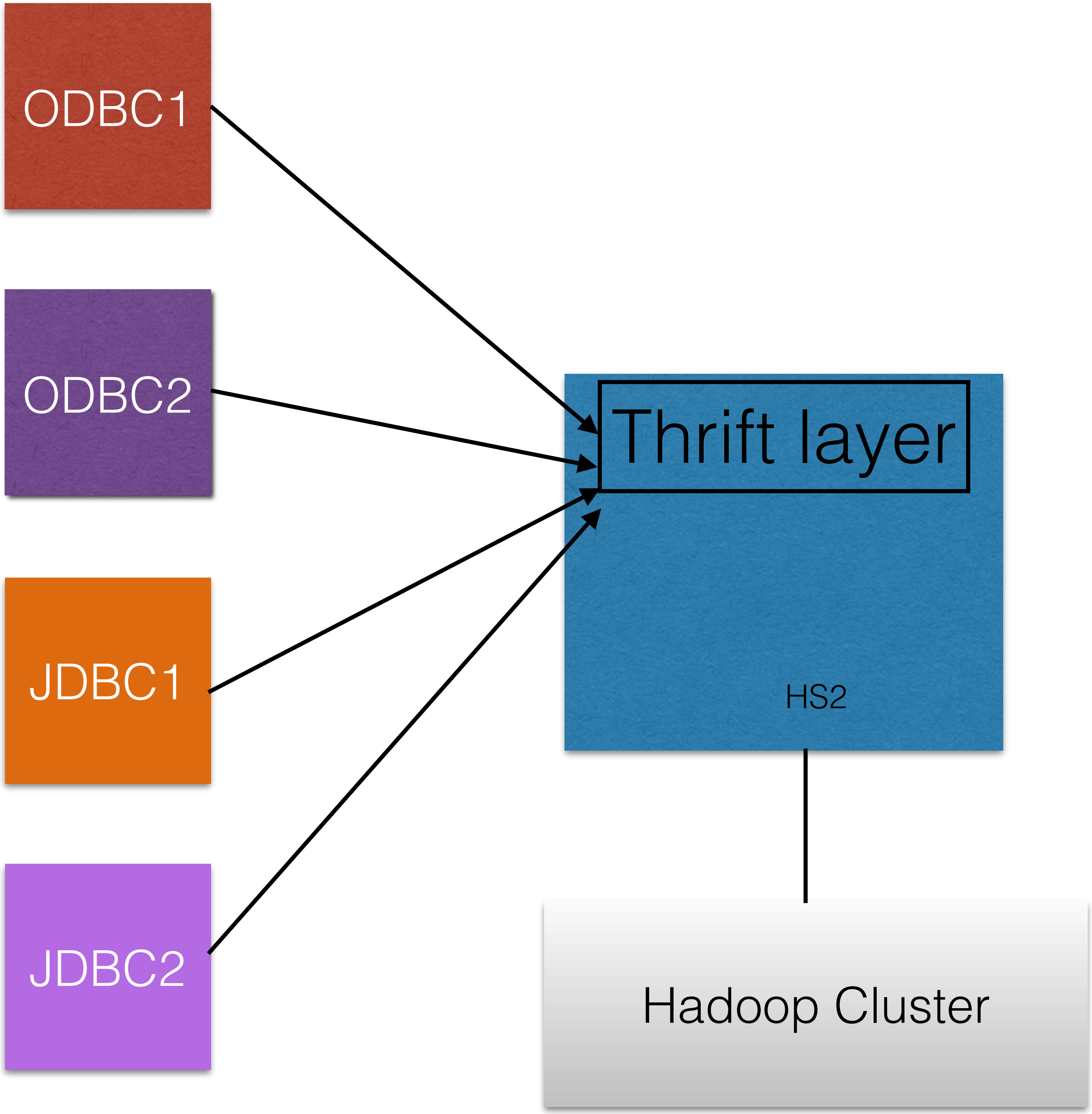
Small, Smaller, Smallest: ResultSet Compression in Apache Hive

Simba Technologies

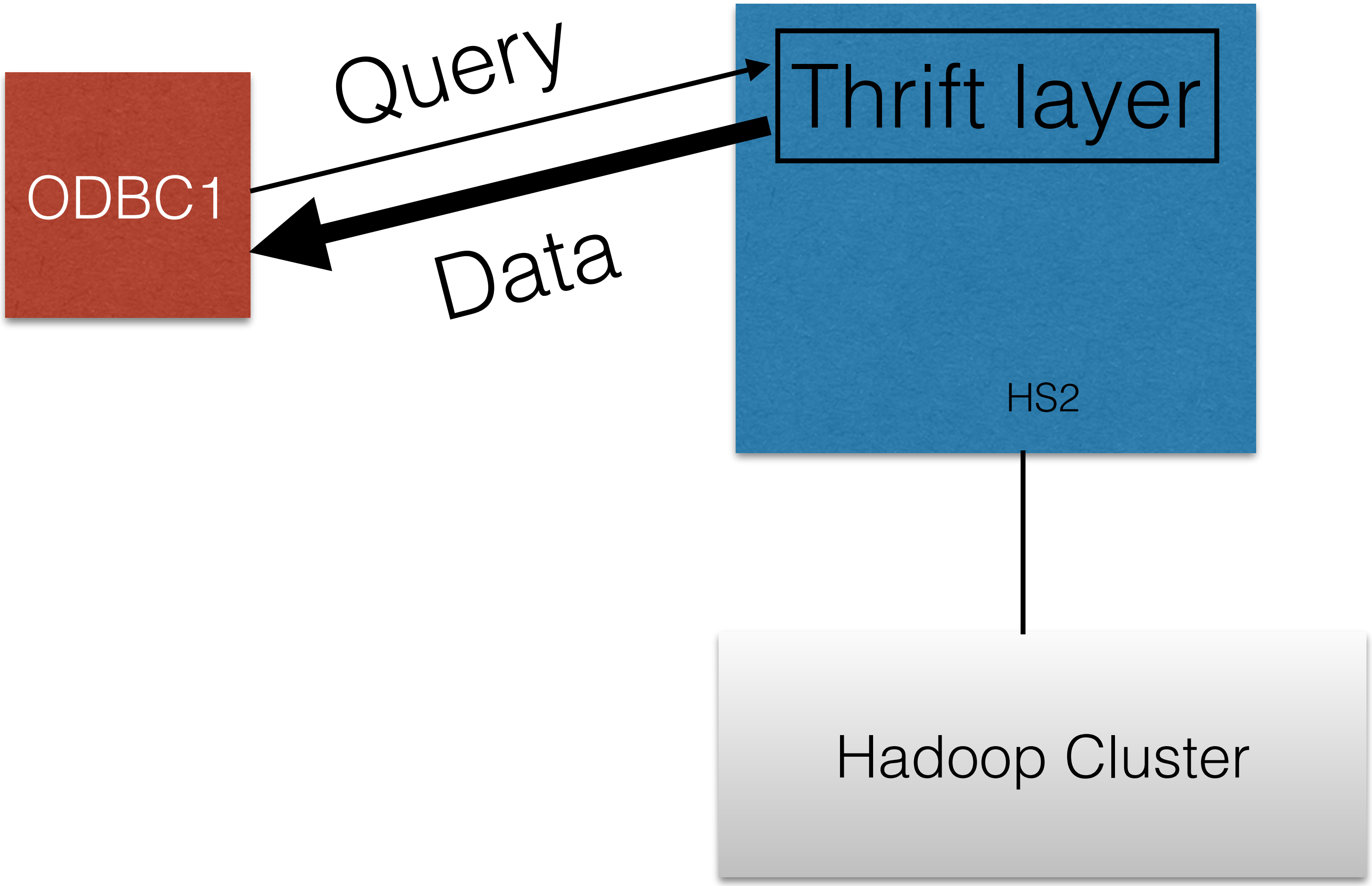
HIVE-10438

- We filed this JIRA today!
- This talk will discuss what HIVE-10438 is about
- ResultSet Compression, plugin architecture
- Results
- Challenges

HiveServer2 (HS2)



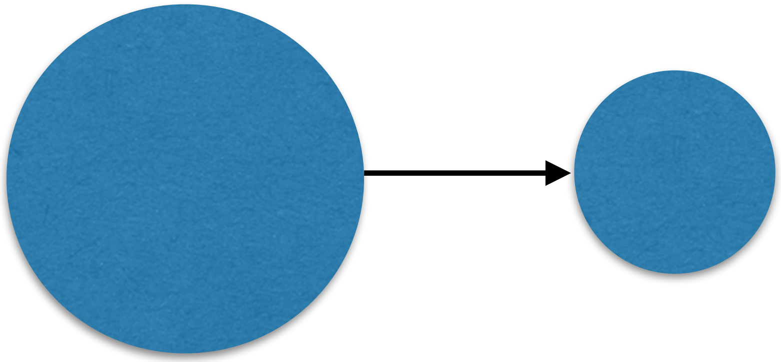
Client query example

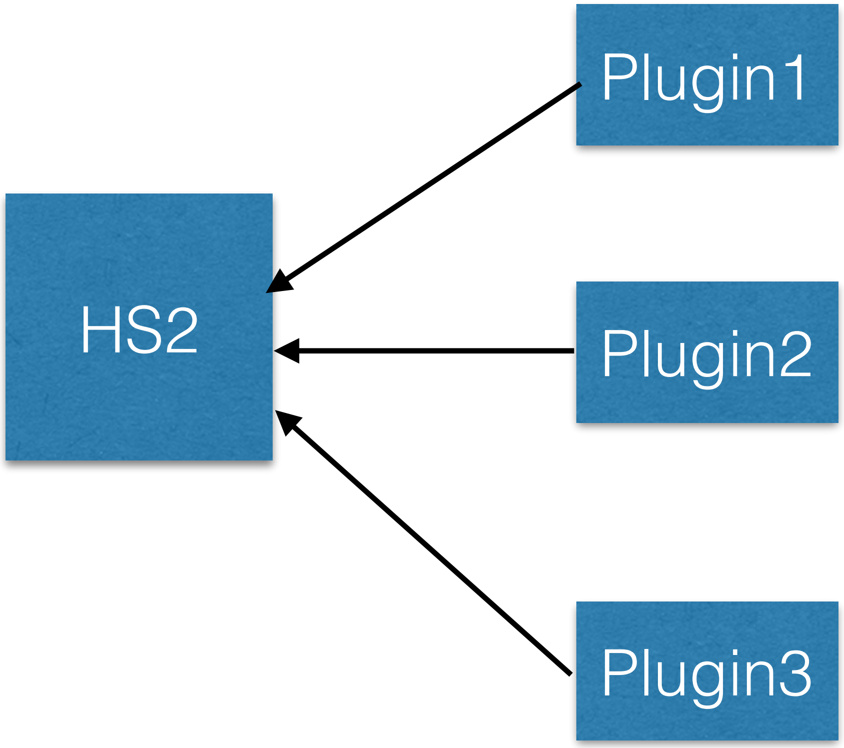


How can we compress ResultSets and improve performance?

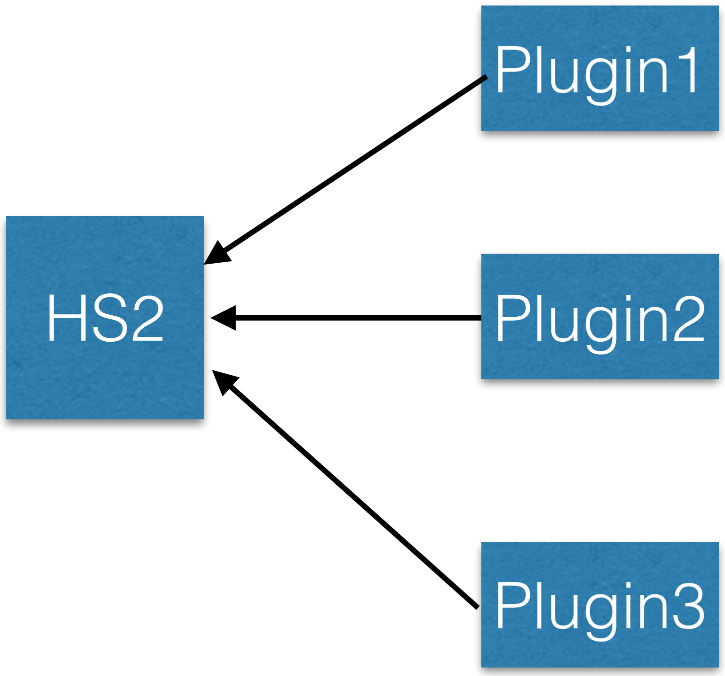
Compression - wishlist

- A compression library should
  - **Compress more, Consume less**
    - High compression ratios
    - High performance
  - **Just Plug it!**
    - Allow extensibility
    - Make compression run-time option



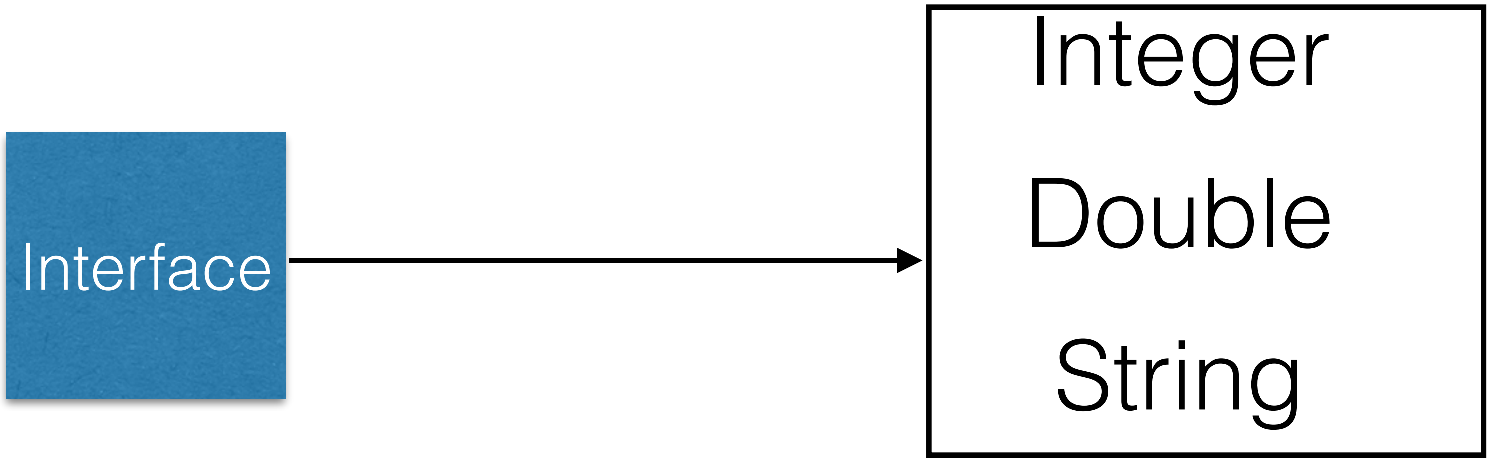


Just Plug it!



- Make compression a runtime option
- Allow everyone to write their own compressors
- Multiple plugins should be simultaneously usable
- Allow activation/deactivation of compression and compressors
- Allow client to choose which ones to use

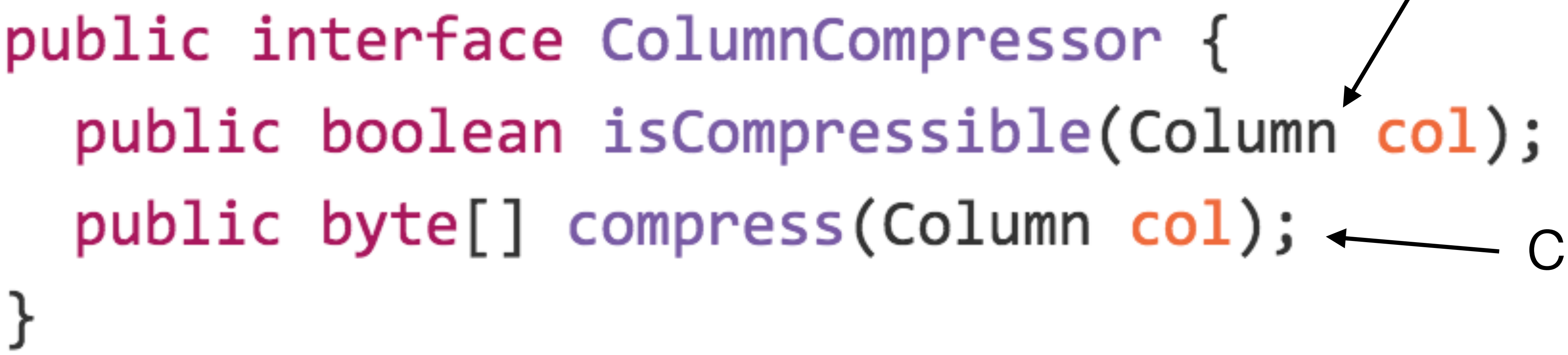
Plugin Architecture



- They all implement an interface, present in Hive
- Each compression technique in it’s own class
- Anyone can implement the interface and plug their own

ColumnCompressor

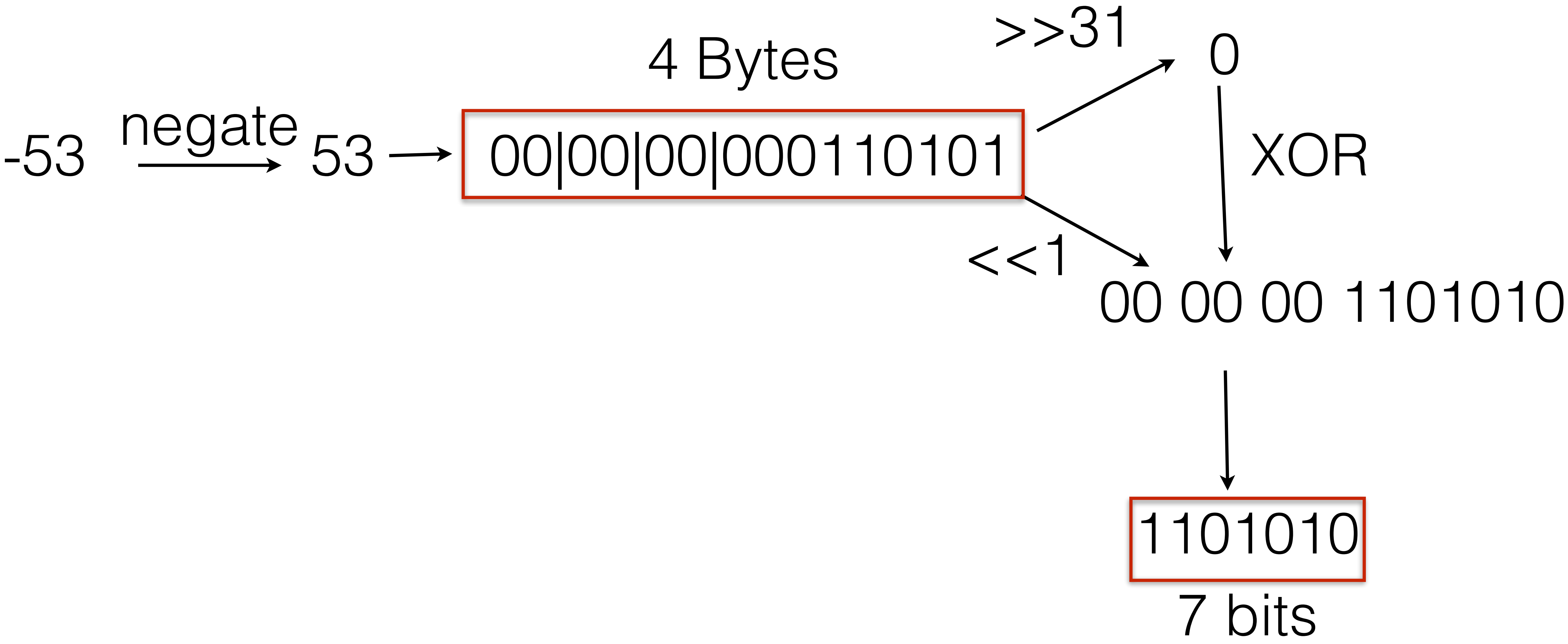
Can you compress this?



```
public interface ColumnCompressor {
  public boolean isCompressible(Column col);
  public byte[] compress(Column col);
}
```

compress this

Integer Compression



-53

negate

53

4 Bytes

00|00|00|000110101

>>31

0

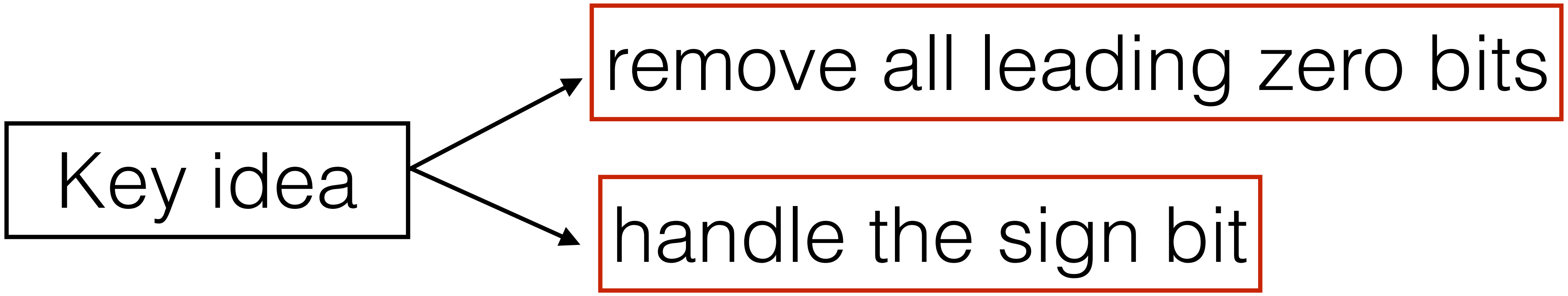
XOR

<<1

00 00 00 1101010

1101010

7 bits

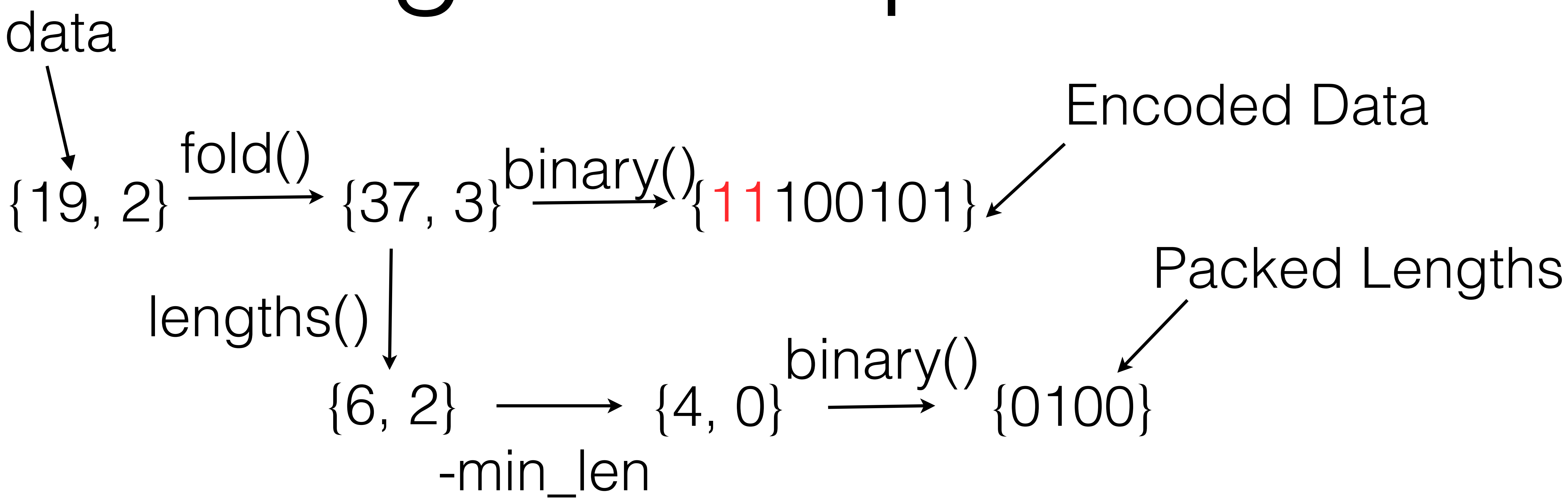


Key idea

remove all leading zero bits

handle the sign bit

Integer Compression



data

{19, 2} fold() {37, 3} binary() {11100101}

Encoded Data

lengths()

Packed Lengths

{6, 2} -min_len {4, 0} binary() {0100}

Send encoded data tightly packed

Send packed length data to help decode

size(encoded Data) + size(packed Lengths) < size(data)

Evaluation Dataset

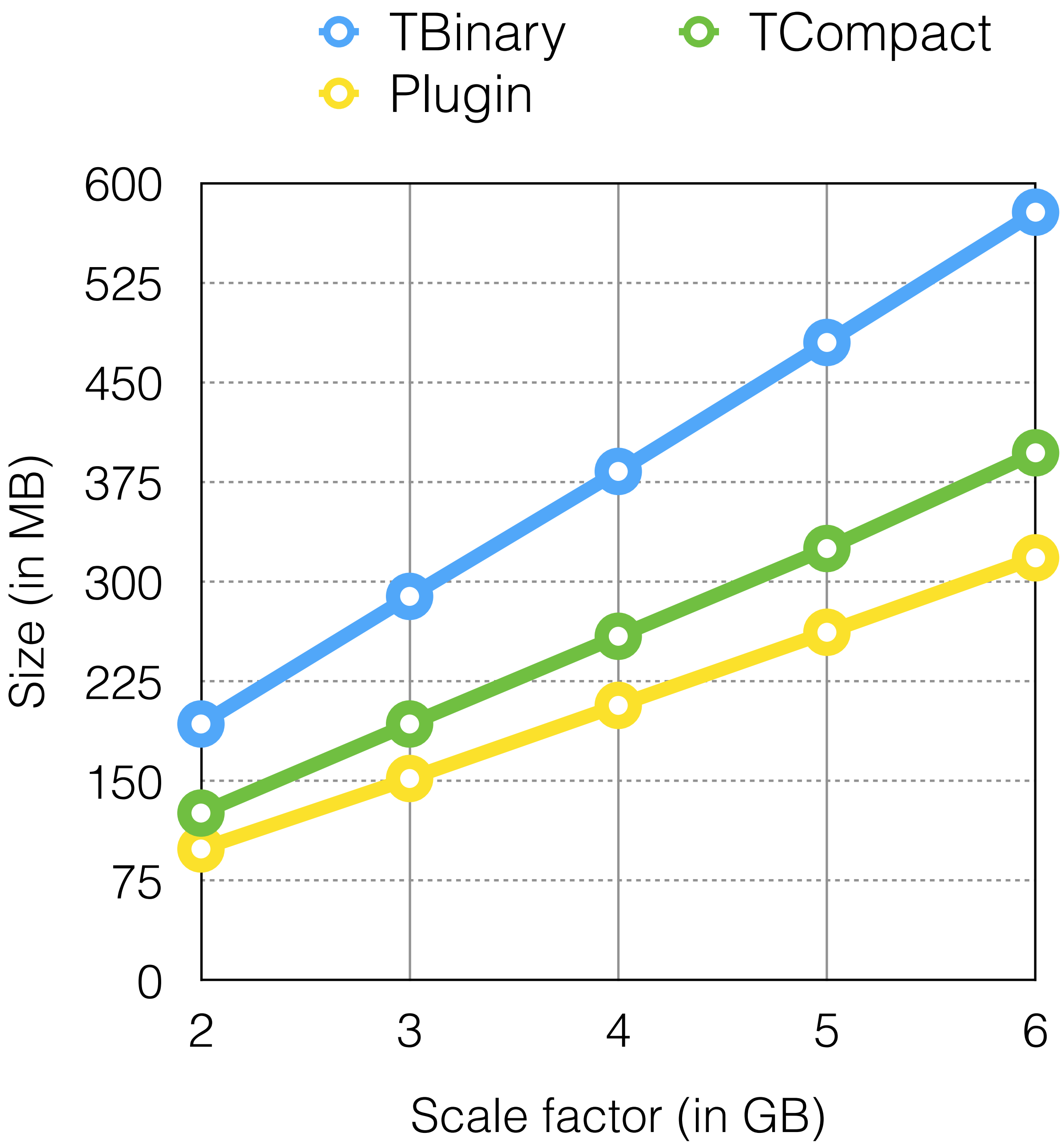
- *Lineitem* table from TPC-H
- Scale factors {2, 3, 4, 5, 6} (GB)
- Query is “select * from lineitem”
- Using an AWS cluster (10 nodes)
- Tested with integer, double and string compressors
- **Objective**: compare performance {TBinary, TCompact, Simba compression} protocol

Setup

- ODBC client running in the same EC2 zone
- For each scale factor, query is run 3 times
- **tcpdump** used to measure amount of data transfer
- Internal tool used to validate tcpdump results
- Average of tcpdump measurements reported

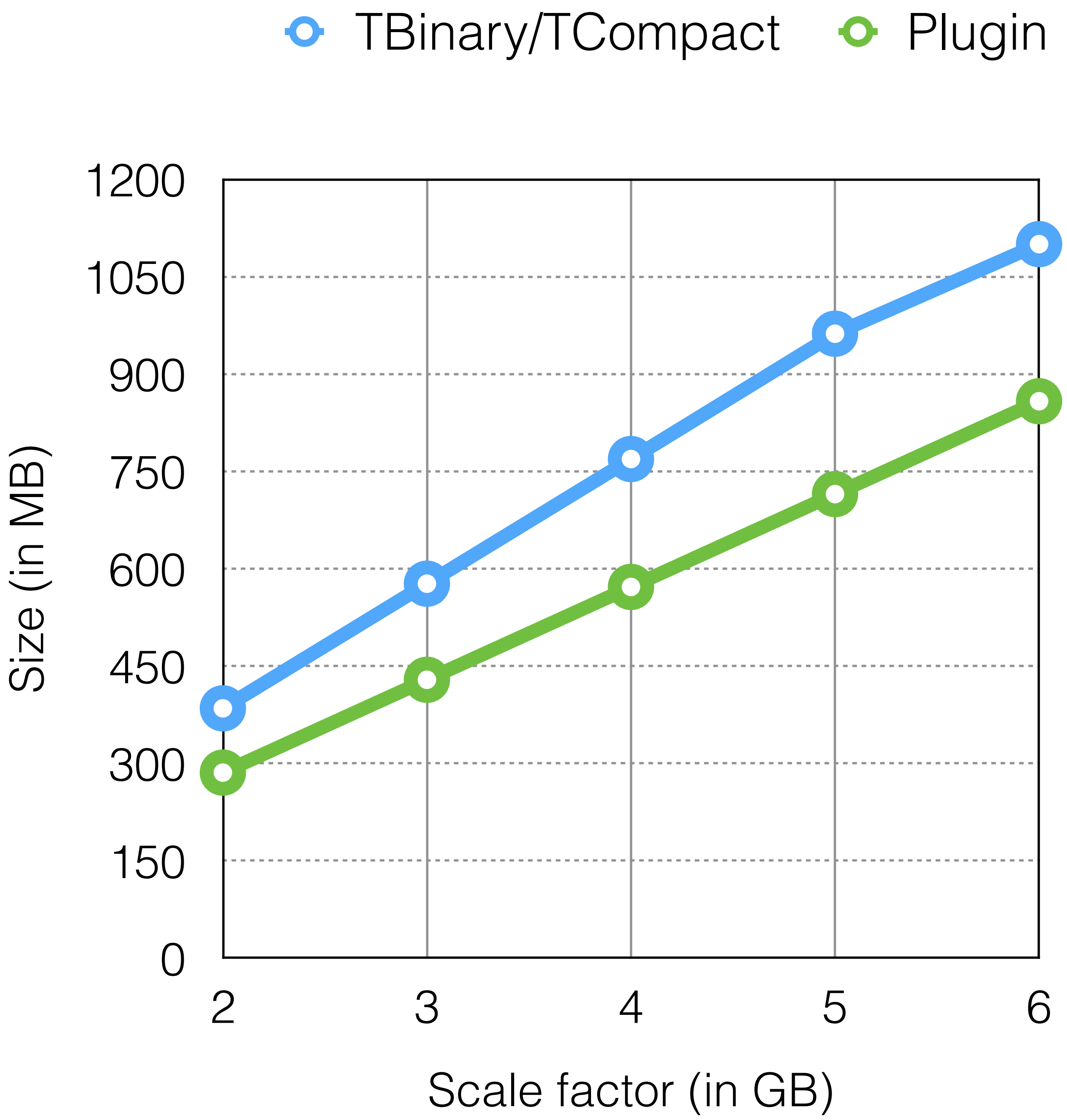
Integer Results

Scale	TBin	TCom	Plug
2G	193	126	99
3G	289	193	152
4G	383	259	207
5G	480	325	262
6G	578	397	318



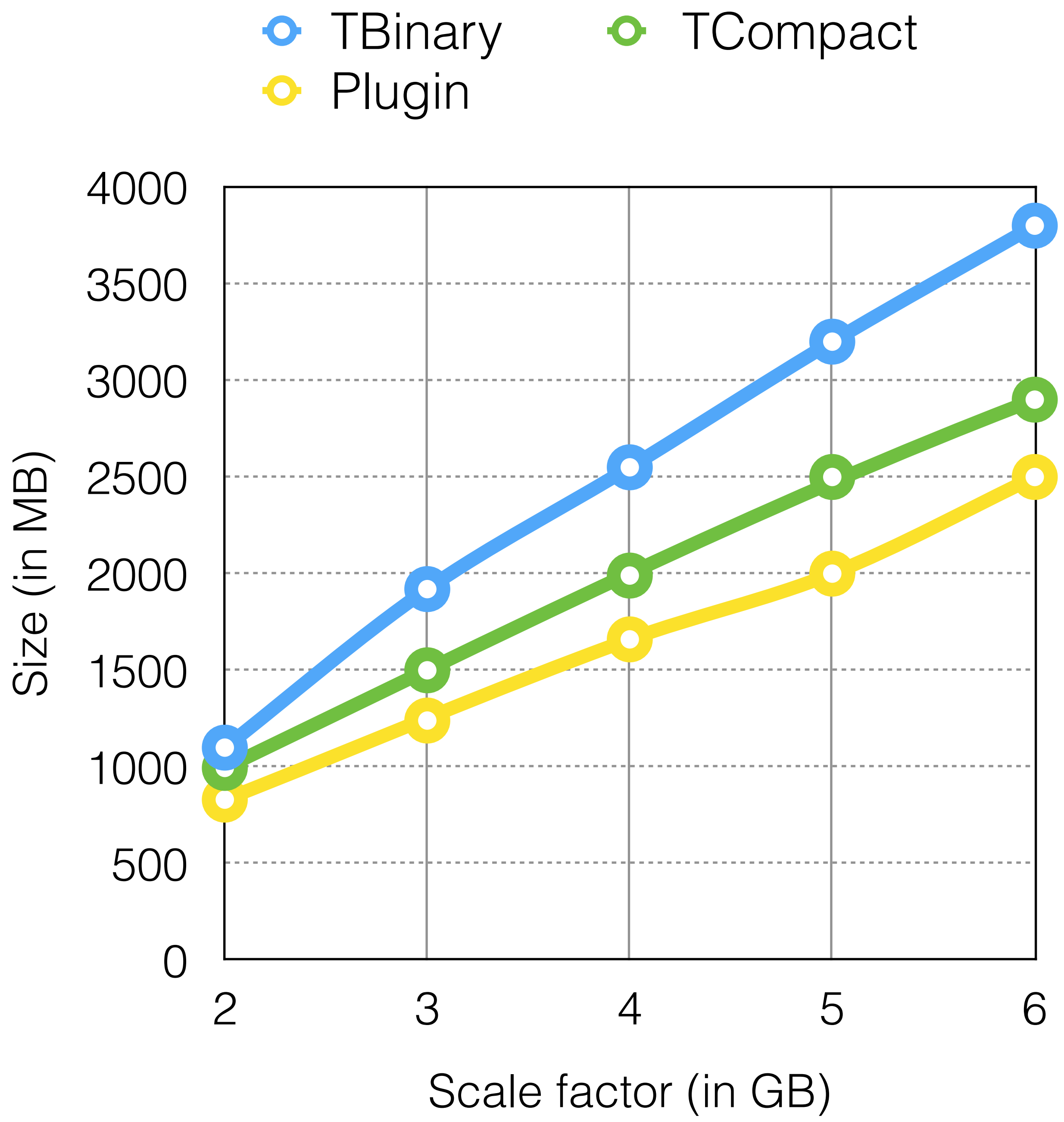
Double Results

Scale	TBin/TComp	Plugin
2G	385	286
3G	577	429
4G	769	572
5G	962	715
6G	1.1G	858

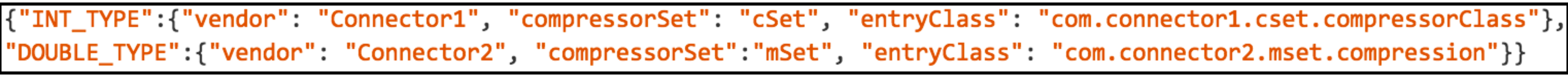


String Results

Scale	TBin	TComp	Plugin
2G	1.1G	995	831
3G	1.92G	1.5G	1.24G
4G	2.55G	1.99G	1.66G
5G	3.2G	2.5G	2.0G
6G	3.8G	2.9G	2.5G



Plugin Configuration - Client



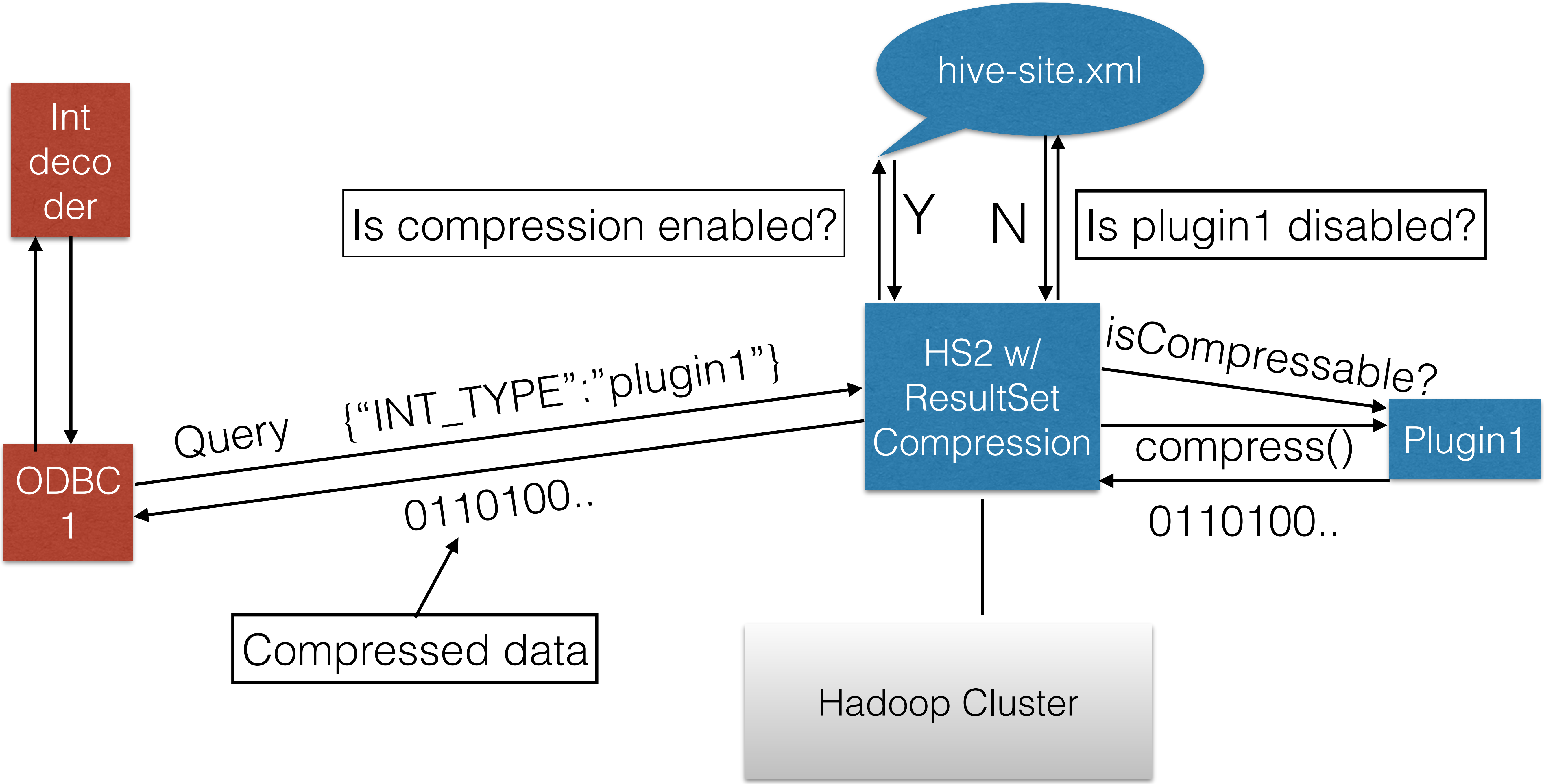
```
{"INT_TYPE":{"vendor": "Connector1", "compressorSet": "cSet", "entryClass": "com.connector1.cset.compressorClass"},
"DOUBLE_TYPE":{"vendor": "Connector2", "compressorSet":"mSet", "entryClass": "com.connector2.mset.compression"}}
```

- Client can use a JSON string to inform server which client to use
- Key by data type
- Can choose different compressor Sets for different types

Plugin configuration - server

- **hive.resultSet.compression.enabled** -> activate ResultSet compression
- **hive.resultSet.compressors.disable** -> comma-separated list of compressors which will *not* be used for compression
- Allows activating/deactivating both compression and compressors

Query Execution



Is compression enabled?

Is plugin1 disabled?

Compressed data

Writing your own Compressors

- For everyone to write their own compressors, they would need a client with a decoder
- To make it easier to observe the end-to-end functionality and write their own compressors, we are also releasing a C++ query submitter
- It has minimal dependencies and can be run on any platform

Status

- We are proposing a plugin architecture for Hive ResultSet Compression as part of **HIVE-10438**
- Code changes: it proposes one new interface and one new class and two configuration options as part of hive-site.xml
- A query submitter that helps for writing and testing new compressors

What about latency?

- We have observed that at scale factors 6 and above, latency numbers reported by tcpdump have high variability
- Although we observed 10 to 15% less round trip time, it was variable
- Reason could be congestion control on AWS
- And/or the default ports on m1.large machines
- We are working on resolving this
- Ideas?

Questions?