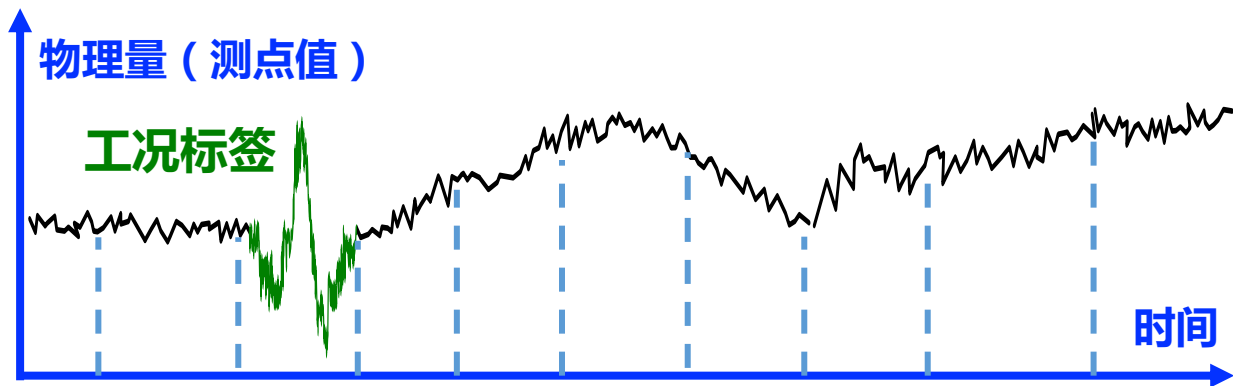


**面向工业互联网的高性能轻量级  
时序数据库系统软件 IoTDB**

# 工业物联网时序数据管理是工业互联网应用落地的基础

工业物联网时序数据是工业设备物理量的数字化记录，是带时间戳的数据，蕴含丰富的工业语义

某风电场001号风机风叶处实时风速记录



← 内涵 价值 ↓

国际工业互联网领导者通用电气 (GE) 公司2012年指出：“充分利用海量时序数据驱动工业创新、竞争和成长，是大数据技术为新工业革命带来的历史性机遇。”

The Rise of Industrial Big Data



## 海量序列存储

据IEC61400-25标准，风机每年运行7500小时将产生 6TB 数据

## 复杂元数据管理

C919试飞中有超过8万测点；测点数量等元数据随设备升级而改变

## 丰富的查询需求

- 数据降采样
- 数据空值填充
- 数据乱序处理

## 边云协同

- 端侧数据管理能力
- 云侧数据分析能力

挑战 →

# 面向工业互联网的高性能轻量级时序数据库系统软件

Apache  
IoTDB

“清华数为”  
IoTDB

2011年  
863计划攻关

2017年3月  
在Github公开

2018年11月经投票  
进入Apache孵化器

2020年9月经投票  
晋升为Apache顶级项目

始于中国高校、历练于工业用户、  
成长成熟于开源社区

C. Mohan, 数据库事务恢复技术奠基者, 美国工程院院士  
“Apache IoTDB是中国高校首个达到国际顶级开源标准的数据库项目, 具有美好前景”



The screenshot shows the Apache IoTDB website homepage. At the top, there is a navigation bar with links for Document, System design, Download, Community, Development, Blog, ASF, Languages, and GitHub. The main heading is "Apache IoTDB" with the subtitle "Database for Internet of Things". Below this, there is a brief description of the database's capabilities and a "Download" button.

The screenshot shows a blog post from the Apache Software Foundation. The title is "The Apache Software Foundation Announces Apache® IoTDB™ as a Top-Level Project". The date is "WEDNESDAY SEPTEMBER 23, 2020". A red box highlights the text "Apache顶级项目官方认证". The main text of the blog post states: "Open Source Internet of Things-native database integrates with the Apache Big Data ecosystem for high-speed data ingestion, massive data storage, and complex data analysis in the cloud, in the field, and on the edge." Below this, there is a quote from Prof. C. Mohan.

用户覆盖德国、美国、中国一批工业互联网厂商和智能制造厂商

# 产品形态：灵活适配“端-边-云”计算环境

部署在嵌入式终端设备的时序  
“数据文件”



**终端（端）**

**为时序数据而生的zip文件**  
支持高性能写入，高压缩比存储，支持简单查询

部署在工控机等边缘计算设备的  
时序“数据库”



**场控（边）**

**高效丰富的时间序列查询引擎**  
提供增删改查，以及聚合查询  
时序对齐等高级功能

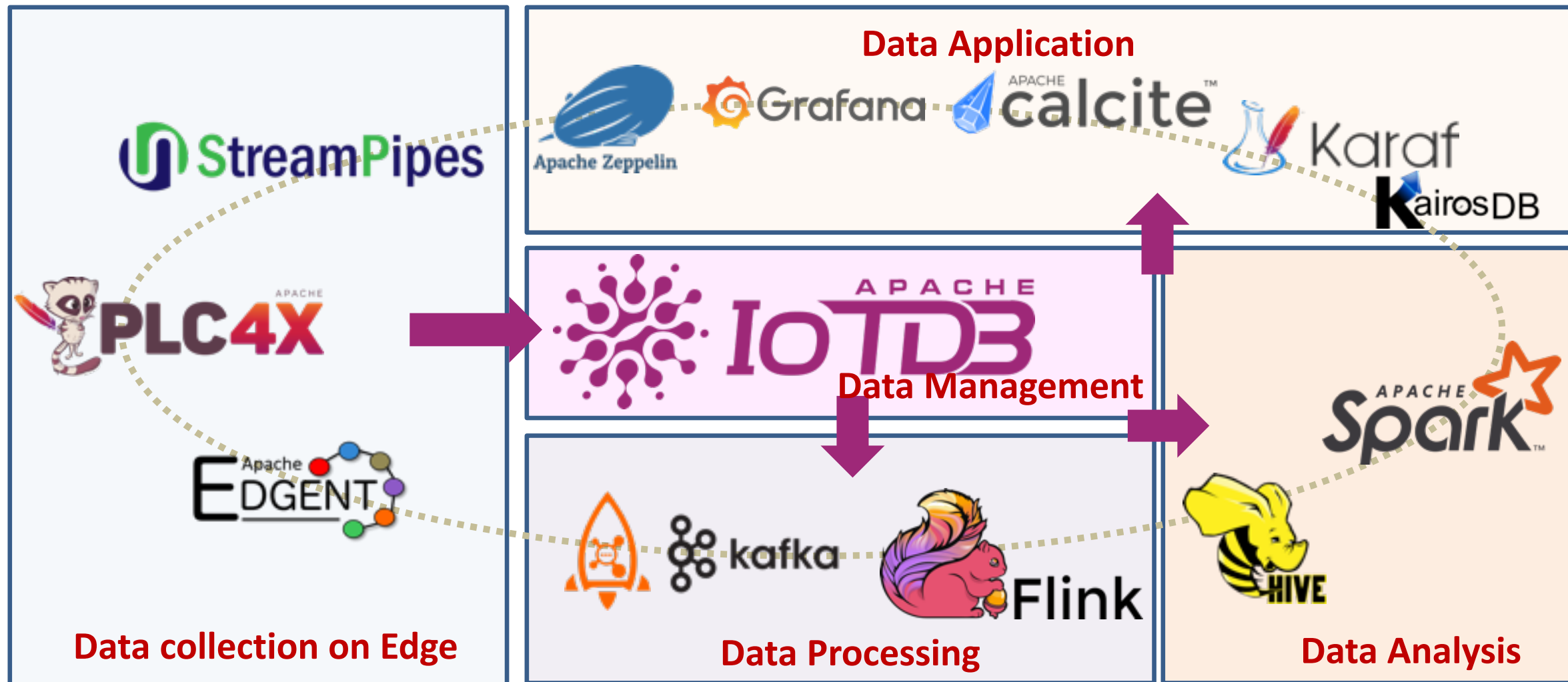
部署在云端数据中心的  
时序“数据仓库”



**数据中心（云）**

**与大数据分析框架无缝集成**  
支持时序数据处理，挖掘分析  
与机器学习

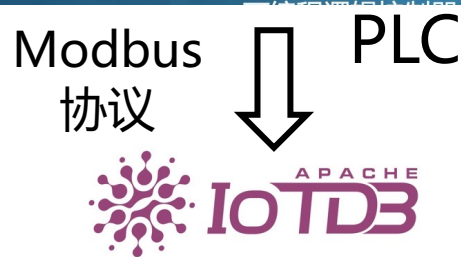
# 产品生态：覆盖IoT+AI全生命周期



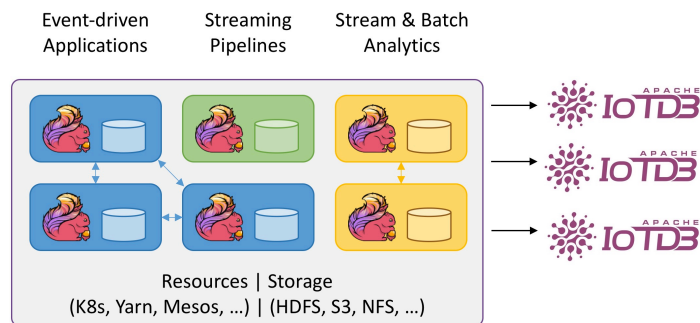
# 与IoT生态集成



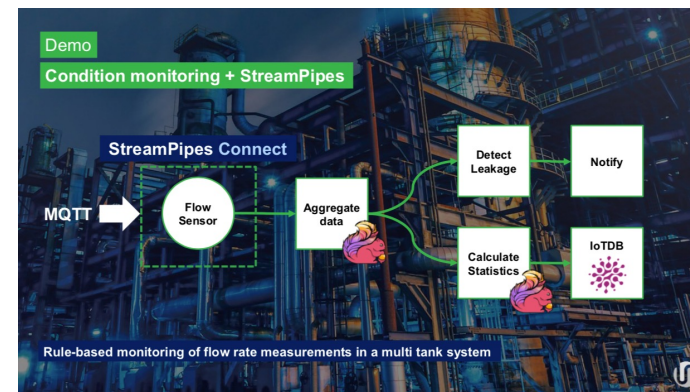
Apache PLC4X  
工业通信协议通用适配器



Apache Flink  
流数据与批处理计算框架



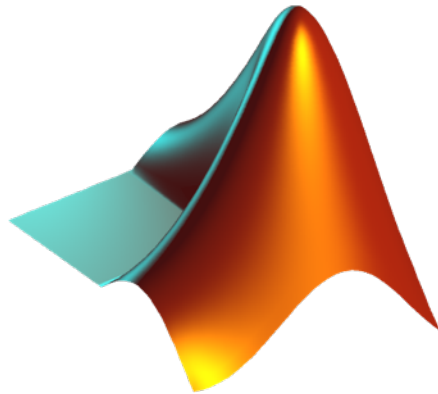
Apache StreamPipes  
面向非技术人员的IoT数据工具箱



# 与AI生态集成



数据可视化



Matlab 时频分析  
(小数据集)

```
datasource = 'jdbc'; username = 'root'; password = 'root';
driver = 'cn.edu.tsinghua.jdbc.TsfileDriver';
url = 'jdbc:tsfile://211.159.153.193:6667/';

conn = database(datasource,username,password,driver,url);
data = select(conn,'select cpu0temp from root.yanmoji.shenzhen.d1 where time > 2017-08-07T15:02:26.759');

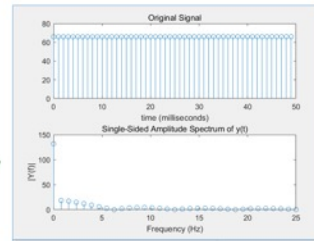
sz = size(data);
Fs = 50; % Sampling frequency
T = 1/Fs; % Sample time
L = sz(1); % Length of signal
t = (0:L-1)*T; % Time vector

y = data.root_yanmoji_shenzhen_d1_cpu0temp;

subplot(2,1,1)
stem(Fs*t(1:50),y(1:50));
title('Original Signal')
xlabel('time (milliseconds)')

NFFT = 2^nextpow2(L); % Next power of 2 from length of y
Y = fft(y,NFFT)/L; % (1)
f = Fs/2*linspace(0,1,NFFT/2+1);

subplot(2,1,2)
stem(f,2*abs(Y(1:NFFT/2+1)))
title('Single-Sided Amplitude Spectrum of y(t)')
xlabel('Frequency (Hz)')
ylabel('|Y(f)|')
```



大数据分析

Import tsfile lib

```
import cn.edu.thu.tsfile._
import org.apache.spark.sql.Session
```

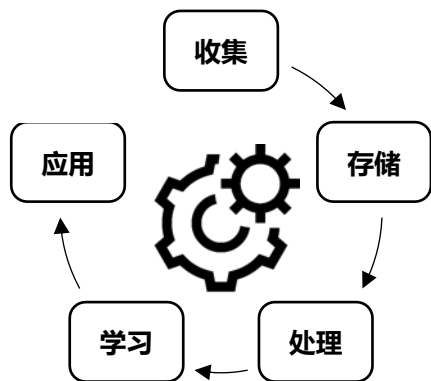
```
val spark = SparkSession.builder().master("local").getOrCreate()
```

```
spark.sql("create temporary view TsFile using cn.edu.thu.tsfile
options(path = \"test.tsfile\")")
```

Spark SQL

```
spark.sql("select * from TsFile where sensor_1 > 1.2").show()
```

# 产品特性：高吞吐、高压缩、低延迟、丰富查询



## 覆盖数据全生命周期

- 多种服务形式
  - 终端
  - 本地控制器
  - 数据中心
- 灵活部署
  - 开箱即用
- 支持数据实时写入/写出
- 支持便捷友好的分析



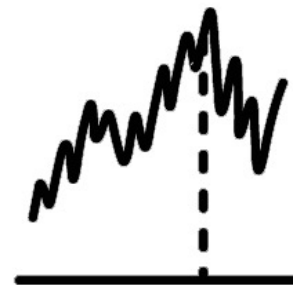
## 高效的数据持久化

- 高速写入
  - 每秒3000万点 (单机)
- 高压缩比
  - 1.37bits/点 [1]
  - 有损和无损压缩
- 支持百万时间序列



## 丰富/低延迟的数据查询

- 快速过滤数据
  - TB级数据百毫秒查询
- 聚合查询数据
  - 十亿点数十毫秒查询



## 扩展的时间序列操作

- 时序分段
- 时序表达
- 子序列匹配
- 时频转换
- 可视化



## 与现有生态系统集成

- MatLab
- Spark
- MapReduce
- Grafana
- Kafka



# 用户交互友好：工作台、命令行、可视化

工作台

The screenshot shows the Zeppelin Notebook interface with a workflow titled 'IoTDB-Zeppelin-Demo'. It consists of four stages:

- Set IoTDB Schema:** A SQL script executed successfully. The script creates a schema for IoTDB with columns for temperature, status, and hardware.
- Insert Time Series:** A SQL script that inserts 10 rows of time-series data into the 'root.in.wf01.wt01' table.
- Query: Table:** A SQL query that retrieves the data from the table, displaying it in a table format.
- Query: Line Chart:** A SQL query that retrieves the data, which is then visualized as a line chart showing temperature and hardware status over time.

Time	root.in.wf01.wt01.temperature	root.in.wf01.wt01.status	root.in.wf01.wt01.hardwa
1	1.1	false	4
2	3.2	true	3
3	2.3	false	2
4	4.4	false	1
5	3.5	false	0

命令行

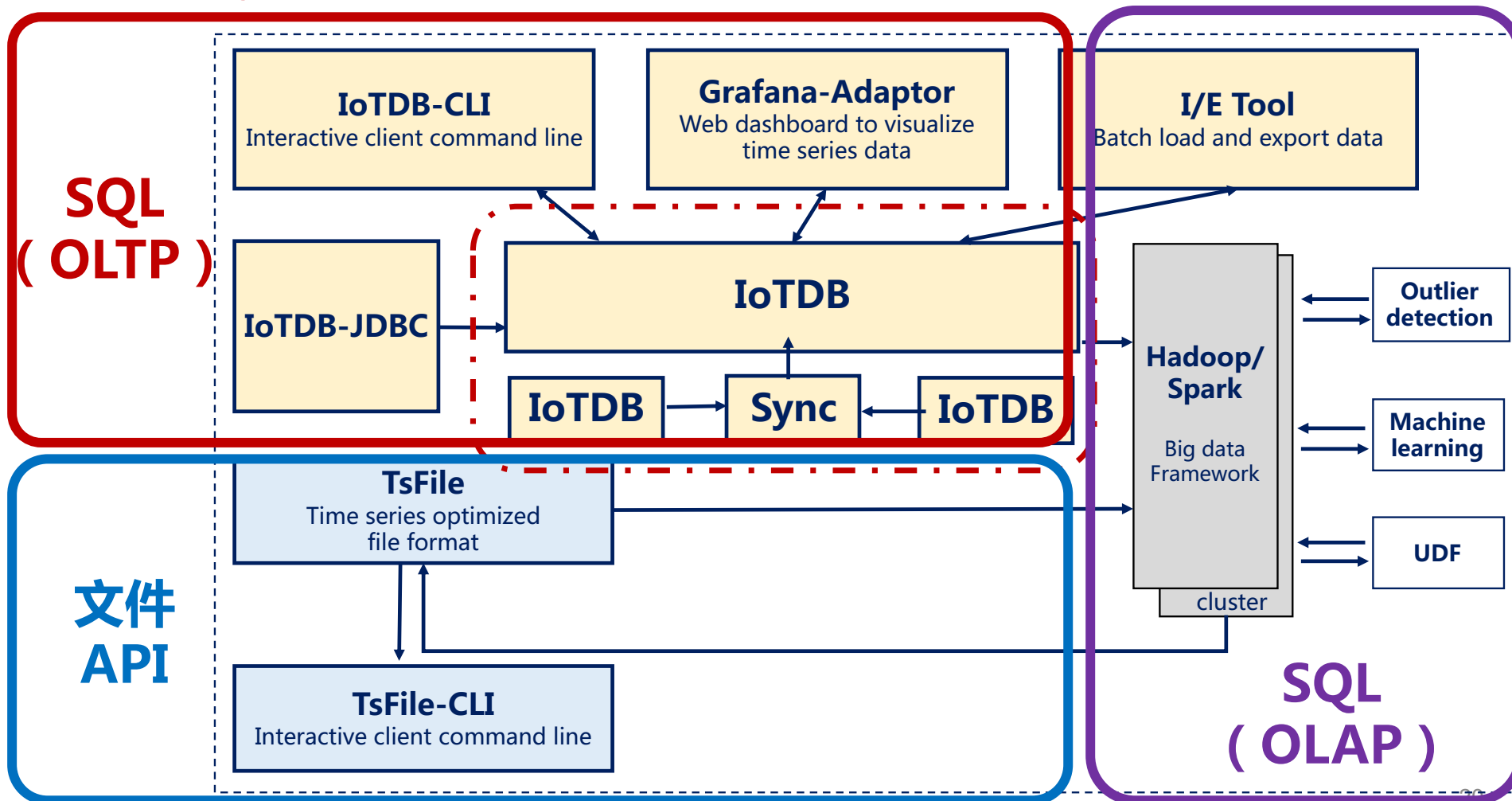
```
IoTDB> login successfully
IoTDB>
```

可视化



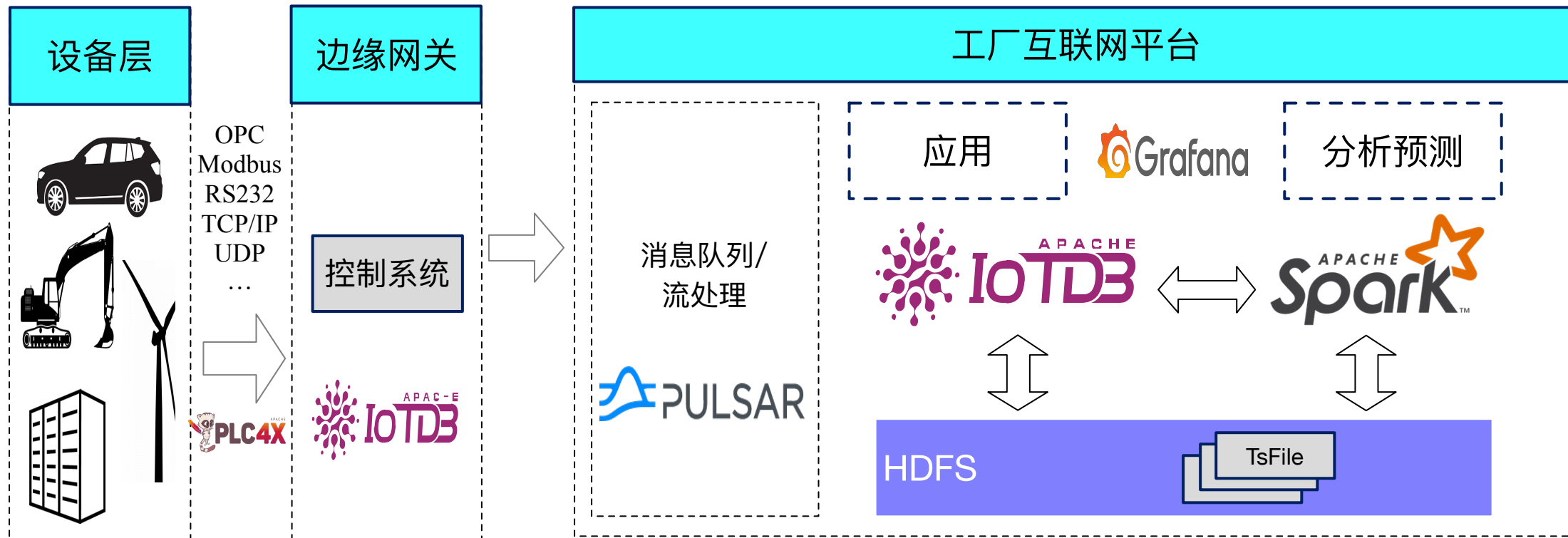
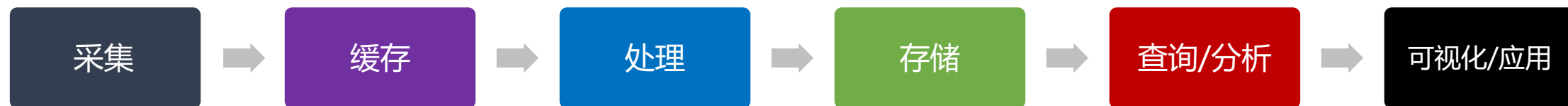
# 编程接口丰富：文件API与SQL语言

像使用MySQL一样存储和查询数据



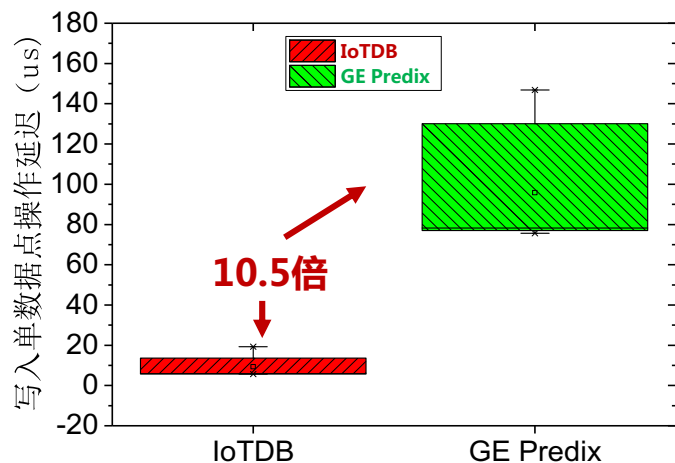
像使用CSV一样存储和访问数据

# 基于IoTDB的应用：端、边、云协同处理

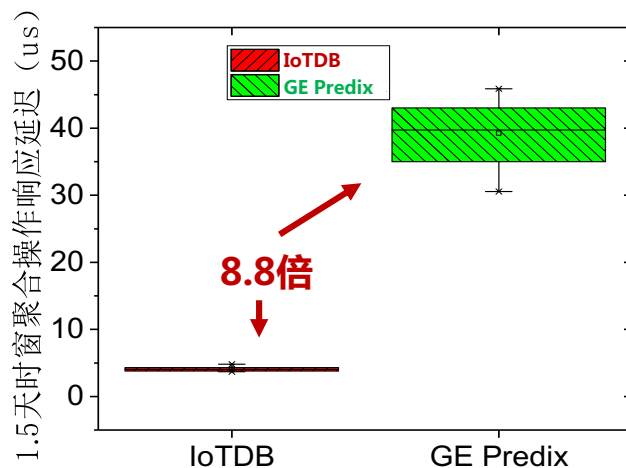


# 性能表现

## 第三方测试

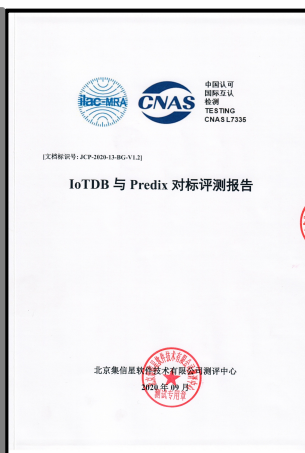


**(写入延迟：越小越好)**

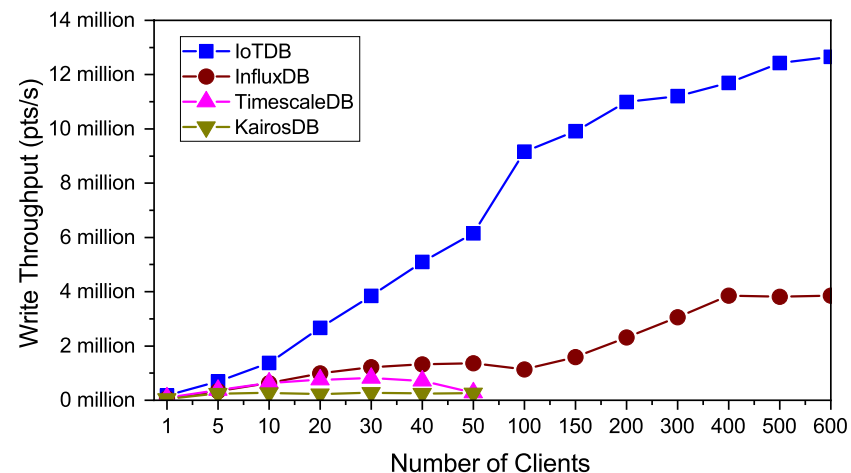


**(响应延迟：越小越好)**

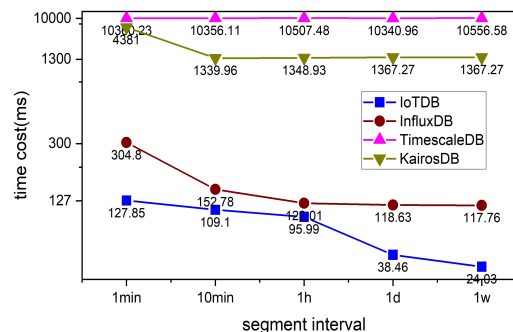
**北京集信星测试报告：与国际工业互联网平台领导者 GE Predix 相比，IoTDB 数据写入延迟小10.5倍，聚合访问延迟小8.8倍；**  
**中国软件评测中心测评中心测试报告：IoTDB 的读写性能优于InfluxDB等其他时序数据库**



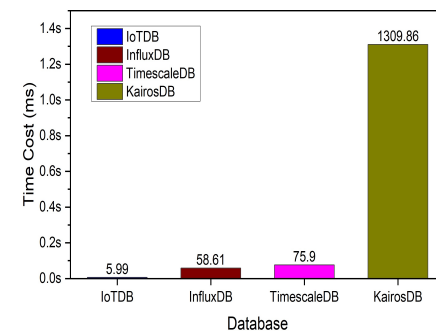
## 与多种时序数据库详细对比



**写入吞吐对比优于竞品3倍以上**



**降采样数据访问  
优于竞品5倍以上**



**实时数据查询  
优于竞品9倍以上**

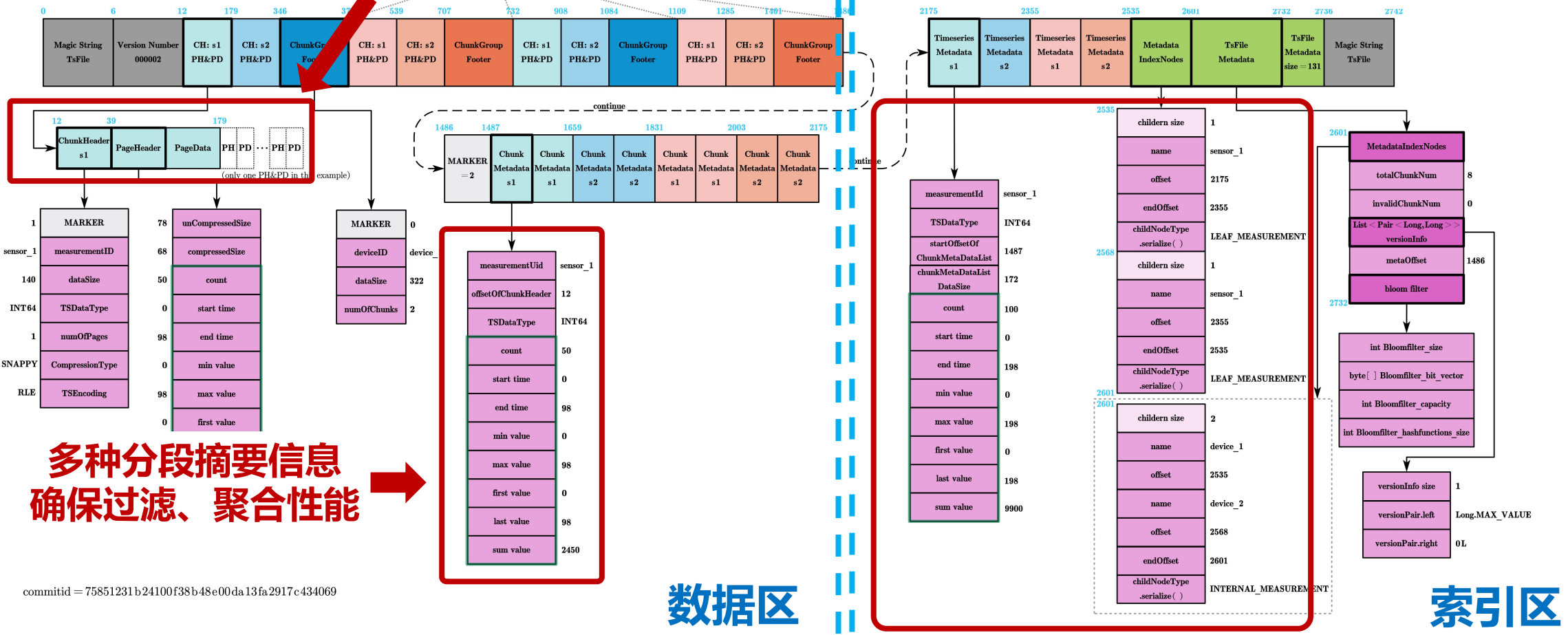
# 部分核心技术

# 紧致存储文件格式TsFile，解决高通量、降采样问题

This TsFile contains two devices: d1,d2.  
 Each device contains two measurements: s1,s2.  
 Four timeseries in total.  
 Each timeseries contains 2 chunks.

**列式存储+  
高压压缩率编码算法**

**文件级索引确保  
海量序列高效检索**



数据区

索引区

# 紧致存储文件格式TsFile，解决降采样、低质量、对齐问题

SELECT

```
last_value(left_temp)  
last_value(right_temp)
```

FROM

```
root.device1
```

GROUP BY

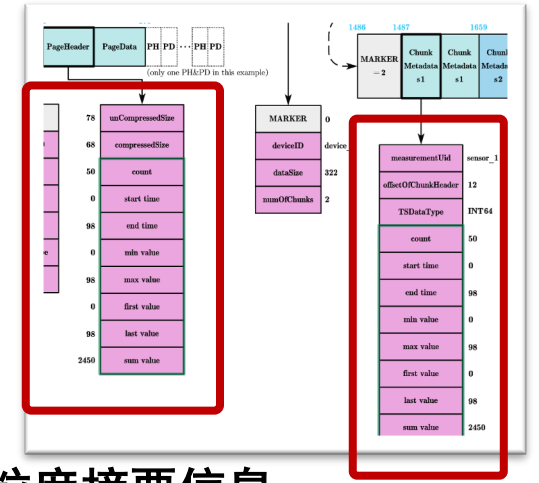
```
( [ now() - 10h, now() ), 1m )
```

FILL

```
( ALL [ PREVIOUS, 5m ] )
```

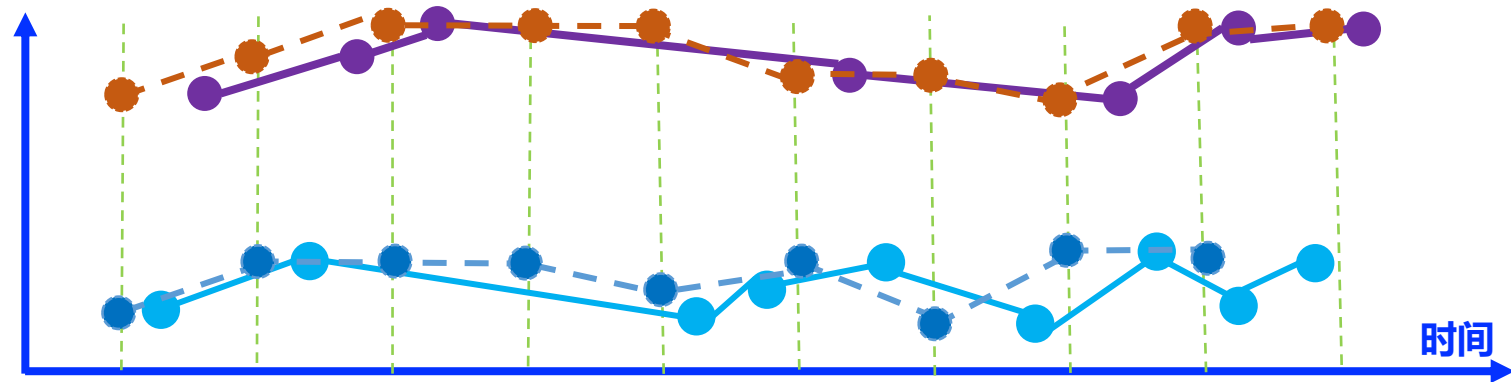
采用分段摘要信息将 ● ●

1. 降采样成每分钟1个数据点
2. 按每分钟进行数据对齐
3. 修补缺失的数据



细粒度摘要信息

粗粒度摘要信息









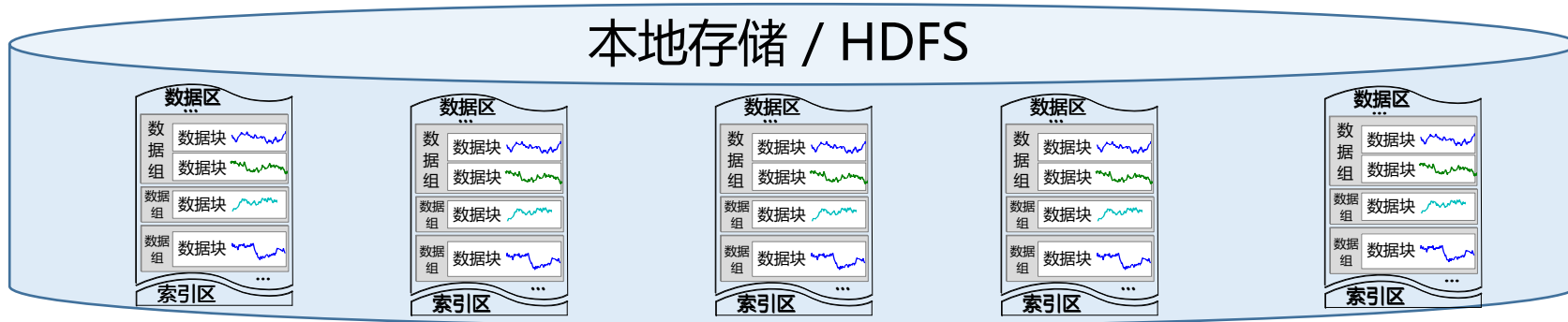
# 支持Spark、Hive等大数据分析平台



```
> SHOW TABLES;  
> SELECT * FROM table1
```



本地存储 / HDFS





# 相关研究论文

## 系统概况：

• [Apache IoTDB: time-series database for internet of things](#), Chen Wang, Xiangdong Huang, Jialin Qiao, Tian Jiang, Lei Rui, Jinrui Zhang, Rong Kang, Julian Feinauer, Kevin A. McGrail, Peng Wang, Jun Yuan, Jianmin Wang, Jiaguang Sun. VLDB 2020

## 乱序/索引技术：

• [Dual-PISA: An index for aggregation operations on time series data](#), Jialin Qiao, Xiangdong Huang, Jianmin Wang, Raymond K Wong. IS 2020

• [KV-match: A Subsequence Matching Approach Supporting Normalization and Time Warping](#), Jiaye Wu and Peng Wang and Chen Wang and Wei Wang and Jianmin Wang. ICDE 2019.

• [Matching Consecutive Subpatterns over Streaming Time Series](#), Rong Kang and Chen Wang and Peng Wang and Yuting Ding and Jianmin Wang. APWeb/WAIM 2018.

• [PISA: An Index for Aggregating Big Time Series Data](#), Xiangdong Huang and Jianmin Wang and Raymond K. Wong and Jinrui Zhang and Chen Wang. CIKM 2016.

## 分布式架构：

• [Apache IoTDB的分布式架构设计](#), Tianan Li, Jianmin Wang, Xiangdong Huang, Yi Xu, Dongfang Mao, Jun Yuan. 中国科学：信息科学 2020

# 相关专利



知识产权具体名称	授权日期
数据异常检测方法和装置	2020.06.02
一种错误数据容忍的虚警过滤方法和装置	2020.07.17
油耗量测定方法及装置、油箱	2020.07.17
一种识别时间序列的数据模式的方法及装置	2019.05.14
一种基于密度的缺失数据填补方法及装置	2019.02.05
一种数据中的异常数据点的检测方法及系统	2019.02.05
一种基于近邻的数据修复方法	2018.12.28
一种基于速度约束的多维时序数据修复方法及装置	2018.12.11
时序数据的编码方法及装置	2020.07.17
时间序列数据库的模式转换方法及装置	2020.06.23
基于向量时钟的分布式键值数据库的副本修复方法与装置	2020.03.13
一种支持处理单元在线切换的流数据处理方法	2020.02.07
分布式计算机存储系统中数据副本异构存储与访问方法	2019.11.08
一种基于滑动窗口的数据可视化方法和系统	2019.06.18
一种基于弹性表模型的多源异构数据存储方法和系统	2019.06.18
一种基于多层次列式存储结构的时序数据存储方法	2019.04.05



# 相关专利

知识产权具体名称	授权日期
多分段多阈值约束的时间序列模式查找方法和系统	2019.02.05
一种基于双层列表结构的容量扩充方法及系统	2019.02.05
一种基于时域和频域的时序数据存储方法和系统	2018.12.28
文件存储方法及系统	2018.12.28
一种时序数据列式存储、查询方法及系统	2018.12.11
一种时序数据乱序操作处理方法和装置	2018.12.11
一种自适应编码长度的时序数据存储的方法和装置	2018.12.11
一种列式存储下多时间序列的连接查询方法及系统	2018.12.11
基于弹性表模型的大规模多源异构数据持久化方法	2018.12.11
一种基于频域特征的子序列检索方法和系统	2018.10.30
一种基于频域特征的时序数据存储的方法和装置	2018.10.30
一种分布式存储系统中多维数据的存储模式转换方法	2018.04.20

# 部分证书

## 第七期优秀大数据产品、解决方案和应用案例及第四期优秀 AI 产品测评结果

### 优秀大数据产品

清华大学	清华数为物联网数据库 IoTDB V1.0
北京东方金信科技有限公司	海盒数据资产管理平台 V3.0
北京东方金信科技有限公司	数据实时同步系统 SMG V3.0
杭州中奥科技有限公司	CDMP V2.5.2
北京明朝万达科技股份有限公司	Chinasec (安元) 安全集中监控与审计系统 V2.0
成都智审数据有限公司	大数据审计分析监管平台 V1.0



## 第二届中国优秀开源项目

# 荣誉证书

清华大学

经审核, 贵单位/团队“Apache IoTDB”项目荣获一等奖。

特发此证。

中国开源云联盟  
二〇一九年十二月十七日



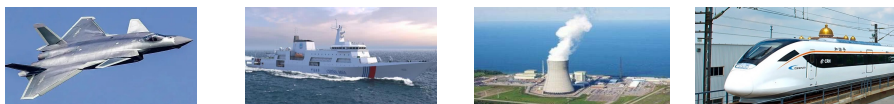
# 部分应用情况

# 国内推广应用情况

根据部分用户证明统计，服务规模以上工业企业**超过200家**。

## 中央骨干企业

中航成飞 中船黄埔 中核集团 中车四方



中国铁塔 湖南中烟 国网河北 湖南大唐



## 北京工业龙头企业

金风科技 北汽集团 京东方



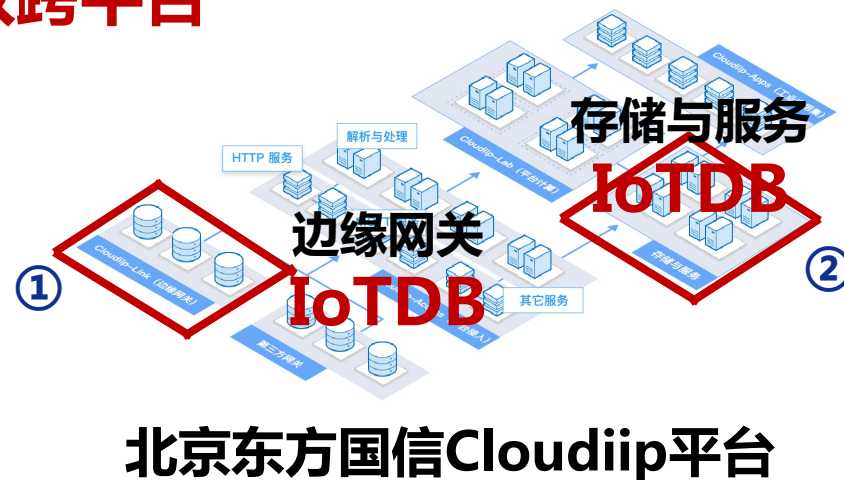
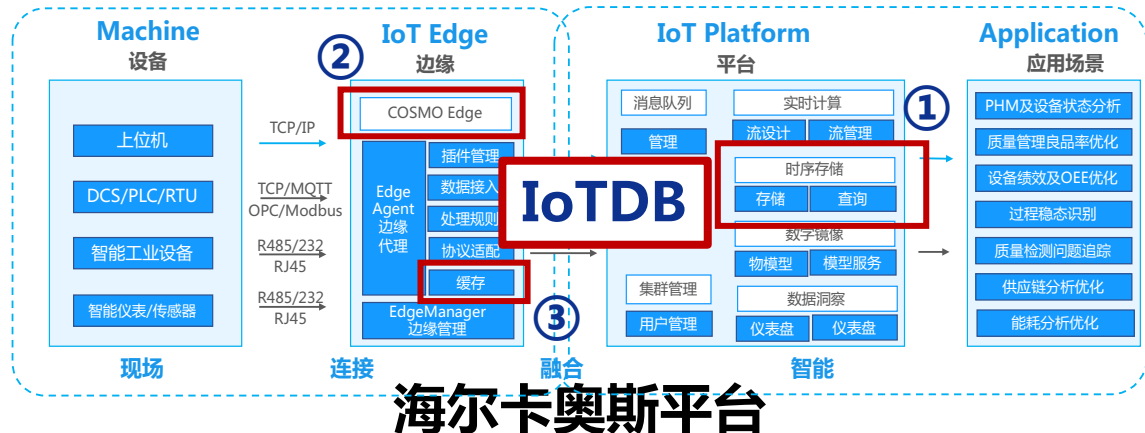
联想北京 神州数码 四维图新



## 间接服务企业名录

- 徐工集团，中联重科
- 东风汽车，猎豹汽车
- 江淮汽车，中钢集团
- 华菱钢铁，湛江钢铁
- 长飞光纤，武汉石化...

## 国家工业互联网双跨平台



# 辐射全球用户

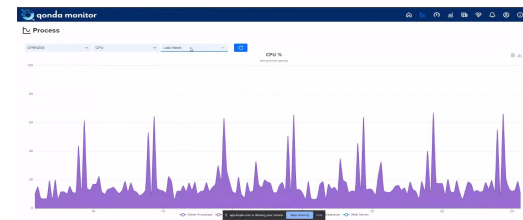


新西兰



Hancock Natural Resource Group

使用IoTDB管理  
50台服务器数据



巴西



BIX TECNOLOGIA

使用IoTDB

存储分析风机数据

From: Igor Gets <...@gmail.com>  
Subject: New here - Python client to pandas  
Date: 2021/02/01 22:23:14  
List: dev@iotdb.apache.org  
Hi everyone,

I am in a project to store and analyze wind turbines data and iotdb seems to be a great fit.

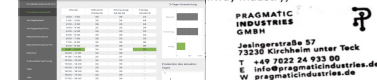
Igor Gets

普戈曼公司的Digital Cockpit**集成IoTDB**服务德国制造业企业

The following list shows the companies that use the Digital Cockpit which is backed by IoTDB. The companies are from various industries, including automotive, railway, aerospace and machine manufacturing.

- BMW AG (Automotive)
- Daimler AG (Automotive)
- Inacore GmbH (Automotive)
- Laempe Mössner (Automotive)
- Fritz-Winter Eis (Automotive)
- The Gill Corporation (Automotive)
- Schütz GmbH & Co. KGaA (Aerospace Industry)
- Heermann Maschinenbau GmbH (Machine Manufacturing)
- Hugo Beck GmbH (Machine Manufacturing)

PRAGMATICMINDS  
普戈曼 (德)  
德国能源部  
资助成立



工程装备大数据平台  
**集成IoTDB**  
服务小松全球市场

2018 年清华大学ソフトウェア学院

日本小松等国際龙头企业の全球市場。

データ管理と分析技術で我々工場のアクセス、高効率特徴検索、技術難問を有効に解決して、

探索使用IoTDB  
升级WinCC产品





**开源社区**



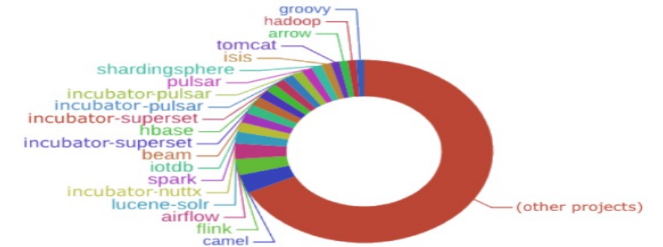
# Apache IoTDB开源社区建设

项目  
活跃度

2020年在Apache  
全球351个项目中  
排名第七

Top Repos by Commits

other projects	176549
camel	8488
flink	7407
airflow	7179
lucene-solr	5396
incubator-nuttx	4466
iotdb	4240
beam	4081
incub-superset	3875
hbase	3715
incub-superset	3610
incubator-pulsar	3500
incubator-pulsar	3500
pulsar	3491
shardingsphere	3412
isis	3090
tomcat	2860
arrow	2779

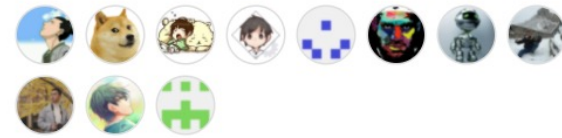


数据来源：最新Apache基金会年报

代码  
贡献者

来自中美德英澳  
等国151人

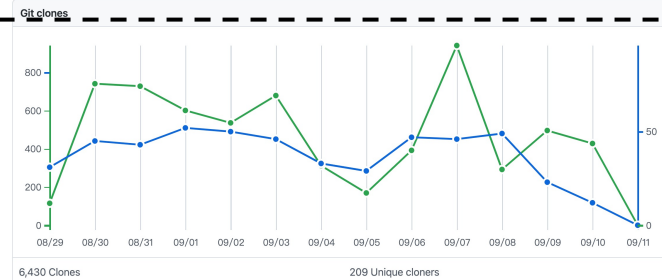
Contributors 151



信息来源：Github统计

用户  
下载量

源码下载量次数  
每周超6000次



信息来源：Github统计

# 开源影响力日益凸显

成为工业物联网生态中的重要一环

国际最热的IT监控系统Prometheus  
将IoTDB作为输入源之一  
(巴西开发者提供)

国内龙头企业以IoTDB  
设置招聘岗位

Section Home Apache IoTDB (Incubating)

工业PLC数据采集系统与IoTDB对接

工业数据传输协议与IoTDB对接

Prometheus EXPORTERS AND INTEGRATIONS

Third-party exporters

- IoTDB exporter

阿里云智能事业群-IoT大数据开发专家 (基于Apache IoTDB)

阿里巴巴 东方国信 云智慧

Apache 二十周年大会上被Apache委员报告



国外工业界用户 (德国) 开展培训



国际著名数据库软件库收录

DB-Engines Ranking of Time Series DBMS

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

This is a partial list of the complete ranking showing only time Series DBMS.

Read more about the method of calculating the scores.

Rank	DBMS	Database Model	Score
1.	InfluxDB	Time Series	26.25 +0.07 +4.68
2.	Kdb+	Time Series, Multi-mode	7.78 +0.18 +2.40
23.	M3DB	Time Series	0.21 +0.01 +0.13
24.	VictoriaMetrics	Time Series	0.14 +0.11 +0.07
25.	Apache IoTDB	Time Series	0.14 +0.03
26.	Blueflood	Time Series	0.12 +0.03 +0.02
27.	Warp 10	Time Series	0.10 +0.03 +0.05
28.	SiriDB	Time Series	0.04 +0.00 +0.01
29.	SiteWhere	Time Series	0.01 0.00 +0.00

开源中国年度最受欢迎开源项目

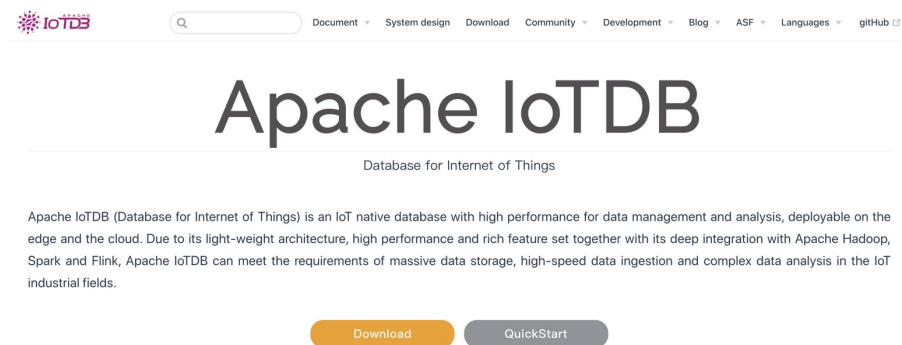


# 融入开源社区

- 期待各位加入
  - qq交流群：659990460
  - 微信群：添加好友 tietouqiao
- mail list:
  - 订阅信息，发送邮件至：  
[dev-subscribe@iotdb.apache.org](mailto:dev-subscribe@iotdb.apache.org)
  - 讨论与反馈，发送邮件至：  
[dev@iotdb.apache.org](mailto:dev@iotdb.apache.org)
  - BUG反馈，提交至：  
<https://issues.apache.org/jira/projects/IOTDB/issues/IOTDB>



- 官方网站：<https://iotdb.apache.org>



- 微信公众号：Apache IoTDB

