



数据技术嘉年华

// Data Technology Carnival

开源 · 融合 · 数智化 — 引领数据技术发展 释放数据要素价值

DM openGauss PolarDB PostgreSQL MongoDB Hive HBase Teradata
OceanBase GreenPlumCassandra MariaDB DB2 SOLite

Memcached Sybase HANA

Aurora

MySQL SQL Server Redi
OSCAR Claims X-DB iBASE Haisql-lemcach
skysDS Kingwon TrendDB Cedar DragonBas
PDW HotDB Server OushuDB Gridsum ZETA
TalDB GeminiDB TDengine Argodb
MogDB Shentong Megawise TeleDB SinoDB
GreatDB KingDB LongDB ChronusDB RadonDB
UXDB CloudTrable TSDB HUABASE HighGoDB
ESgynDB AnalyticDB SequoiADB ArkDB
GoldenDB AIsQL CynosDB OpenBASE QuantumDB
Base Kingbase TimesTen

Oracle MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen

MySQL SQL Server Redi

OSCAR Claims X-DB iBASE Haisql-lemcach

skysDS Kingwon TrendDB Cedar DragonBas

PDW HotDB Server OushuDB Gridsum ZETA

TalDB GeminiDB TDengine Argodb

MogDB Shentong Megawise TeleDB SinoDB

GreatDB KingDB LongDB ChronusDB RadonDB

UXDB CloudTrable TSDB HUABASE HighGoDB

ESgynDB AnalyticDB SequoiADB ArkDB

GoldenDB AIsQL CynosDB OpenBASE QuantumDB

Base Kingbase TimesTen



中国DBA联盟
All China DBA Union

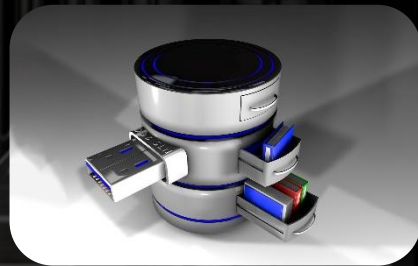


墨天轮

Memcached Sybase HANA
DM openGauss PolarDB PostgreSQL MongoDB DB2 SOLite
OceanBase GreenPlumCassandra MariaDB Hive

概述

光大银行以金融科技创新为支撑，深化金融科技赋能，不断提升数据资产管理能力，自2019年启动数据仓库国产化升级改造，历时三年实现核心数据仓库迁移，完成架构模型转化，成为国内首家实现国产化数据仓库平台大集中的股份制商业银行。本次演讲回顾光大银行数据仓库国产化之路，介绍国产化数据仓库平台建设背景、体系架构、建设成果及推广应用情况。



目录

CONTENTS

01

背景及目标

02

规划与建设

03

成果与创新

04

推广应用





核心技术是国之重器，最关键最核心的技术要立足自主创新、自立自强。市场换不来核心技术，有钱也买不来核心技术，必须靠自己研发、自己发展。

——习近平



深化金融数据要素应用，夯实数字基础底座，发挥数据要素倍增作用。

——中国人民银行《金融科技发展规划》



以金融科技创新为支撑，深化金融科技赋能，提升数据资产管理能力，摆脱对国外产品的依赖，提升我行对数据仓库技术的自主可控能力与敏捷交付能力。

——光大银行

01

全栈国产

- 国产数据仓库体系
- **9**大国产化数据仓库建设要素
- 国内科技力量建设

02

自主可控

- 研发可控
- 过程可控
- 结果可控
- 成本可控

03

数据驱动

- 数据服务于业务
- 十大主题模型整合
- **40+**集市数据管理
- **PB**级数据资产

04

敏捷创新

- 一站式开发平台
- 一站式迁移
- 一站式互联互通
- 一站式调度

05

生态融合

- **MPP**技术生态
- 数据仓库体系生态

技术自主可控，构建国产化数据仓库平台体系



开发敏捷高效，一站式了解数据模型建设



集群分布优化，覆盖数据应用多场景



SLA机制建设，加速银行数字化转型



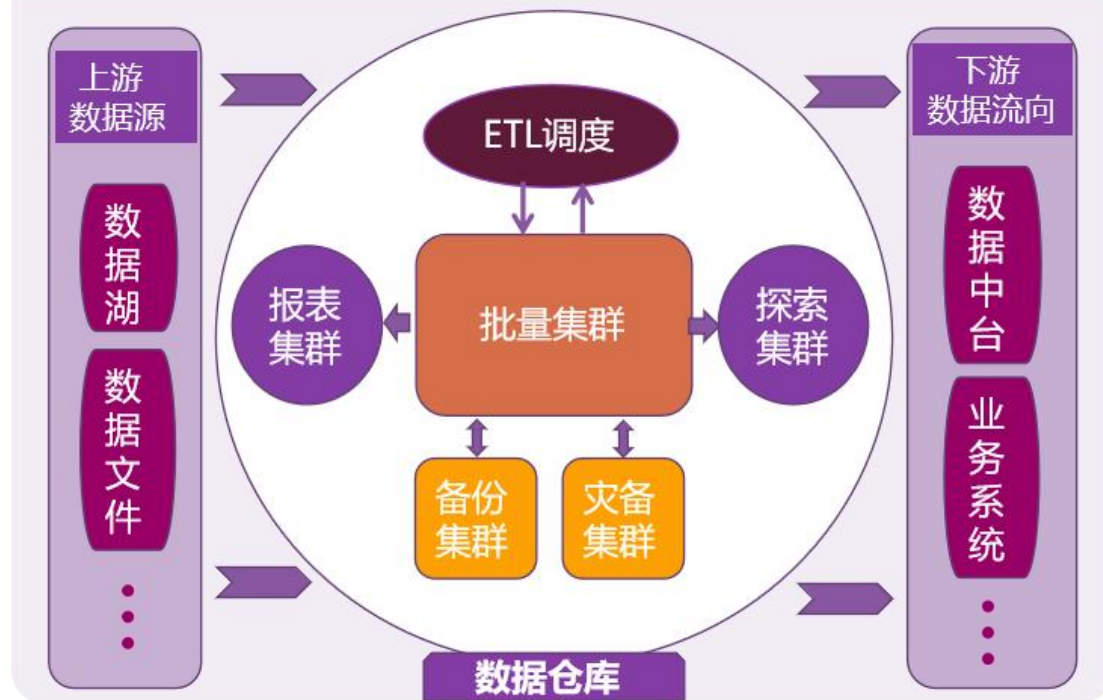
平台统一管理，降低数据分析技术准入门槛



迁移透明平滑，业务无感知



国产化数据仓库全景图



数据仓库功能定位

ETL调度	数仓数据导入、导出、分析加工、数据处理管控中枢。
批量集群	核心数仓，数据加工分析处理核心，各功能集群数据源。
探索集群	支持数据分析处理，灵活查询，业务自定义数据财富挖掘。
报表集群	数仓分析加工结果集，快速响应业务需求。
备份集群	实现数仓逻辑备份，满足开发测试数据需求等。
灾备集群	实现数仓离线物理灾备，保证数据资产安全。

规划-时间规划

两步走战略：先集市后平台、先GreenPlum后Teradata

第一步

2019年12月

GaussDB平台上线

新建试点应用集市云缴费大数据智能分析系统、网络融资智能分析系统。

2021年4月

完成报表集群上线工作

2021年7月

实现Greenplum平台成功下线

完成20余个应用系统的迁移工作，共涉及数据量数百TB，上万张表。

第二步

2021年9月

搭建采用固态硬盘（SSD）的新集群

实现对高斯批量集群的升级替换，一次性成功完成数百TB数据，百万文件迁移切换

2021年11月

探索集群上线

2022年2月

备份集群上线

搭载鲲鹏服务器、麒麟V10操作系统，实现全栈国产化，

2022年3月

Teradata平台成功切换至GaussDB平台

全行40余个数据平台及分析系统基于GaussDB建设，实现国产化数据仓库平台大集中。

建设-逻辑架构

数据仓库平台系统逻辑架构

应用层

- 支持国产**PC**和操作系统
- 兼容手机等移动端用户使用

服务层

- 前端**MIS**和**FRQ**报表工具
- 引擎服务
- 联机数据服务**API**接口

数据层

- 使用**GaussDB**
- 承载模型层加工
- 共性、集市应用数据加工

基础组件层

- 底层为**Suse Linux**操作系统
- 高斯集群使用华为存储
- 网络配备万兆光纤

应用层

POIN前端查询

MIS/FRQ前端
查询

客户端查询工具

ES接口访问

服务层

报表服务
Mis、Freeqry

应用引擎服务

连接数据服务
API

Web服务
REST /MyBatis

数据层

模型基础数据

各平台数据

总分行应用
系统

GaussDB

统一调度
知识库

互联互通工具
知识库

Weblogic

Oracle、Mysql

基础组件层

Suse
Linux

Python

JAVA

操作系统

华为高端
存储

万兆网

X86服务器

硬件、网络、存储

专用物理集群

私有云

数据仓库平台系统物理架构

网络 存储层

- **F5**负载均衡的架构
- 提供用户访问界面

前端 用户层

- 统一的访问接口
- 具有横向扩张能力的应用数据

后端 应用层

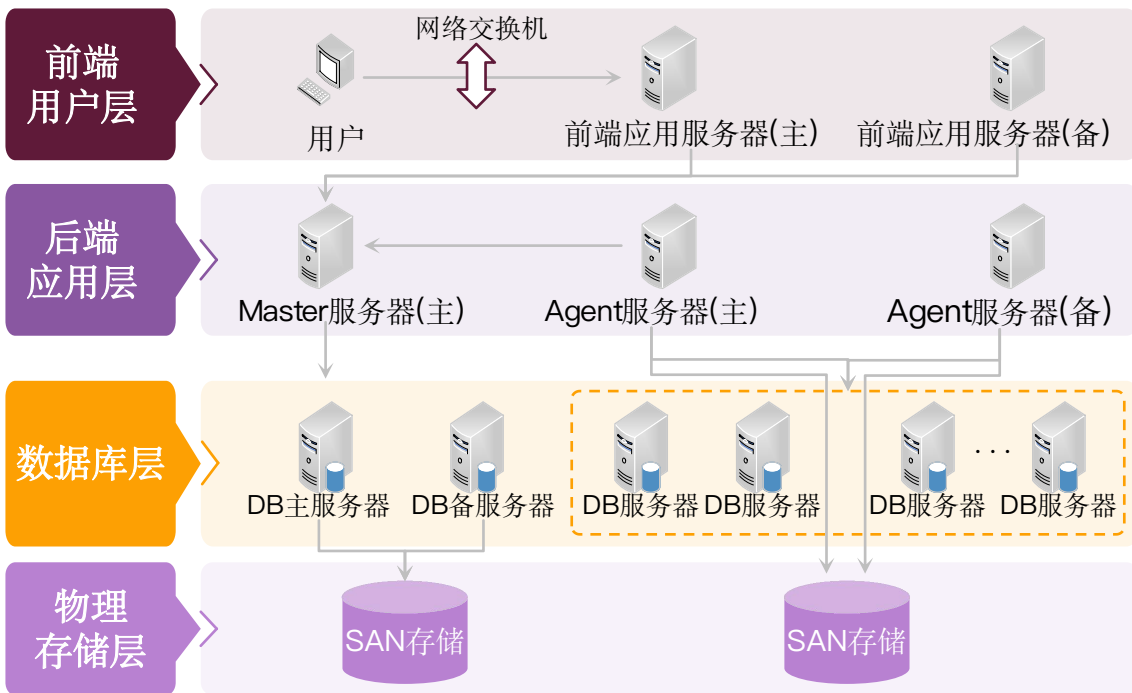
- 采用统一调度体系

数据库 层

- 支撑数据仓库平台、各集市的高斯平台，具有高可用性

物理 存储层

- 利用**SAN**高端存储数据文件



数据仓库平台物理部署示意图

数据仓库平台应用架构

基础层数据集

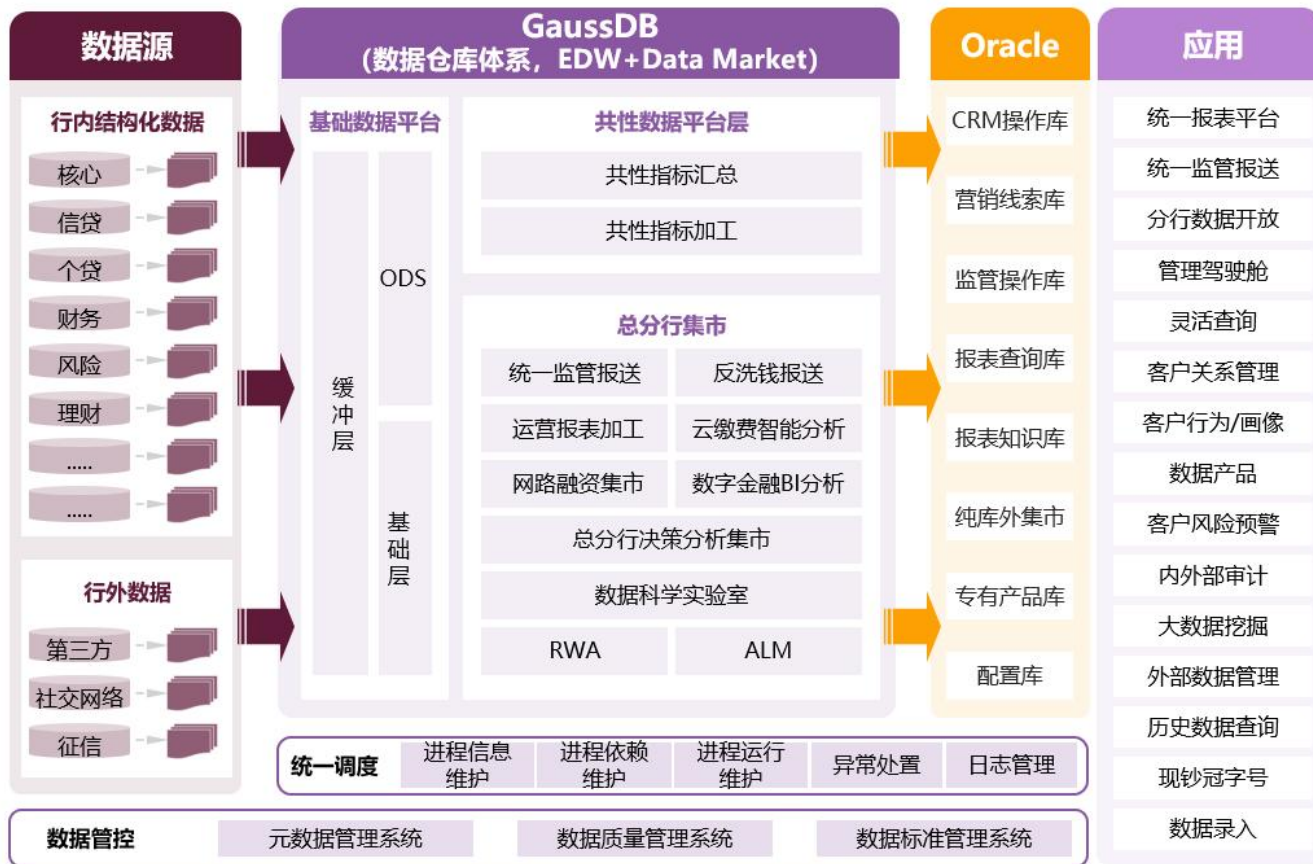
- 源系统数据进行集成和汇总
- 形成十大主题数据

共性平台层数据集

- 对基础层数据进行预联接、预计算、预聚合等操作

集市层数据集

- 对接40余个集市的上百个交易应用



数据仓库平台应用架构示意图

GreenPlum→GaussDB

20余个应用系统，数据
量数百TB，上万张表。



国产化硬件改造

上百节点，数百TB数据，
百万级别文件迁移切换。

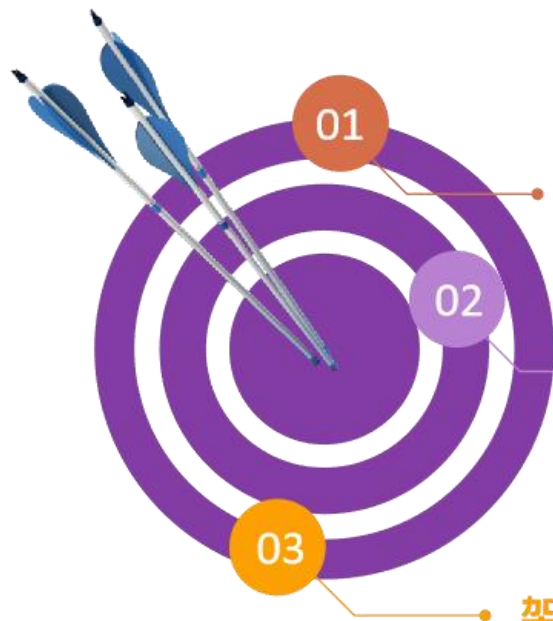


Teradata→GaussDB

近万张表，3000余个脚本，
2000余个数据模型，千亿
条数据的迁移工作，实现
了国内首家国产化数据仓
库平台大集中。



以95%自研代码的GaussDB为基础的国产化软件环境



01 国产化数据平台

建立国产化数据仓库运行平台，建立符合此平台技术特点的开发管理方法和规范，实现通用的模板化程序。

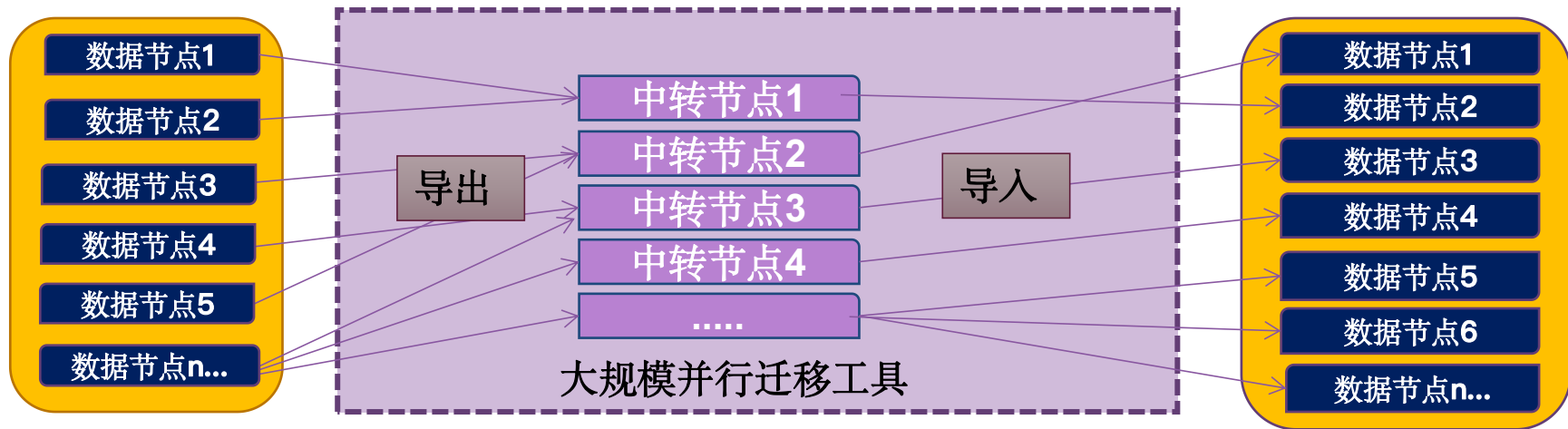
02 技术自主可控

解决我行在MPP数据库方面对于国外产品的依赖，有效降低了信息安全风险，有利于实现技术自主可控的目标。

03 架构优化

逐步将Teradata平台和Greenplum平台上的应用进行迁移，实现成本总体可控的同时减少数据仓库体系的复杂性。

成果2-并行多集群迁移工具



大规模迁移解决方案

1. 支持GP或高斯同类型集群间的数据迁移，支持GP和高斯跨不同类型数据库的数据迁移
2. 迁移效率可以随着迁移服务器的增加而线性扩展
3. 迁移范围可配置化，可以灵活变更迁移内容
4. 可以并发调度，并且可以灵活的控制并发数
5. 每次数据量60TB，支持总共迁移数据量数百TB
6. 支持冷、热数据的分别迁移，分批次迁移

成果3-全智能化迁移验证工具

打造全智能自动化数据迁移验证平台

迁移验证



通过**自动化验证程序**，对新老集群中的记录数分表统计，实现自动化比对。

数据完整性验证



抽取一定数量的表用
checksum的方式进行验证。

功能验证



结合**业务需求**进行**功能匹配验证**及跑批性能验证。

持续验证**6个月**，验证正确率始终在**98%以上**。

成果4-多集群互联互通工具

在高斯平台的主集群、报表集群、探索集群之间，以高效率实现批量自动化、一次性批量任务的导出和加载，解决MPP与Hadoop集群间同步链路长的问题。

互联互通工具的功能

开发与运维

抽取加载任务

生成投产包

页面操作

模版批量操作

GDS集群

异常处理

自动化处理异常信息

任务的创建和管理

运行任务监控

等待任务临时同步

调度工具的松耦合管理

互联互通工具结构

前端界面

任务开发

临时同步

申请审批

状态查询

互联互通工具

用户管理

日常任务开发

权限管理

批量任务导入

日志管理

导出投产包

状态监控

申请及审批

告警管理

临时数据同步

统一调度工具

GDS集群

MPP集群

Hadoop集群

主集群

报表
集群

探索
集群

互联互通工具

Hadoop实验
室集群

成果5-自主研发调度工具

具有“专场”“抗疫”新模式的自主研发的调度工具

- **支持多场景需求调度**

统一调度工具，支持多平台、多批次、多种触发规则的并发调度，支持并发数设置、优先级设置、自定义日历
- **高效率调度引擎机制**

采用内存计算引擎，智能判断当前ETL运行情况，支持海量并发任务调度，不受数据库负载的影响
- **并行可扩展调度架构**

支持多工作站执行任务，任务并行执行可最大化利用各种资源，可线性扩展，架构清晰，便于部署、维护和扩展
- **智能化实时告警效率**

任务失败告警、任务到点未启动告警、任务到点未完成告警等情况下，可支持邮件及短信的告警信息发送
- **专场和抗疫模式**

专场模式保障特殊任务能及时完成，抗疫模式在数据发生错误时可有效阻断错误扩散



成果6-运维体系



推动国产数据仓库全方位应用





统一技术路线

- 国产化数据仓库平台完成了数据技术的统一，降低技术应用成本和管理复杂度，有效管控技术风险。



提升数据质量

- 国产化数据仓库平台通过数据集中整合、数据质量检核引擎和检核规则配置，深化数据治理，构建数据质量闭环，持续提升数据质量水平。



敏捷数据服务

- 敏捷数据服务：构建起开发敏捷、投产敏捷、运行敏捷、服务敏捷的研发机制，加速数据需求迭代，提升交付运行能力，提升数据支撑能力。
- 财务、监管、信用卡、零售、风险、云缴费、公司业务、数金

欢迎关注



金融科技干货分享
长按二维码关注我们
每周一篇成长快乐

DM openGauss PolarDB PostgreSQL MongoDB DB2 SQLite
OceanBase GreenPlumCassandra MariaDB Hive HBase Teradata

Memcached Sybase HANA
Aurora
Oracle Redshift CouchDB
MySQL SQL Server RedisDB H2 LevelDB Percona
Tbase Kingbase TimeTen
GoldenDB AlisQL CynosDB OpenBase QuantumDB
EsgyDB AnalyticDB SequoiaDB Arcton

GreenDB AlisQL CynosDB OpenBase QuantumDB
HBase Kingbase TimeTen
MySQL SQL Server RedisDB H2 LevelDB Percona
Oracle Redshift CouchDB
AuroraHive HBase Teradata MogDB
Memcached Sybase HANA
DM openGauss PolarDB PostgreSQL MongoDB DB2 SQLite
OceanBase GreenPlumCassandra MariaDB Hive

谢谢观看

THANKS FOR WATCHING



中国DBA联盟
All China DBA Union



墨天轮

DM openGauss PolarDB PostgreSQL HIVE HBase Teradata
OceanBase GreenPlumCassandra MariaDB DB2 SQLite
Memcached Sybase HANA
Aurora
MySQL SQL Server Redis
Oracle Claims X-DB iBASE HaisqlMemcached
SiyiTSDB Kingrow TrendDB Cedar DragonBase
PDW HotDB Server OushuDB Gridsum ZETA
TaIDB GeminiDB TDengine ArgonDB
PicoHBase Teradata
ShenTong TeleDB DolphinDB SinoDB
GreatDB KingDB LongDB ChronusDB
HashData Huayisoft
UXDB CloudTable TSDB HUABASE HighGoDB
EgynDB AnalyticDB SequoiADB ArkDB
GdIncDB AlISQL CynosDB OpenBase QuantumDB
Base Kingbase TimesTen
MySQL SQL Server RedisTSDB H2 LevelDB Percona
Oracle RedisDynamoDB Gbase Redshift CouchDB
AuroraHive HBase Teradata MogDB
Memcached Sybase HANA
DM openGauss PolarDB PostgreSQL MongoDB DB2 SQLite
OceanBase GreenPlumCassandra MariaDB HIVE HBase Teradata