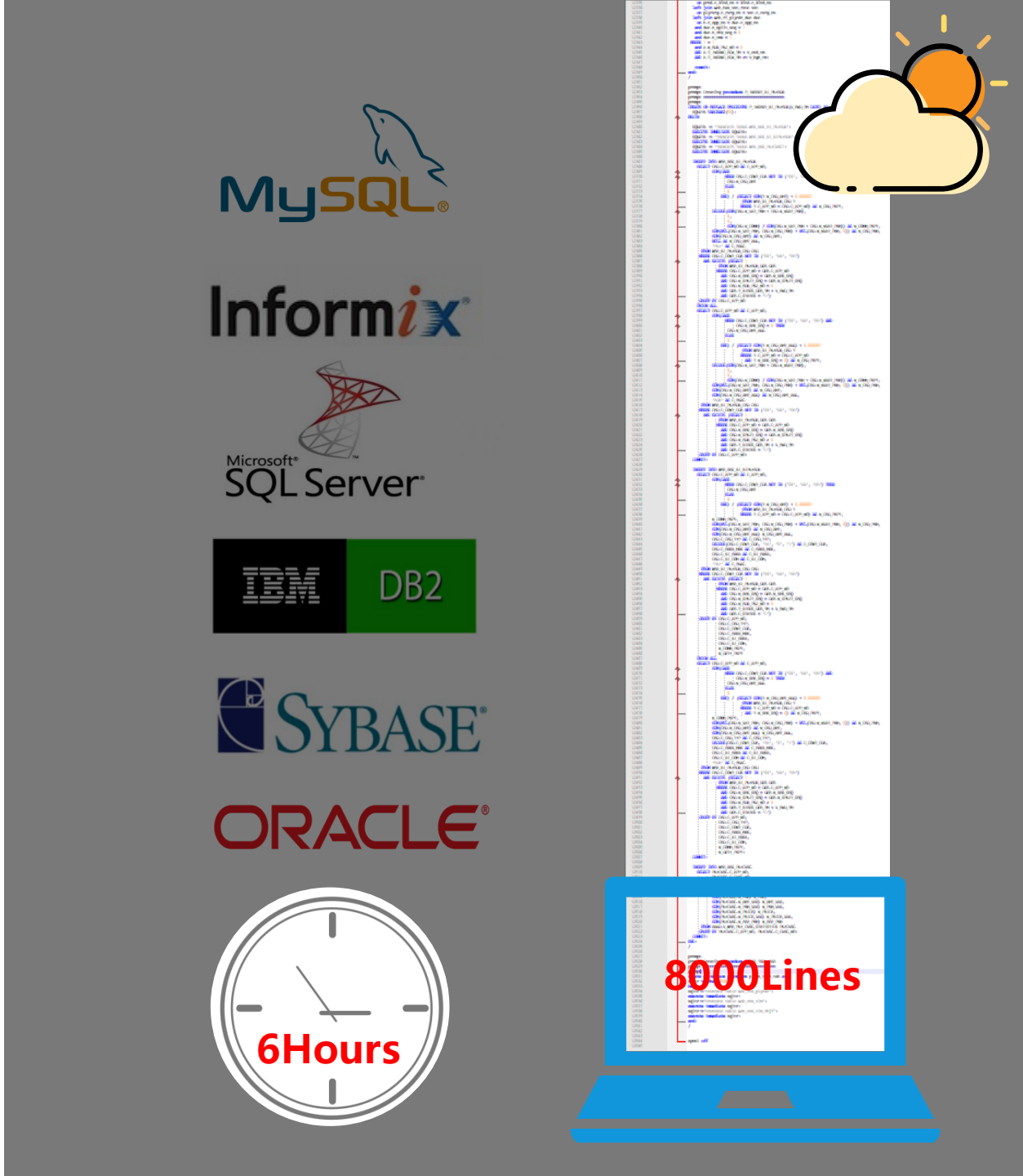
Off-line Batch Running

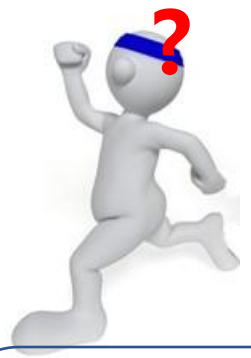
Write thousands of lines of stored procedure, and the batch running is very slow! It took more than six hours to finish, and the business was affected the next day.

Barely finished running, afraid of making mistakes, and there is no time to run again!

Even if I can deal with it this year, the volume of data will become larger and larger, and I do not know whether it will make do next year!



Off-line Batch Running




Running batches requires external source data, the arrival time of which is uncontrollable, and the window time may be reduced a lot.

Time is spent on the inbound and outbound of the database, but the batch running can only be done with database, and we dare not risk running batches on production database.

Business people expect to be able to run data for several months, but because the batch running is too slow, they can only run data for a few days.

SPL Base Speed Up for Off-line Batch Running

Replacing Storage Procedures with SPL Better Algorithms

SPL supports parallel queries

Provide optimized algorithms for the join, traversal and parallel of batch running, and the performance is better than stored procedure.

Direct file-based computing eliminates the time spent on the inbound and outbound of database

SPL data files are mainly oriented to data analysis and calculation

Binary Compressed Column Storage, Double Increment Segmentation, Arbitrary Parallel Technology

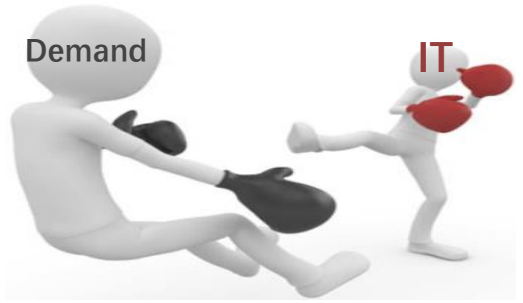
Make full use of hardware computing ability to improve data storage and computing performance

Online Query

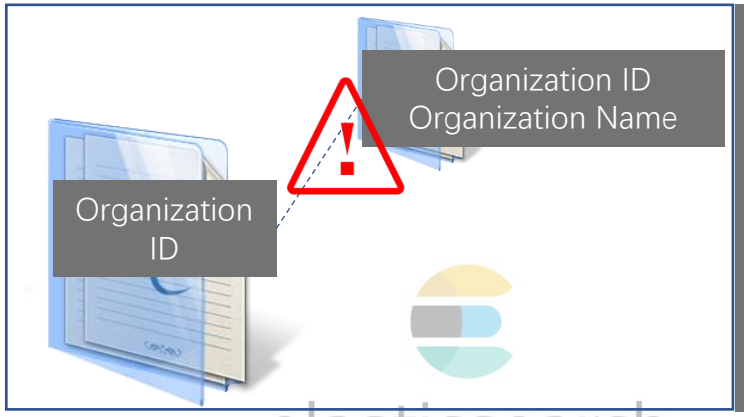
There are too many data tables that need to be joined, and have to wait half of an hour for the result of the report. The result has not come out when you've finished a cup of tea, and the business people are very angry.

There is no problem with single query, but in case of batch queries, there will be big problem when the concurrencies are high.

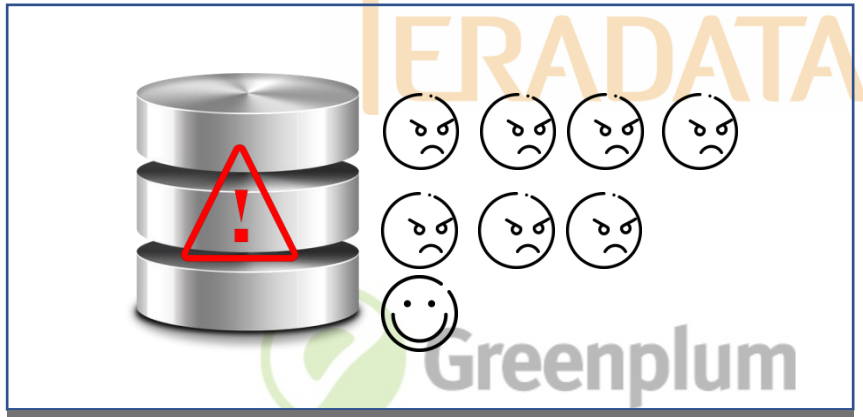
To guarantee the speed we can only suppress demand. The business department wants to check the result for one year, and we can only provide result for one month.



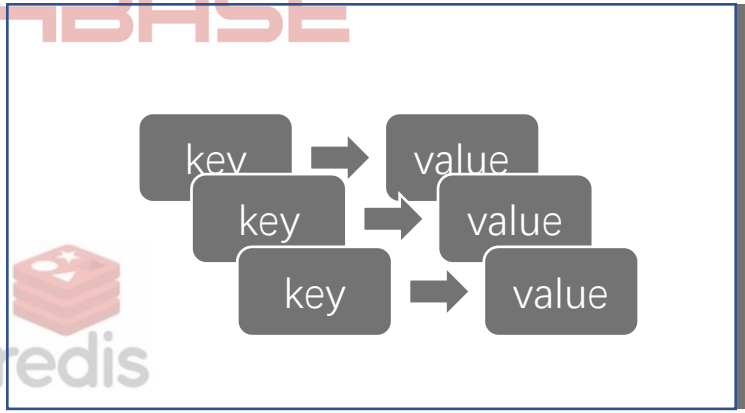
Online Query



Using search engine to achieve query, the performance is acceptable, but it can not achieve subsequent calculation. With UDF written in large quantities, it is difficult to maintain.



When we need to query the full amount of data, the pressure is concentrated on the central data warehouse, and it can not support concurrent queries.



Using KV database to achieve query, It's OK with single task, with batch query it is difficult to support. It can not handle complex logic, and UDF will be flying all over the sky.

SPL Base High Performance Online Computing Solution

Orderly compressed storage, high-speed positioning and batch reading in query

Free column storage compression, effectively reducing disk IO time

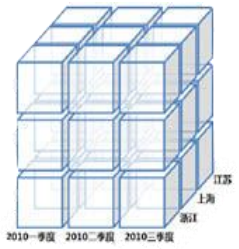
SPL code is simple and easy to implement complex logic

Deliberate distribution reduces data transferring in cluster nodes

Pre-association computing technology to speed up JOIN operations to constant time

Multidimensional analysis

The drag response of multidimensional analysis is very slow, and it makes you think that the drag is not effective. If you make several drags, the system will collapse.



The professional OLAP server is not slow, but does not support nested SQL queries dragged out from BI front-end.

With pre-aggregation method, the speed is improved, but 100T hard disk is not enough, and can only aggregate a dozen of dimensions in advance.

SPL Base Multidimensional Analysis Solution

Innovative pre-aggregation Technology

Pre-aggregate some dimensional combination

Pre-aggregate at a higher dimension level

High Performance Storage and Parallel Computing

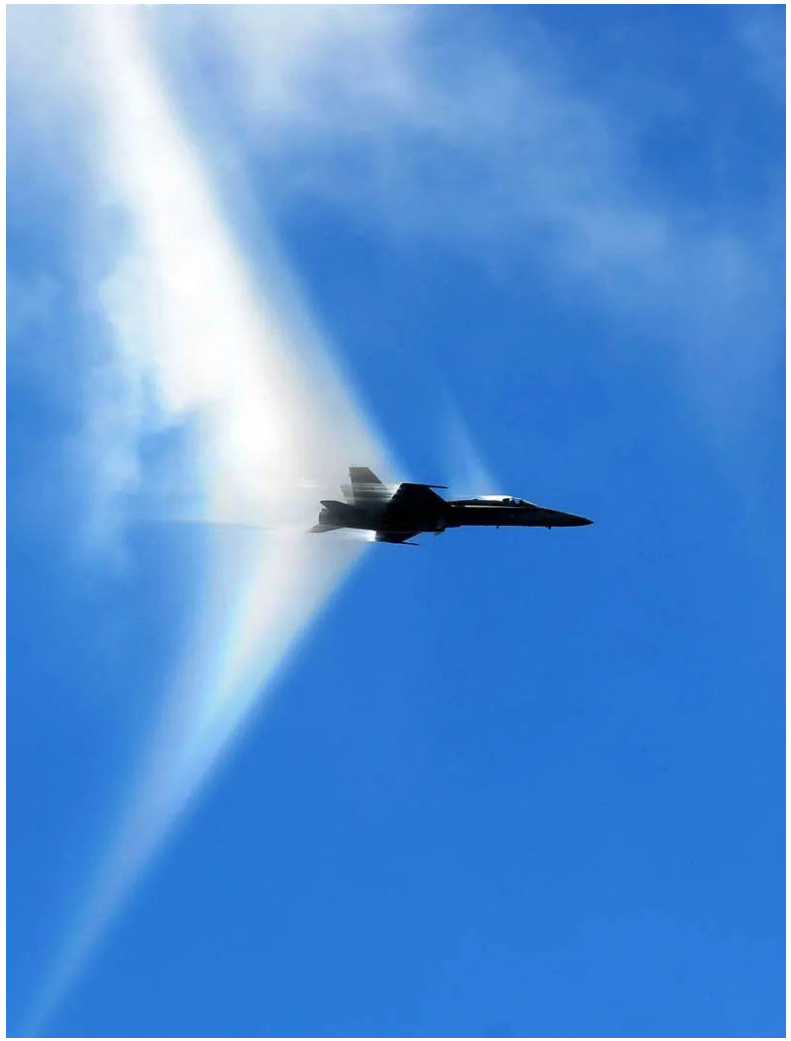
Ordered Compression and Free Column Storage

Multithread Parallel Computing

Algorithm optimization

Pre-association speed up JOINs

Deep optimization of complex SQL



Software and hardware environment

The memory database has come on-line, how can the system still be so slow?

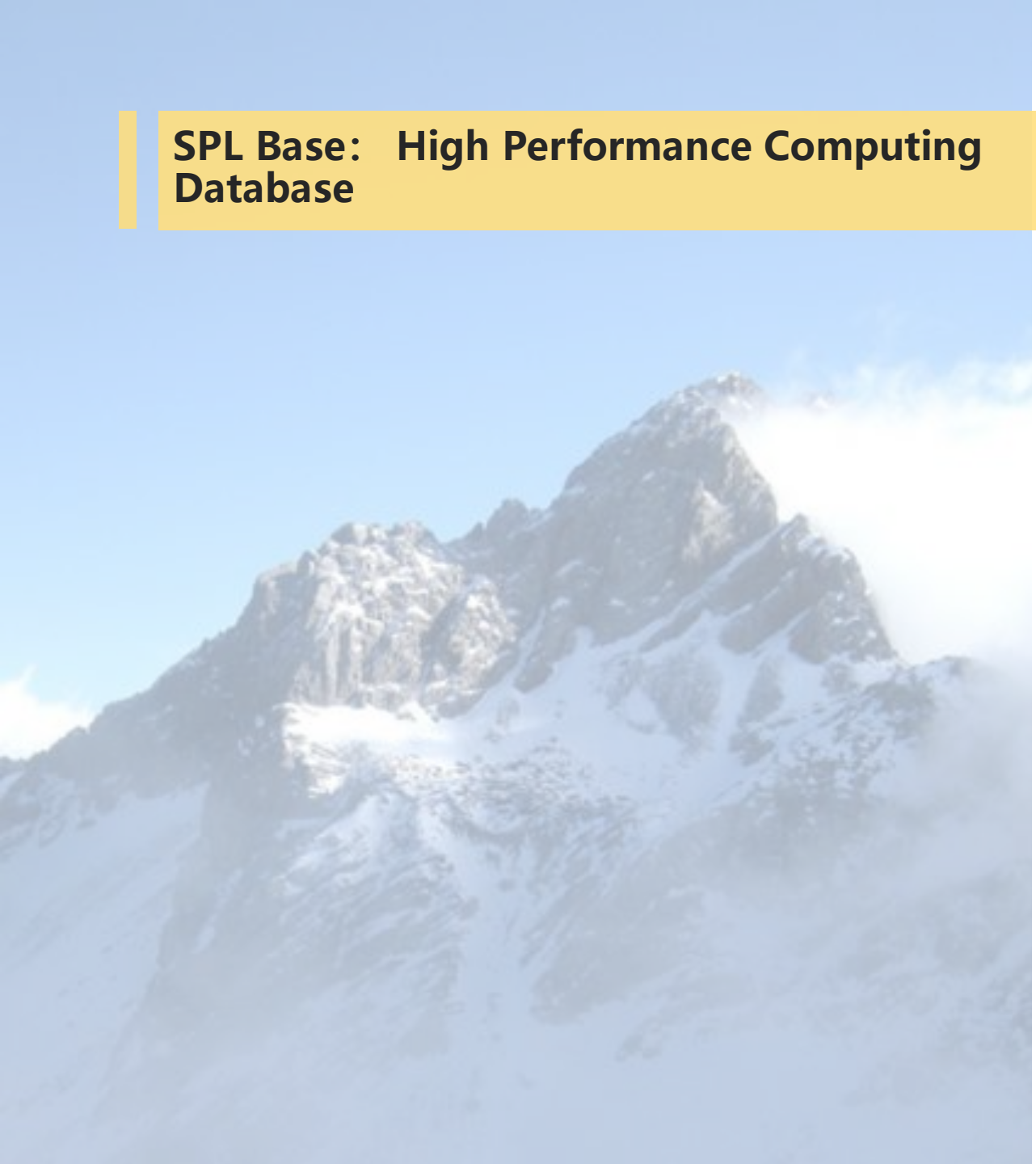
Memory has been added to 1T, why do we still have memory overflow error?

The number of MPP nodes has reached the limit. What if the pressure is still increasing?

There are already dozens of Hadoop nodes, and why is the performance still poor?

The integrated machine is so expensive. Will it solve the problem?

SPL Base: High Performance Computing Database



High Performance Computing Concept

Software can't change the speed of hardware
High Performance Computing = High Efficiency Algorithmic Design + High Efficiency Algorithmic Implementation

SPL Innovated Computing System

The SQL of relational database is like an arithmetic system with only addition, while SPL invented multiplication.

SPL High Performance Computing Mechanism

Efficient traversal, efficient Joins, high performance storage, parallel computing