



润乾

RAQSOFT



集算器

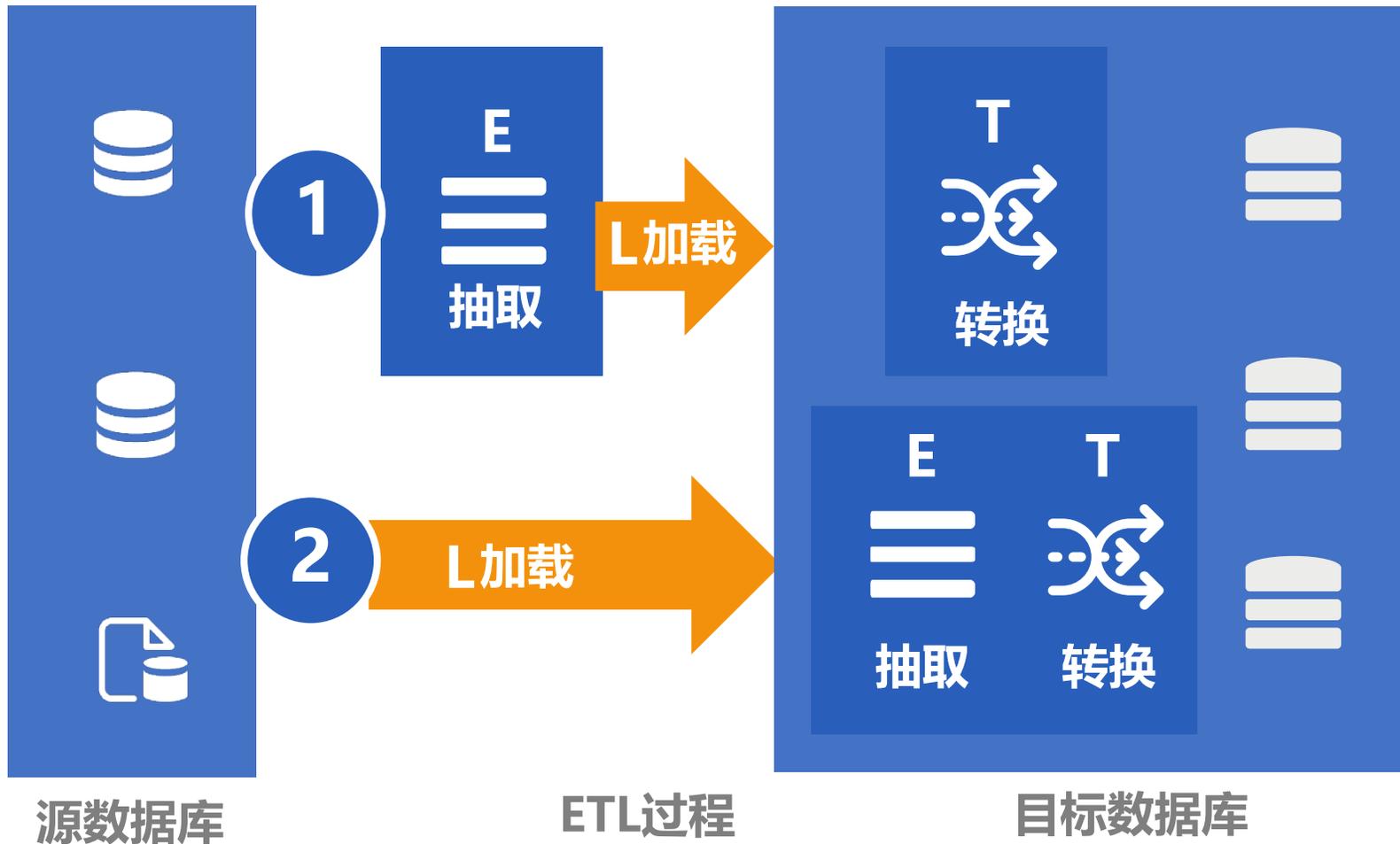
创新大数据计算引擎

ETL中间件

润乾软件出品



传统ETL过程体系结构



缺乏库外计算能力
扰乱ETL过程!

• ETL? ELT? LET?



ETL常常做成ELT甚至LET



01

为什么会这样？

- ETL涉及很多复杂计算和多样数据源，而目前只有关系数据库有足够的计算能力，所以先入库再计算

02

带来的恶果！

- 数据库臃肿、负担重，数据管理混乱
- 性能低下，时间成本高（数据库写入很慢），影响ETL窗口时间

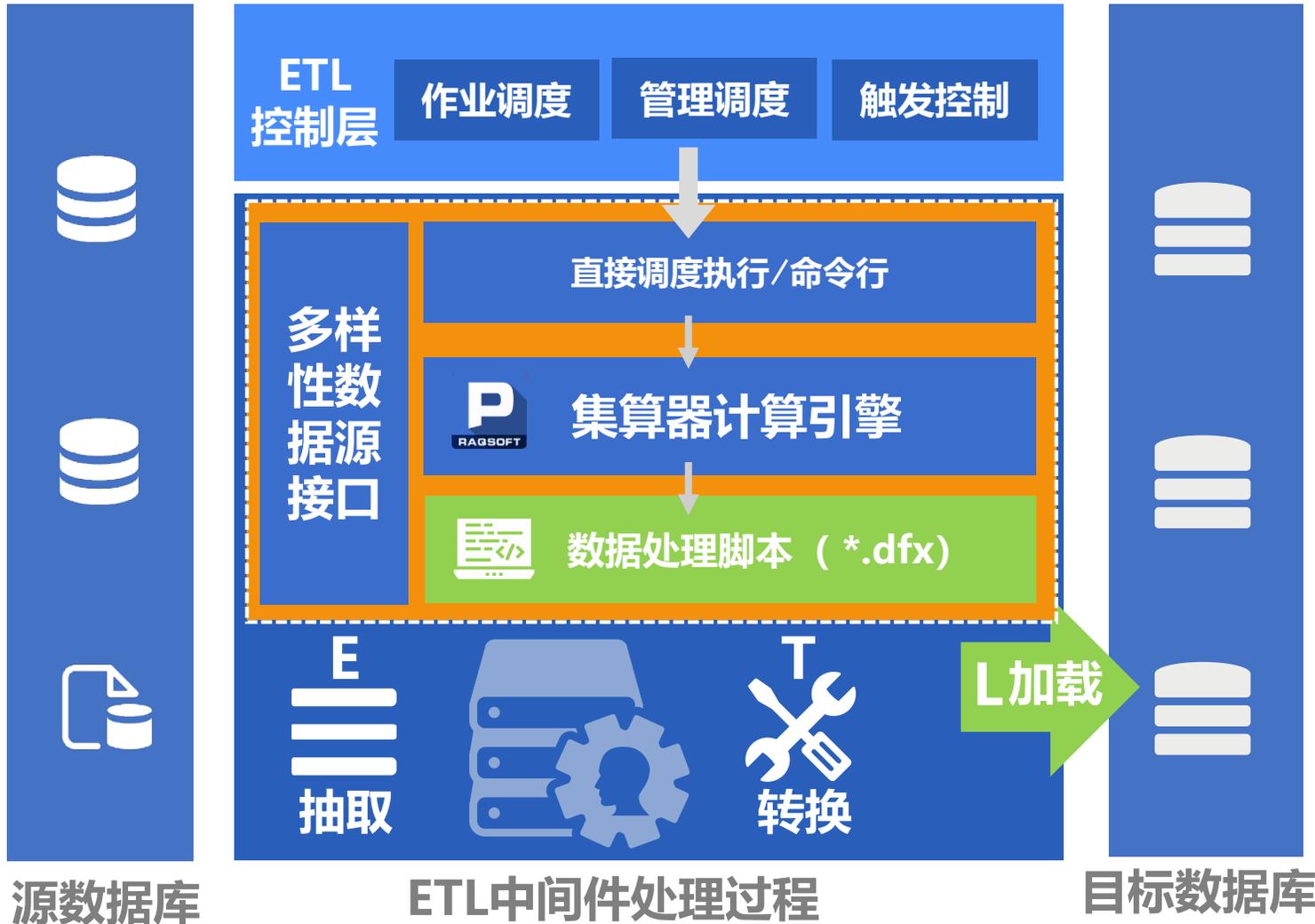
P
RAGSOFT

解决方案

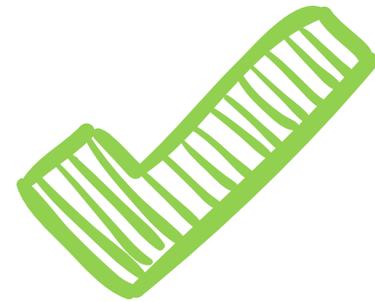
引入库外计算引擎，直接计算多样数据源，不依赖于数据库，真正做成ETL



ETL中间件-库外计算引擎-集算器



库外计算实现合理的ETL





集算器丰富的计算类库

专门针对（准）结构化数据设计，让开发更简单、高效

	A	B	C
1	=esProc.query("SELECT 订单ID AS 合同,订购日期 AS /读取销售记录表		
2	=A1.group(销售)		
3	=create(销售,今年销售额,去年销售额,客户数,大客户数)		
4	for A2	=A4(1).销售	
5		=A4.select(year(日期)==年份).sum(金额)	
6		=A4.select(year(日期)-年份-1).sum(金额)	
7		=A4.group(年份).sum(金额)	
8		=A7.count()	
9		=B7.count(>=10000)	

SPL 分组、循环

	A	B	C
1	=esProc.query("select * from 员工表")		
2	=A1.select(性别=="男")		
3	=A1.select(出生日期>=date("1970-01-01"))		
4	=A2^A3	/交运算，统计晚于1970年出生的男员工	
5	=A2&A3	/并运算，统计男员工或者晚于1970年出生的员工	
6	=A2\A3	/运算，统计早于1970年出生的男员工	
7	=A4.sum(工资)		
8	=A5.avg(年龄)		
9	=A6.sort(出生日期)		
10		/集合作为基本数据类型	
11			

SPL 集合运算

	A	B	C
1	=file("交易记录.txt").import@t0		
2	=A1.sort(客户编码,交易日期)		
3	=A2.select(车辆型号=="捷达" 车辆型号=="迈腾").dup@t0		
4	=A3.derive(interval(交易日期[-1],交易日期):间隔)		
5	=A4.select(车辆型号[-1]=="捷达" && 车辆型号=="迈腾" && 客户编码==客户编码[-1])		
6	=A5.avg(间隔)		

SPL 排序、过滤

	A	B	C
1	=esProc.query("select * from 员工表")		
2	=A1.sort(入职日期)		
3	=A2.pmin(出生日期)	/出生最早的员工的记录序号	
4	=A2(to(A3-1))	/直接用序号访问成员	
5	=esProc.query("select * from 股价表 where 股票代码='000062'")		
6	=A5.sort(交易日期)		
7	=A6.pmax(收盘价)	/收盘价最高的那条记录的序号	
8	=A6.calc(A7.收盘价/收盘价[-1]-1)		
9		/直接用序号访问成员	
11			

SPL 有序集合



集算器多样性数据源支持

- RDB: Oracle、DB2、MSSQL、MySQL、PG...
- TXT/CSV、JSON/XML、Excel、LOG...
- Hadoop: HDFS、Hive、HBase、Spark
- MongoDB、Redis
- HTTP、ALI-OTS
- Ftp、WebService、SAP
- 阿里云、Cassandra
- ES、Kafka
- 内置接口，即装即用

使用集算器，可以直接访问多样数据源进行混合ETL数据计算



ETL中间件-集算器的优势

擅长处理复杂的ETL过程的中间件，不用再硬编码，省时省力省人工！



实现简单灵活



提高处理性能



缩短时间窗口占用



允许多源混合处理

扫码关注

润乾软件 公众号



- ✓ 润乾软件最新动态
- ✓ 产品应用场景、案例
- ✓ 了解润乾，联系润乾

《数据蒋堂》 公众号



- ✓ 技术干货分享
- ✓ 每周一期
- ✓ 微信直播交流