Alibaba Cloud

AnalyticDB-V
A Hybrid Analytical Engine Towards Query Fusion for Structured and Unstructured Data

Chuangxian Wei, Bin Wu, Sheng Wang, Renjie Lou,
Chaoqun Zhan, Feifei Li, Yuanzhe Cai

Chuangxian Wei

Director, Alibaba Cloud
August, 2020

Alibaba Cloud

1. Background

2. System design

3. Optimization

4. Evaluation

Alibaba Cloud

1. Background

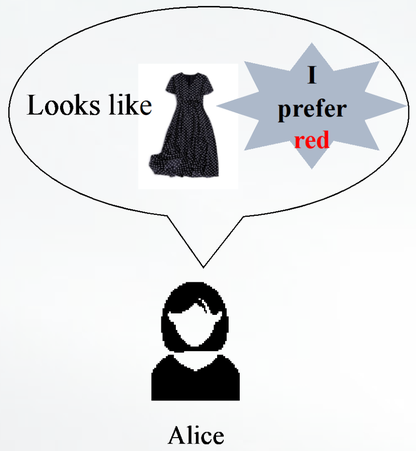
2. System design

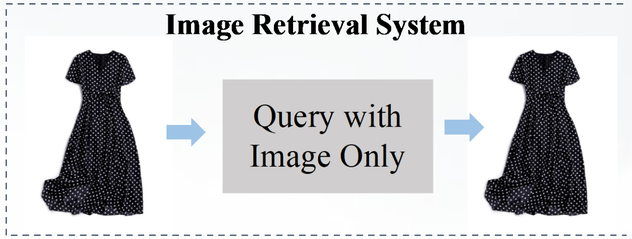
3. Optimization

4. Evaluation

Alibaba Cloud

1.1 Background





	Price	Shipment	Rating	Color	Style
	< $100	free-shipping	> 4.5 (of 5)	Red	Movie star style
Image Retrieval System	✗	✗	✗	✗	✓
DBMS	✓	✓	✓	✓	✗

阿里云 aliyun.com

盒马 鲜·美·生活

淘

Alibaba Cloud

1.2 Background

AI

High dimension

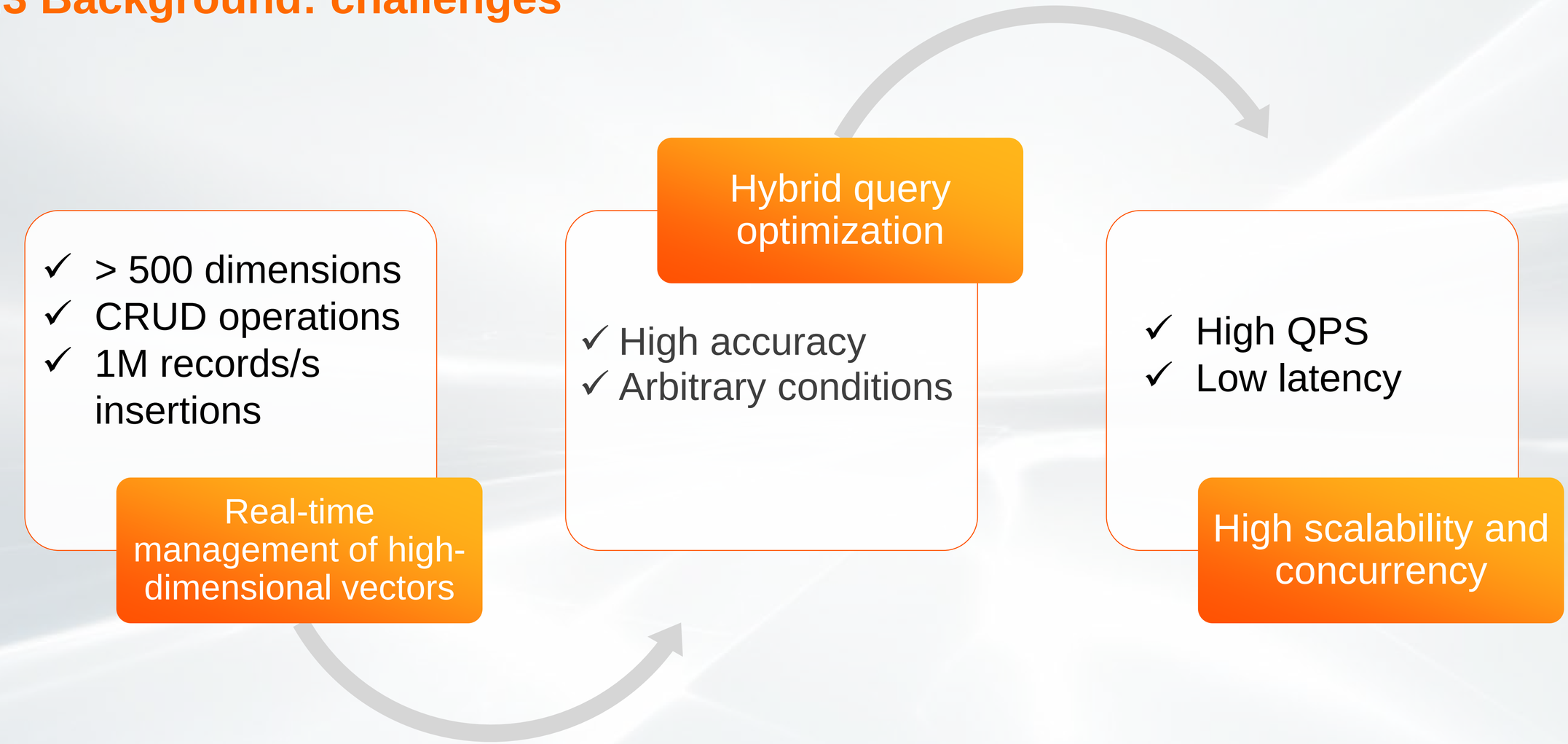
Hybrid analytics

SQL syntax

AnalyticDB-V

Alibaba Cloud

1.3 Background: challenges



Alibaba Cloud

1. Background

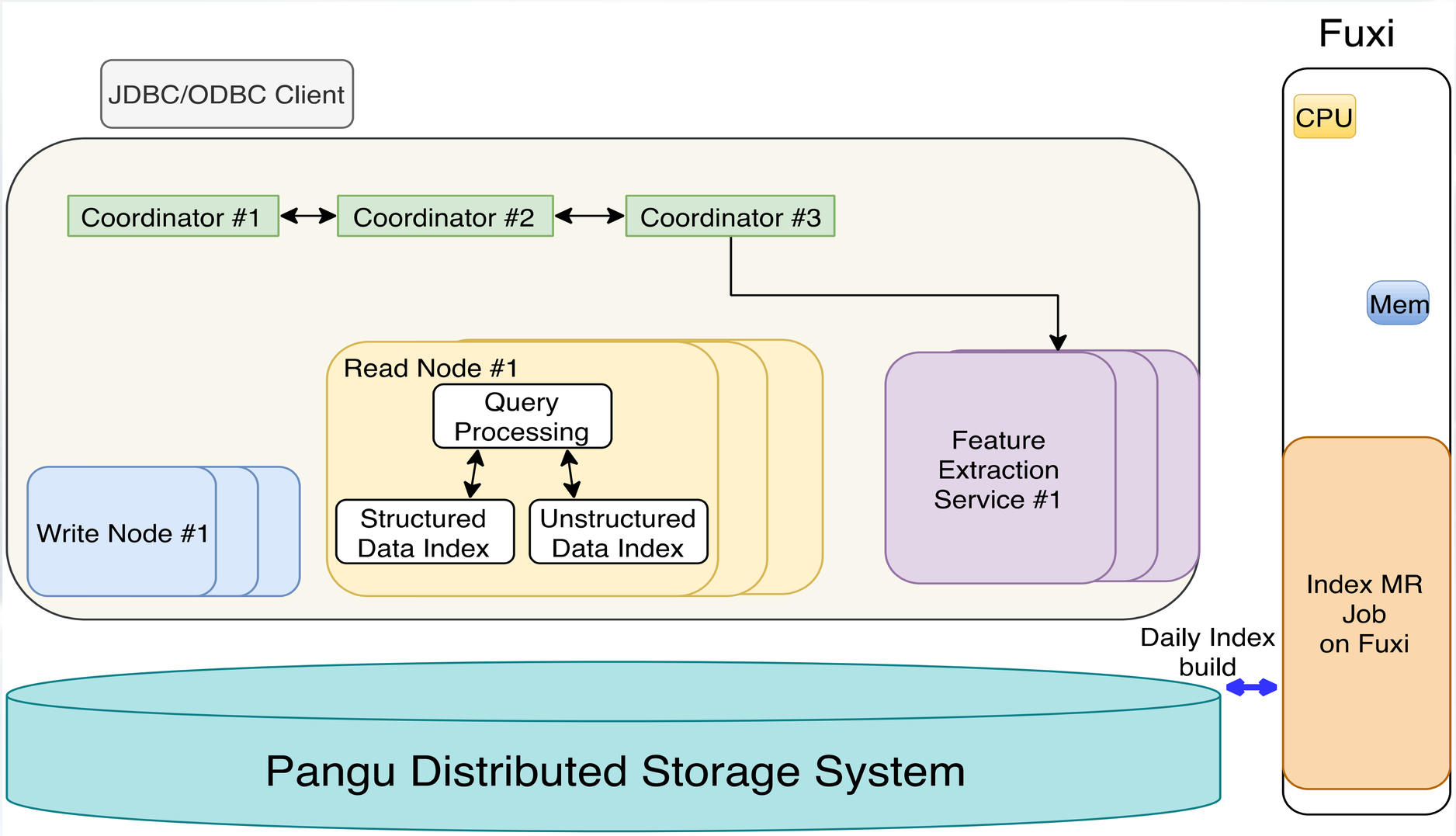
2. System design

3. Optimization

4. Evaluation

Alibaba Cloud

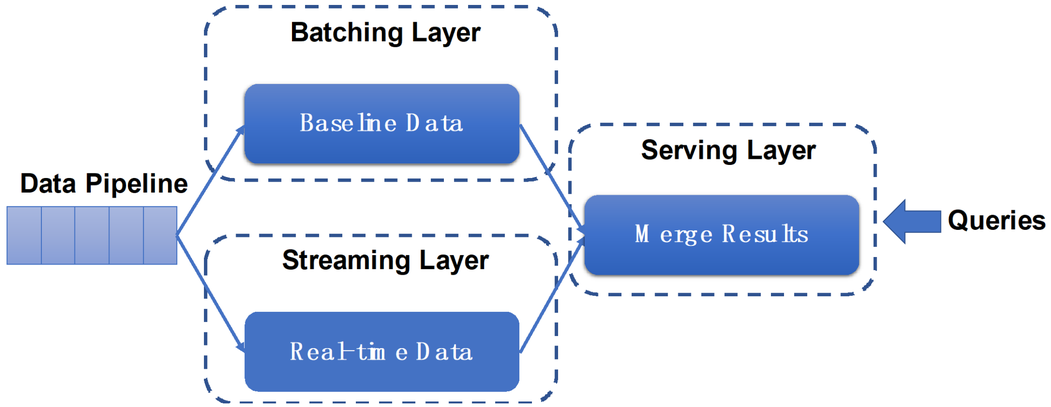
2.1 System design: overview

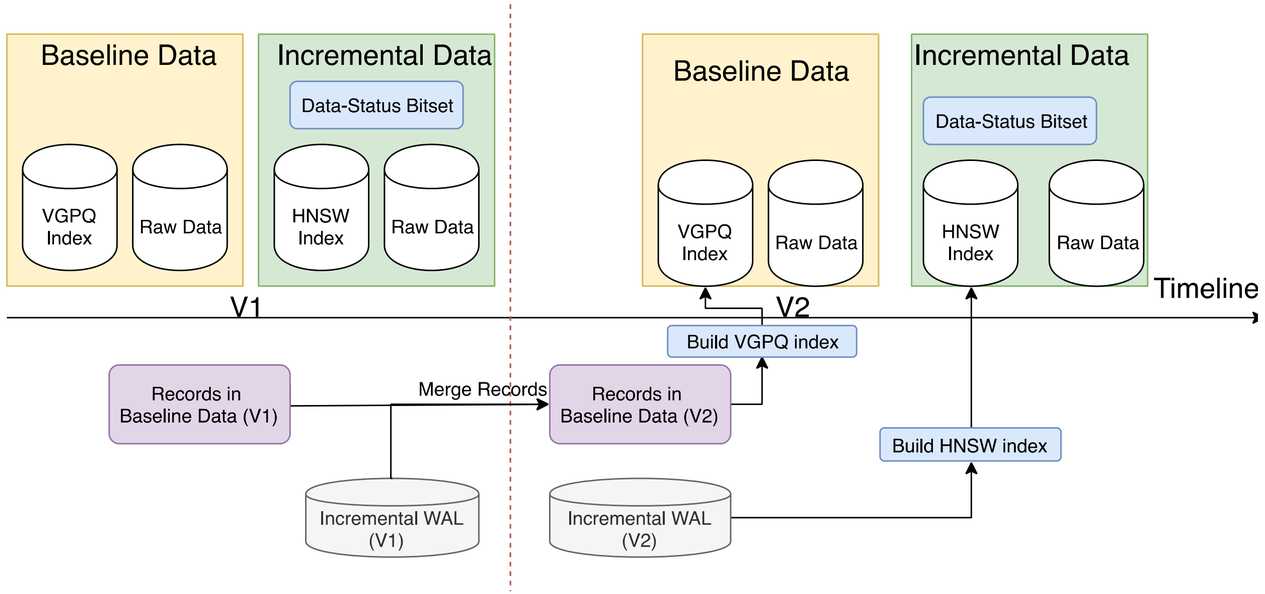


Alibaba Cloud

2.1 System design: Lambda framework

- **Query**
  - Query is running on both baseline data and incremental data.
  - VGPQ index is used for baseline data for low memory consumption.
  - HNSW index is used for incremental data for real-time index building.
- **Manipulation**
  - Support INSERT, UPDATE, and DELETE
  - Delete bitset is used to mark deleted rows.
  - Update is treated as delete + insert.
- **Merge**
  - Old baseline + incremental -> new baseline data
  - Data vacuum, remove deleted data.
  - Rebuild VGPQ index

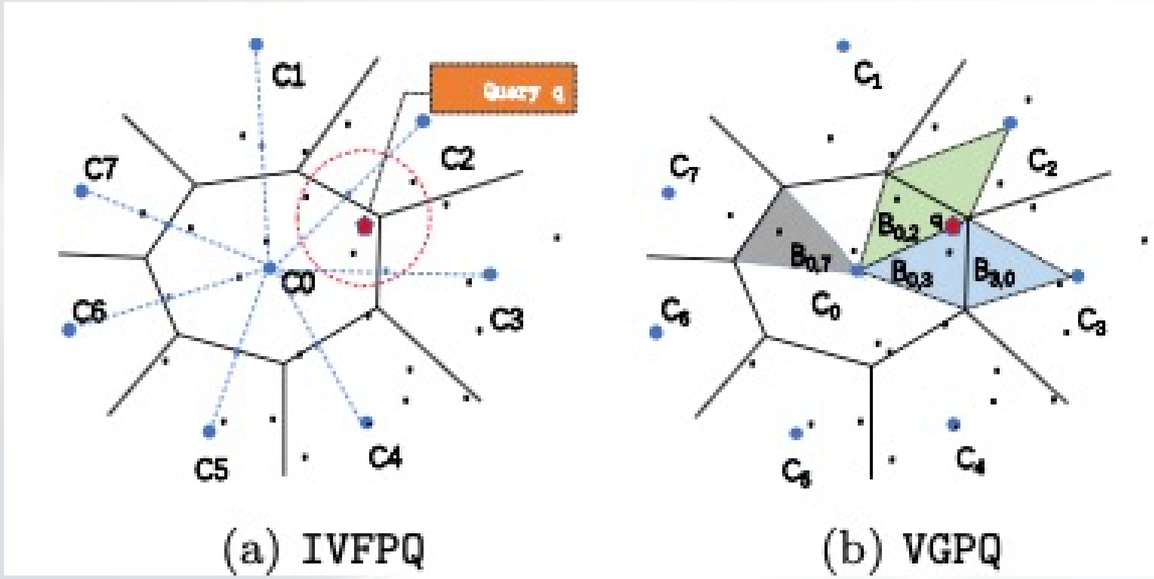




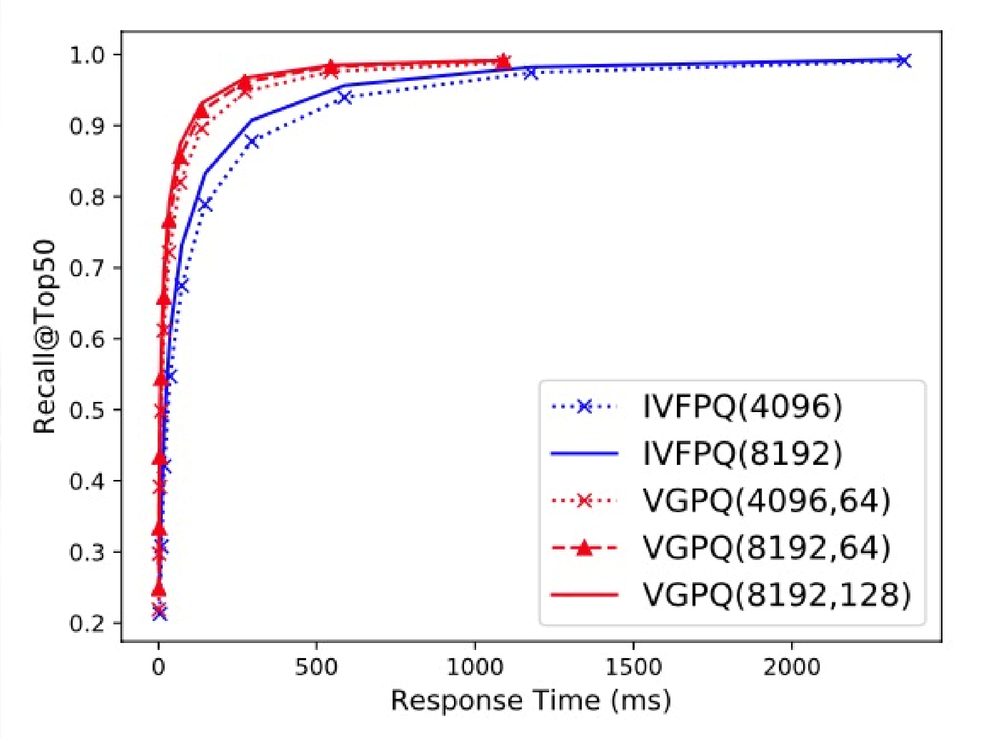
Baseline Data and Incremental Data in ADBV

Alibaba Cloud

3.1 Optimization: Voronoi graph product quantization(VGPQ)



Motivation of VGPQ



Performance of VGPQ

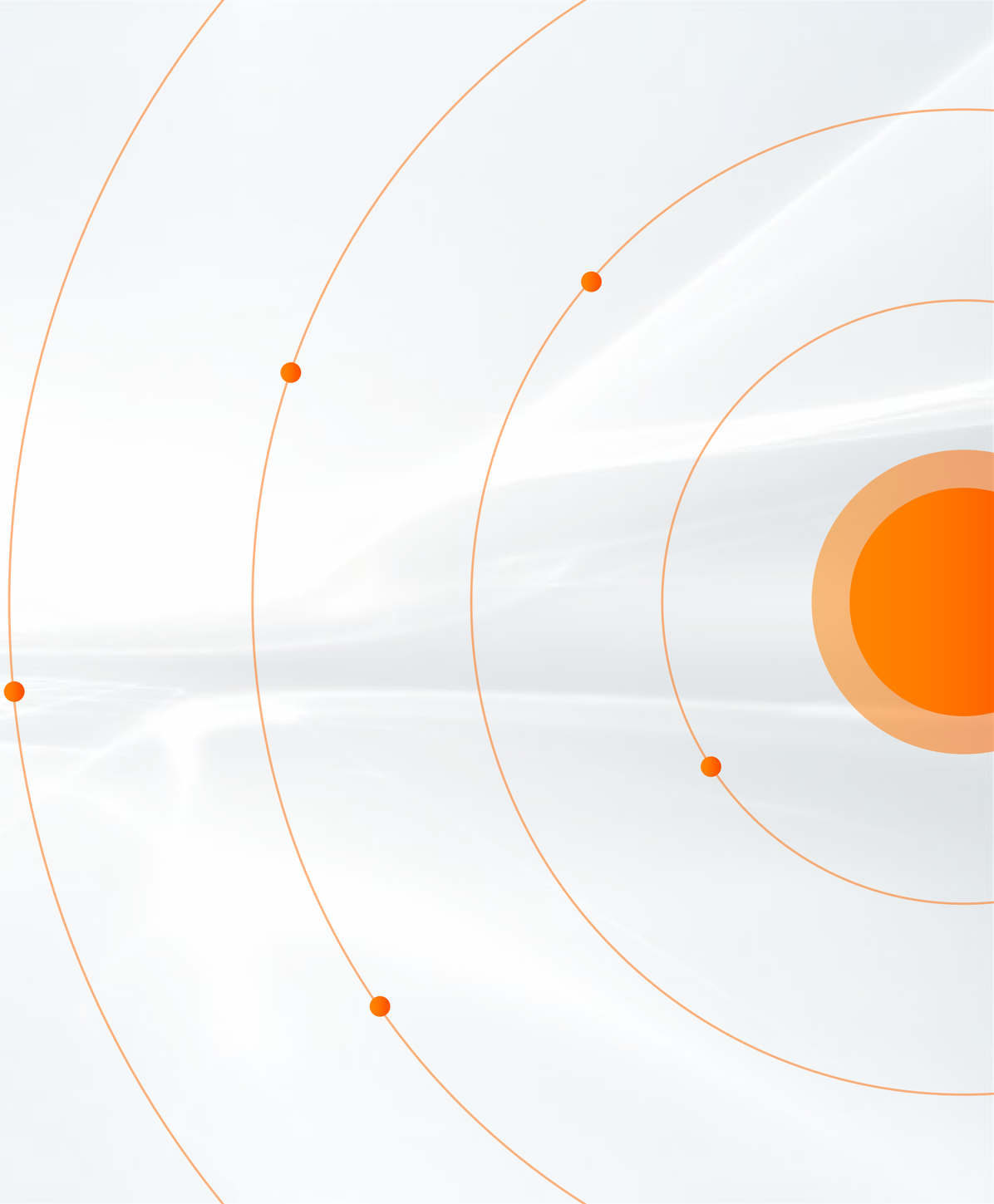
Alibaba Cloud

1. Background

2. System design

3. Optimization

4. Evaluation

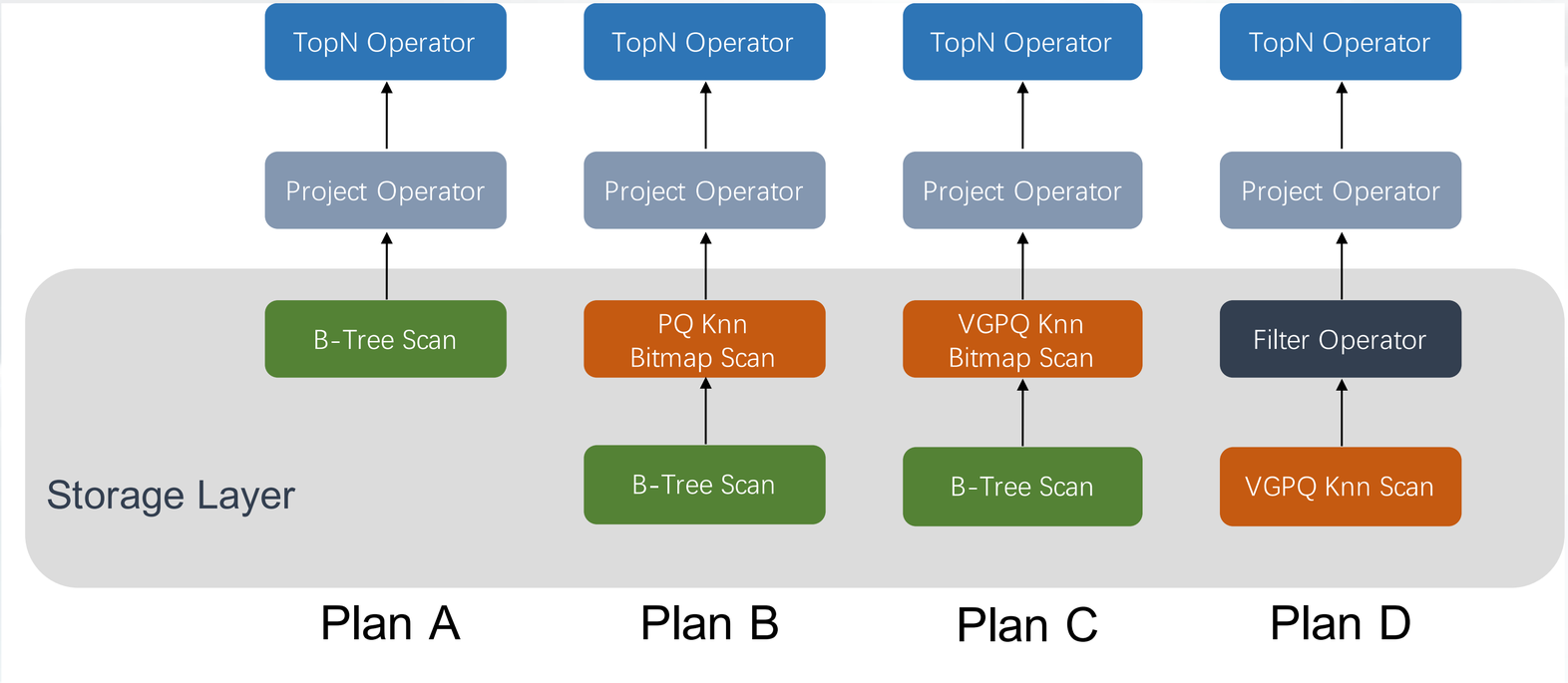


Alibaba Cloud

3.2 Optimization: Hybrid query optimization

```
1       SELECT id, DISTANCE(f, FEATURE_EXTRACT('img'))
2                as distance
3       FROM T
4       --structured predicates
5       WHERE T.c >= p1 and T.c <= p2
6       ORDER BY distance,
7       -- return top-k closest tuples
8       LIMIT k;
```

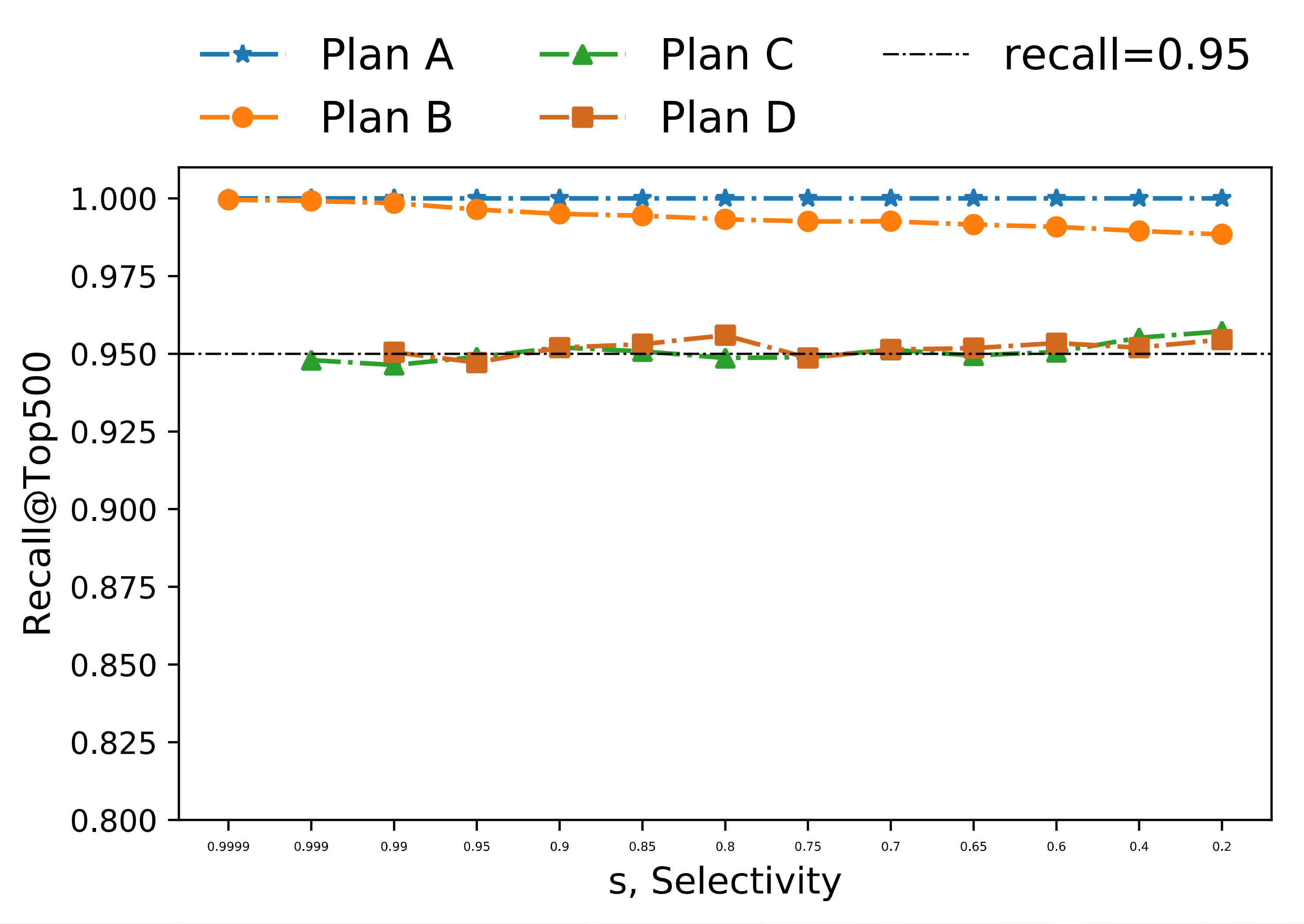
Hybrid Query Optimizer

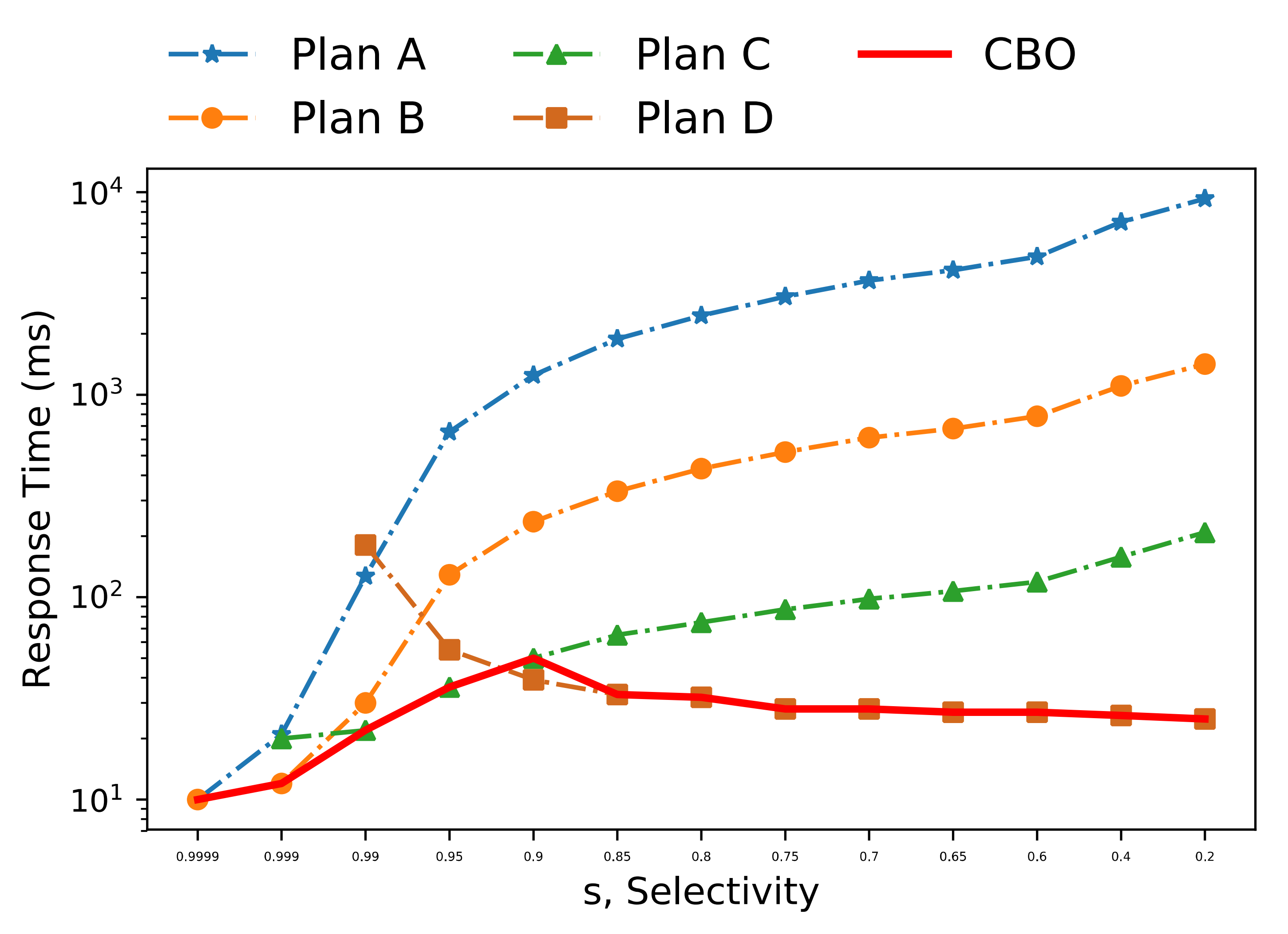


Alibaba Cloud

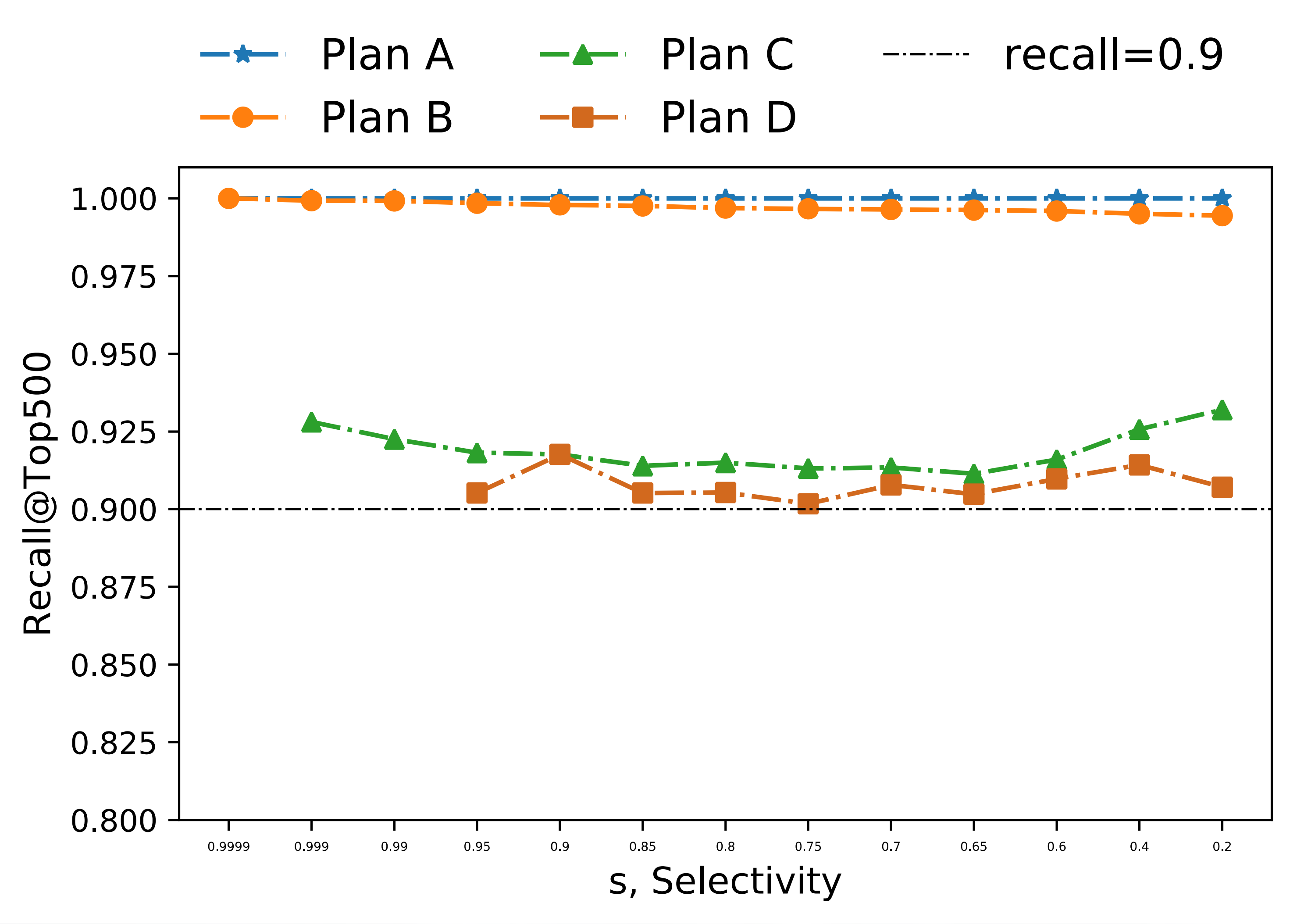
3.2 Optimization: Hybrid query optimization

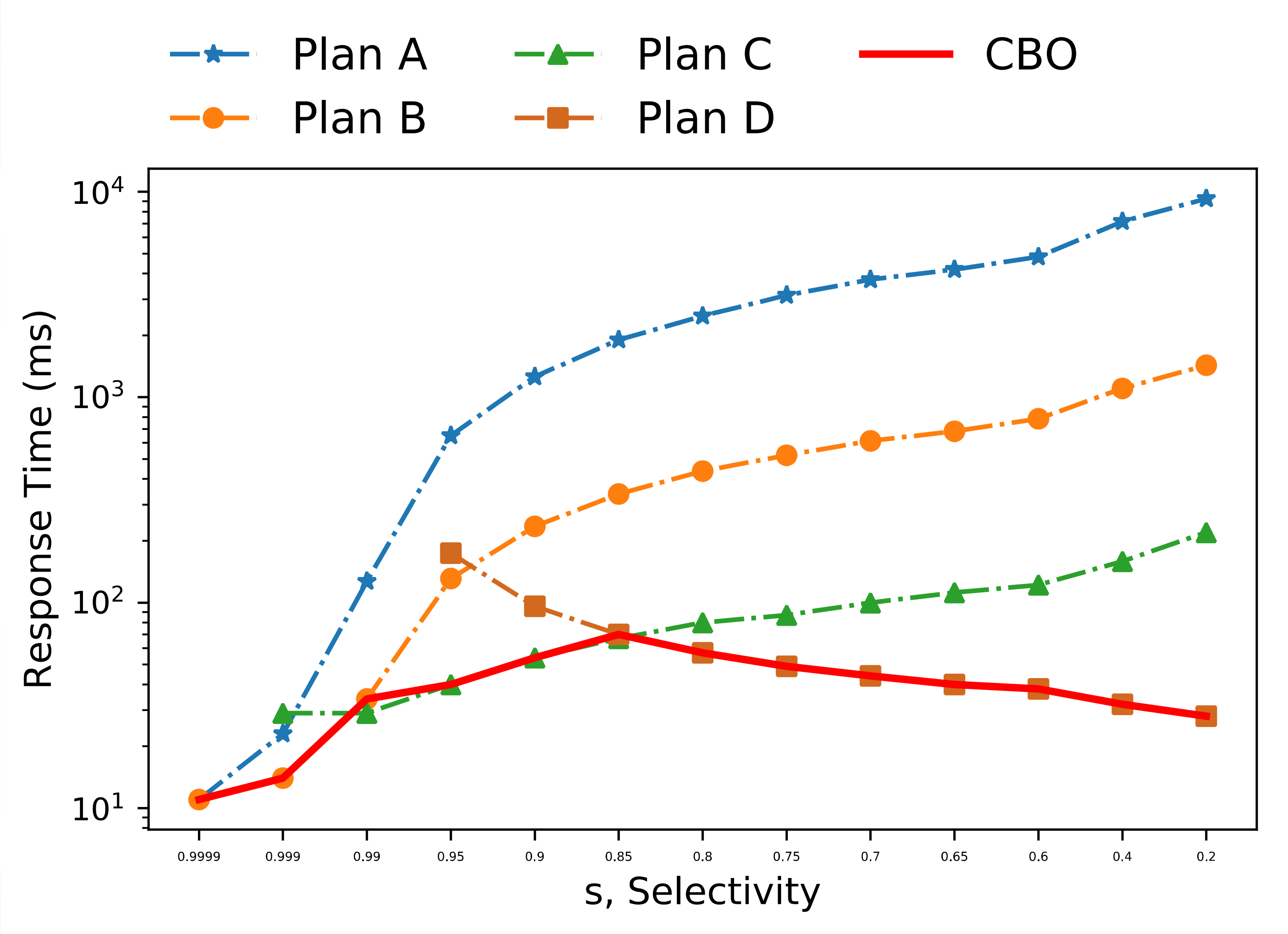
Topk=50 recall=95%



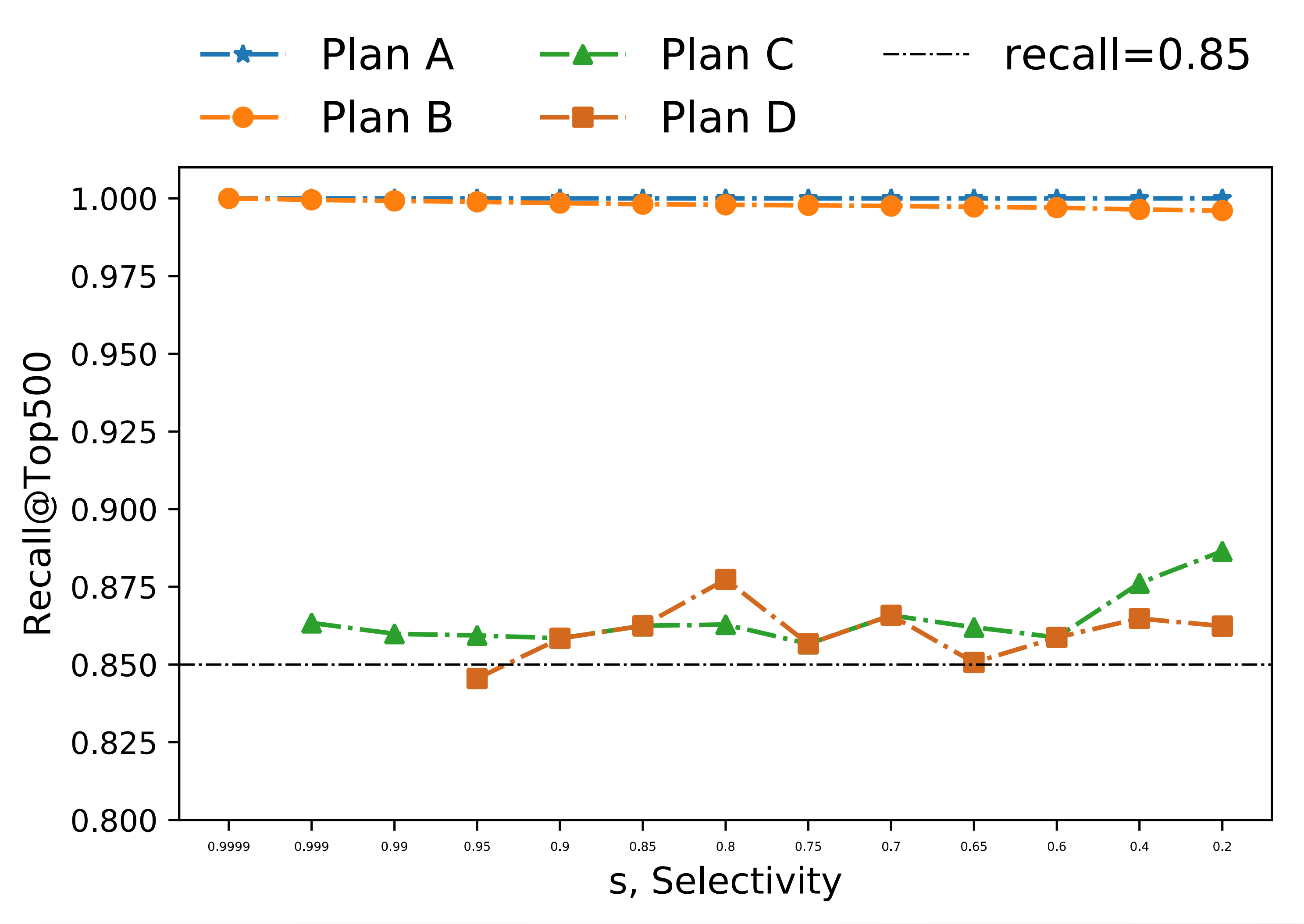


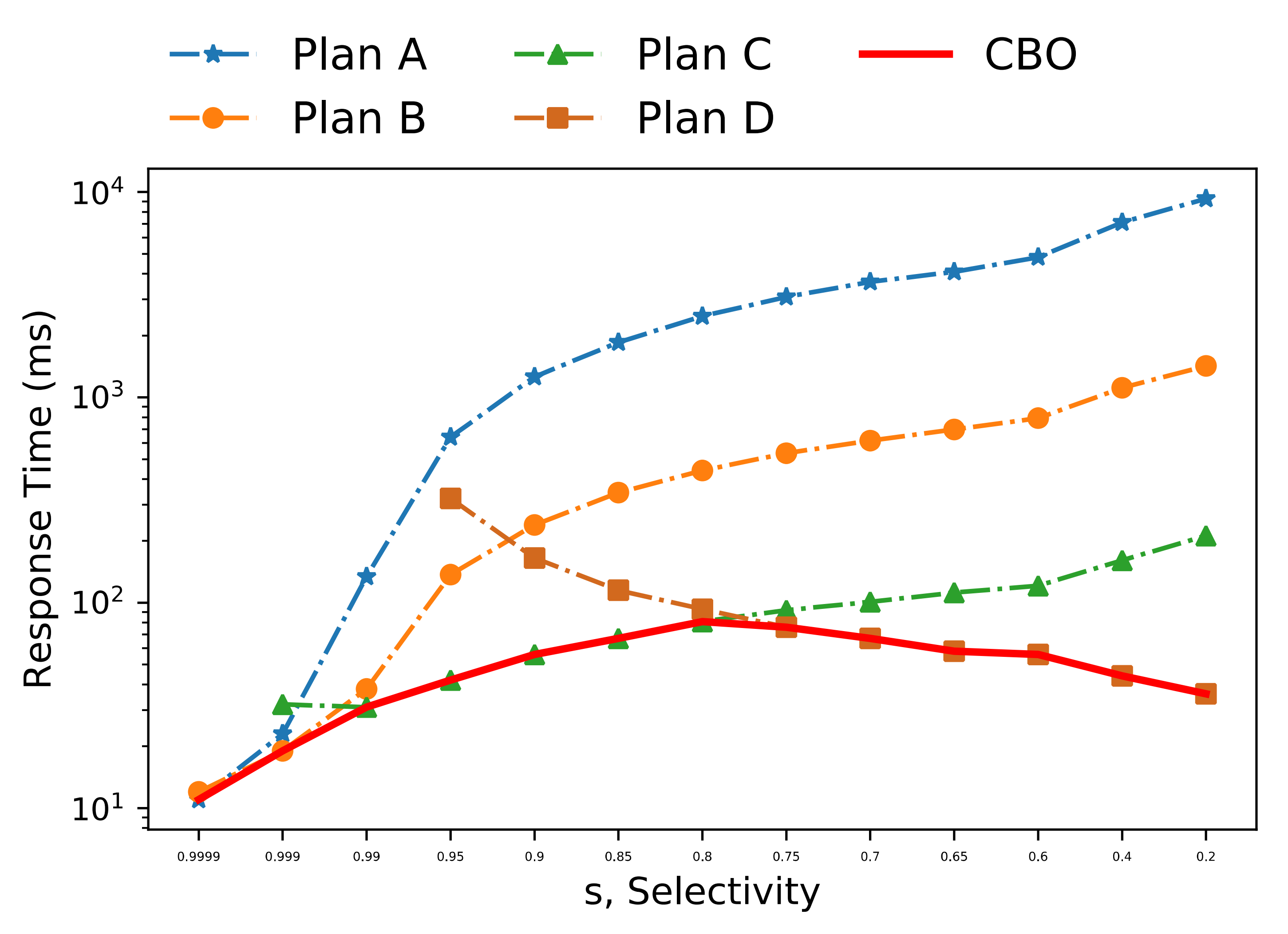
Topk=250 recall=90%





Topk=500 recall=85%





Alibaba Cloud

1. Background

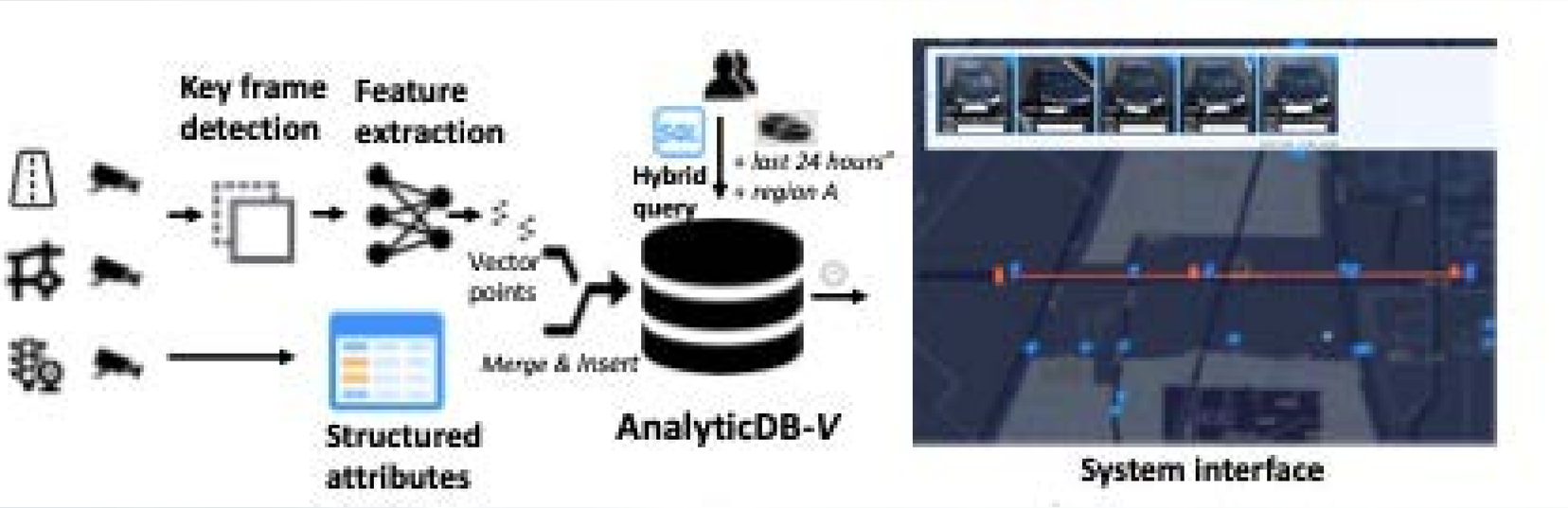
2. System design

3. Optimization

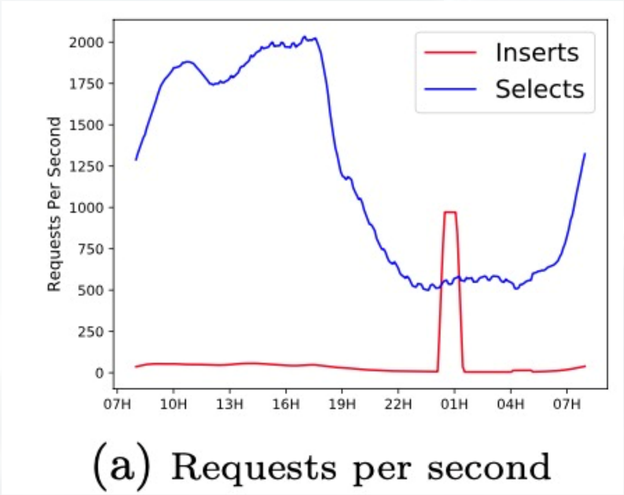
4. Evaluation

Alibaba Cloud

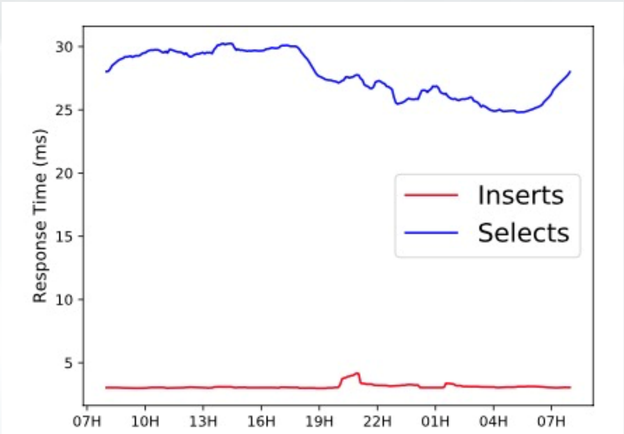
4 Evaluation



Use case study



(a) Requests per second



(b) Average Response Time per operation

Alibaba Cloud

Thank you