源软件技术优胜奖》



Oushu Lava AI Cloud
AI云平台



Oushu LittleBoy
人工智能组件

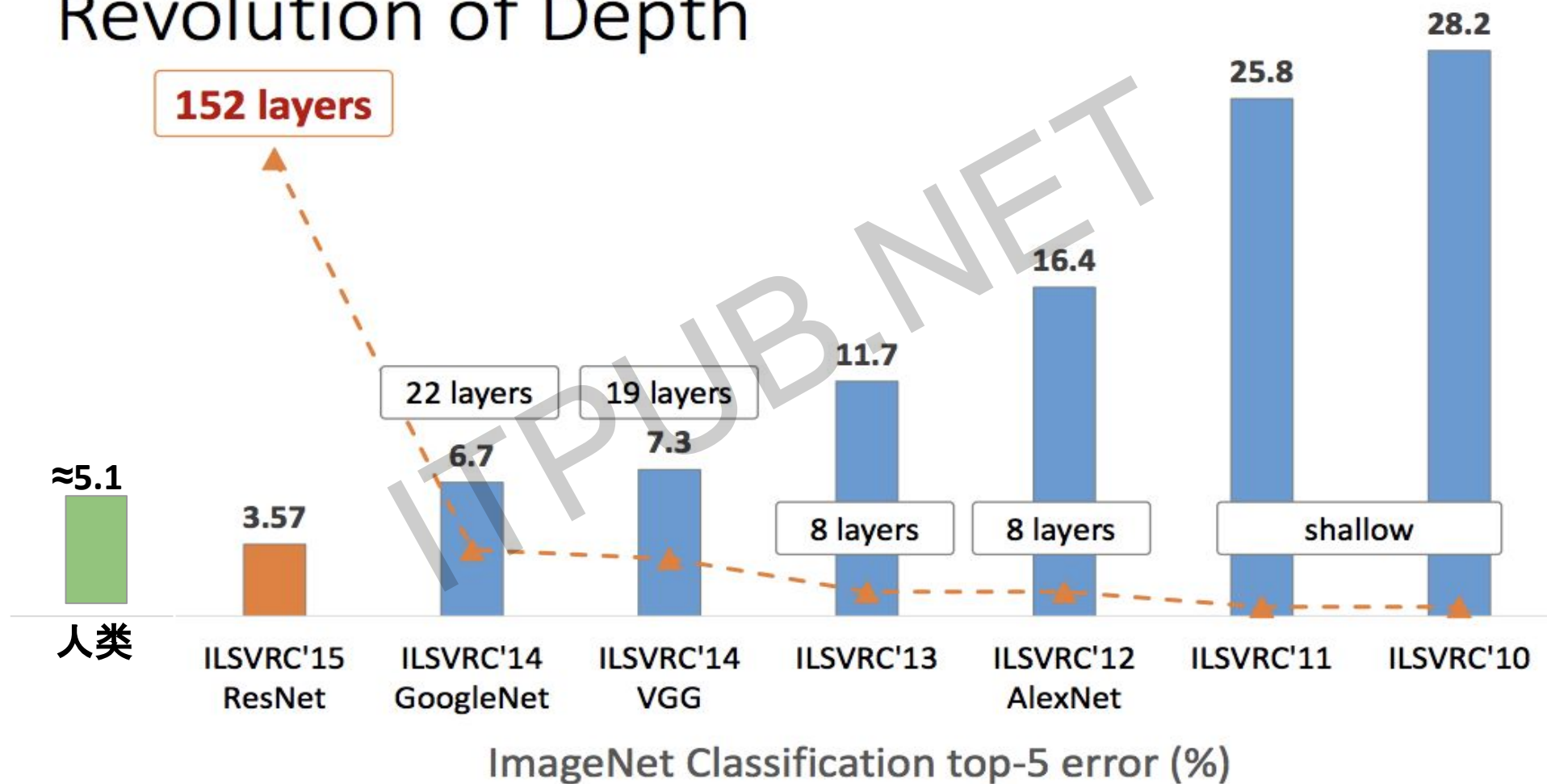


Oushu Database
世界上最快的新一代数据仓库



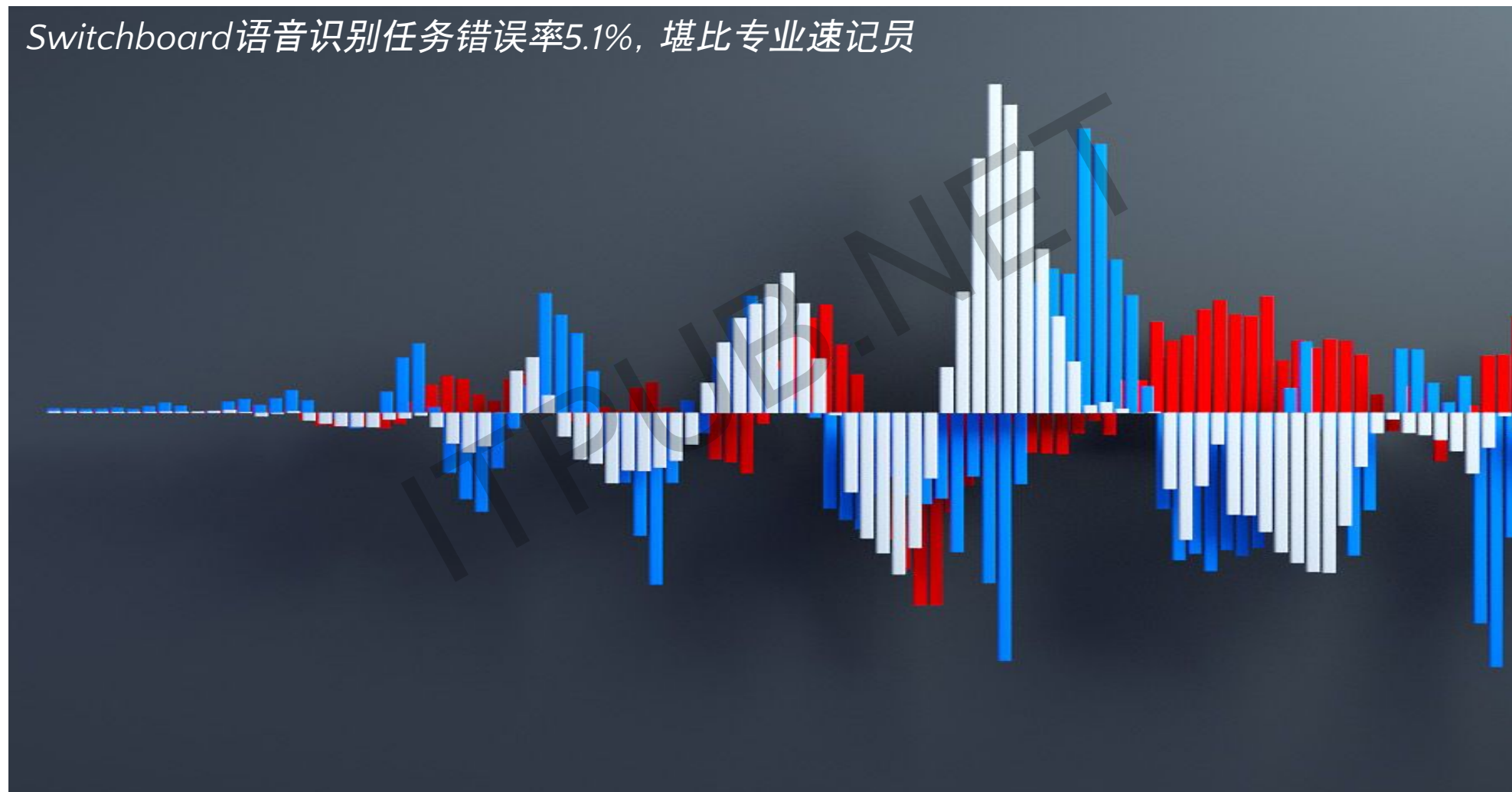
AI深度学习的技术进展

Revolution of Depth



AI深度学习的技术进展 (续)

Switchboard语音识别任务错误率5.1%，堪比专业速记员



AI版图

AI+行业

AI+安防

AI+金融

AI+医疗

AI+工业

AI+家居

机器人
自动驾驶

基础AI技术

计算机视觉

语音识别

自然语言处理

机器学习

数据平台技术

数据仓库

流处理

图数据管理

数据治理

AI硬件

CPU

GPU

FPGA

NPU

Sensor

人工智能大规模应用的挑战

AI与商业应用场景的探索



AI+行业人才短缺

AI研发和应用的门槛过高，体验较差



降低AI应用使用门槛，
提升体验

偶数AI产品及解决方案 (金融为例)



Oushu Lava AI云平台: 支持公有云和私有云



高易用性,
一键创建集群



云数仓服务



云AI服务



按使用量计费



零运维



- 控制面板
- OushuDB
- LittleBoy
- 集群管理
- AI工作室
- 流程管理
- 模型管理
- 用户管理
- 账单
- 系统报警
- 定价
- 文档
- 技术支持
- 审核
- 私有云版本下载

主页 / 控制面板



¥ 99885.67

账户余额



2

用户人数



1

集群数量



0

系统报警

集群信息

集群名称	节点数	状态	操作
dbtest-yao	1	Running	▶ 启动 ■ 停止 🗑 删除

用户信息

用户名	创建时间	上次登录时间	所属用户组	角色 ?	状态
oushu	2017-11-11 00:00:00	2018-07-31 22:27:16	无	oushuadmin	Active
parker@oushu	2018-07-29 12:39:03	2018-07-29 12:40:18	default	dbadmin	Active

OushuDB云数据仓库



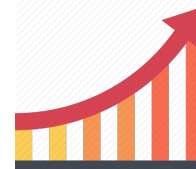
兼容SQL标准



PB级交互式查询



Hadoop生态



支持第三方工具

核心优势



极速性能
比同类产品快5-10倍



新一代架构
动态扩容



对AI无缝支持

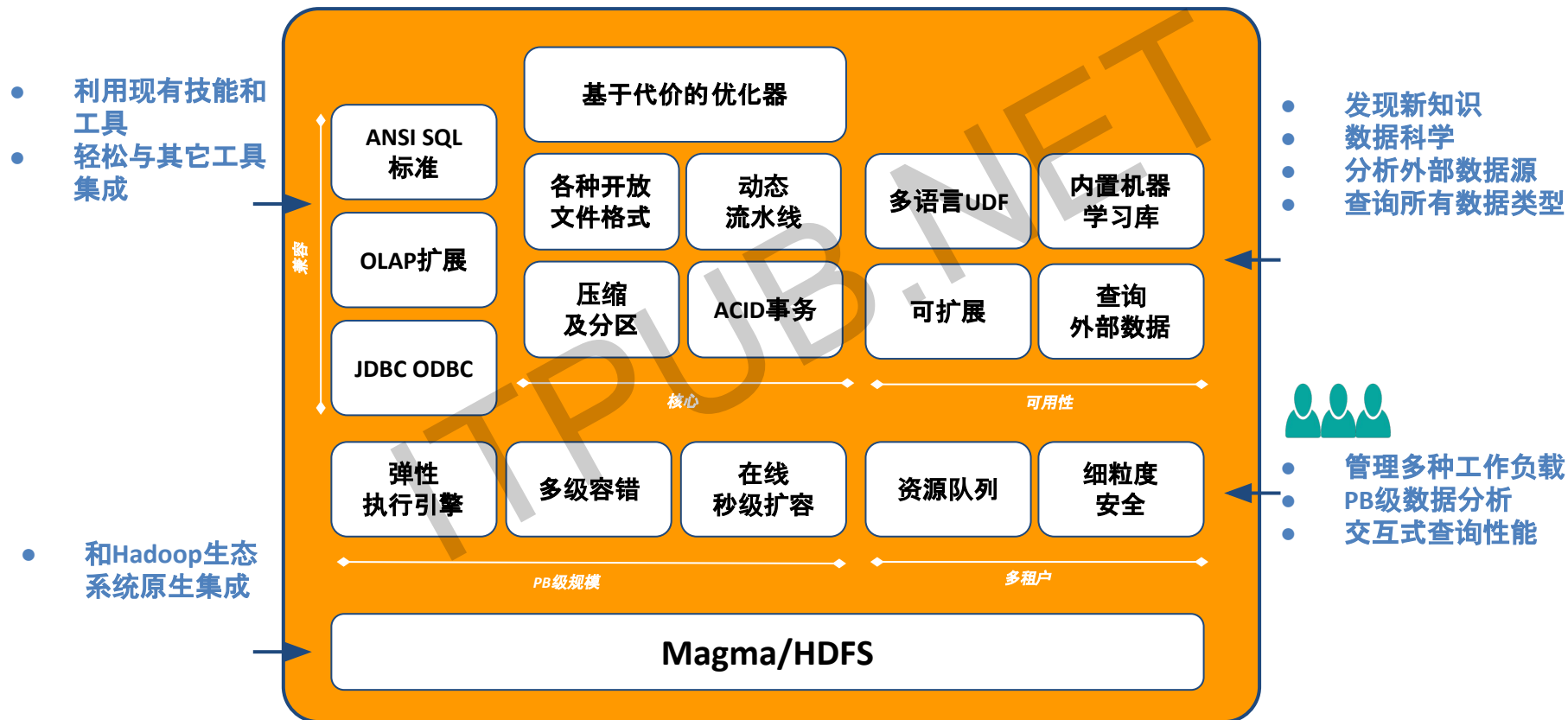


存储可插拔
支持行存和列存

数据仓库引擎比较



Oushu Database主要功能



OushuDB vs 其它数据仓库

特征	Oushu	GPDB	Teradata	Oracle Exadata	DB2	Impala	Hive	SparkSQL
为云平台优化	yes	no	no	no	no	no	no	no
性能	top	high	middle	middle	middle	middle	low	low
面向AI	yes	no	no	no	no	no	no	no
可扩展性	high	middle	middle	low	low	middle	high	middle
支持开放硬件平台	yes	yes	no	no	yes	yes	yes	yes
应用易迁移性	high	high	high	high	high	no	no	no
Share-Nothing架构	MPP++	old MPP	old MPP	no	no	old MPP	no	no
支持OLAP扩展	yes	yes	yes	yes	yes	partial	partial	partial
高级负载均衡	yes	yes	yes	no	yes	no	no	no
在线系统扩容	yes	no	no	no	no	yes	yes	yes
高级资源管理	yes	no	no	no	no	yes	yes	yes

OushuDB vs 其它数据仓库

特征	Oushu	GPDB	Teradata	Oracle Exadata	DB2	Impala	Hive	SparkSQL
基于代价的优化器	yes	yes	yes	yes	yes	weak	weak	weak
支持SQL2011	yes	yes	yes	yes	yes	no	no	no
高速互联协议	yes	yes	no	no	no	no	no	no
可插拔存储	yes	no	no	no	no	no	no	no
表分区	yes	yes	yes	yes	yes	yes	yes	yes
GP/Oracle兼容	yes	yes	no	yes	yes	no	no	no
PL/SQL等	yes	yes	yes	yes	yes	no	no	no
易用性	top	high	middle	low	low	low	low	low
自动弹性伸缩	yes	no	no	no	no	no	no	no
列式/行式存储	yes	yes	partial	partial	yes	yes	yes	no

OushuDB v3.0+

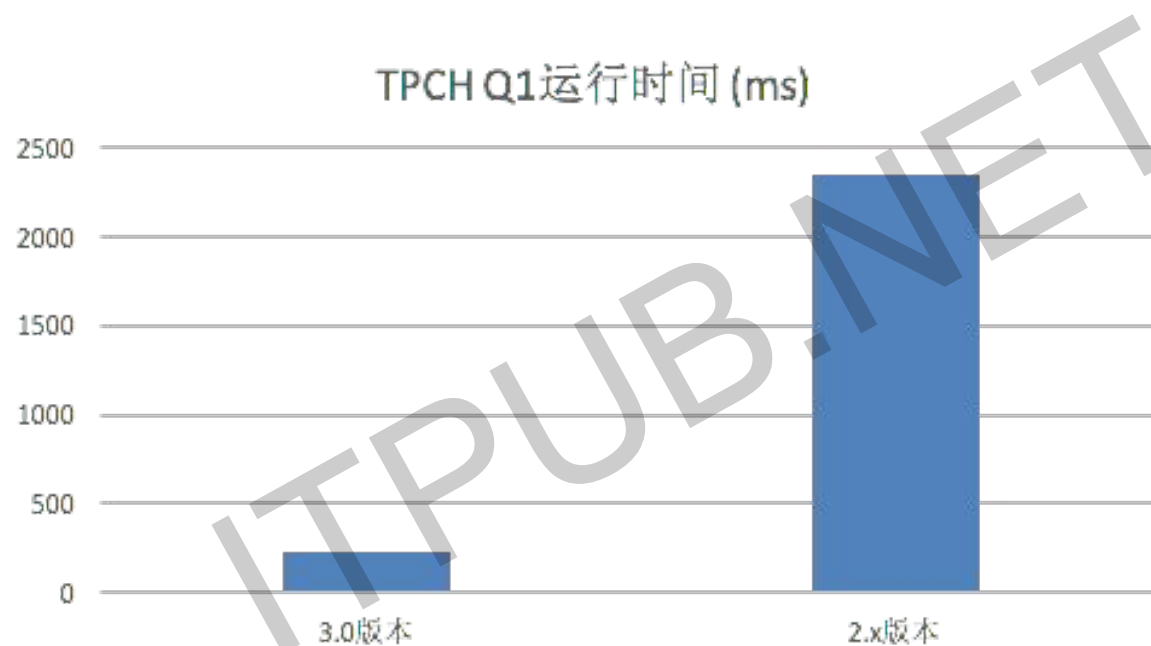
- 全新SIMD执行器，性能提升约10倍。
 - 满足客户SLA，实现交互式查询的一大步
- 支持ORC外部存储格式，外部存储性能提升10-50倍
 - 访问外部数据和内部数据性能类似，可节省客户1倍存储空间
- 支持新一代可插拔存储框架，
 - 用户可以自己快速扩展新的外部数据源，比如 S3等
- 提供了存储、查询和修改空间数据的能力
 - 对需要处理空间数据的客户很重要（比如：海军，空军等）

10 倍左右性能提升

世界上最快的执行器



3.0版本性能提升:TPCH基准测试



OushuDB基于SIMD的新执行器秒杀其他产品

单位 (毫秒ms)	OushuDB	Spark	Ratio
select l_orderkey, count(*) from lineitem group by l_orderkey;	14314.14	OOM	NAN
select l_partkey, count(*) from lineitem group by l_partkey;	4127.98	29299	7.10
select l_suppkey, count(*) from lineitem group by l_suppkey;	1142.61	18181	15.91
select l_linenum, count(*) from lineitem group by l_linenum;	363.51	9570	26.33
select l_quantity, count(*) from lineitem group by l_quantity;	370.15	11367	30.71
select l_extendedprice, count(*) from lineitem group by l_extendedprice;	4929.78	29736	6.03
select l_discount, count(*) from lineitem group by l_discount;	392.41	10371	26.43
select l_tax, count(*) from lineitem group by l_tax;	352.99	10371	29.38
select l_shipmode, count(*) from lineitem group by l_shipmode;	630.63	11371	18.03
select l_linestatus, count(*) from lineitem group by l_linestatus;	329.30	11217	34.06
select l_shipdate, count(*) from lineitem group by l_shipdate;	638.51	16077	25.18
select l_commitdate, count(*) from lineitem group by l_commitdate;	642.31	16161	25.16
select l_receiptdate, count(*) from lineitem group by l_receiptdate;	647.12	15649	24.18
select l_shipinstruct, count(*) from lineitem group by l_shipinstruct;	823.09	11539	14.02
select l_shipmode, count(*) from lineitem group by l_shipmode;	630.63	11371	18.03
select l_comment, count(*) from lineitem group by l_comment;	39032.16	OOM	NAN
AVERAGE(除去spark OOM 语句)	1138.30	15161.07	21.66

34.06倍

平均快15-20倍

不同数据类型的列的分组聚集

标准及第三方工具兼容

- SQL-92, SQL-99, SQL-2003, OLAP扩展, 与Postgresql和GPDB语法兼容
- 标准JDBC, ODBC连接
- 存储过程 (PL/pgSQL, PL/python et al)
- 支持与Apache Hadoop兼容的任何发行版本
- 支持各大BI工具
- 支持informatica等主流ETL工具
- Hadoop标准ODPi发起者之一



LittleBoy人工智能



数据加载



特征工程



自动模型训练



模型发布

核心优势



零门槛
无需AI背景



AutoML
自动调优



深度学习
准确率高



行业应用广泛



大规模
分布式训练

控制面板

» OushuDB

» LittleBoy

集群管理

AI工作室

流程管理

模型管理

用户管理

账单

系统报警

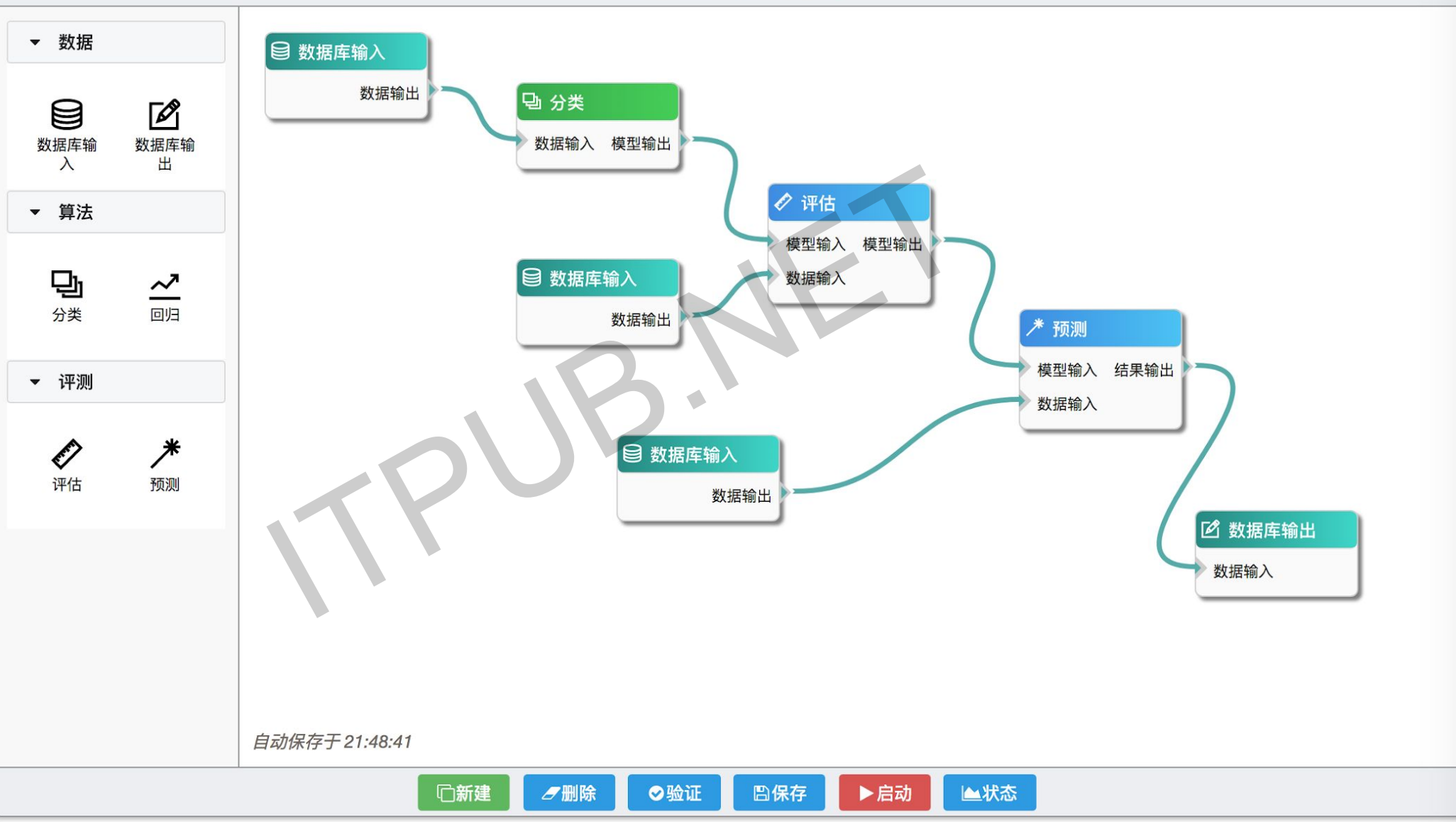
\$ 定价

文档

技术支持

私有云版本下载

AI工作室 信用卡反欺诈模型



某银行反洗钱AI应用

项目背景

- 各金融机构按照央行规定的可疑交易标准报告可疑交易
- 该银行的反洗钱规则引擎可将交易进行过滤和标记,并推送给业务人员进行调查、判断和确认。
- 反洗钱规则引擎被标记的总案件量较多,业务人员在此种情景时下“大海捞针”,往往要耗费大量的时间和精力。
- 在本次的测试环境中统计发现:每天触发规则引擎的账户约225,而最终认定并成功报送的大约2户,成功报送比例约1%。

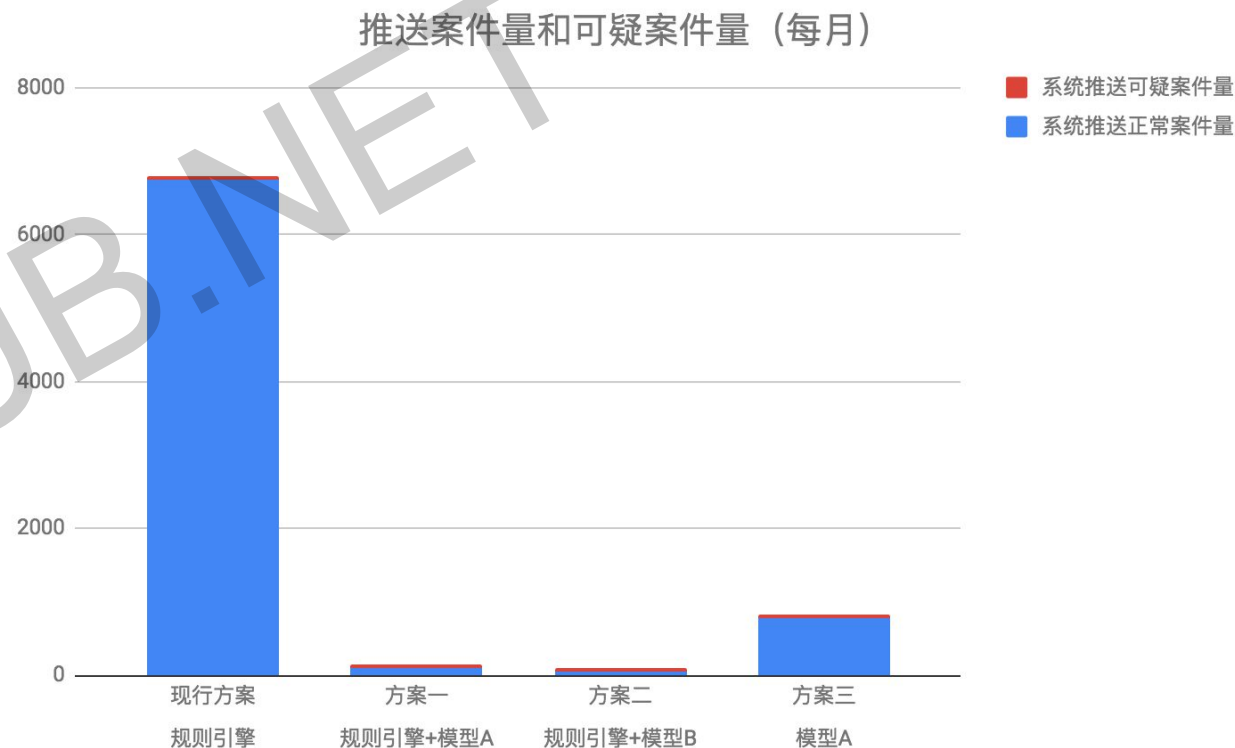
项目目标

- 1) 减少人工参与
- 2) 挖掘更多特征
- 3) 探索创新应用



项目成果

- 通过AI建模, 给出多个可行模型方案
 - 现有方法: 查全率100%, 查准率约1%
 - 偶数模型A: 查全率100%, 查准率51.43%
 - 偶数模型B: 查全率86.11%, 查准率92.08%
- 可以**上百倍**的减少推送可疑案件量和人工排查工作量





偶数科技

让数据更简单，让企业更智能