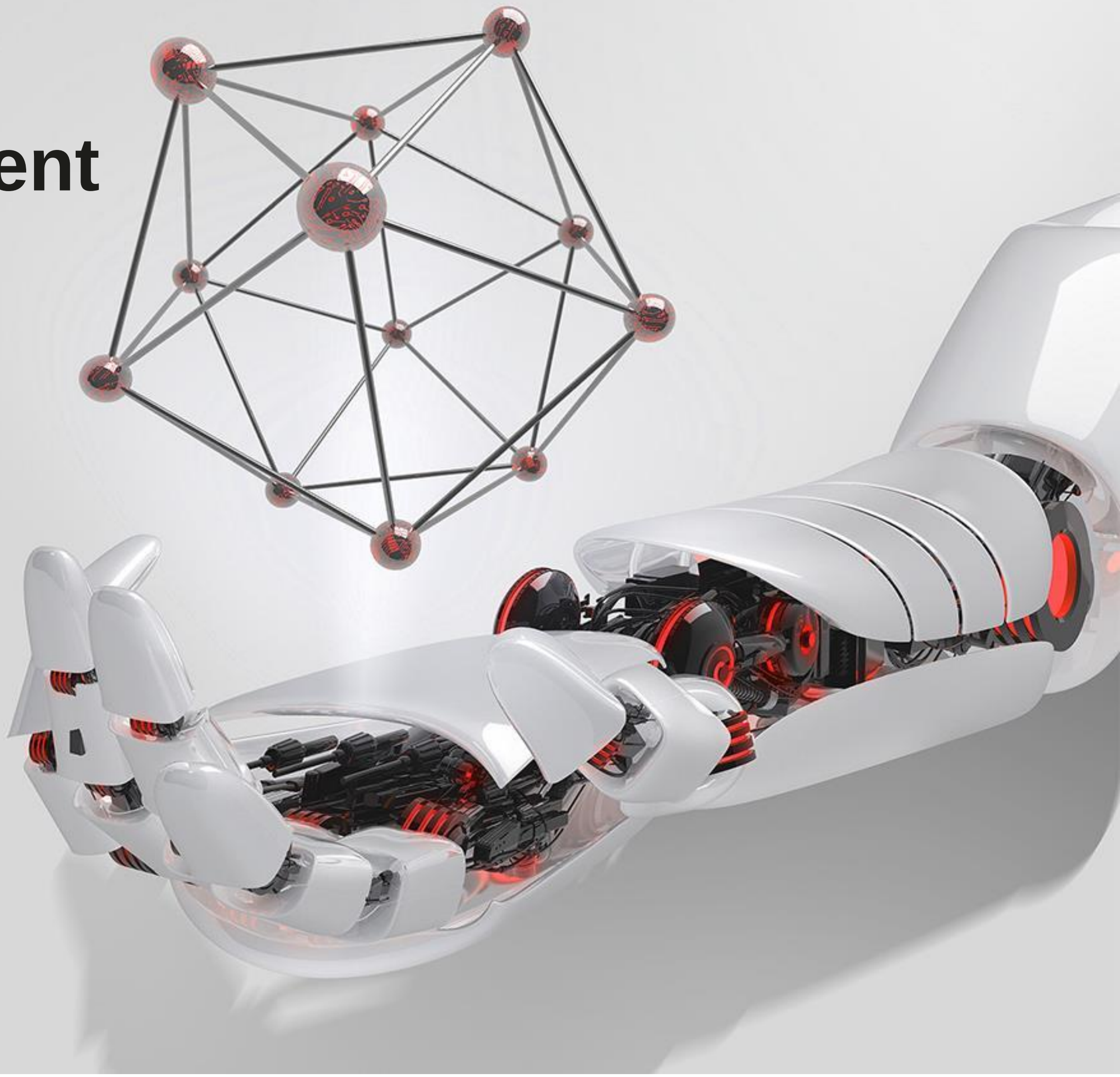
Data Management at Huawei: Recent Accomplishments and Future Challenges



Dr. Jianjun Chen
Technical VP
Huawei America Research

ICDE 2019, Macau

HUAWEI

Disclaimer

Data Management at Huawei: Recent Accomplishments and Future Challenges [ICDE 2019]

- ***Jianjun Chen*** *(Huawei America Research); Yu Chen (Huawei America Research); Zhibiao Chen (Huawei Technologies Co., Ltd.); Ahmad Ghazal (Huawei America Research); Guoliang Li (Huawei Technologies Co., Ltd.); Sihao Li (Huawei Technologies Co., Ltd.); Weijie Ou (Huawei Technologies Co., Ltd.); Yang Sun (Huawei America Research); Mingyi Zhang (Huawei America Research); Minqi Zhou (Huawei Technologies Co., Ltd.)*

Other paper and talk referenced in this talk:

1. FusionInsight LibrA: Huawei's Enterprise Cloud Data Analytics Platform, ***VLDB 2018***, *Huawei America Research*
2. Are Databases Ready for the Cloudification of Telco Systems?, ***ICDE 2016 Keynote by Dr. Götz Brasche***, *Huawei European Research*

Agenda

1. Brief Overview on Huawei Data Management
2. Recent Accomplishments: Huawei MPPDB (IT)
3. Recent Accomplishments: GMDB (CT)
4. Future Challenges: Edge Data Management
5. Conclusions

HUAWEI

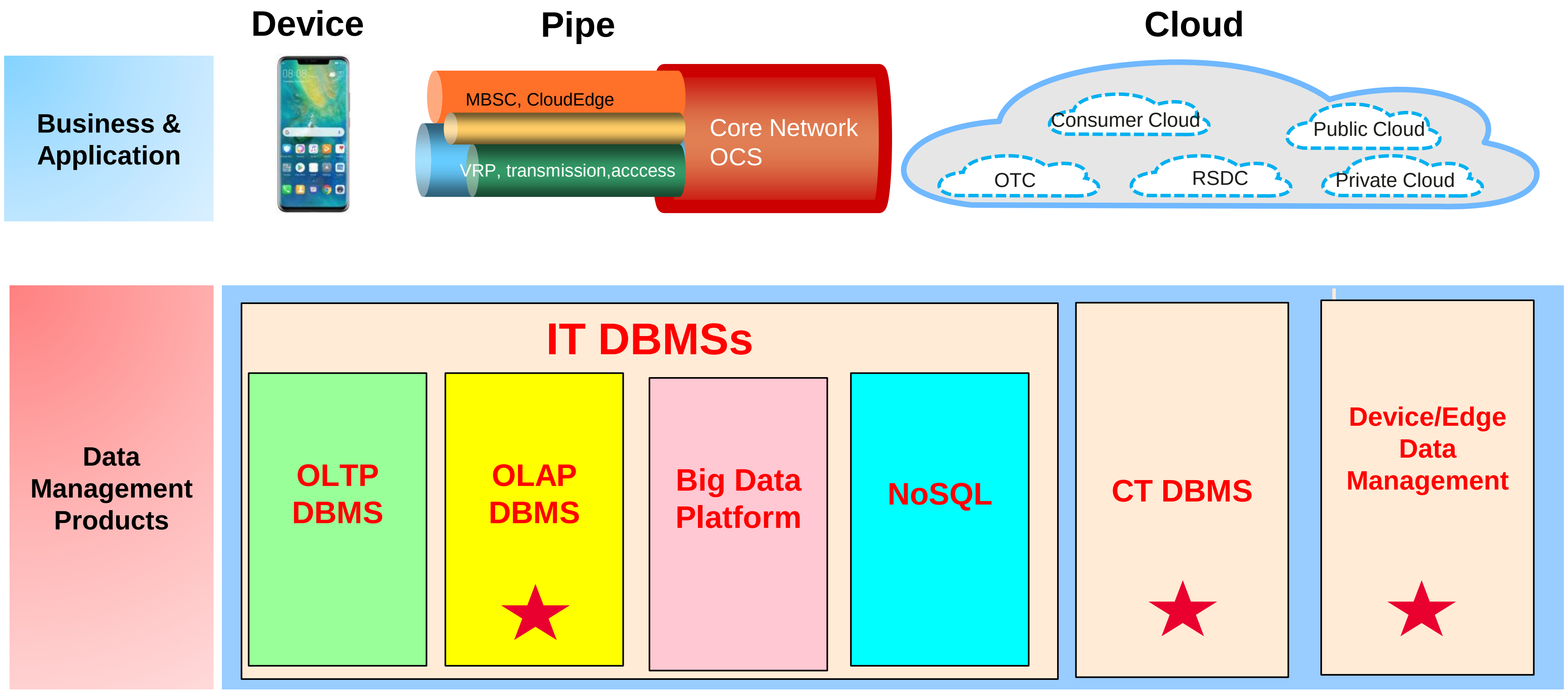
A Glimpse of Huawei

- Founded in 1987, a leading global provider of *information and communication technologies* (*ICT*) infrastructure and *smart devices*
- Provide solutions across four key domains: *Telecommunication networks, IT, Smart devices and Cloud services*
- One of the top 3 global smartphone makers in 2018, with about half of the revenue in 2018 from smart devices
- *Gauss*, Huawei’s database department, belongs to 2012 Lab, Central Software Institute(CSI) and has multiple global research labs

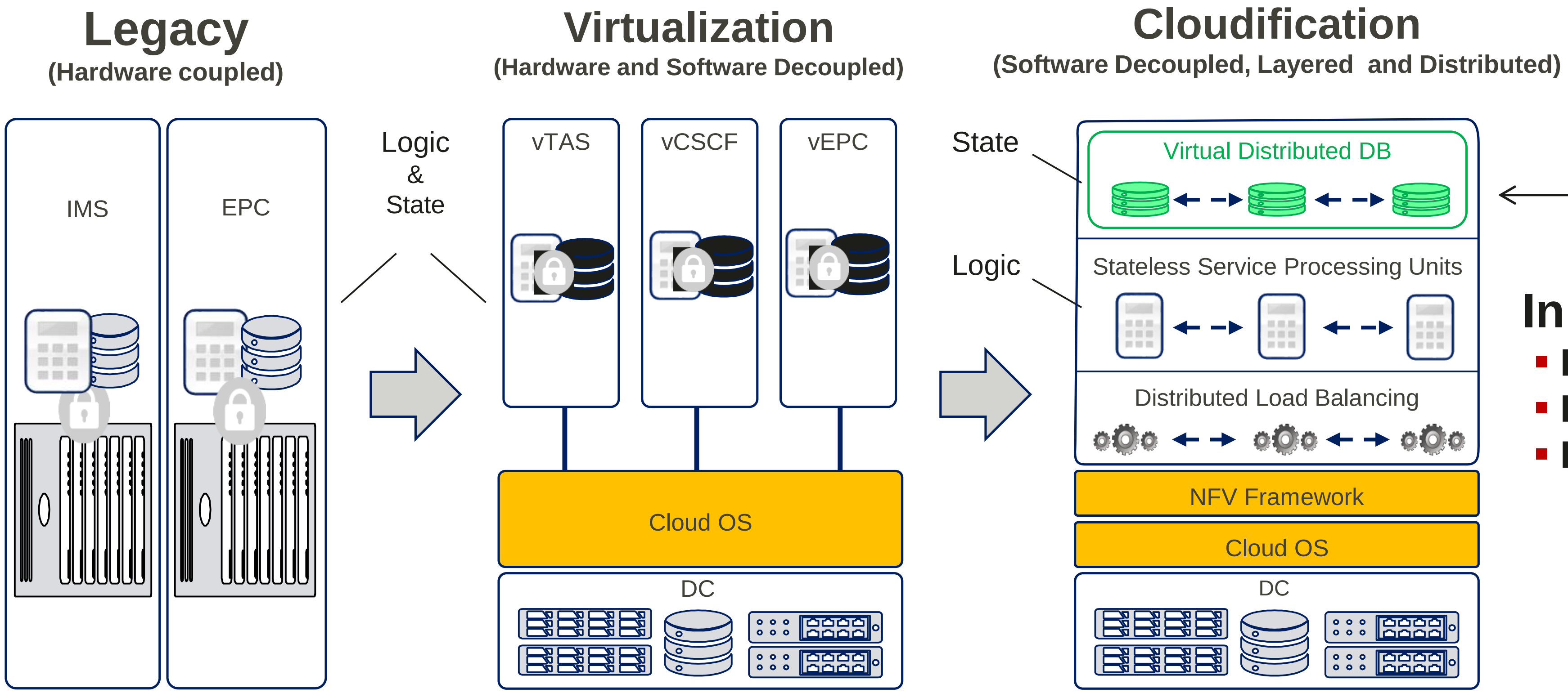
HUAWEI

Huawei Data Management 10000 Foot View

Bring digital transformation to every person, home and organization for a fully connected and intelligent world



Telecom Infrastructure Cloudification

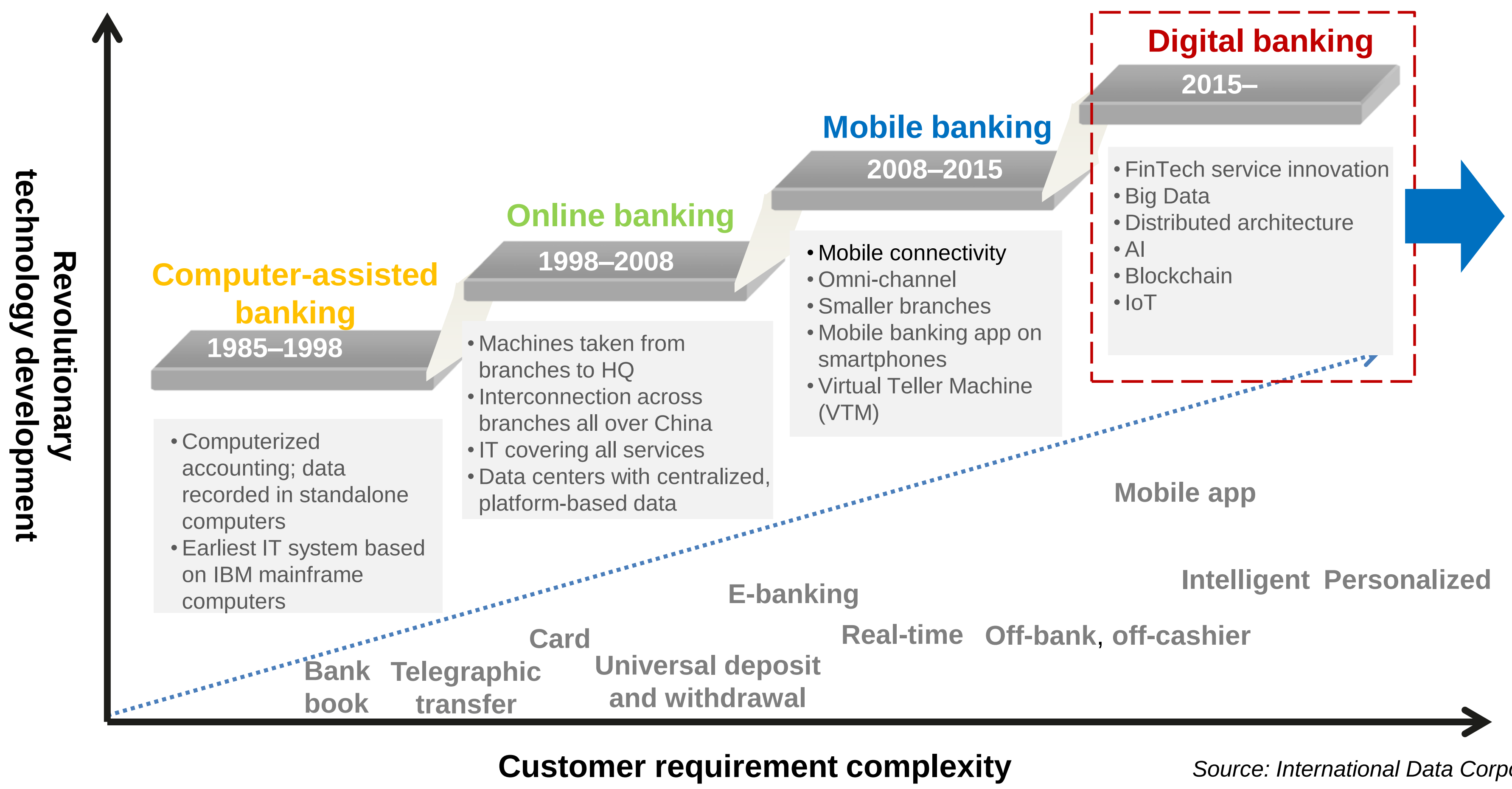


In a cloudified system:
- **Logic & state are separated**
- **Each layer is distributed**
- **Each layer is scalable**

Virtualization is not enough
- **Application and its state tightly coupled**
- **Hard to scale**

HUAWEI

IT Development Trend – China's Banking Industry Has Begun Comprehensive Digital Transformation



Banks' large, centralized IT architecture is becoming **distributed** over commodity hardware

Benefits:

- High available across multiple DCs
- Launch businesses faster
- Lower cost

Source: International Data Corporation (IDC) and Gartner

Global Mobile Traffic Growth Forecast

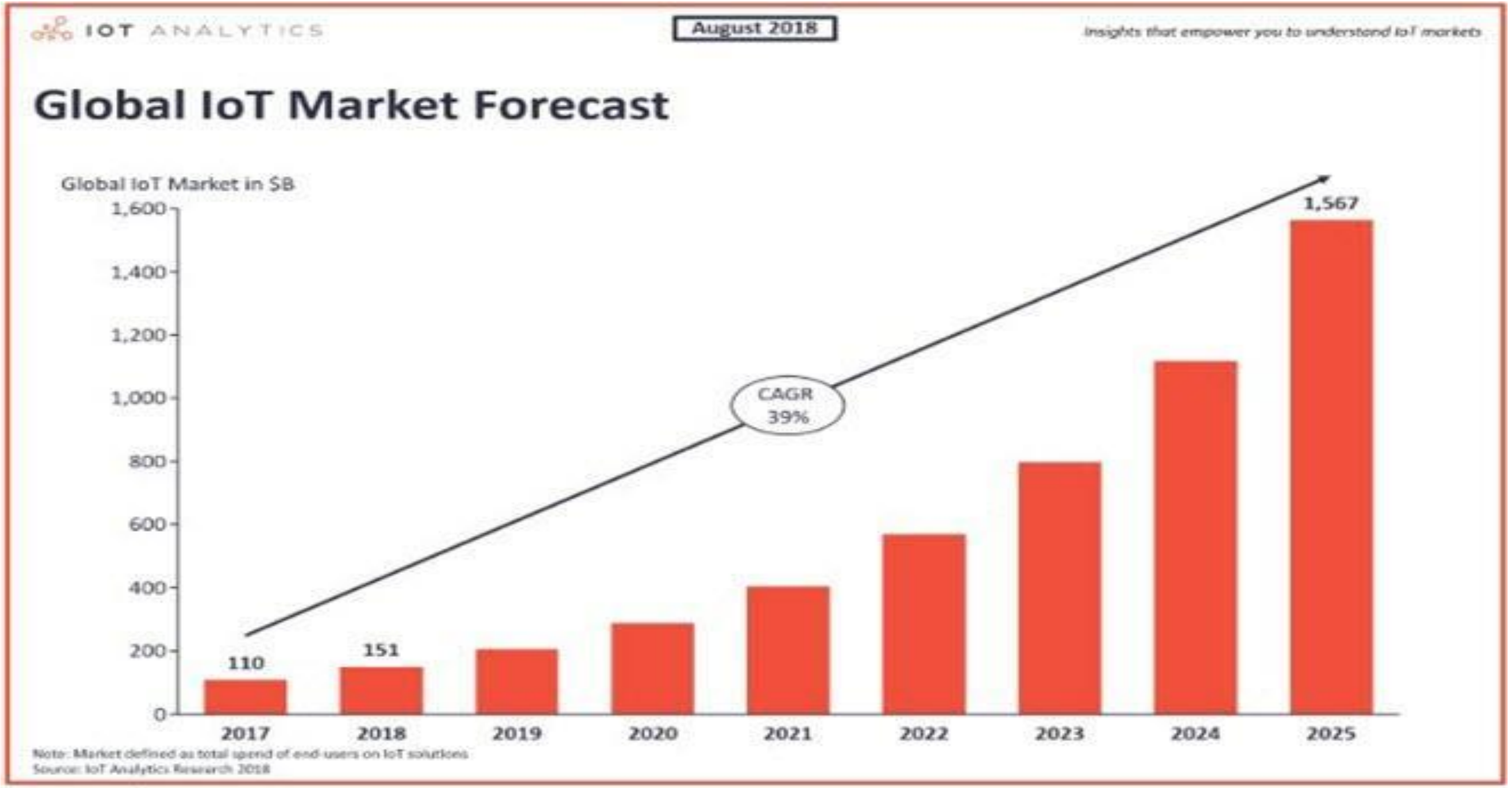
(1 exabyte = 1,000 PB)

- CAGR: compound annual growth rate
- M2M: machine-to-machine

Source:

Cisco virtual networking index: Global data traffic forecast update, 2017-2022 White Paper (https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html#_Toc953327)

Global IoT Market Forecast



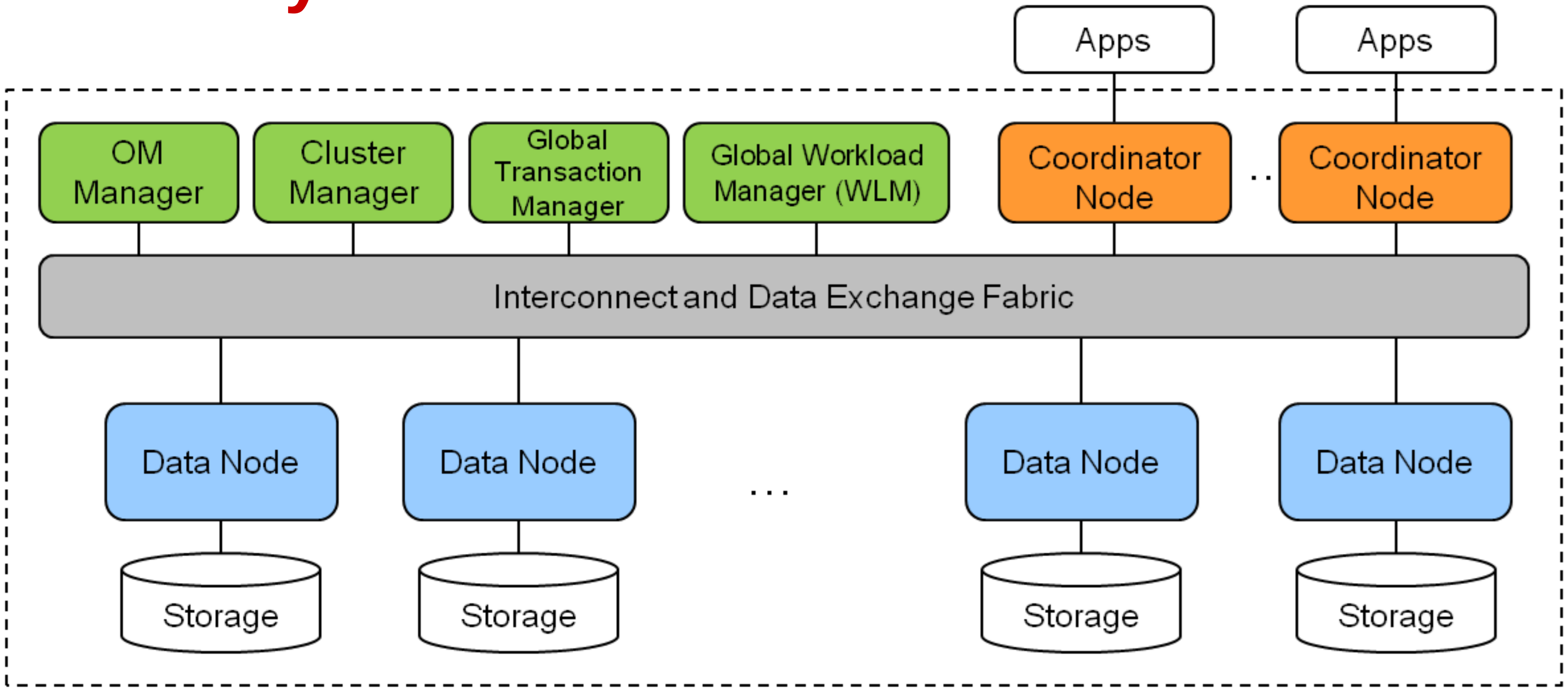
Source: DBS Asian Insights, Internet of Things The Pillar of Artificial Intelligence, June 28, 2018

Agenda

1. Brief Overview on Huawei Data Management
2. Recent Accomplishments: Huawei MPPDB (IT)
3. Recent Accomplishments: GMDB (CT)
4. Future Challenges: Edge Data Management
5. Conclusions

HUAWEI

Huawei MPPDB System Overview



- Shared-nothing architecture with partitioned data nodes
- Support both row and columnar stores
- Support both on premise (FusionInsight MPPDB) and on cloud (FusionInsight LibrA)
- **Global transaction manager(GTM)** and 2PC for cross-partition transactions
    - ✓ Currently support Read Committed isolation level

HUAWEI

Recent Accomplishments in Huawei MPPDB

- **GTM-Lite: scalable distributed transaction processing for HTAP workload**
- Multi-modal analytics: unified analysis over multiple data models
- Auto tuning: adaptive query optimization with ML algorithm

Note many recent projects not covered in this talk

  - *E.g. Industrial-strength OLTP using main memory and many cores [HardDB & Active 2019]*

HTAP Motivation

- Huawei HTAP (hybrid transaction analytical processing) scenarios:
  - Real-time operational reporting
  - Bank fraud detection
  - Campus security monitoring
- Benefits of HTAP
  - Enable quick decision on recent data
  - Reduce operational cost by only maintaining one system
  - Eliminating data movement across OLAP and OLTP

GTM-lite Motivation

Problem: GTM can become system bottleneck over HTAP workload

- Each transaction requires multiple rounds of interactions with GTM, including
  - Generate global XID and acquire global snapshots (visibility check at DN)
  - Transaction complete
- Not designed for executing a large number of concurrent simple OLTP queries

Observations

- Many customers' OLTP transactions only access single shards
  - ✓ E.g. Users of Internet banking most time access their own data
- Irrelevant transactions don't care about the ordering

GTM-lite strategy: bypass GTM for single-shard transactions

HUAWEI

GTM-lite Design and Challenges

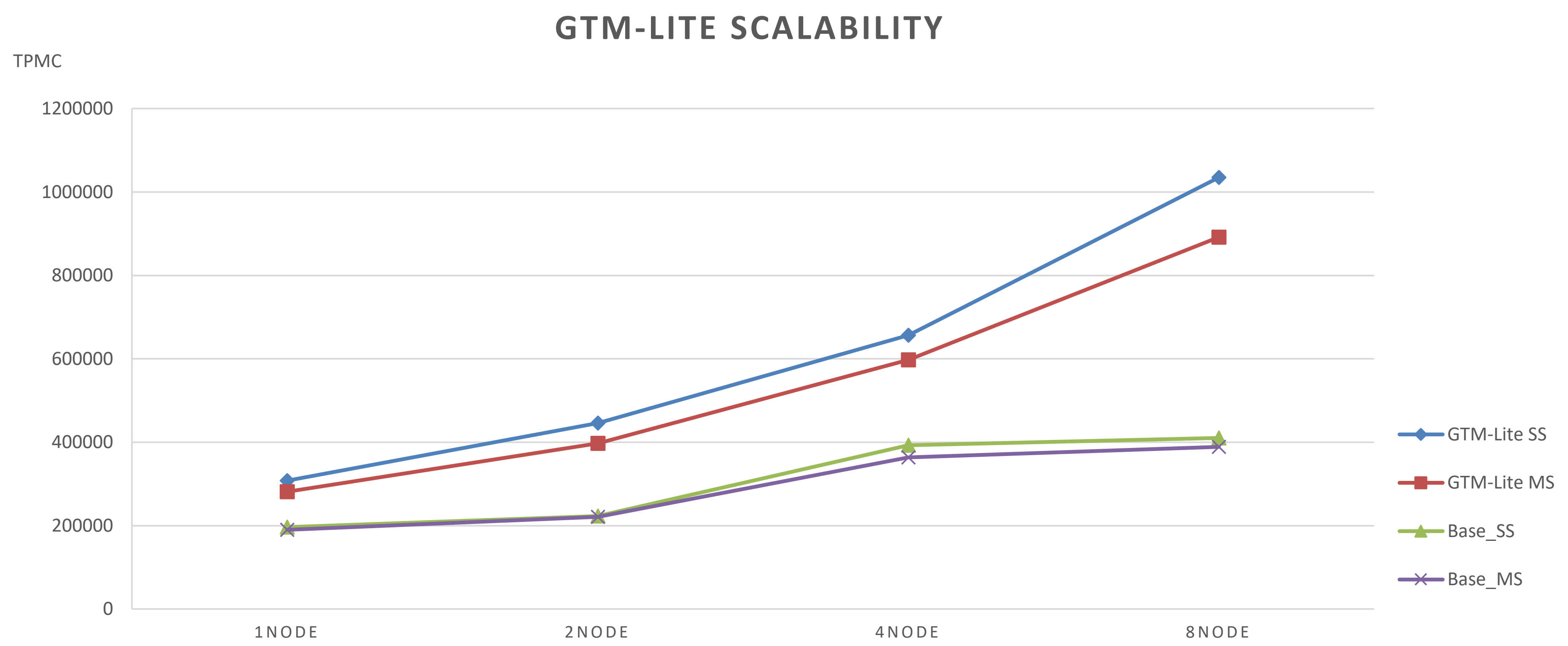
- DN performs visibility check using local snapshots instead of global snapshots
- Single-shard transactions only acquire local snapshots on DNs
- Multi-shard transactions still acquire global snapshots as before

Challenges: Local snapshots and global snapshots can have conflicting views of transaction ordering

  - For multi-shard transactions, DN need to adjust their local snapshots based on global snapshots
  - Two anomalies and their solutions presented in [ICDE 2019]

HUAWEI

GTM-lite Performance Results



- Modified TPC-C workload to control % of single shard transactions
  - SS: 100% single-sharded
  - MS: 90% single-sharded, 10% multi-sharded

HTAP/GTM-Lite Summary

HTAP allows data analysis over fresh data and reduces system maintenance overhead

GTM-Lite provides scalable distributed transaction processing for HTAP workload

- Bypass GTM for single shard transactions
- Work well for workloads with majority of single shard transactions

Future work:

- Comprehensive GTM-Lite performance study
- HTAP support in other DB kernel components
  - e.g. workload management, query optimization etc

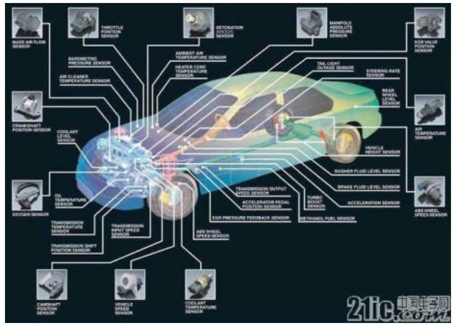
HUAWEI

Recent Accomplishments in Huawei MPPDB

- GTM-Lite: efficient distributed transaction processing for HTAP workload
- **Multi-modal analytics: unified analysis over multiple data models**
- Auto tuning: adaptive query optimization with ML algorithm

Multi-modal DBMS Motivation

GPS & Map

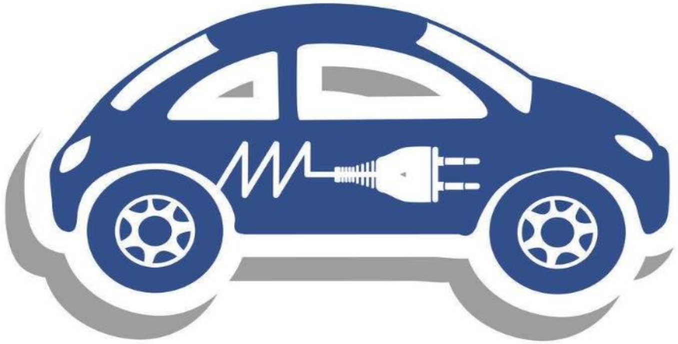


Sensors

Autonomous Vehicles

Locations & Road Networks

Video and Objects



Running Status

Shape & Distance of Objects

Camera

Radars & Lidars

Self-driving cars need to collect & process information from many sources

- E.g. external sensors (e.g. cameras, radars and lidars) and car status (e.g. battery and speed)
- Need to perform **unified** analysis over multiple data sources

HUAWEI

Multi-modal Query Example

```
With Obstacles(id, name, size) as  gvideo('select id, name, size from objects where
pixel(size) > rectangle(x,y)')
With Obstacle_distance(id,name,distance) as gpointcloud('select id, name, distance,
where pointcloud where pixel(size) > rectangle(xy)')
select
        rule.id, rule.action
from
        rule,
        (select
                map.grid, map.road
        from
                map,
        where current.loc = map.grid ) as road,
        gtimeseries('select minute, avg(speed) from speed group by minute having
minute = current_minute()' as speed(min, speed_avg)
where
            Obstacle.size > rectangle (a,b)
            and Obstacle_distance.distance / speed. speed_avg < 10
            and rule.name =obstacle.name
```

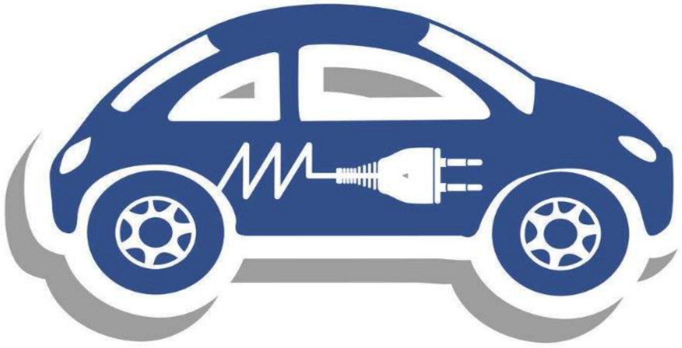
Find the obstacle based on the video and pointcloud data

Find the current avg speed of the the car

Find the self-driving car

Find the self-driving rule based on:

1. Obstacle in front of the car
2. The road info based on the map and current location
3. The speed of the car



Advantages of a Multi-modal Database System

- Simplify application development and database maintenance
  - Only one API to learn and one system to maintain
- Fast query processing through integration
  - Avoid unnecessary materializing intermediate results and data passing
  - Provide opportunities for generating a globally optimized plan

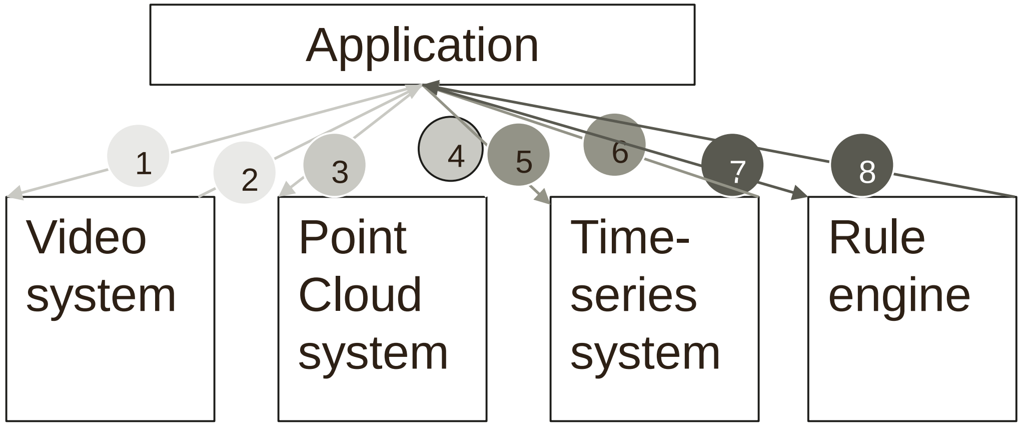
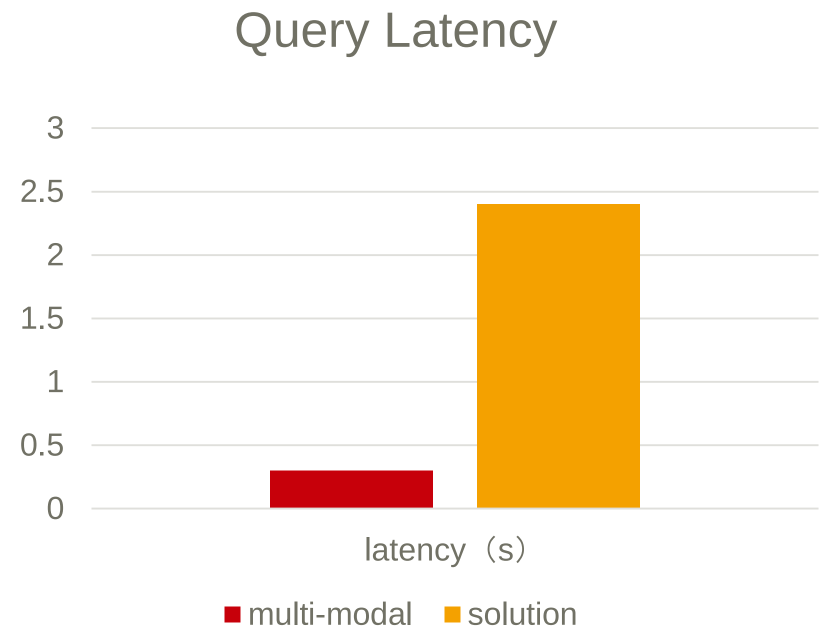
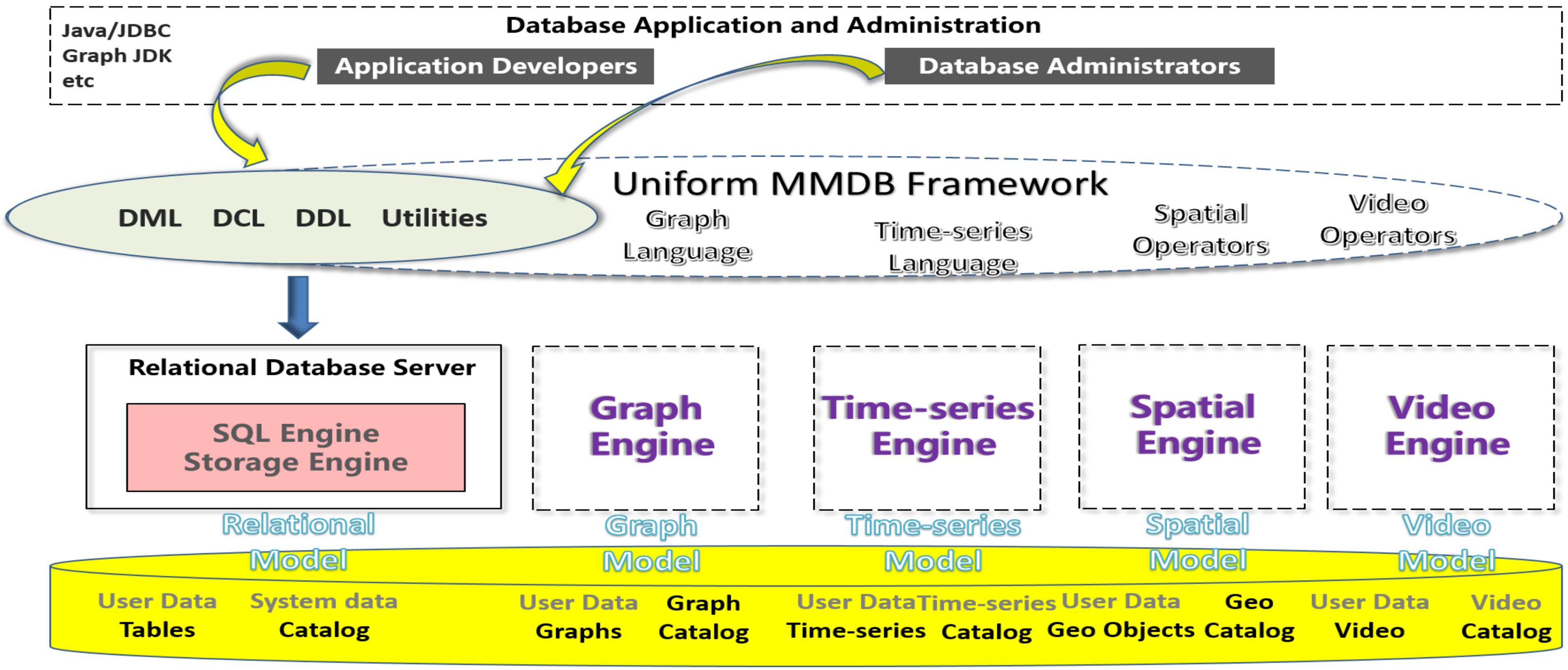


Illustration of a typical solution-oriented architecture over a set of separated systems



HUAWEI

Architecture of Huawei MPPDB with Multi-modal Extension



Extended MPPDB to support multiple processing engines

- Unified query language interface
- Engines are dynamically pluggable, e.g. Time-series, Spatial, Graph
- Unified data store, currently using RDBMS

HUAWEI

Multi-modal analytics Summary

Provides an unified interface with pluggable engine support

- More efficient query processing
- Less overhead in overall system maintenance

Future work

- Global query optimization across data models
- Understand better of the limit and trade-offs for adding new models

Recent Accomplishments in Huawei MPPDB

- GTM-Lite: efficient distributed transaction processing for HTAP workload
- Multi-modal analytics: unified analysis over multiple data models
- **Auto tuning: adaptive query optimization with ML algorithm**

Motivation for an Auto Tuning Query Optimizer

Performance tuning is **hard**

- ❑ Require significant expertise: sending senior engineers to customer sites
- ❑ Could take days for tuning customer workload at initial deployment time

Selectivity estimation is one key factor to generating good plans

- ❑ Calculated based on statistics on base tables and models
- ❑ Estimation error can be large, due to inaccurate statistics, skew, predicate correlation etc

Our goals

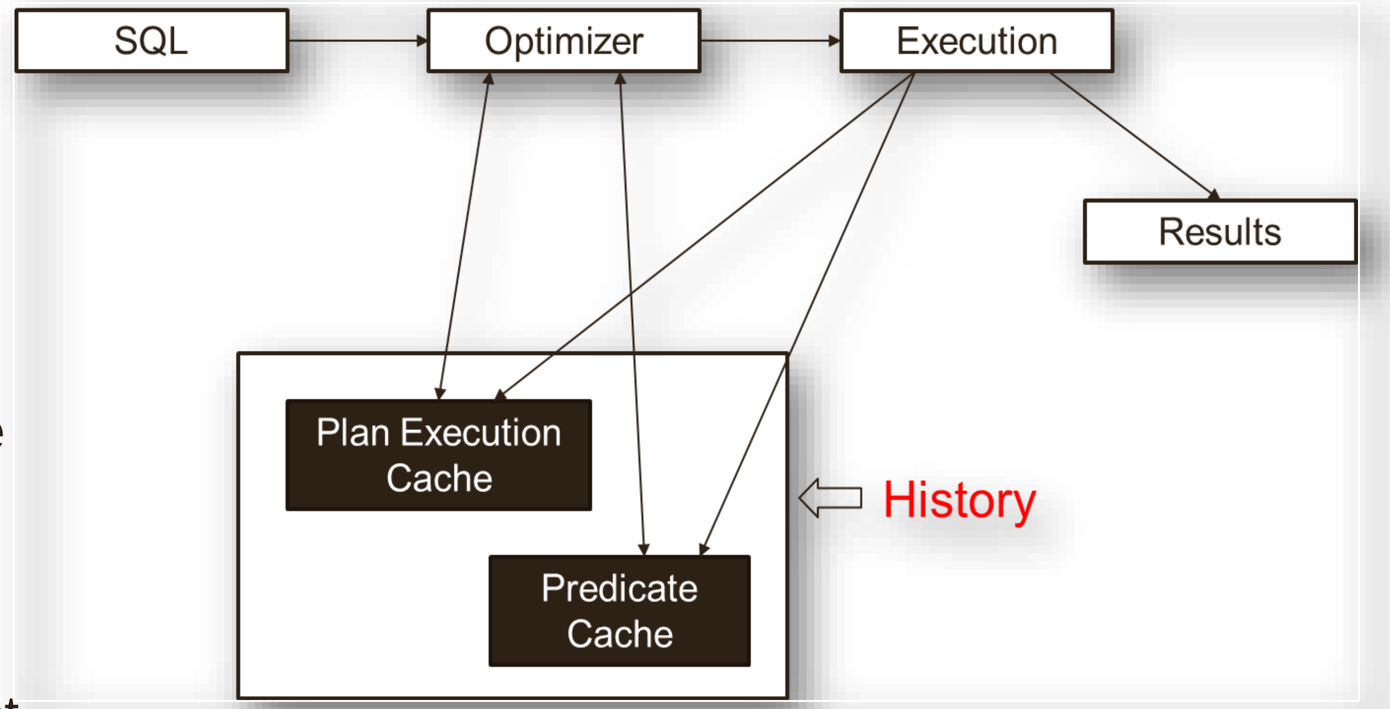
- Build an auto tuning query optimizer based on execution history
- Reduce the initial tuning time from days to hours

HUAWEI

Adaptive Selectivity Estimation with Machine Learning

Key Ideas

- Store actual selectivity from execution for future query optimization
- Matching based on logical plan signature



Plan Execution Cache

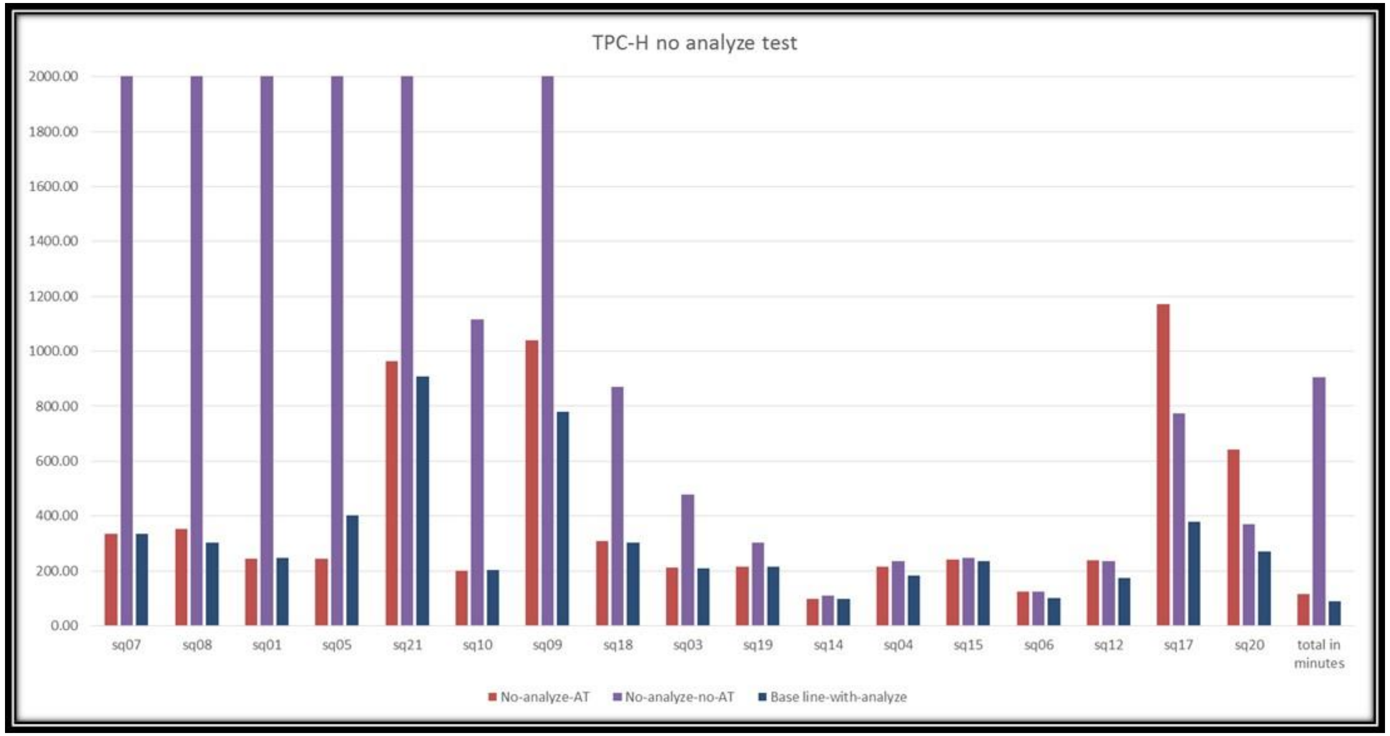
- Exact plan match
- Effective for reporting workloads
- Currently handles select, join and aggregate etc

Predicate Cache

- Use of KNN (K nearest neighbor) to find best match over similar predicates
- Currently only handles selection predicates

HUAWEI

Experimental Results: plan execution cache



- 1TB TPCH without collecting stats over the Lineitem table to show the effectiveness of adaptive optimization
- 16X improvement overall in total runtime

Experimental Results: predicate cache [VLDB 2018]

Single parameter

Cache Size	Cache Hits	KNN Error
100	218	1.2%
50	118	2.0%
25	65	3.6%

Two parameters

Cache Size	Cache Hits	KNN Error
100	4	3.1%
50	1	4.0%
25	1	6.0%

- **Predicate cache**
  - K=5
  - One parameter (constant) experiment
    - Early line items : l_receiptdate <= l_commitdate – c days, c is between 1 and 80
    - Late line items: l_receiptdate > l_commitdate + c days, c is between 1 and 120
    - **Error with KNN less than 4%**
  - Two parameters (constants) experiment
    - Combine early and late lineitems as range
    - l_receiptdate between l_commitdate – c1 and l_commitdate + c2
    - Where C1 range from 1 to 80 and C2 range from 1 to 120
    - **Error with KNN less than 6%**

HUAWEI

Auto Tuning Summary

Adaptive query optimization with ML algorithm

- Greatly improves the selectivity estimation accuracy
- Significantly reduces the labor cost of performance tuning

Future work - a lot of them!

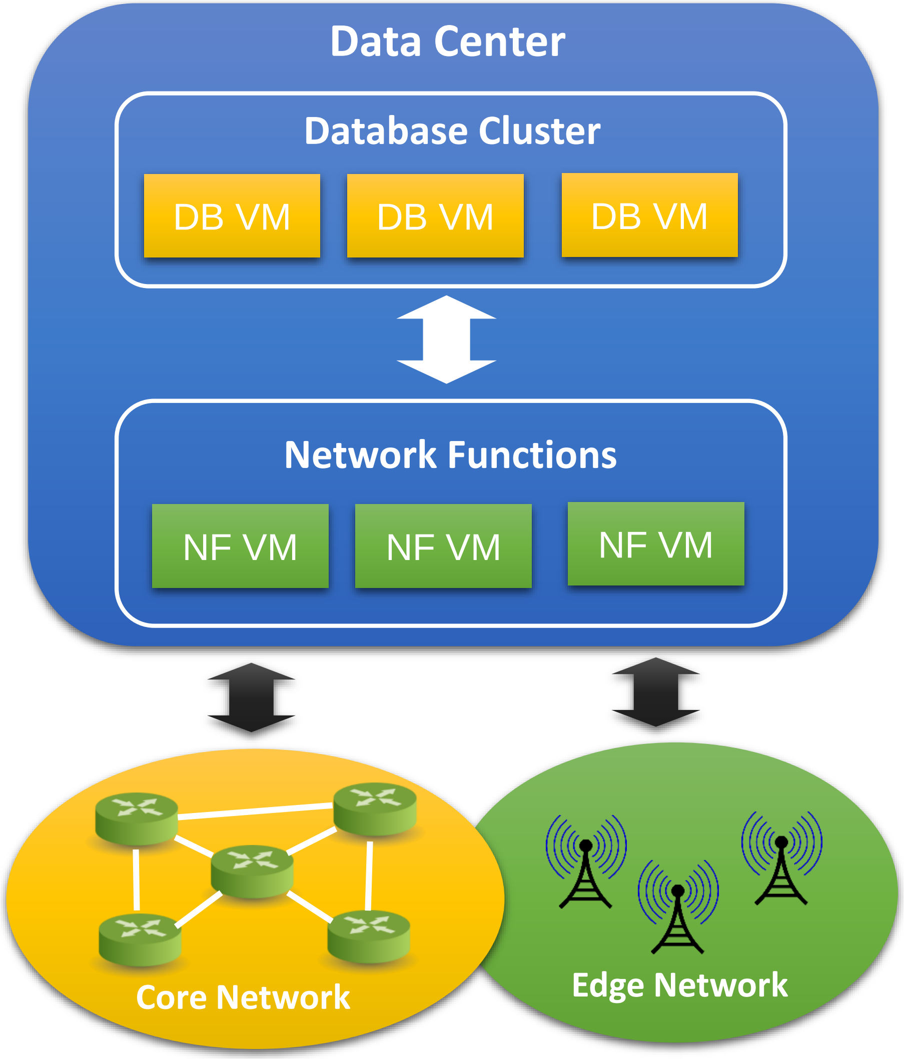
- How to guarantee auto-tuning process converges
- Manage plan regressions
  - Sometimes better selectivity estimation may not result in better plans
- Extend beyond query optimization
  - E.g. predict workload pattern for better workload management

Agenda

1. Brief Overview on Huawei Data Management
2. Recent Accomplishments: Huawei MPPDB (IT)
3. Recent Accomplishments: GMDB (CT)
4. Future Challenges: Edge Data Management
5. Conclusions

HUAWEI

Network Function Virtualization (NFV)



- Transformation in Telecom:
  - Specialized hardware with monolithic software ==> virtualized network functions on commodity hardware
- Scalability
  - Independent scalability database and network functions
- High availability (> 99.999%)
- Cost saving
  - Using commodity hardware
  - More efficient utilization of resources

HUAWEI

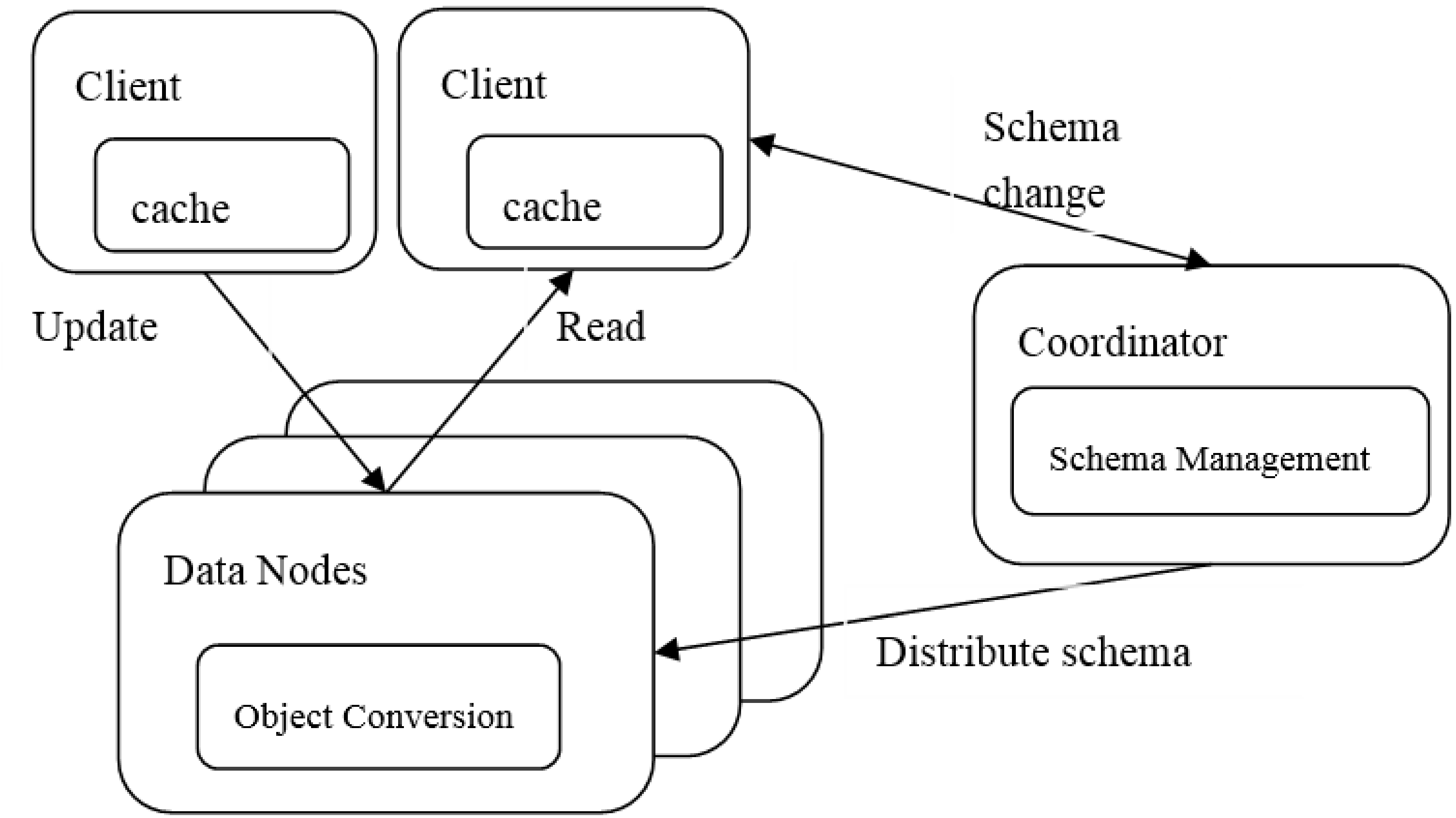
A Telecom DB Differs from an IT DB in Many Aspects

- **Non-Relational Data Model, Hierarchical Objects**
  - Best modeled as tree-structured objects
- **Extreme low latency for simple operations (i.e. microseconds)**
  - Most are simple access by keys
  - SQL is nice to have for analytics
- **Relaxed, configurable ACID**
  - Multi-object transaction not supported
  - Controllable data durability: volatile – checkpoint only, async. / sync. logging
- **Low resource consumption**
  - Used in different hardware scenarios (e.g. edge/core networks)
  - Telecom business pretty cost sensitive
- **High availability (> five 9's) and scalability**
- **Publish/Subscribe system for data change notifications**

⇨ **GMDB:** **a** distributed in-memory KV store

HUAWEI

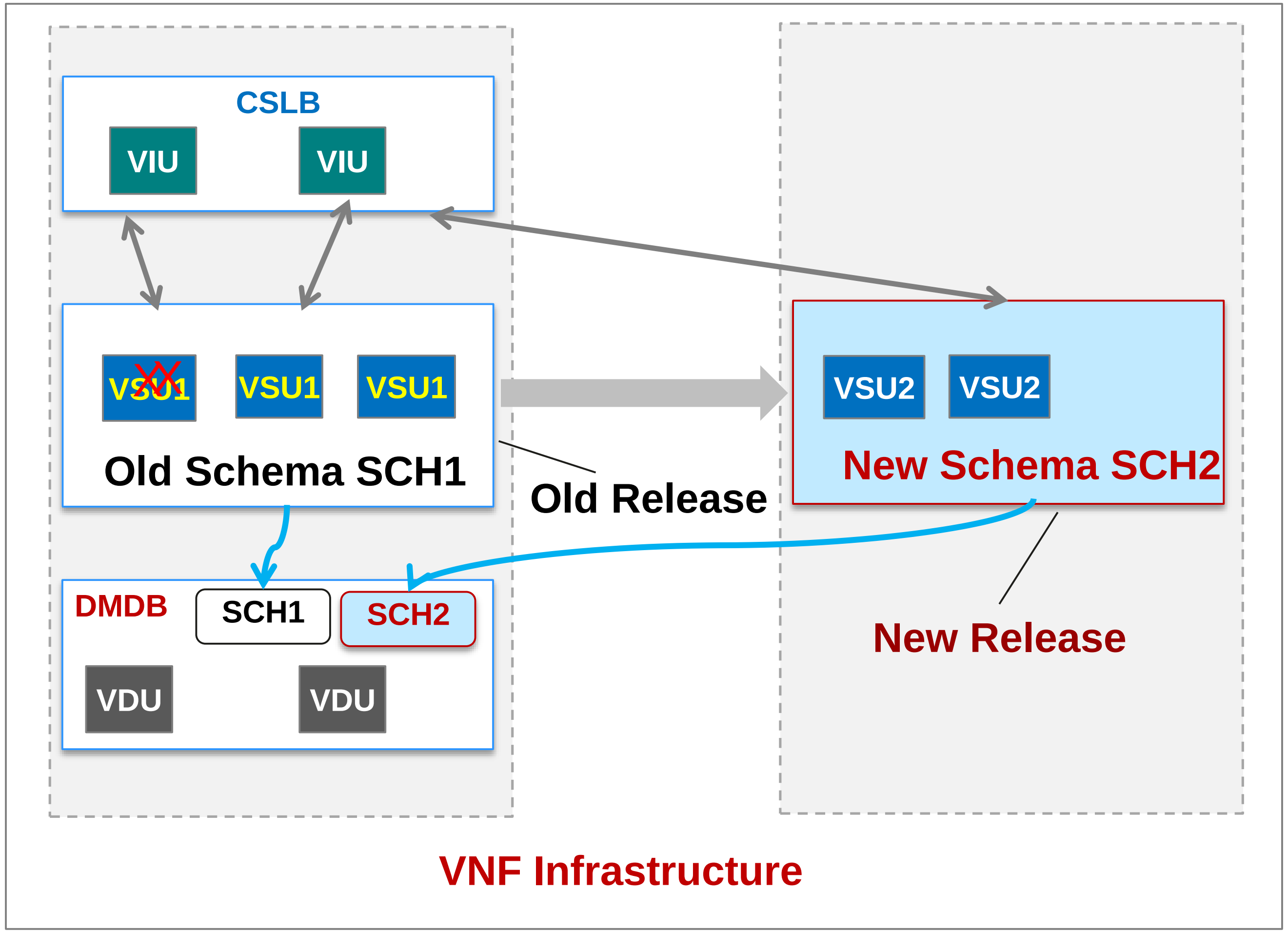
GMDB: Huawei's Distributed In-memory DBMS for Telecom



- **Coordinators** handle system metadata and app schema changes
- Data are partitioned and stored in **Data Nodes** (in-memory KV store)
- Each partition allows one writer and multiple readers
- Client maintains a **local cache** that subscribes to relevant data changes

Online Schema Evolution

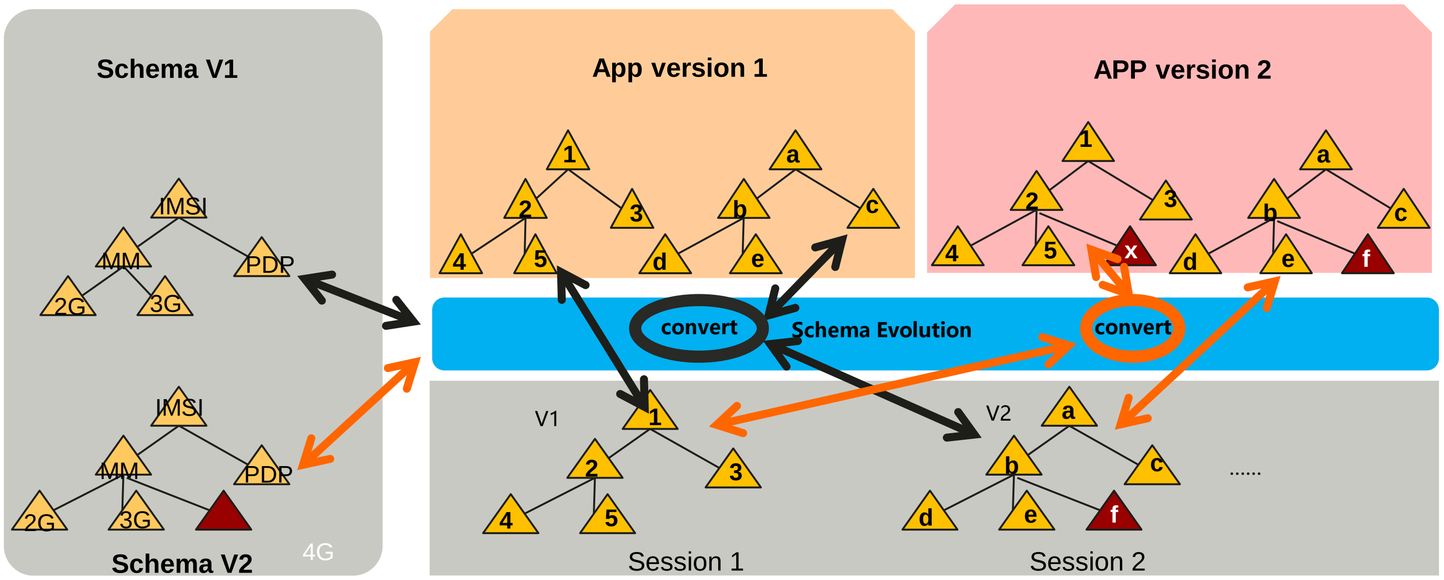
- New versions of an application may require new schemas
  - E.g. adding new fields and nodes
- Several versions of the schema can co-exist in the database
- Goal: no system downtime during application upgrades



HUAWEI

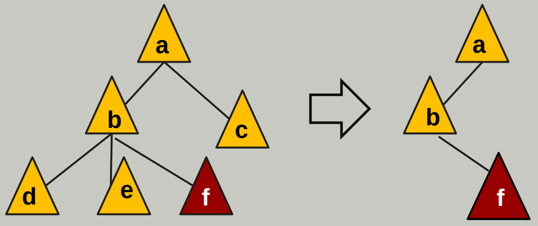
Dynamic Data Conversion at Data Nodes

- Data nodes only keeps one copy of data: no data redundancy
- System maintains schema history information
- Data nodes automatically converts data based on versions when serving requests
- Only delta data changes are synced across clients and database



Pub/sub based on delta：

- **Transfer data is about 15% of original data size**
- **E2E delay is about 2ms**



GMDB Summary

GMDB: a distributed in-memory KV database for CT

- High availability
- Extreme low latency
- Relaxed ACID
- Low resource consumption

Online schema evolution

- Support online application upgrade
- Dynamic convert data with different schemas at Data Nodes

Future work:

- Support more types of online schema changes
- Hardware acceleration to further reduce latency and CPU consumption

HUAWEI

Agenda

1. Brief Overview on Huawei Data Management
2. Recent Accomplishments: Huawei MPPDB (IT)
3. Recent Accomplishments: GMDB (CT)
4. Future Challenges: Edge Data Management
5. Conclusions

Motivation of Edge Computing

Cloud computing paradigm

- Both data and computation are in cloud
- Cloud computing benefits: high available, high elastic, no/low management overhead, pay per use etc

IoT devices become pervasive in our lives

- E.g. smart phones, video surveillance cameras, autonomous vehicles etc
- Producer/consumer of huge amount of data

Advantages of edge computing over cloud computing

- Low latency
- Reduce network bandwidth consumption
- Better privacy by avoid sending sensitive data to cloud

HUAWEI

Integration between Cloud Computing and Edge Computing

Motivating example using MBaaS(Mobile Backend as a Service):

- ❑ Existing MBaaS only supports device to device data sync through cloud
- ❑ Benefits for smart devices in vicinity to collaborate using local network
  - Lower latency
  - Not dependent on the Internet
  - Better privacy, e.g. users may not want to upload some data to cloud
- ❑ We are building an MBaaS system that supports both
  - Cloud data sync
  - Direct data sync in a wireless ad hoc network

Many application scenarios:

- ❑ Community video surveillance systems
- ❑ Smart homes
- ❑ Autonomous vehicles

HUAWEI

Vision: Ubiquitous Computing

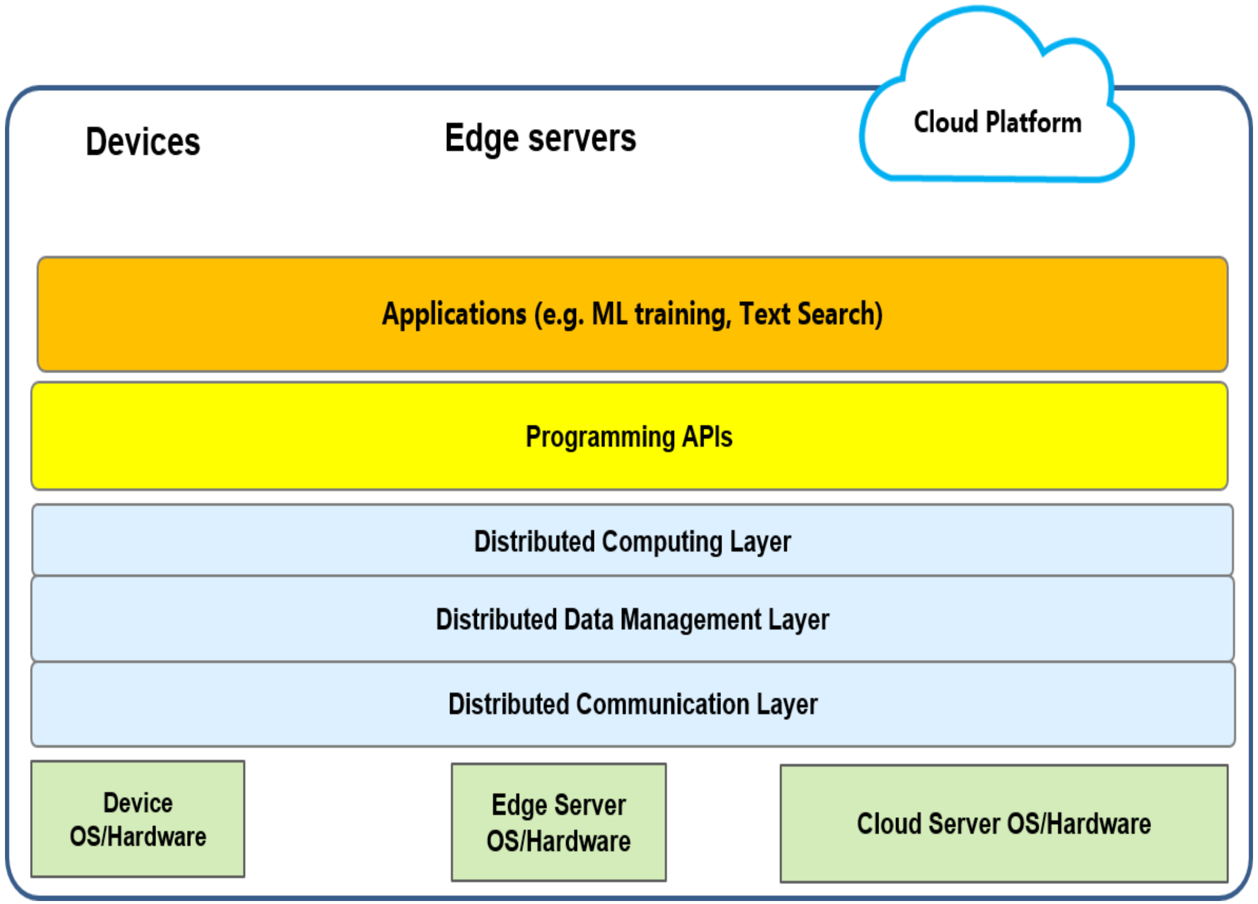
Distributed computing across device, edge and cloud

- Entities treated equally as nodes in the system
- Unified programming API
- Distributed networking, data and computing layers

System automatically decides

- Data location/replicas
- Program scheduling and execution strategy
- Right communication protocols

Can be provided as a service



HUAWEI

Challenges for Ubiquitous Data Management

Ubiquitous data access

- Data can be in different format and moved around
- Devices can go offline any time and be dynamically added and removed

Heterogeneous:

- Nodes can have very different hardware capabilities
- Nodes can run different operating systems, networking protocols etc

Real-time:

- Huge amount of data generated at edges, e.g. AR/VR, autonomous vehicles
- 5G can help, but many existing system assumptions may need to change

Secure:

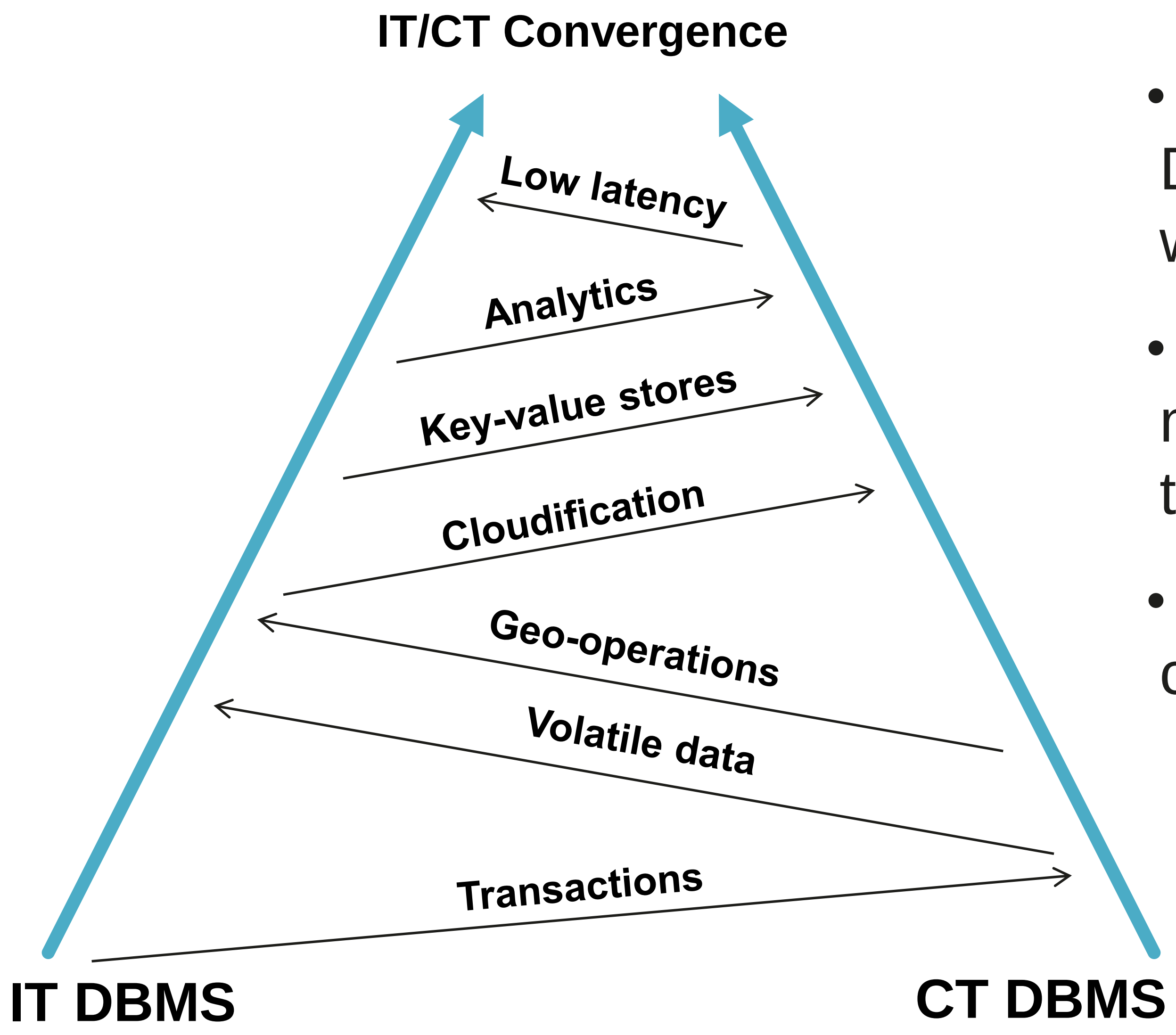
- Given user data can be distributed in many places, how to guarantee no privacy violation?
- Given user data owned by different entities, how to build a trustworthy system?

HUAWEI

Agenda

1. Brief Overview on Huawei Data Management
2. Recent Accomplishments: Huawei MPPDB (IT)
3. Recent Accomplishments: GMDB (CT)
4. Future Challenges: Edge Data Management
5. Conclusions

Takeaway I: IT/CT Convergence also in DBMS



- With the constant advancement of DB technologies used in IT and CT, we see a clear convergence
- The convergence is actually a mutual borrowing and improvement of technologies
- Thereby increasing further the chances to borrow and reuse

HUAWEI

Takeaway II: Server DBMS Transitioning to Ubiquitous Data Management

Device and Edge becoming more and more important

- Cloud centric => Ubiquitous across device, edge and cloud

Application scenarios become much more versatile

- The boundary between DBMS and applications becomes more vague
- New abstraction and common middle tiers can be useful

Core DBMS concepts and technologies still useful

- E.g. declarative APIs, distributed transactions, query optimization etc

Many DBMS research problems may become 100X harder in the ubiquitous computing environment

- E.g. heterogeneity, elasticity, scalability, security etc

HUAWEI

Takeaway III: Huawei Data Management Progress and Future

1. Centralized databases	Distributed databases
2. On-premise, box software	Cloud database services
3. Individual analytics	Multi-modal analytics
4. Labor intensive DB management	Auto tuning with ML
5. Server databases	Ubiquitous data management

Recent Huawei database related publications/talks:

1. ***Data Management at Huawei: Recent Accomplishments and Future Challenges,*** ***ICDE 2019***
2. ***FusionInsight LibrA: Huawei's Enterprise Cloud Data Analytics Platform,*** ***VLDB 2018***
3. ***Fiber-Based Architecture for NFV Cloud Databases,*** ***VLDB 2017***
4. ***Are Databases Ready for the Cloudification of Telco Systems?,*** ***ICDE 2016 Keynote by Götz Brasche***

HUAWEI

Thank you!
Q&A