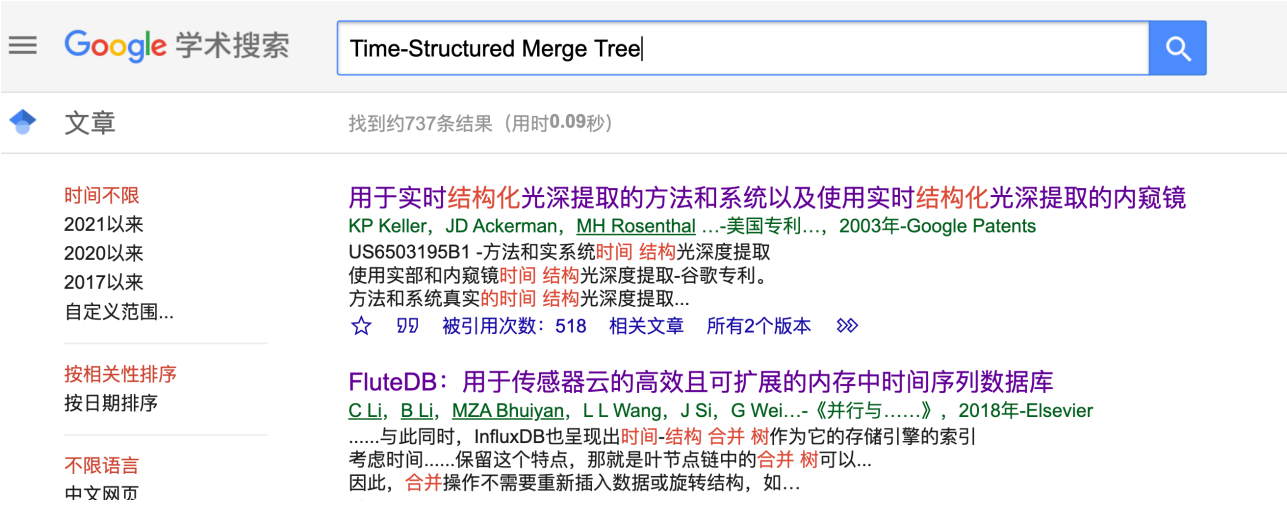
“Time-Structured Merge Tree”谷歌学术相关论文收集

一、搜索



二、论文列表（论文名为链接）

1. Time Series Databases and InfluxDB 2017

influxDB 1.x版本的论文，也算产品报告，对于存储引擎的选用有说明，但很简略，常作为其它论文讲述关于influxDB的相关工作时的引用

2. Suitability Of Influxdb Database For Iot Applications 2019.08

influxDB相关，作为相关工作的引用

3. IoT Database’ Technologies: Research Review 2020

总结了物联网场景下三种存储技术和三个具体的时序数据库，技术介绍和优劣分析，没有提出什么新的内容

4. TritanDB: Time-series Rapid Internet of Things Analytics 2018.01

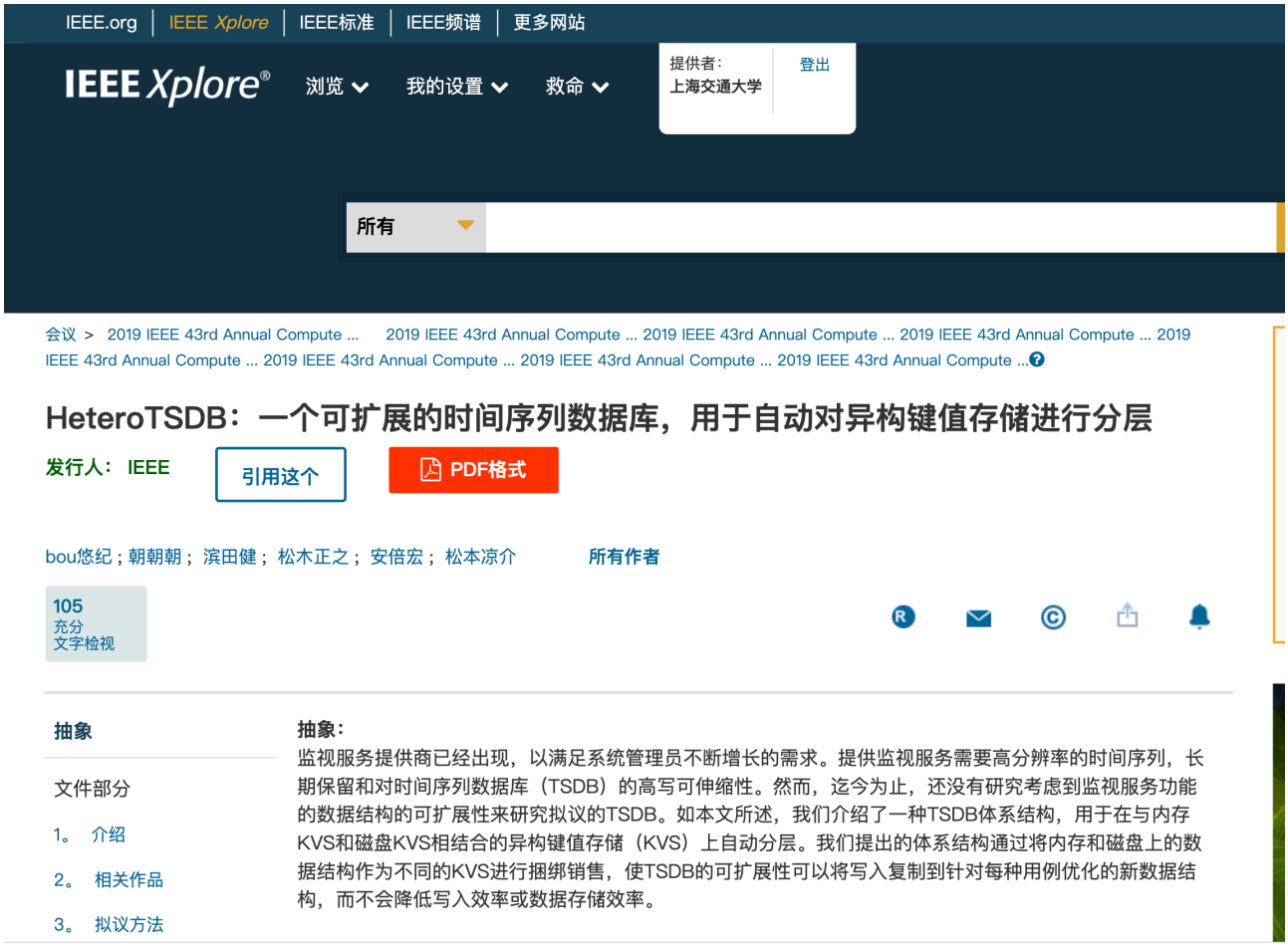
提出物联网下数据管理的两大困难：“异构事物集成”“云数据库接收和查询性能”

对公共物联网数据进行调查，发现其中大多数显示出独特的平坦，宽泛和数字化的特征，并混合了均匀和不均匀间隔的时间序列。将调查结果与压缩策略和存储数据结构相结合，研发带有新特征的数据库，支持预测。

全文38页，有点长。

5. HeteroTSDB: An Extensible Time Series Database for Automatically Tiering on Heterogeneous Key-Value Stores 2019

如图，通过交大机构登录获取全文



抽象中提出异构键值存储，

关注问题：对于物联网场景的大量数据存在异构事物集成的困难

6. FluteDB: An Efficient and Dependable Time-Series Database Storage Engine 2017.09

无法获取全文

FluteDB: An efficient and scalable in-memory time series database for sensor-cloud 2018.07

是上文的拓展

如图，通过交大机构登录获取全文

强调

- FluteDB是用于传感器云的高效且可扩展的时间序列数据库。
- FluteDB中的索引为时间序列数据提供了灵活的存储技巧。
- FluteDB通过根据数据温度调整磁盘访问来提高效率。
- FluteDB优化了其数据封装和容错策略。

抽象

最近，随着大规模传感器网络的广泛使用，时间序列数据大量生成，需要进行处理。但是，这些传统数据库在处理云中的如此大的流数据时显示出它们在存储上的局限性，甚至它们的实际可靠性和可用性也难以得到保证。为了解决这个问题，本文提出了FluteDB，这是一种高效且可扩展的传感器云内存时间序列数据库。我们充分分析时间序列数据及其相关操作的独特特征，以在效率，可伸缩性，资源消耗，可靠性和可用性之间取得适当的平衡。具体来说，基于对持续出现的时间序列问题的根本原因进行的综合分析，FluteDB Target针对内存和物理存储中的关键操作优化了策略，但以部分可接受的数据精度和一致性为代价。不同数据类型的压缩算法。对所有子模块的验证表明，我们改进的策略在实时序列环境中的性能明显优于现有方法。全局实验结果还表明，集成的FluteDB将查询延迟减少了17倍，将写入速率提高了98倍，并节省了约47％的存储资源。在实际和模拟故障中，FluteDB的平均可用服务时间，恢复率和程度与最新的可靠性和可用性策略相比，这表明FluteDB可以提供高度稳定的大规模数据云服务。

较为具体的说明了所提出的新数据结构TTSM的实现，性能表现上，与传统的B+树和LSM树对比得到了抽象所述结果。

7. Timon: A Timestamped Event Database for Efficient Telemetry Data Processing and Analytics 2020

Publication: SIGMOD '20: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data • June 2020 • Pages 739–753 • https://doi.org/10.1145/3318464.3386136

ABSTRACT

With the increasing demand for real-time system monitoring and tracking in various contexts, the amount of time-stamped event data grows at an astonishing rate. Analytics on time-stamped events must be real time and the aggregated results need to be accurate even when data arrives out of order. Unfortunately, frequent occurrences of out-of-order data will significantly slow down the processing, and cause a large delay in the query response. Timon is a timestamped event database that aims to support aggregations and handle late arrivals both correctly (i.e., upholding the exactly-once semantics) and efficiently. Our insight is that a broad range of applications can be implemented with data structures and corresponding operators that satisfy associative and commutative properties. Records arriving after the low watermark are appended to Timon directly, allowing aggregations to be performed lazily. To improve query efficiency, Timon maintains a TS-LSM-Tree, which keeps the most recent data in memory and contains a time-partitioning tree on disk for high-volume data accumulated over long time span. Besides, Timon supports materialized aggregation views and correlation analysis across multiple streams. Timon has been successfully deployed at Alibaba Cloud and is a critical building block for Alibaba cloud's continuous monitoring and anomaly analysis infrastructure.

SIGMOD的论文，但抽象有点看不懂了

8. Taurus Database: How to be Fast, Available, and Frugal in the Cloud 2020

9. Gorilla: a fast, scalable, in-memory time series database 2015