

REPORT REPRINT

Keeping its modular strategy, PingCAP adds columnar storage engine for analytics

MAY 08 2019

By James Curtis

In addition to releasing a columnar-based storage to better enable analytics for hybrid workloads, the company has rolled out a new Kubernetes-centric operator to enable stateful workloads in a cloud environment. PingCAP continues to see traction in China, where it was founded, while Southeast Asia, Europe and North America are also growth areas.

THIS REPORT, LICENSED TO PINGCAP, DEVELOPED AND AS PROVIDED BY 451 RESEARCH, LLC, WAS PUBLISHED AS PART OF OUR SYNDICATED MARKET INSIGHT SUBSCRIPTION SERVICE. IT SHALL BE OWNED IN ITS ENTIRETY BY 451 RESEARCH, LLC. THIS REPORT IS SOLELY INTENDED FOR USE BY THE RECIPIENT AND MAY NOT BE REPRODUCED OR RE-POSTED, IN WHOLE OR IN PART, BY THE RECIPIENT WITHOUT EXPRESS PERMISSION FROM 451 RESEARCH.



Introduction

Distributed database vendor PingCAP continues to see traction in China, where it was founded, while Southeast Asia, Europe and North America continue to be growth areas as well, perhaps fueled by a series C funding round it landed in September 2018. A greater focus on hybrid workloads is also contributing to the company's traction, including the release of a new columnar-based storage engine that can easily be slotted into the PingCAP TiDB environment. Moreover, the company has released a Kubernetes-friendly operator for deploying and managing TiDB containers.

451 TAKE

PingCAP has picked up a number of enterprise clients in Southeast Asia, but the US and European markets are also targets – the company reports a collection of POCs in the works. PingCAP has its eye on the hybrid processing market, which is evident by the company's release of a new columnar storage engine, called TiFlash, that can be added to the company's component-centric TiDB database. The company has been working on so-called 'cloud native,' as well, having released a Kubernetes-centric operator to enable stateful workloads of the company's TiDB database. With new developments that enable hybrid workloads and capabilities to run stateful workloads in the cloud, PingCAP is positioning itself well to not only raise its profile, but also to see stronger traction in North America.

Context

Founded in China in 2015, PingCAP follows a modular design for its distributed SQL-based database known as TiDB, which is provided as open source; the startup also provides some enterprise functionality. The company's TiDB database was inspired in part by the Google Spanner paper, which asserts the importance of maintaining relational standards of SQL and ACID (consistency) while being able to employ a distributed architecture. While Spanner is positioned as a globally distributed database capable of handling ACID transactions, PingCAP is moving more toward targeting hybrid workloads capable of handling transactions and analytics, even though PingCAP can likewise be deployed to handle globally consistent transactions.

With its roots in China, the company continues to see traction in that country. Southeast Asia is also a growth area for the company, specifically in Singapore, India and Vietnam, where it reports landing a number of large enterprise customers. The US and Europe are likewise a focus area for the startup, where PingCAP reports a number POCs in play. PingCAP does not publicly report customer count, but does state that it has 'a handful' of paying customers. Furthermore, the company has approximately 140 employees that are mostly based in Asia, with a team of 10 residing in North America.

Meanwhile, in September 2018, PingCAP landed a series C round of funding for \$50m that was led by FOSUM and Morningside Venture Capital. To date, the startup has secured \$71.6m in total funding.

Products

In a previous report on PingCAP, we noted the company's interest with hybrid workloads as a focus area. The company's TiDB database, which is built with a collection of components and offered as open source, can handle globally consistent transactions, but the company points out that many of its customers want the ability to do analytics on those transactions. TiDB leverages the company's TiSpark component (based on Apache Spark) to carry out analytics on the incoming transactional data, but the company has furthered its approach to hybrid workloads, and recently released a new columnar-based storage engine called TiFlash to drive better performance.

The new storage engine sits alongside TiKV, the company's storage engine for transactions based on a key-value store, and holds an exact replica of the data, enabled by the use of the Raft consensus protocol. Sitting above TiFlash is TiSpark, to facilitate processing, and beside that is the company's MySQL-compatible SQL component that serves as a query optimizer and directs queries to either TiFlash (OLAP) or TiKV (OLTP) for processing. Besides improving analytical performance, the company points out that a separate storage engine for analytics eliminates the so-called 'noisy neighbor' issue that can be caused when analytical queries can impact incoming transactional data.

Besides a new columnar storage engine, PingCAP is addressing cloud-centric deployments and is specifically looking to enable stateful workloads. As such, the company has released a new Kubernetes-centric operator – a type of container toolset – to handle all of the running, deploying, maintaining, and backup and recovery of TiDB in a container environment. Because TiDB is component-based, the TiDB Operator is able to effectively handle all of the resources of each component, including the TiDB MySQL-compatible SQL layer, TiKV, TiFlash and TiSpark, such that the components do not interfere with each other.

Competition

Given its distributed SQL and ACID capabilities, we place PingCAP among the NewSQL vendors. However, many of these vendors also promote hybrid transactional and analytical capabilities. Vendors include NuoDB, HarperDB, FaunaDB, VoltDB, MemSQL, InterSystems' IRIS Data Platform, Actian X and SAP HANA. Esgyn and Splice are leveraging Hadoop and open source tools to address hybrid workloads.

The public cloud vendors also pose competitive challenges to PingCAP, specifically Google's Cloud Spanner, which can be paired with Google BigQuery. AWS's Amazon Aurora can be paired with Redshift for mixed workloads.

With PingCAP's compatibility with MySQL and modular design, MariaDB is potentially a direct competitor. MariaDB recently rolled out its X3 Platform, which combines the MariaDB TX (transactions) and MariaDB AX (analytics) products into a single platform. It also takes a multiple-storage-engine approach, utilizing such engines as InnoDB or MyRocks for transactions, ColumnStore for analytics, and Spider (a virtual storage engine used for sharding).

Many of the traditional relational vendors are providing offerings for hybrid. Oracle, for instance, has its In-Memory Column Store product that can be paired with row storage for handling dual workloads. IBM has its Shadow Tables functionality that can be deployed within Db2 for hybrid, and Microsoft leverages what it calls Columnstore Indexes, as well as the ability to enable advanced analytics with SQL Server R Services.

Lastly, many NoSQL vendors can offer mixed workload capabilities, such as DataStax, MongoDB, MarkLogic, Aerospike, Couchbase, Redis Labs and FairCom.

SWOT Analysis

STRENGTHS

The company's modular-based distributed SQL database means that customers can configure it to suit their needs, including the ability to handle transactional and analytical workloads.

WEAKNESSES

While the company is having success in China and Southeast Asia, it is still working to establish a foothold in the US, where the competition is quite strong.

OPPORTUNITIES

Hybrid is a good strategy for the company because many enterprises start with transactional workloads and then move toward analytics on those transactions. PingCAP will also see potential customers since it supports MySQL, providing an avenue for customers to switch when moving MySQL applications.

THREATS

The distributed SQL (NewSQL) market, which tends to also attract vendors driving hybrid workloads, is a highly competitive space – so much so that some vendors can get left off enterprises' lists, thus the need for companies such as PingCAP to distinguish their differentiation, particularly as they work to raise their profiles.