

# DTCC

## 2013中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2013

大数据 数据库架构与优化 数据治理与分析

SequeMedia  
盛拓传媒

IT168.com

ITPUB

ChinaUnix

Database  
BDaaS  
flowingdata  
DB2  
NoSQL MySQL  
Oracle Big Data



# 数据质量管理

——金融行业实践





## 侯松

- 新浪微博：**@麻袋爸爸**
- 拥有9年以上的数据库研发管理经验，7年的金融行业从业经验。
- 现任平安集团DBA上海分组经理，曾任中宏保险系统运维经理。
- 擅长项目管理及金融行业数据库应用架构设计。
- 曾担当上海证券交易所交易系统设计，中宏保险两城三中心项目设计。
- 曾任多家培训组织及IT公司的技术顾问。

# 目录



因何而来

设计为始

层级部署

规范设计

# 因何而来



1

法规审计

2

专业管理

3

成本控制

4

性能持续



# 因何而来



# 目录



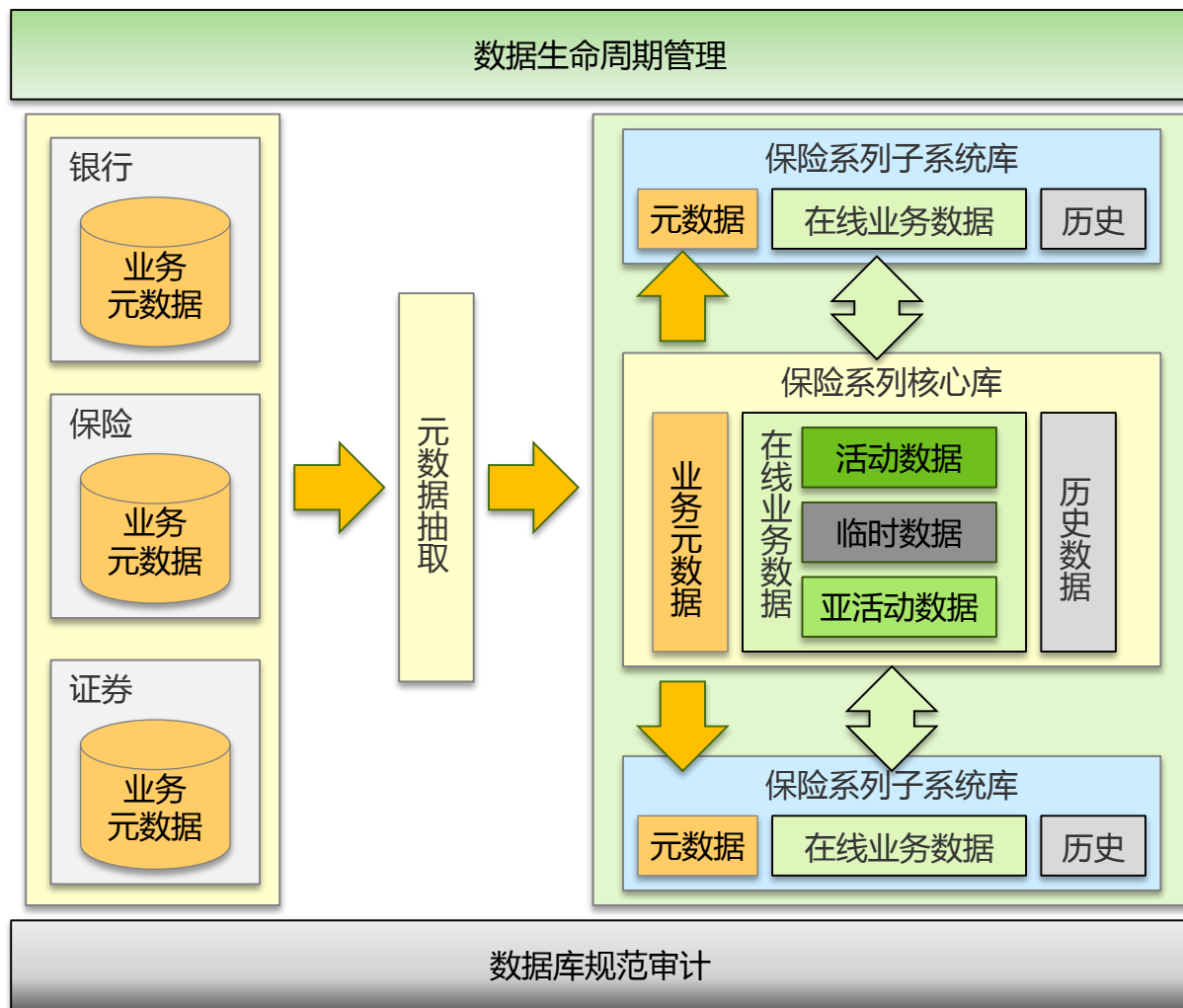
因何而来

设计为始

层级部署

规范审计

# 设计为始



## 数据库建模

- 核心数据库建模
- 元数据、在线业务数据、历史数据分隔表空间
- 业务元数据抽取
- 在线业务数据分类
- 在线业务数据和历史数据赋予生命戳保证细粒度数据质量
- 子系统库参照核心库建模



# 设计为始



## 行级生命戳

- 数据的校验

地区	...	...	...	生命戳
上海	...	...	...	...

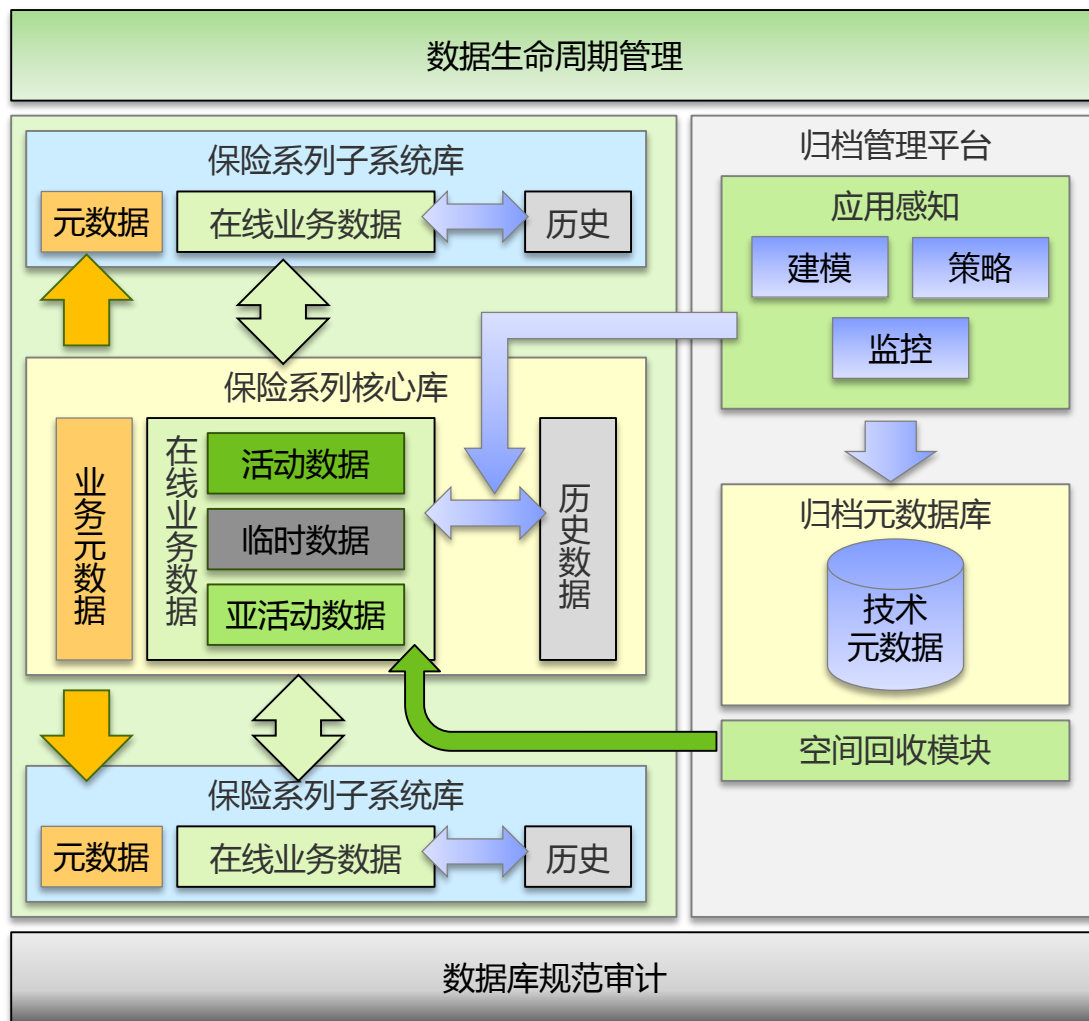
创建人	张三
创建时间	2012-01-01
更新人	李四
更新时间	2013-01-01

- 数据归档的校验

地区	...	...	...	生命戳
上海	...	...	...	...

创建人	张三
创建时间	2012-01-01
更新人	李四
更新时间	2013-01-01
归档人	王二
归档时间	2013-04-01

# 设计为始



## 数据迁移

- 核心库和子系统库通过 GoldenGate 完成数据交互 替代原始 ETL 方式

## 数据剥离

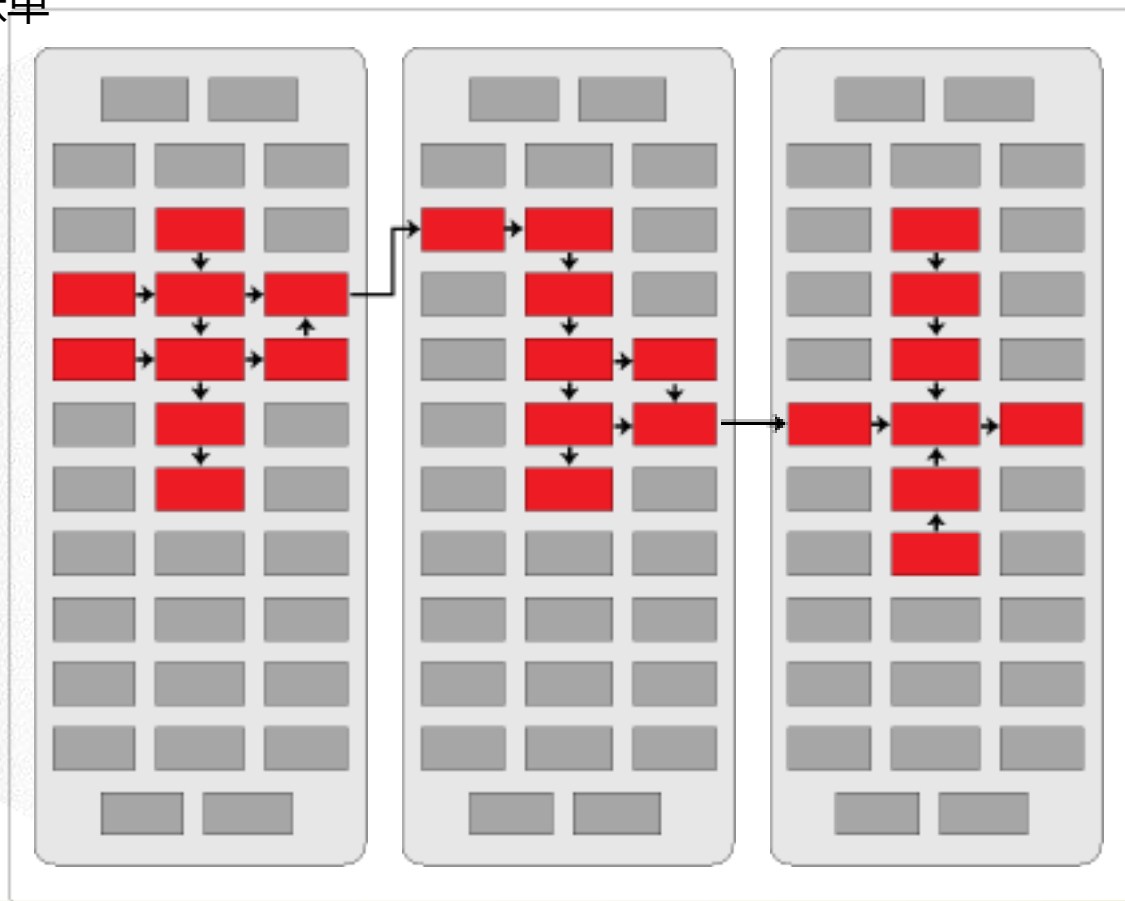
- 借助归档管理平台感知应用模型
- 建立数据归档模型
- 制定剥离策略
- 监控剥离过程并控制负载
- 将剥离信息写入归档元数据库
- 封装包完成在线回收空间

# 设计为始



## 应用感知

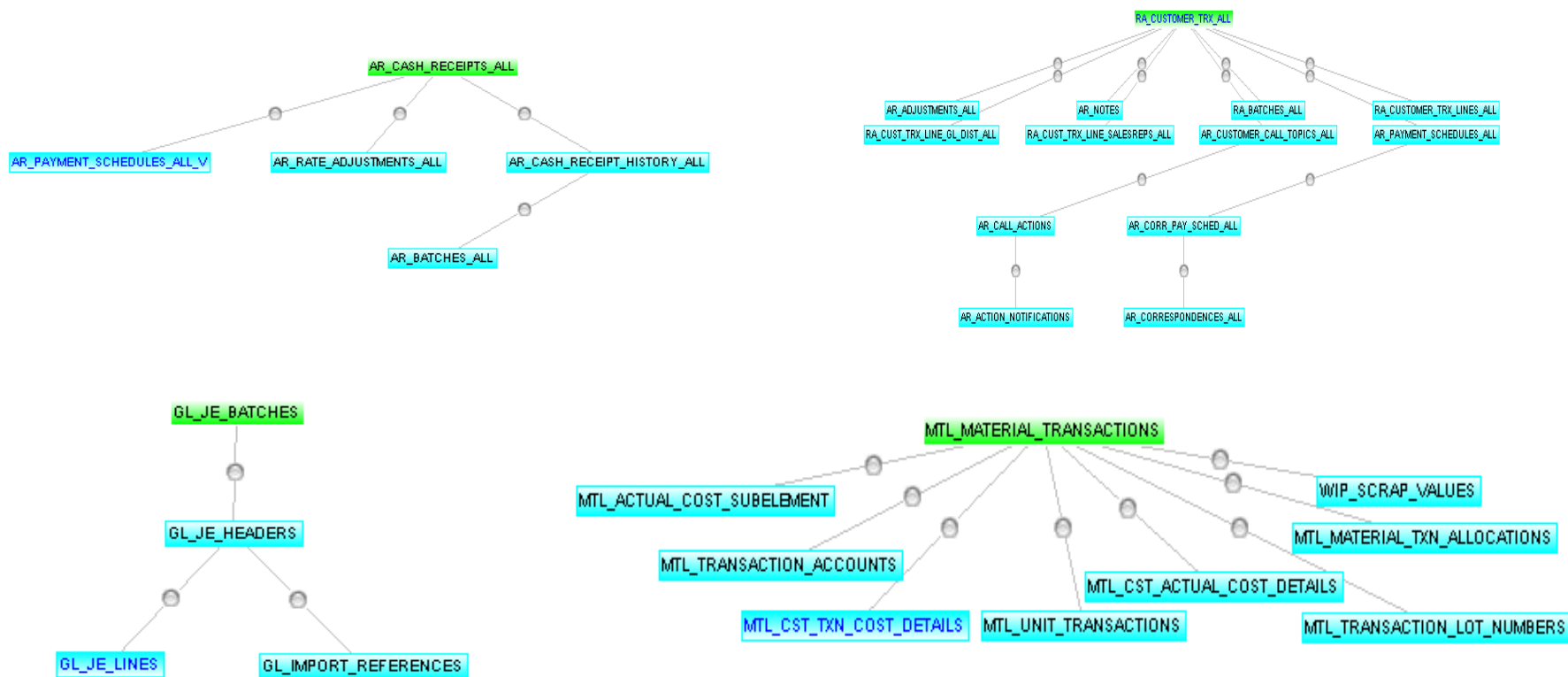
- 识别对象关系：高耦合度业务对象可能跨：表、表空间、分区。
- 举例：用户，销售账单，服务账单



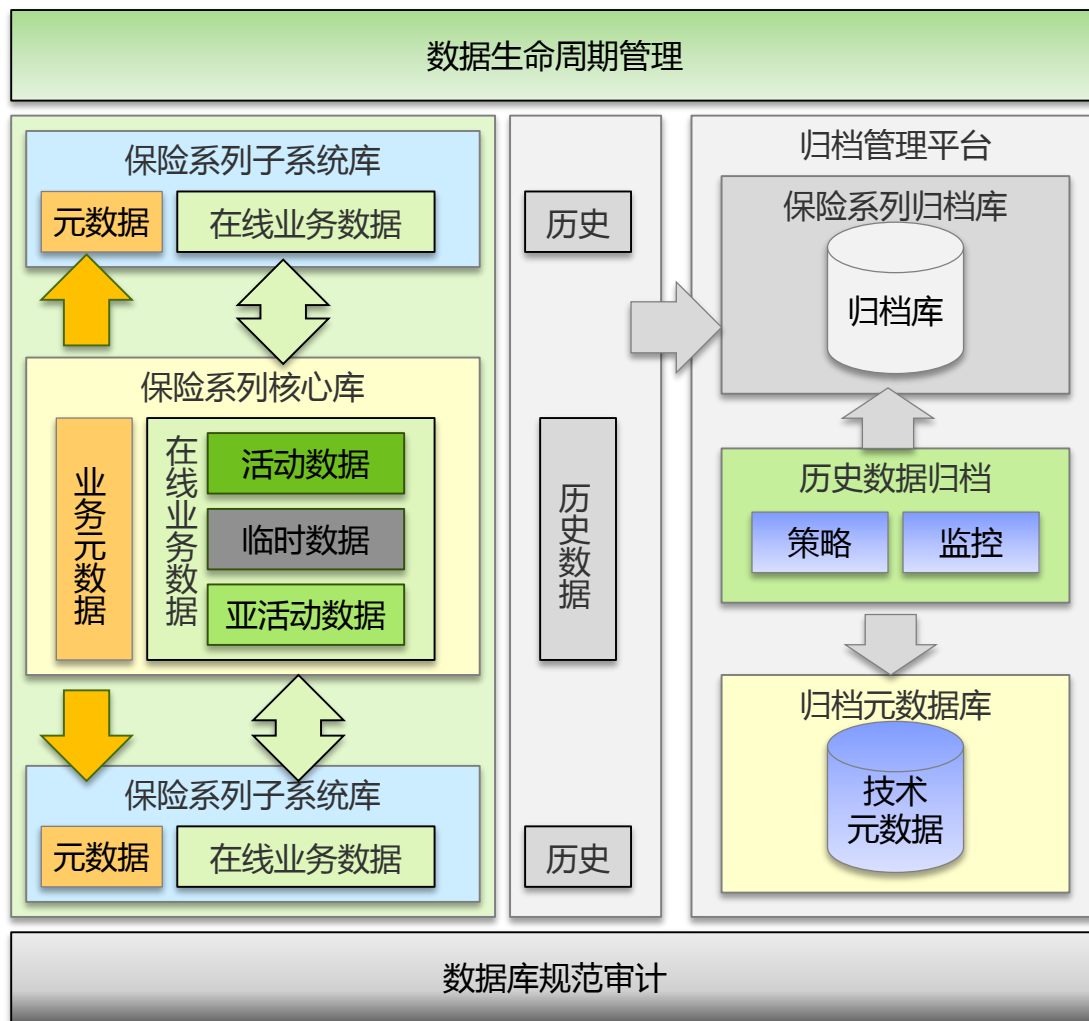
# 设计为始



- 业务对象关系举例——耦合度高



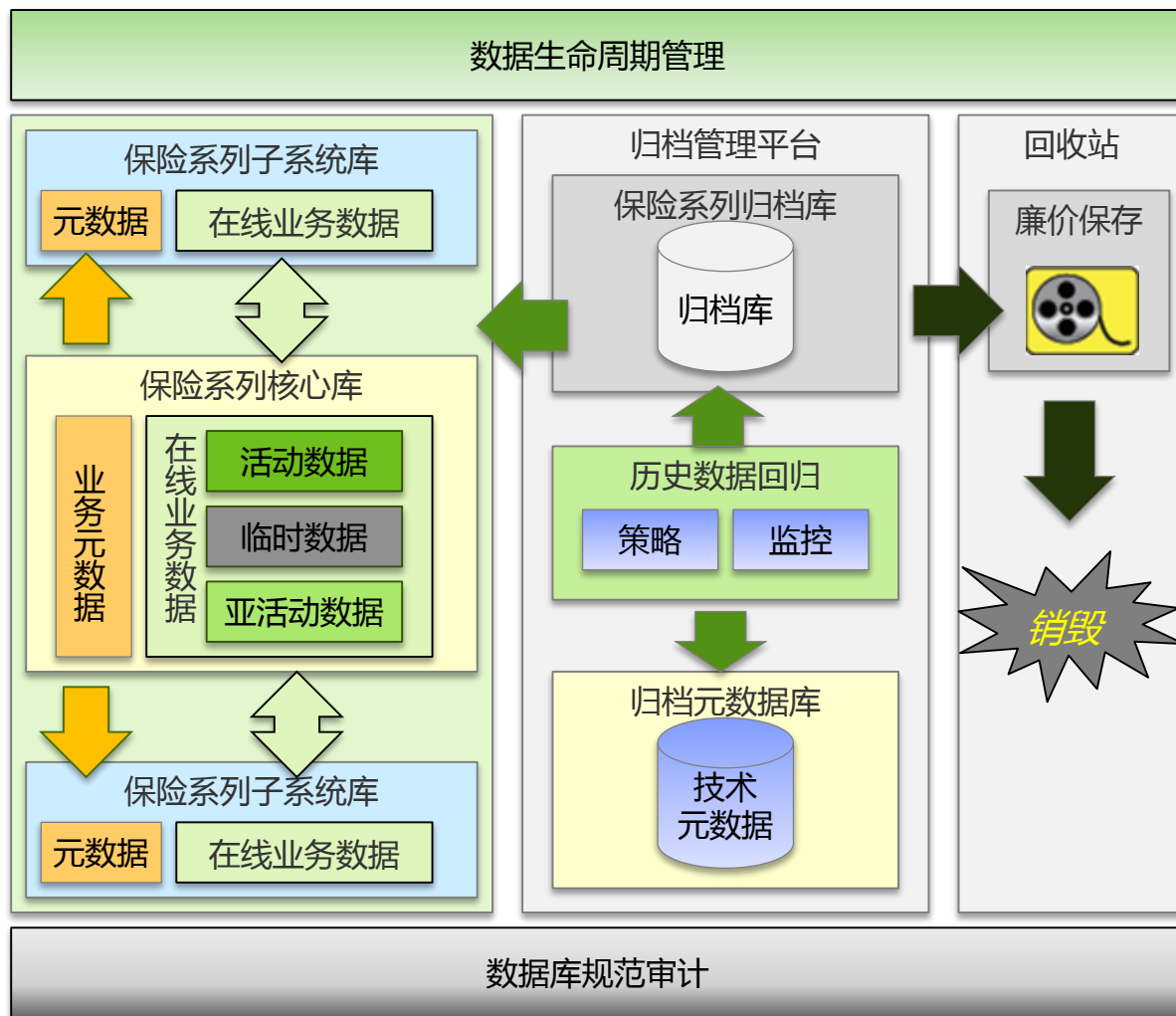
# 设计为始



## 历史数据归档

- 借助归档管理平台分离历史数据
- 历史数据表空间离线
- 迁移到归档数据库
- 将归档信息写入归档元数据库

# 设计为始



## 历史数据回收

数据回归：

- 借助管理平台反归档历史数据
- 将历史数据激活
- 将归档信息写入归档元数据库

数据回收：

- 过期数据廉价保存（如：磁带）
- 第三方存管或销毁

# 目录



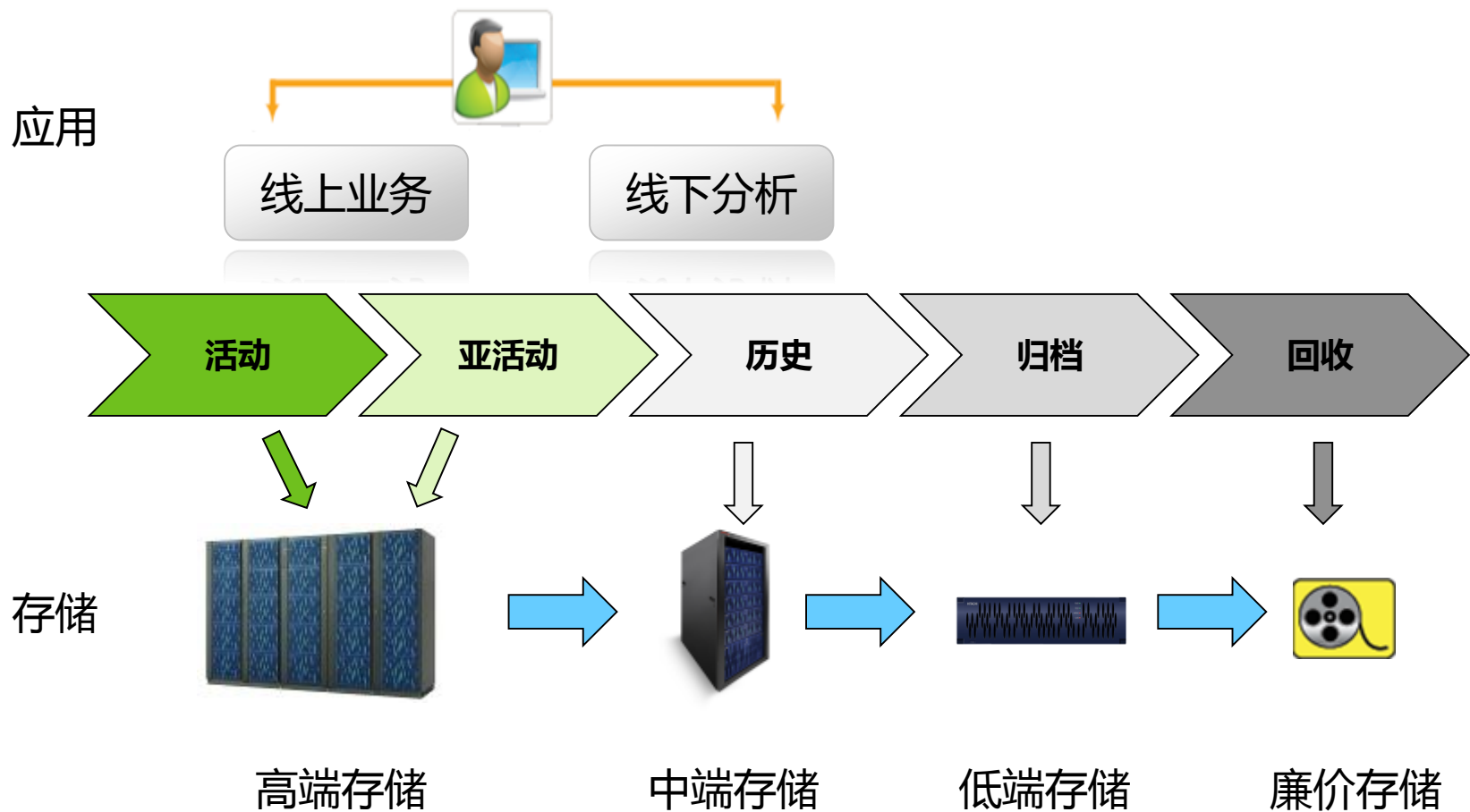
因何而来

设计为始

**层级部署**

规范审计

# 层级部署

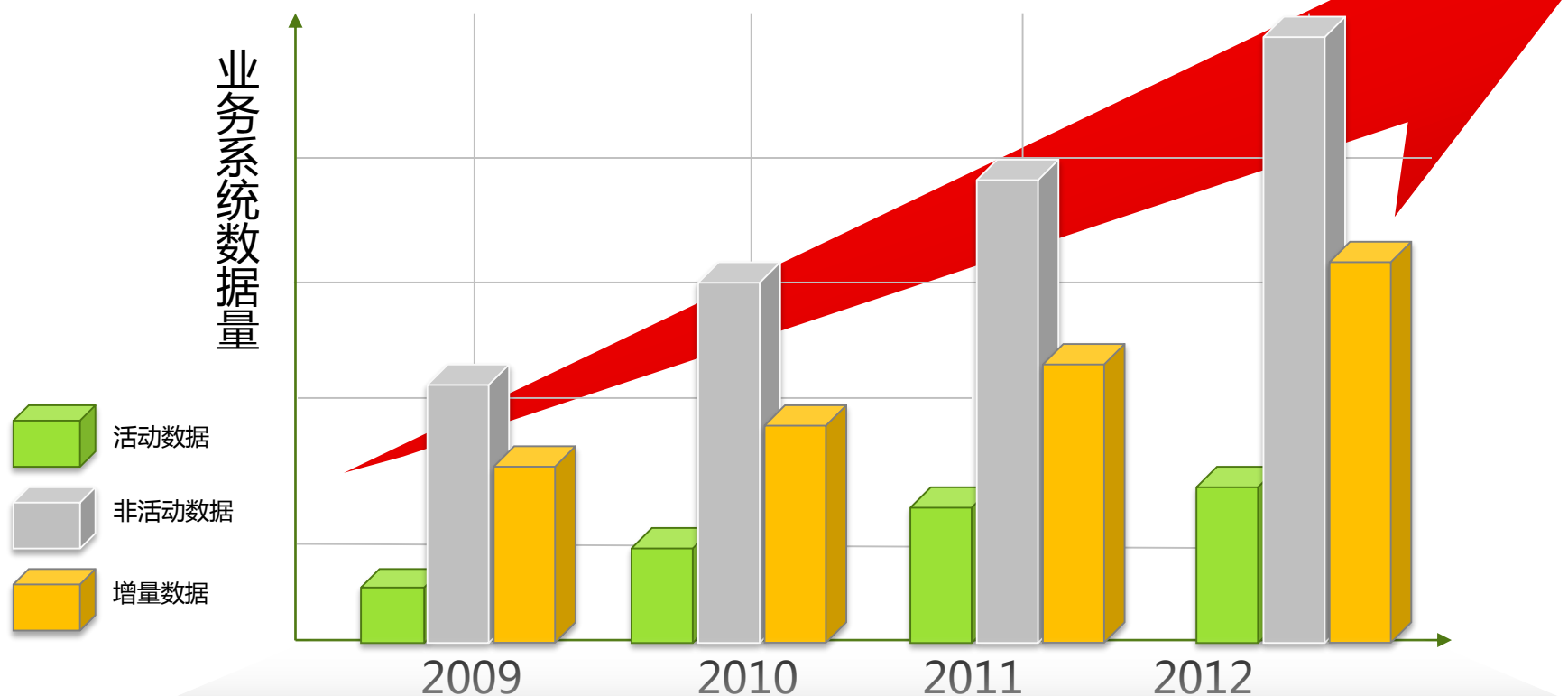




# 层级部署

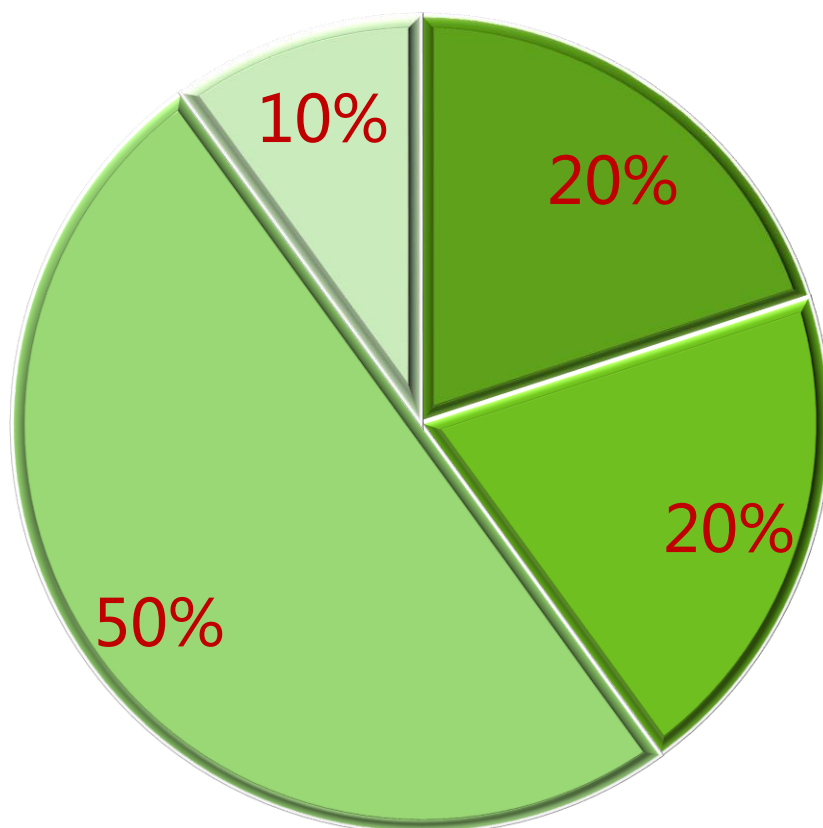


- 业务系统经过多年的运行，产生了海量高价值的数据。
- 60%是非活动数据，20%是活动数据，20%是亚活动数据。
- 非活动数据年平均数据增长量可达30%~40%。





## 数据类型分布率



- 活动数据
- 亚活动数据
- 历史数据
- 临时数据

# 层级部署



## 财务应用题：

一个5TB的数据库，对应一套灾备环境，层级部署带来的收益是怎么样的？

存储类型	单一部署			层级部署		
	容量（GB）	首年费用	次年增长	容量（GB）	首年费用	次年增长
高端	10000	¥4,500,000	¥1,350,000	4000	¥1,800,000	¥540,000
中端	0	¥0	¥0	1000	¥90,000	¥27,000
低端	0	¥0	¥0	4000	¥180,000	¥54,000
廉价	0	¥0	¥0	1000	¥500	¥150
合计	10000	¥4,500,000	¥1,350,000	10000	¥2,070,500	¥621,150

首年费用节省=4,500,000-2,070,500= 2,429,500元

次年增长节省=1,350,000-621,150= 728,850元

# 层级部署



性能应用题：



## 负重胖子 VS 世界冠军

# 目录



因何而来

设计为始

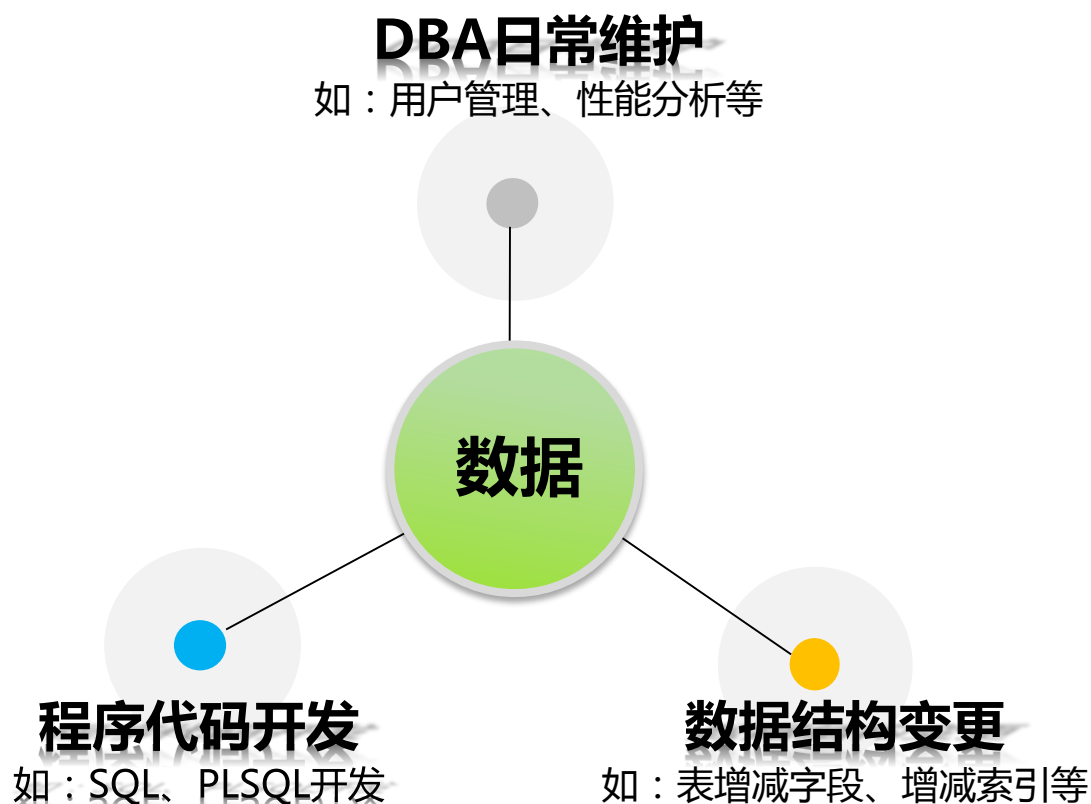
层级部署

**规范审计**

# 规范审计



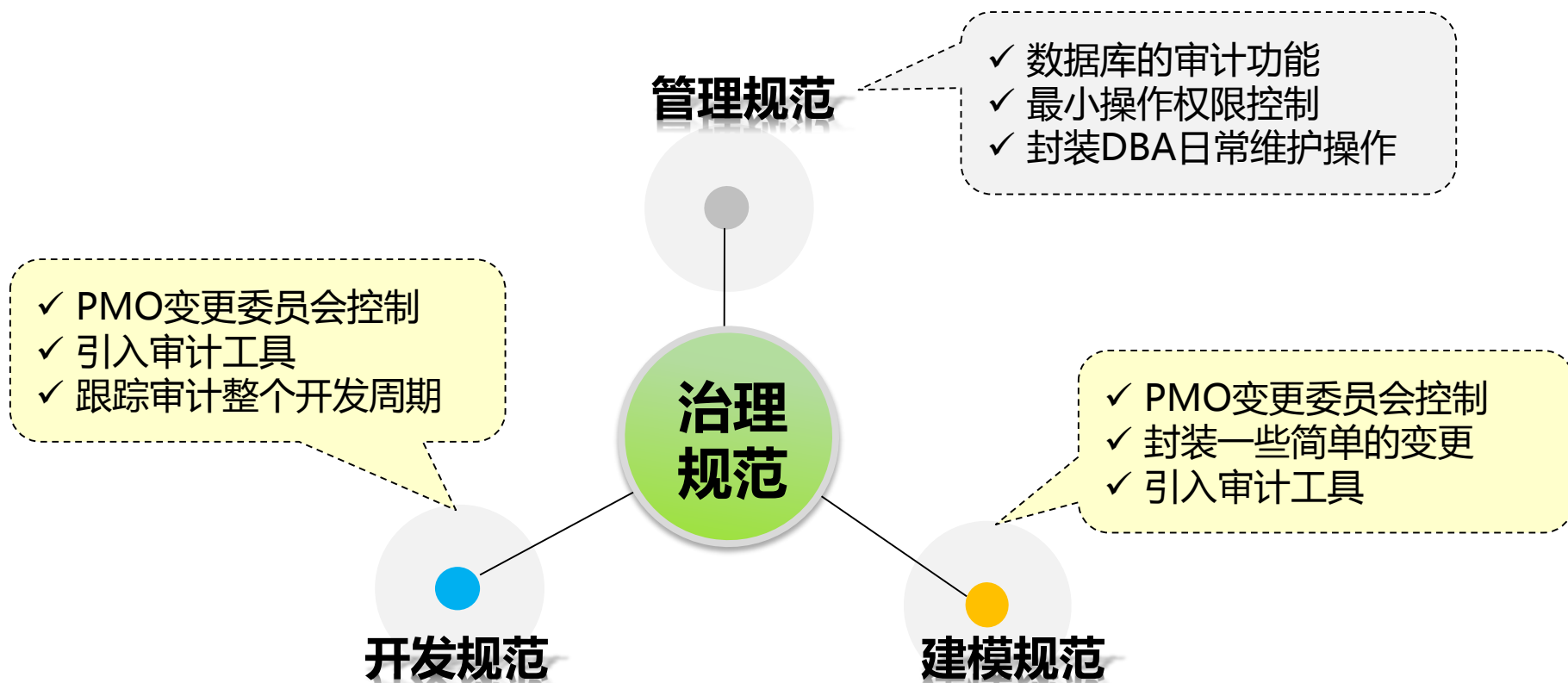
- ◆ 数据接触可以归总为三个维度：DBA日常维护、程序代码开发、数据结构变更。
- ◆ 数据质量的审计维度可以接触跟踪为主线，进行必要的审计和控制。



# 规范审计



- ◆ 制定一套有行业和企业特色的数据治理规范，并工具化保证严格执行。
- ◆ 治理规范从三个维度控制，特别是反映DDL的建模规范、反映DML的开发规范。

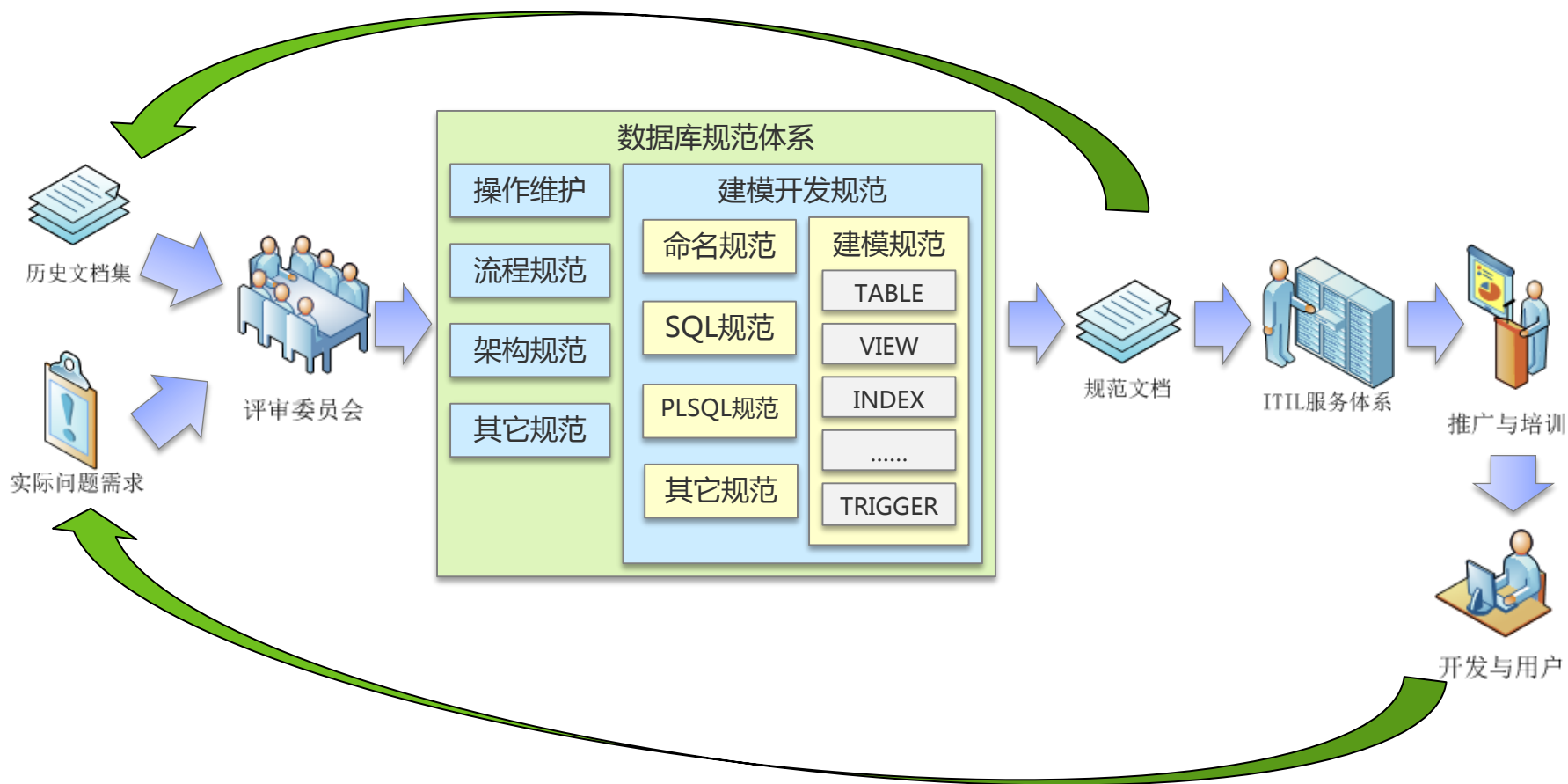




# 规范审计



## ◆ 数据库规范体系建设





# 规范审计



## 1

### DBAUDIT

- 数据库建模及数据结构变更审计工具。
- 主要针对建模规范进行审计，比如：TABLE、INDEX、VIEW等数据库对象的创建、删除、修改。
- 提供开发自主审计接口和上线审计接口，并定期报告绩效。

## 2

### Code Analysis

- 程序代码开发规范审计工具。
- 主要针对程序代码进行审计，比如：SQL语句，PACKAGE、PROCEDURE、FUNCTION代码等。
- 提供开发自主审计接口和上线审计接口，并定期报告绩效。

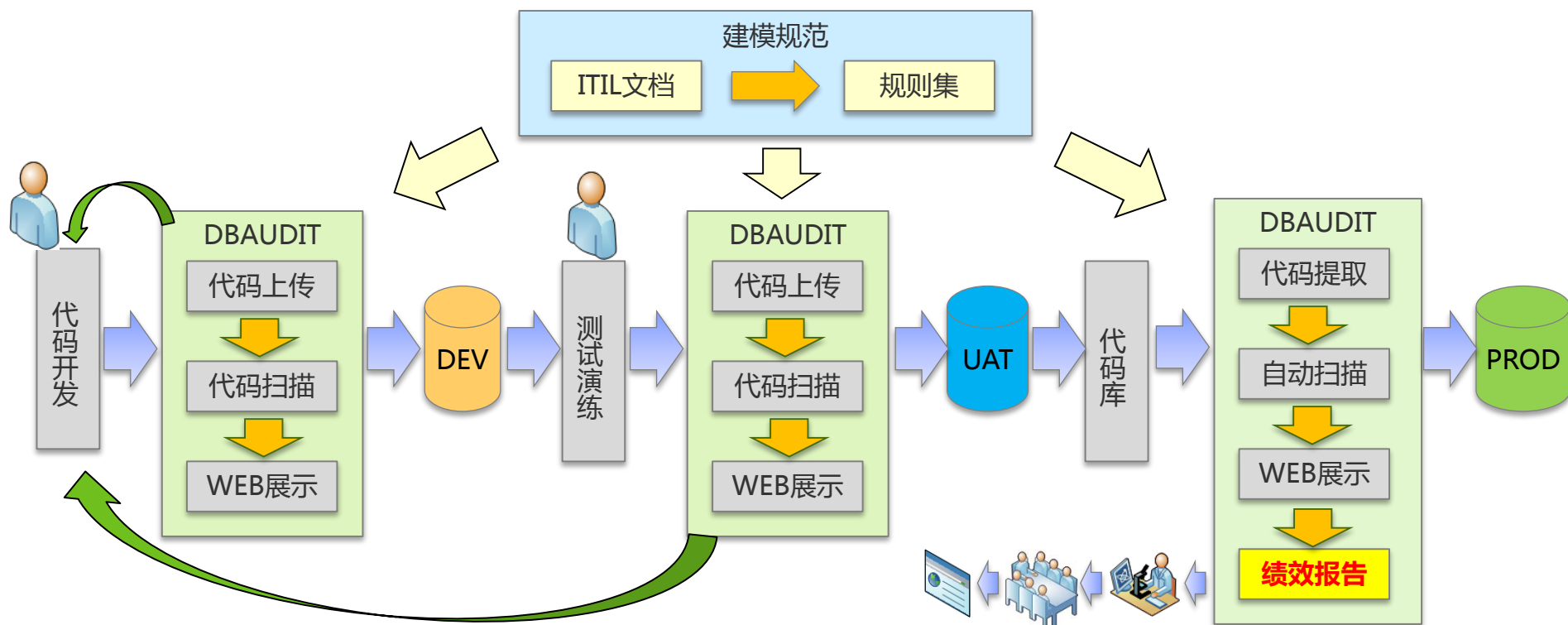


# 规范审计



## 1 DBAUDIT

- ◆ 文档化建模规范，根据规范文档生成相关规则集，并注入到DBAUDIT工具。
- ◆ DBAUDIT工具贯彻开发、测试、上线三个阶段。开发测试阶段，违规项和整改建议将返回再次开发或者申请例外处理；上线阶段，生成当月绩效报告，分析讨论后反馈KPI绩效报告。



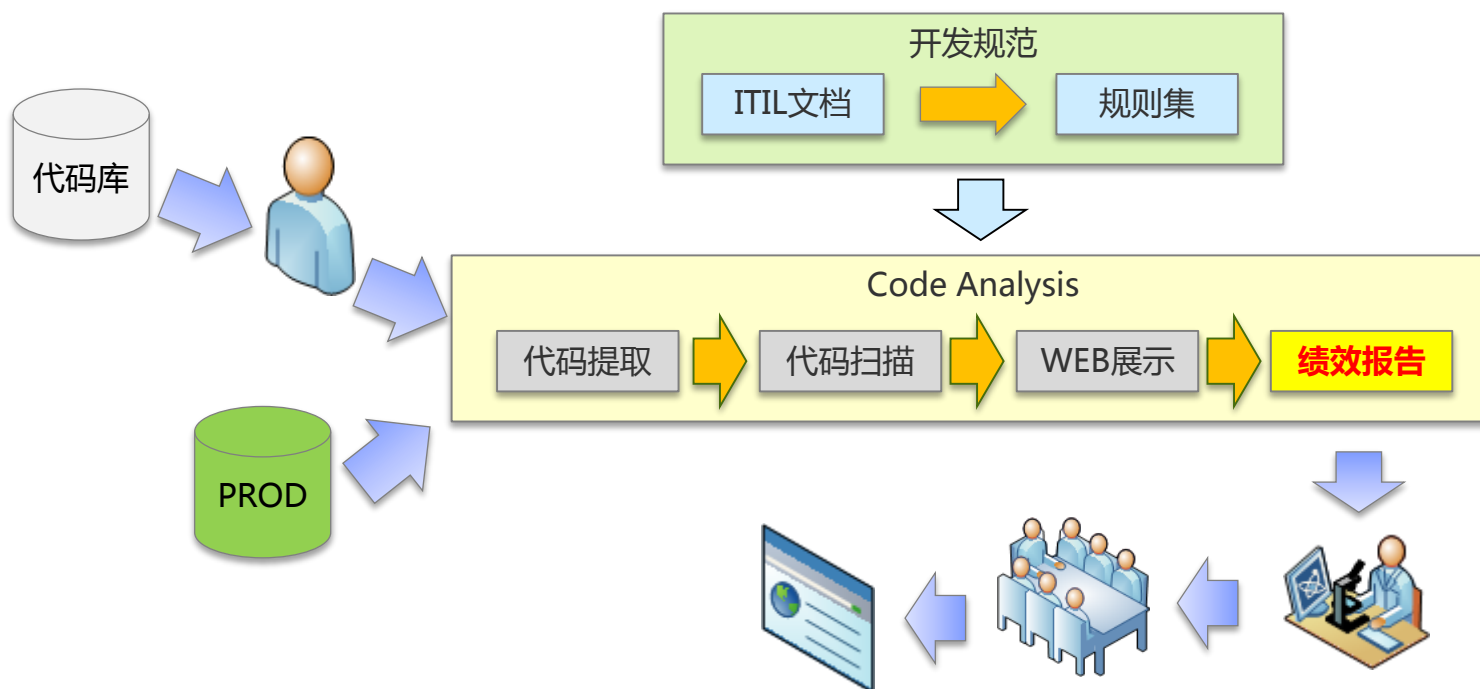
# 规范审计



## 2

## Code Analysis

- ◆ 文档化SQL和PLSQL开发规范，根据规范文档生成相关规则集，并注入到Code Analysis工具。
- ◆ Code Analysis工具主要针对线上代码质量检查，同时也开放接口给开发者自行检查，并生成当月绩效报告，分析讨论后反馈KPI绩效报告。





## 优点

1. 自主研发，个性化定制，直接反映现实情况。
2. 适用性强，每个规则和建议都是经验的沉淀。
3. 避免第三方工具的个性差异的尴尬。
4. 实现了数据生命周期为主线的数据质量管理。

## 缺点

1. 工具开发和维护的长期人力成本较高。
2. 严格的流程控制可能引起开发周期的延长。



# 评委意见



YES



法规审计

YES



专业管理

YES



成本控制

YES



性能持续

# 结语




- 感谢：IT168、ITPUB、DTCC
- 感谢：中国平安DBA团队
- 感谢：在座的每一位伙伴

The top half of the image features a green gradient background with diagonal bands of varying shades, from dark green on the left to light green on the right.

谢 谢

欢迎莅临

2013中国数据库技术大会



Database  
**BDaaS**  
flowingdata  
DB2  
NoSQL MySQL  
Oracle Big Data