

DB Tsai

LinkedIn: www.linkedin.com/in/dbtsai/ • Cupertino, CA, 95014 • dbtsai@dbtsai.com • (650) 666-1122

Summary

- Building a Modern Data Platform to drive Apple's success by bridging the gap between business and technology with a clear strategic vision.
- Experienced Data Platform engineering leader with a proven track record of building world-class teams. Skilled in hiring, developing, and retaining top talent.

Experience

Apache Foundation

Remote

Apache YuniKorn PMC member / Committer

Mar 2019 – Present

- Mentor since YuniKorn entered Apache Incubation.

Apache Spark PMC member / Committer

May 2015 – Present

- My contributions, <https://github.com/apache/spark/commits/master?author=dbtsai>

Apple

Cupertino, CA

Head of Core Data Platform

Oct 2023 – Present

- I lead four areas at Apple: Spark, Native Engine, Flink, and Data Security, growing the teams from 2 to 25 members. It's an honor to lead these award-winning teams, recognized with the ACM SIGMOD Awards in 2022 and 2023, demonstrating our impactful work and leadership in the big data industry.
 - (i) Apple's Spark team, including several PMC members and committers, co-received the 2022 ACM SIGMOD System Award for their Apache Spark contributions. They leverage open-source influence to align developments with Apple's needs, and have contributed to various features in Spark/Kubernetes and Spark/Iceberg integrations. Additionally, they maintain an internal distribution of Spark.
 - (ii) Enhancing faster data insights and reducing compute expenses at Apple's scale is crucial, saving the company millions each month. The native engine team has developed a vectorized native Spark accelerator based on Apache DataFusion that requires no changes to user code and has open-sourced it as Apache DataFusion Comet.
 - (iii) Apple's Flink team, including several PMC members and committers, developed a Java-based Flink Kubernetes operator tailored to Apple's needs and open-sourced it under Apache Flink. Two team members co-received the 2023 ACM SIGMOD System Award for their contributions. Recently, Google retired their Flink K8s operator in favor of ours, highlighting its broader adoption beyond Apple.
 - (iv) At Apple, safeguarding customer data is paramount, particularly in light of GDPR and DMA requirements. Our security team developed Parquet Modular Encryption and Apache Iceberg Table Encryption, enabling scalable columnar data encryption compatible with Spark, Trino, and Flink. This protects hundreds of petabytes of sensitive data in production. We have open-sourced this technology, and its adoption as an industry standard has been gratifying, enhancing accessibility and utilization. This team also designed and implemented fine-grained access control, enabling users to define row-level filtering, column-level access, column masking rules by modifying Spark execution plans. This key component of our data governance strategy prevents data duplication, reduces storage and compute usage, and avoids data staleness.

Senior Engineering Manager

Oct 2019 – Sep 2023

Staff Software Engineer

Apr 2018 – Sep 2019

- Developed Apple Data Platform strategies, enhanced Spark functionalities for Apple's needs, and made significant contributions to Apache Spark. Advocated for Apple's requirements as an Apache Spark PMC member.

Netflix Los Gatos, CA
Machine Learning Researcher / Technical Lead Apr 2015 – Mar 2018
• Built the renowned Netflix Personalization and Recommendation System.

Alpine Data Labs San Francisco, CA
Senior Machine Learning Engineer Apr 2013 – Apr 2015
• Developed scalable ML algorithms like Decision Tree, Information Gain-based Variable Selection, and one-pass Linear Regression with L1/L2 penalty using Spark, contributing them to Apache Spark. These contributions led to my becoming a Spark committer.

KeeKa, StartX 2012 Summer Stanford, CA
Co-founder and CTO Jan 2012 – Mar 2013
• A Social Network Connecting People through Fashion.

Selected Publications

Petabyte-Scale Row-Level Operations in Data Lakehouses Aug 2024
• Anton Okolnychyi, Chao Sun, Kazuyuki Tanimura, Russell Spitzer, Ryan Blue, Szehon Ho, Yufei Gu, Vishwa Lakkundi, **DB Tsai**
• Accepted by VLDB 2024

MLlib: Machine Learning in Apache Spark Apr 2016
• Xiangrui Meng, Joseph Bradley, Burak Yavuz, Evan Sparks, Shivaram Venkataraman, Davies Liu, Jeremy Freeman, **DB Tsai**, Manish Amde, Sean Owen, Doris Xin, Reynold Xin, Michael J Franklin, Reza Zadeh, Matei Zaharia, Ameet Talwalkar
• Journal of Machine Learning Research, Vol 17 (34) pp. 1-7

Quantum Zeno and anti-Zeno effect of a nanomechanical resonator Mar 2010
• Po-Wen Chen, **DB Tsai**, Philip Bennett
• Physical Review B 81, 115307

Optimal control of the silicon-based donor-electron-spin quantum computing Jun 2009
• **DB Tsai**, Po-Wen Chen, and Hsi-Sheng Goan
• Physical Review A 79, 060306(R)

Education

Stanford University Palo Alto, CA
ABD in Applied Physics Ph.D. program Sep 2010 – Jun 2012
M.S. in Electrical Engineering Sep 2010 – Jun 2012

National Taiwan University Taipei, Taiwan
M.S. in Physics Sep 2006 - Jul 2008

National Cheng Kung University Tainan, Taiwan
B.S. in Physics Sep 2002 - Jun 2006