# Cloud Bigtable在广告技术中的使用

郭斌 Google Cloud 架构师

What is Cloud Bigtable

Cloud Bigtable is a **petabyte-scale, fully managed** NoSQL database service for use cases where **low latency** random data access, **scalability** and **reliability** are critical.
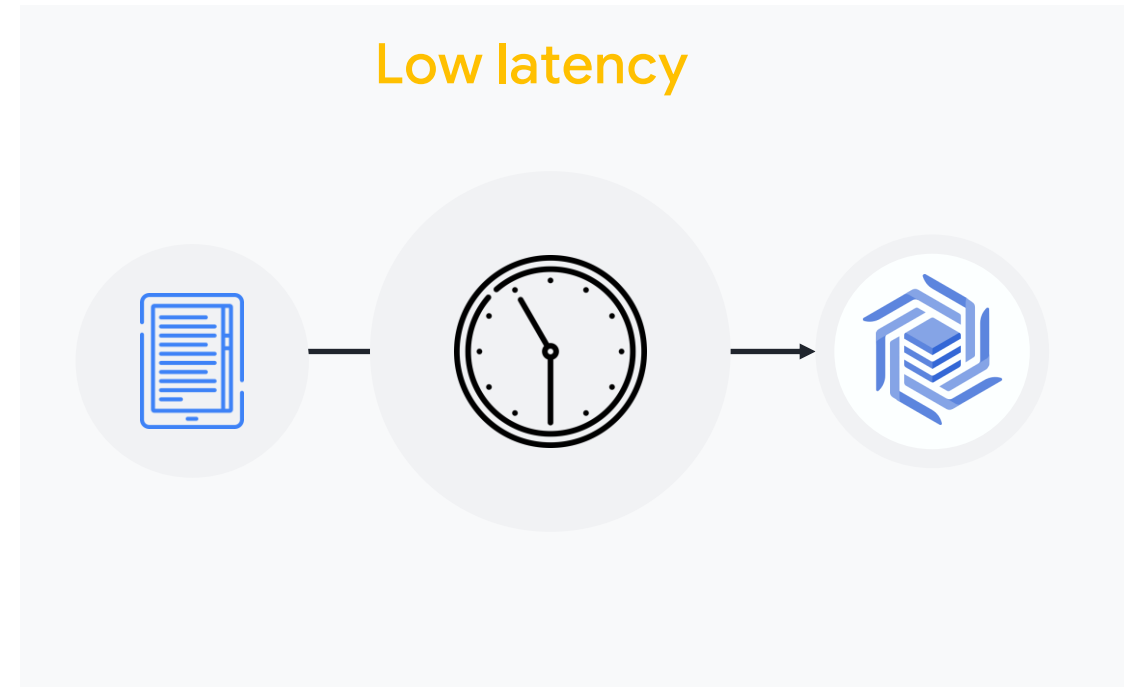
Being fully managed, it can scale seamlessly. It can also **integrate** with the Apache® ecosystem and supports the HBase™ API.

It's the home for demanding persistent **Key-Value** and **Columnar** workloads on Google Cloud

Speed

High throughput

Low latency
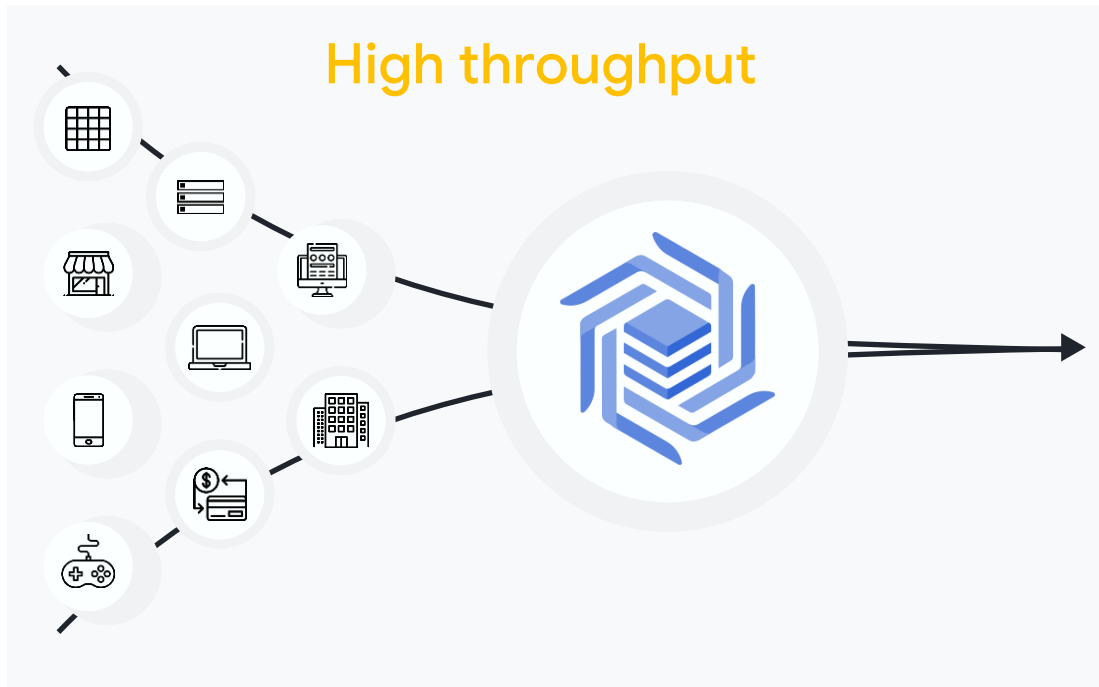
High write throughput per node
for **random access** workloads

Consistent 99th percentile **low single
digit** read and write latency

What's unique about Cloud Bigtable?

Exceptional throughput and storage capabilities

Designed and built for big data

Replication mode

Zero touch, integrated and tuned for real-world HA use cases

Workload isolation

App profile to support different workloads against different regions. For example, batch vs serving path

Built for petabyte-scale consumer internet

Earned its reputation with globally-available Google products like Ads, YouTube, Search, Maps, and Google Analytics

Apache HBase

- Implemented in Java (heap and gc tuning)
- Peer replication similar to multi-cluster deployments with eventual consistency (manual backfill)
- Custom client connection handling for multi-cluster
- Custom machines and tuning for HBase cluster
- Requires management of tables, regions, balancing, and optimizations
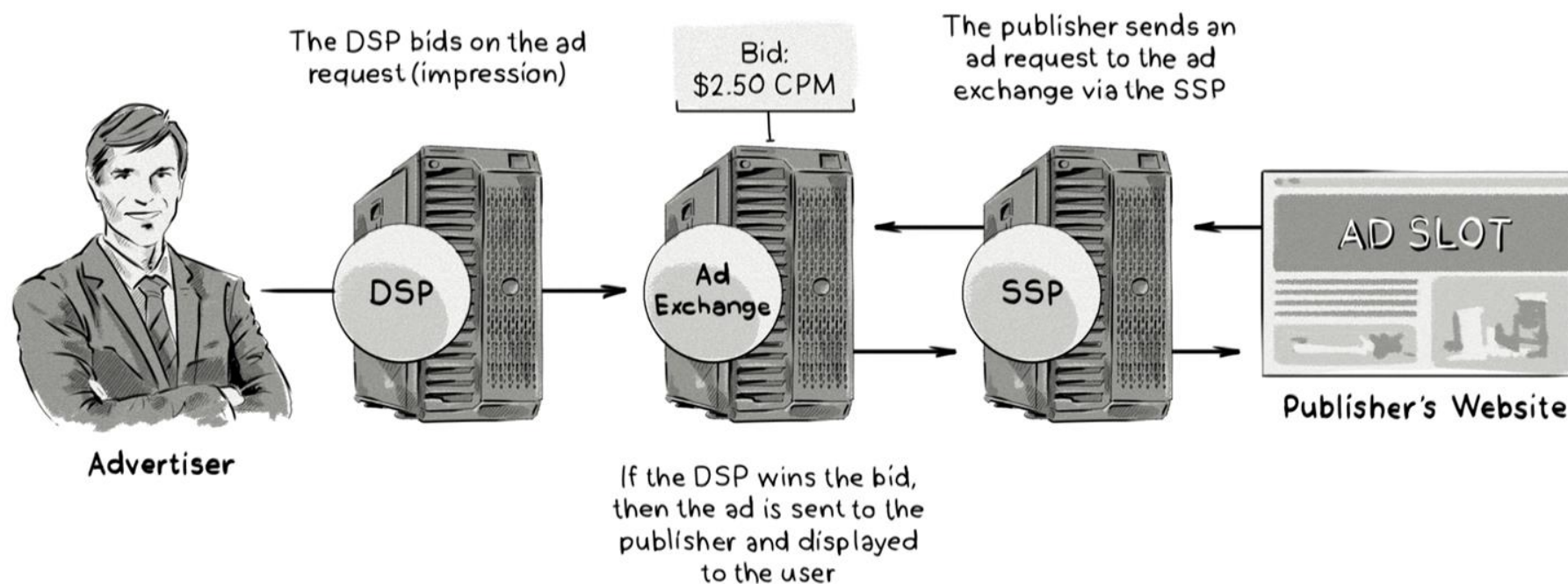- Support coprocessors, custom HBase builds

VS

Cloud Bigtable

- Implemented in C++
- Multi-cluster deployments with eventual consistency
- App-profiles - multi-cluster routing
- No customizations to Bigtable nodes (CPUs, Memory, Disk IO)
- No customizations to Memtable, Blockcache
- "Hands-free" managed platform
- No access to tablet servers for coprocessors, no server customizations

From developer's point of view:
Difference between HBase and Cloud Bigtable

	HBase	CBT
Access: Java	HBase authentication	IAM(Java Bigtable HBase or Bigtable client libraries)
Access: other programming languages	Apache Thrift	IAM(Bigtable client libraries)
APIs compatibility	HBase APIs	Besides HBase compatible APIs, CBT has native APIs
Shell	HBase Shell	CBT
Test environment	HBase single node	gcloud beta emulators bigtable start

What's real-time bidding?

The DSP bids on the ad request (impression)

Bid: $2.50 CPM

The publisher sends an ad request to the ad exchange via the SSP

DSP

Ad Exchange

SSP

AD SLOT

Publisher's Website

Advertiser

If the DSP wins the bid, then the ad is sent to the publisher and displayed to the user

Source: https://adtechbook.clearcode.cc/media-buying-methods/

Adtech
## Real-time bidding

User matching(用户匹配）

There are different user-identification methods. From web browsers, cookie syncing is used, and match table can be used between SSPs, DSPs, DMPs, etc. Bigtable is well suited for such workloads

User profile store (用户画像)

Fast lookups extract user segments for selecting segments; The store is updated frequently based on the user's interaction with ads, site they visit, or actions they take.

Joining impressions, clicks and conversions

(机器学习）

Perform a key lookup to join data immediately to use real-time data to train models online.
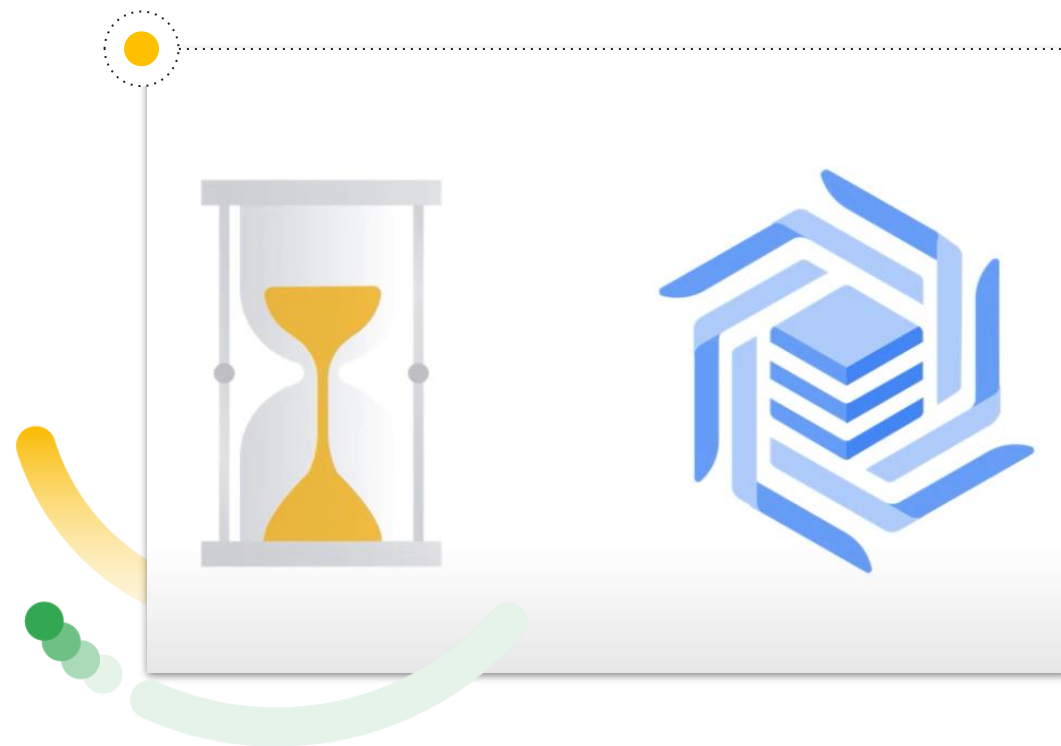
User identification

Web Browsers

AdTech companies use third-party cookie and cookie syncing to identify users across different websites. **Cookie-matching table on Cloud Bigtable** is used for this purpose

Mobile devices

Provide device's advertising ID(e.g. IDFA and AID), but privacy setting allows users to block companies from collecting their advertising iderntifer.
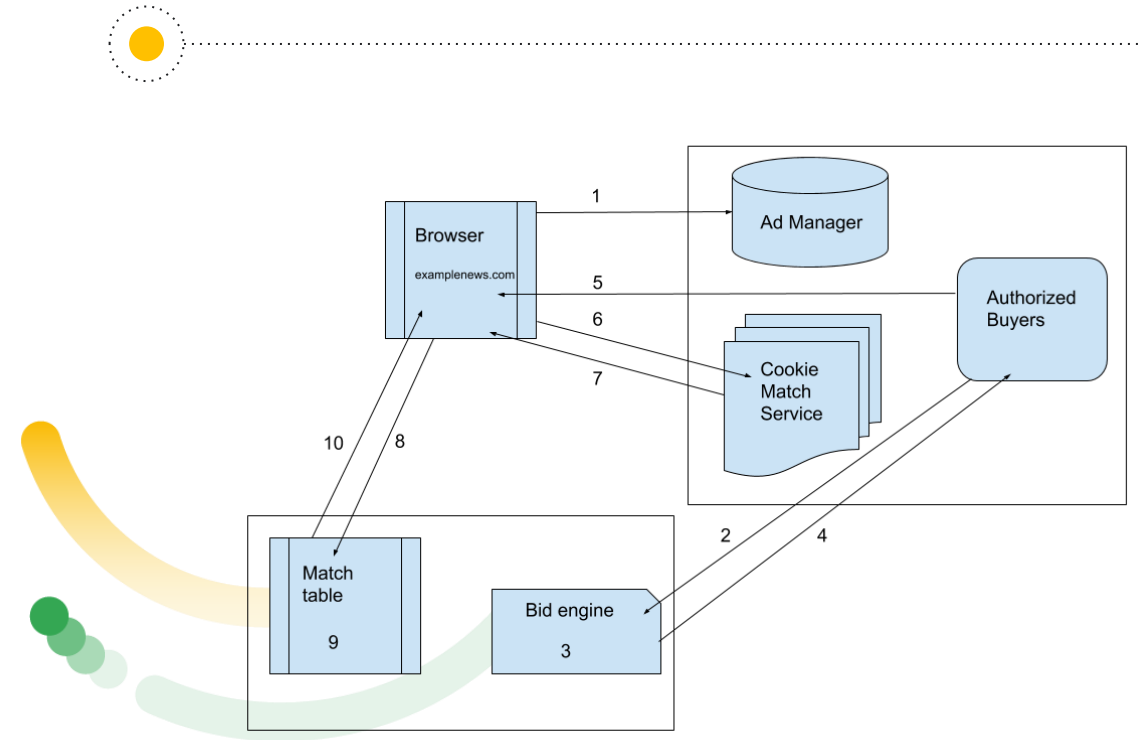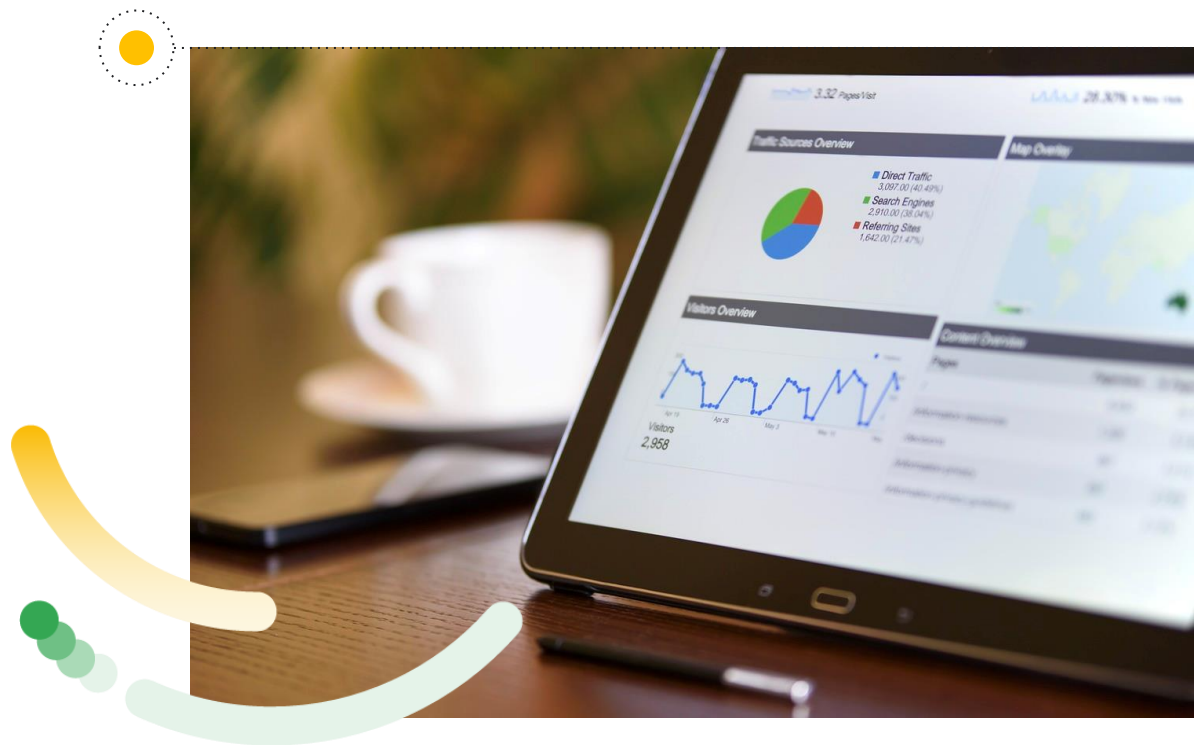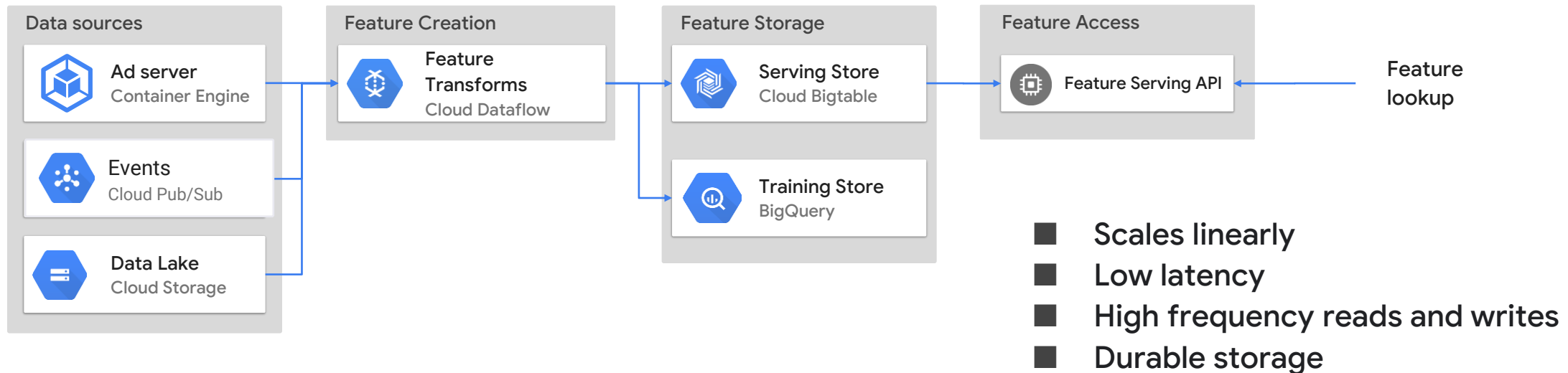
Real-time user profile

Segment or feature

Cloud Bigtable is a distributed low-latency user profile. The user store contains the segment a specific user belongs to, which is used to as inputs for bidding decision logic.

Advertisers use audience segments to target users across the web and mobile platforms.

Real-time prediction

Requests, impressions and clicks are combined with key fields in the request, to be used as input features to the machine learning models. Historical feature values with a longer retention time are stored in Cloud BigQuery for training purposes.

Data sources
- Ad server
  Container Engine
- Events
  Cloud Pub/Sub
- Data Lake
  Cloud Storage

Feature Creation
- Feature Transforms
  Cloud Dataflow

Feature Storage
- Serving Store
  Cloud Bigtable
- Training Store
  BigQuery

Feature Access
- Feature Serving API

Feature lookup

- Scales linearly
- Low latency
- High frequency reads and writes
- Durable storage