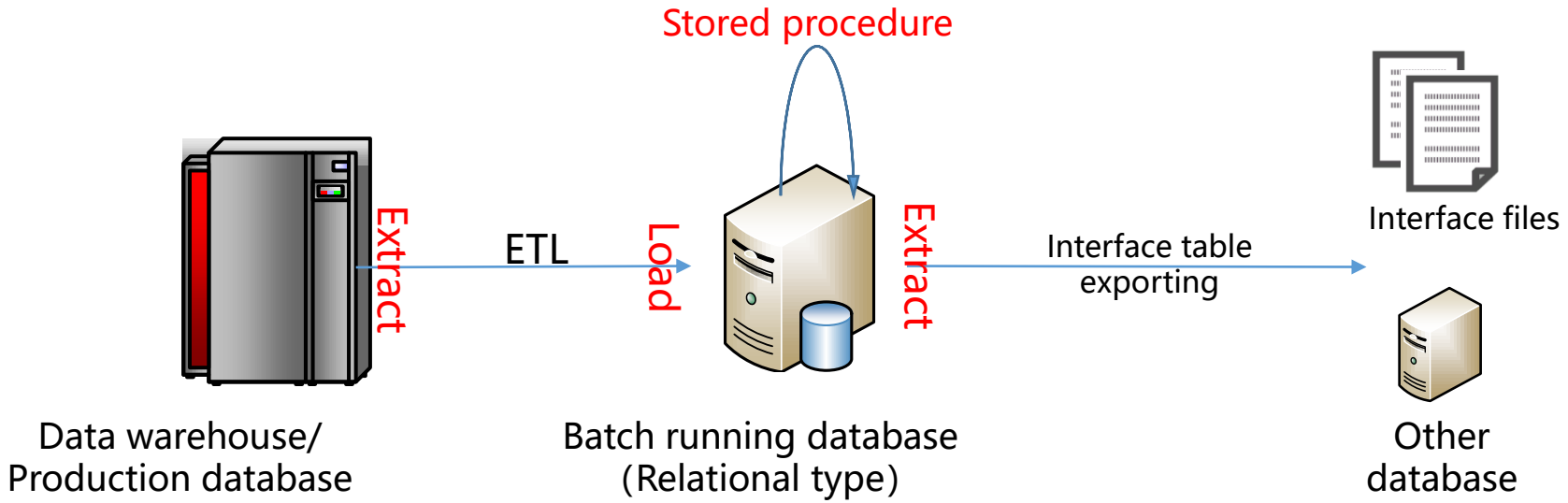
High Performance Batch Running (Batch Computing Task) Solution

esProc SPL Base Application Profile

Current status: Database batch running is too slow and time-consuming



Data Warehouse/Production Database

Oralce, Teradata, Hadoop, etc.

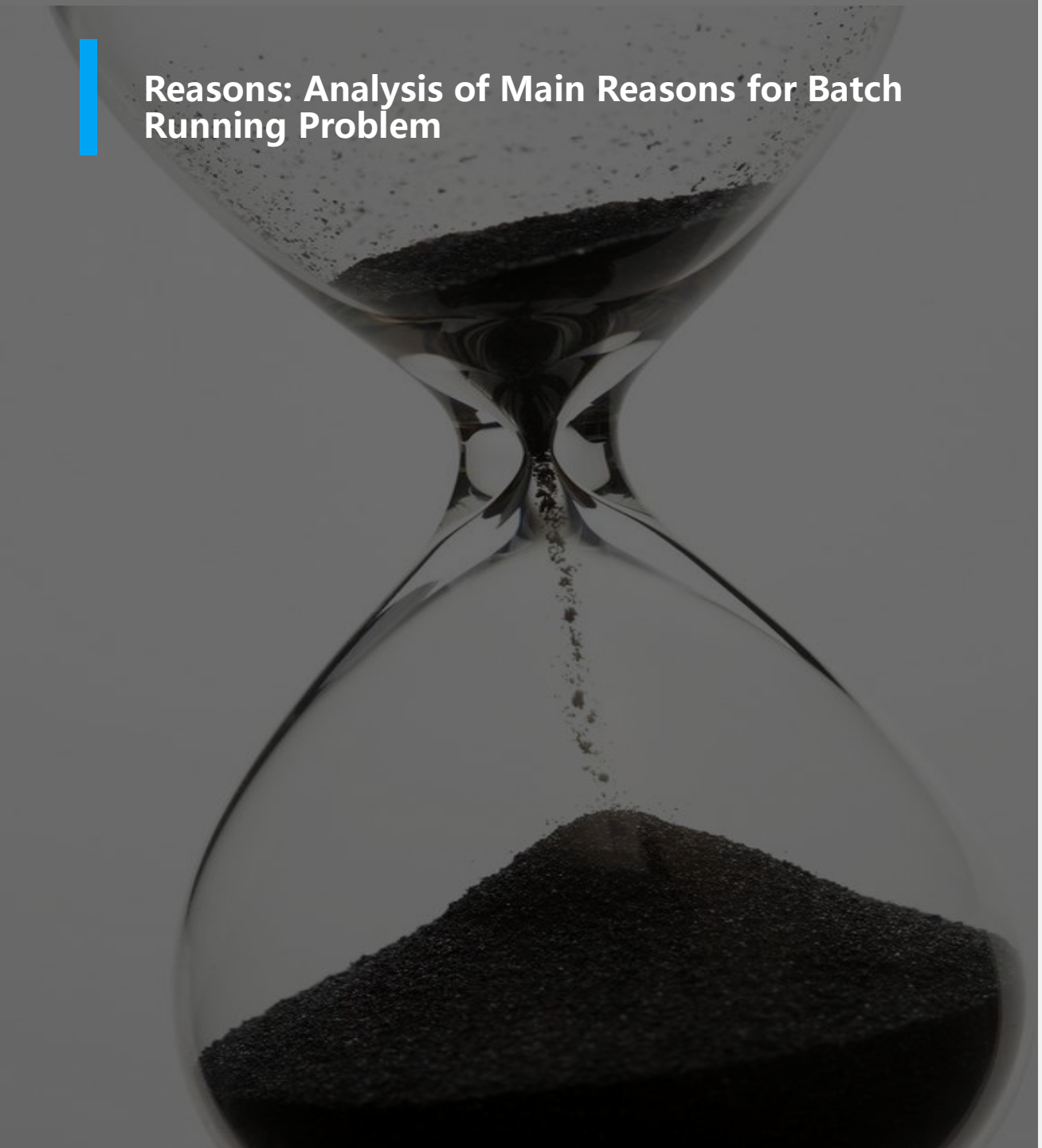
Batch running database

Oracle, DB2, Mysql, etc.

Interface files/Other Database

Text files: other applications
BI database, other application database

Reasons: Analysis of Main Reasons for Batch Running Problem



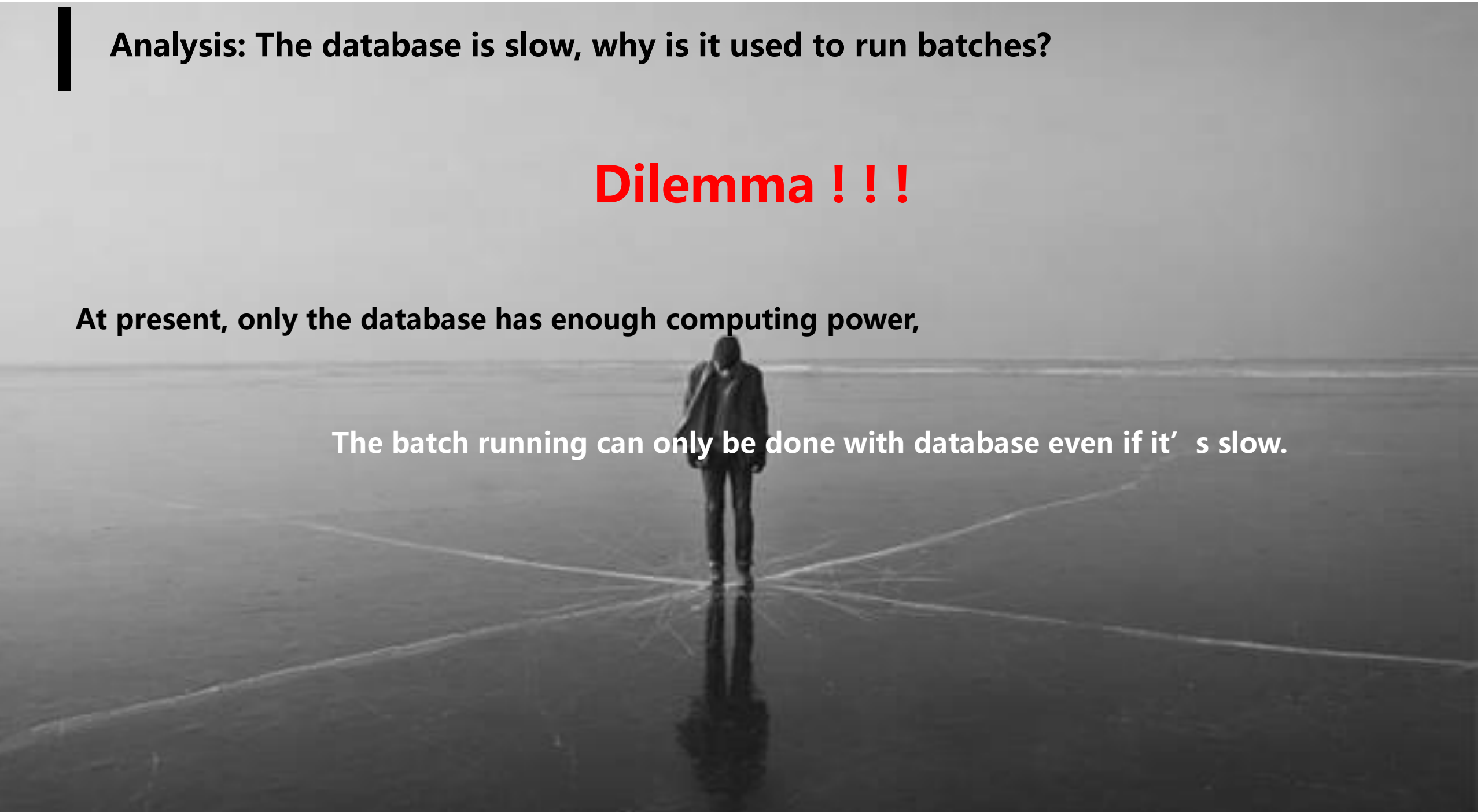
Relational database access is too slow

The storage and calculation ability of database is closed, too much checking and processing are needed for data entry and exit, and it takes too long to import and export a large amount of data.

Poor performance of stored procedures

By the constraints SQL syntax, many efficient algorithms are not supported.
The complexity of the algorithm is high and the calculation of cursor traversal is slow.
Very slow in writing temporary tables.

Analysis: The database is slow, why is it used to run batches?



Dilemma ! ! !

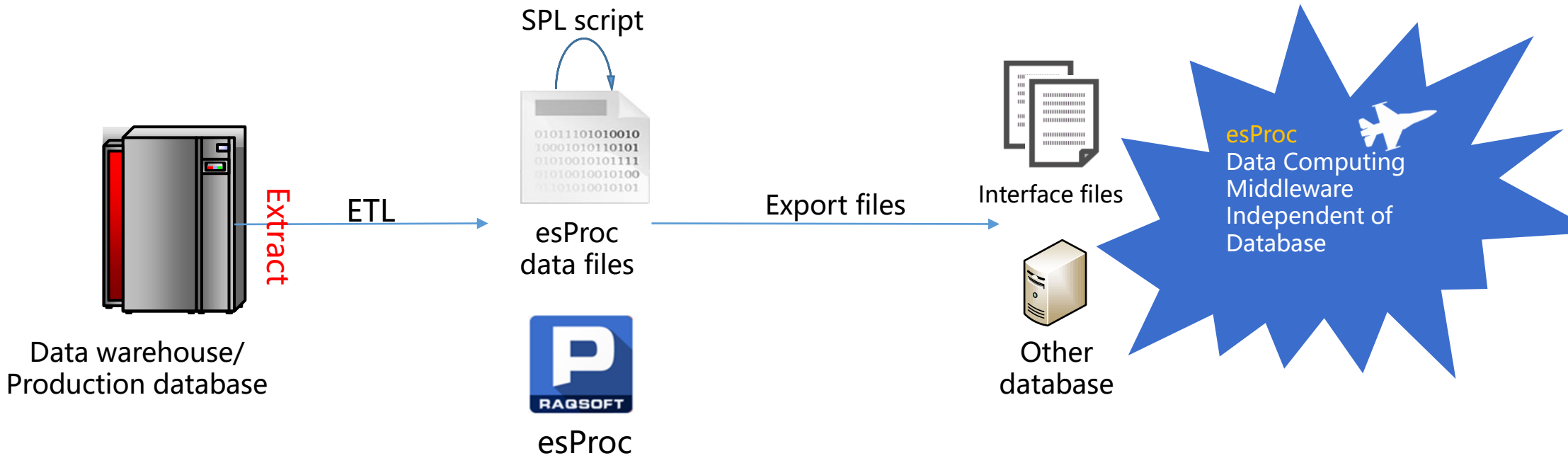
At present, only the database has enough computing power,

The batch running can only be done with database even if it’ s slow.

Analysis: How to break the dilemma?

esProc does not depend on the database, and provides the ability of direct file system calculation and implementation of better algorithm!

New Solution：New system architecture with esProc



ETL

ETL calls esProc SPL script

esProc/data files

High Performance Storage Using esProc Data Files

Interface files/other databases

Text files: other applications
BI database, other application database

Analysis: Main Methods of Speed-up of esProc

Replacing Storage Procedures with SPL Better Algorithms

esProc supports parallel query.
Provide optimized algorithms for join, traversal and parallel of batch running data, and the computational performance is superior to stored procedure.

Direct file-based computing eliminates the time of entry and exit of database

esProc data file is mainly oriented to data analysis and calculation.
Binary, Compression, Column Storage, Double Increment Segmentation, Arbitrary Parallel, and other technologies.
Make full use of hardware computing ability to improve data storage and computing performance.

Automatic conversion: esProc automatically converts SQL into SPL

SQL+

Programmers mark up SQL based on computation and data characteristics

Automatic conversion

esProc automatically converts tagged SQL statements into SPL

Reduce workload

esProc automatically completes 90% conversion and reduces labor costs.

Successful case: a large insurance company batch running project

Optimizing demand

Informix stored procedure is used in running batches. The running time is too long and it needs to be optimized urgently.

Data size

Policy Form: 35 million; Policy Details: 123 million

esProc increase speed by **52.9** times

Scene	Before optimization	After optimization	Increase
Determine risk premium quotation	3600 seconds	68 seconds	52.9 times
Find the last three-year policies	6672 seconds	1020 seconds	6.5 times

Successful case: a bank batch running project

Optimizing demand

The details of public loan agreement contain 48 SQL, which takes 1.5 hours, and becomes the bottleneck of batch running. DB2 optimization is difficult.

Code volume

SQL: 3300 lines; esProc SPL: 500 cells

esProc increase speed by **8.5** times

Scene	Before optimization	After optimization	Increase
Details of Public Loan Agreement	85 minutes	10 minutes	8.5 times

- The End -

THANK YOU