



基于HBase的 企业级大数据平台

目录

- 1 公司及平台简介
- 2 **HBase**为企业级平台提供的解决方案
- 3 企业级业务应用支持



我们是谁？

国内最大的保险公司

数亿客户

十几亿保单

上百万营销员




传统数据管理手段所面临的挑战

- 数据虽已集中，但未被集成和整合。
- 范式化模型带来的查询时需要大量join操作。
- 难以实现以客户为颗粒度的查询和分析
- 没有一个ROI很好的解决方案来实现我们的需要。

通过建立企业级大数据平台，我们期望能做到


- 企业级数据资产的整合和连接
- 基于该平台来支撑企业内部的各类数据应用
 - 面向营销员、客户的精确查询服务
 - 面向以客户为颗粒度的精准分析
 - 实现以客户为中心的风控体系
 -
- 以一个较低的成本来解决我们的各类需求

金字塔大数据平台——现况

			
规模	加工	数据	应用
5TB+高速内存 1PB+存储容量	每日加工上亿条 数据	数亿客户数据 十几亿保单数据	为公司管理层及 百万销售员等提 供数据查询和计 算服务

目录

- 1 公司及平台简介
- 2 **HBase为企业级平台提供的解决方案**
- 3 企业级业务应用支持



The diagram features a central green cloud-like shape containing the text '传统数据库存在越来越多的问题'. Surrounding this central shape are five smaller green cloud-like shapes, each containing a specific problem. These are connected to the central shape by a thick, light green line that forms a continuous loop around it. The problems listed are: '系统繁多' (Many systems), '查询复杂' (Complex queries), '加工缓慢' (Slow processing), '更新困难' (Difficult updates), and '丢失细节' (Loss of details).

传统数据库存在越来越多的问题


查询复杂

加工缓慢

系统繁多

更新困难

丢失细节



插入即更新
允许只更新部
分列

百万并发
毫秒返回

容纳数十亿行
和数百万列

细节全部保留

加工集中

所以我们用HBase

One Giant Table for All System

一张
宽表
存储
全国
系统
保单



容纳上亿行数据
与上百万列

公司原始
数据库

系统A

SQL Server

系统B

Oracal

系统C

MongoDB

系统D

infomix

One Giant Table for All System

国
一
张
宽
表
存
储
全
部
系
统
保
单

HBASE		
	rowkey	
row1	: 业务键+110000+X	列1, 列2, 列3.....
row2	: 业务键+110000+Y	列1, 列2, 列3.....
row3	: 业务键+440000+X	列1, 列2, 列3.....
row4	: 业务键+440000+Y	列1, 列2, 列3.....
.....		

容纳上亿行数据
与上百万列

公司原始
数据库

系统X
系统Y
.....

北京 (110000)

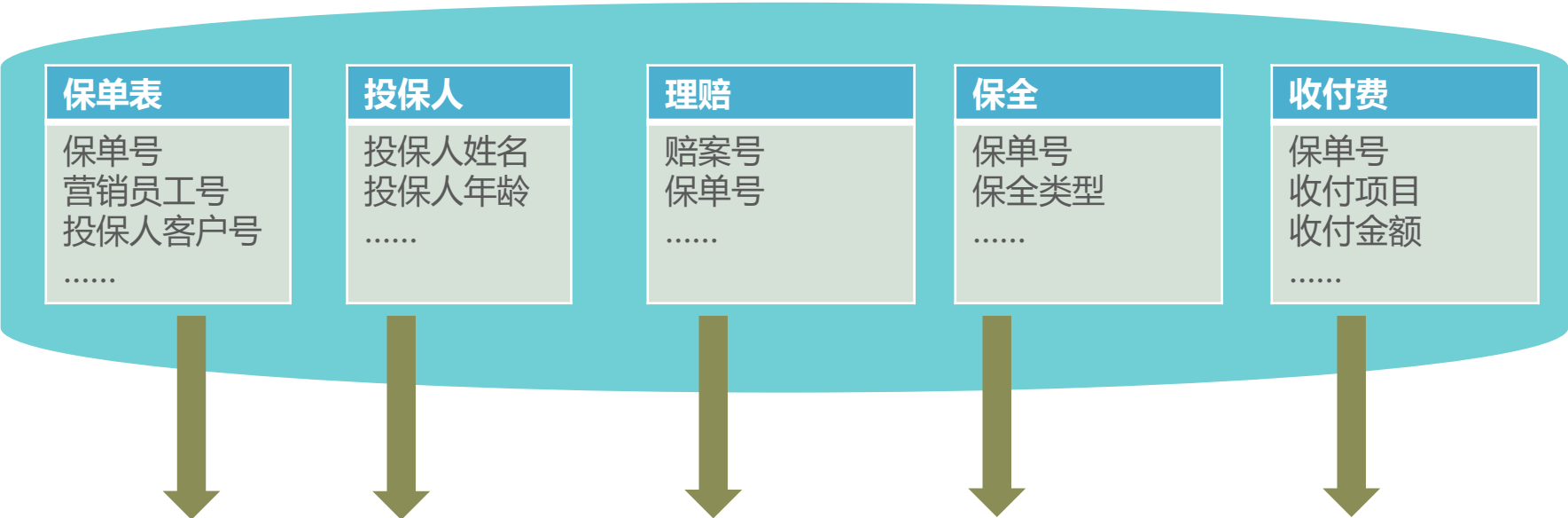
全国各系统分省单独部署

系统X
系统Y
.....

深圳 (440000)

One Row for One Contract

数据按行聚集



Hbase大宽表

保单号 row_key	保单 JSON	投保人 JSON	理赔01 JSON	理赔02 JSON	...	保全01 JSON	保全02 JSON	...	收付费01 JSON	收付费01 JSON
----------------	------------	-------------	--------------	--------------	-----	--------------	--------------	-----	---------------	---------------	-------

提前ONZ，
增加效率

新增数据仅
需添加新列

One Column for One Original Row

一个程序即可计算全部保单指标

原始表

收付费表			
保单号	收付流水号	收付项目	收付金额
001	11	A	100
001	12	B	200
001	13	C	-100

“11” 是收付费表主键值

行内容以JSON的方式进行存储

保单号 row_key	收付费11	收付费12	收付费13
001	{“流水号”: “11” ,”收付项目”: “A”,”收付金额”: “100”}	{“流水号”: “12” ,”收付项目”: “B”,”收付金额”: “200”}	{“流水号”: “13” ,”收付项目”: “C”,”收付金额”: “200”}

宽表

彻底解决更新难、更新慢的问题。

新增记录仅需要在HBase中追加新列即可。

One Json for All Content

原系统表

保单号	字段1	字段2	字段3	字段4	字段5
11	A1	A2	A3	A4	A5
12	B1	B2	B3	B4	B5
13	C1	C2	C3	C4	C5



金
字
塔
大
数
据
平
台

11 column=ci:保单信息, value={保单号:11,字段1:A1,字段2:A2,字段3:A3,字段4:A4}

12 column=ci:保单信息, value={保单号:12,字段1:B1,字段2:B2,字段3:B3,字段4:B4}

13 column=ci:保单信息, value={保单号:13,字段1:C1,字段2:C2,字段3:C3,字段4:C4}

保留细节
谨防丢失

目录

- 1 公司及平台简介
- 2 **HBase**为企业级平台提供的解决方案
- 3 企业级业务应用支持

客户资源管理

客户画像： 支持数亿客户上千标签的高效刷新，精准刻画老客户

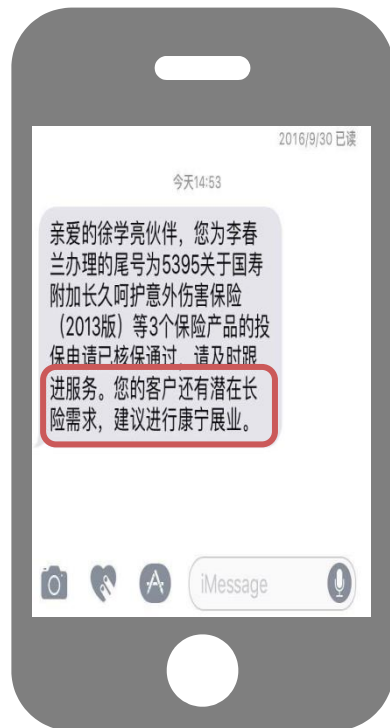


精准营销

购买能力分析

分析已有客户的再次购买能力，
并推送给前台

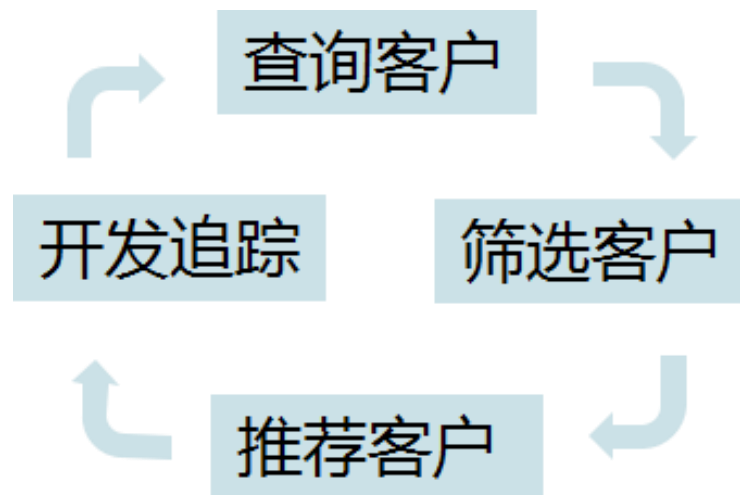
计算客户的购买能力并将结果
推送给营销
员



提高
展业
成功率

客户开发平台

实现对已有客户的分类筛选、定向推送和追踪



保障分析

保障分析：分析客户已有保障，试算客户应有保障

产品介绍

对公司200多总颁险种进行大数据建模分析，从身故、伤残、重疾、医疗、养老5个角度分析客户保障情况。

应用于手机端APP



发现客户保障缺失
提供销售契机



大数据智能核保

风险保额预警：实现以客户为中心的累计风险保额计算

产品介绍

利用大数据平台整合客户、保单数据，进行以客户为中心的风险保额计算，为公司核保系统提供强有力的数据支撑。

两百多个保单粒度风险保额

三百多个客户粒度风险保额

长险寿险净风险保额
长险重疾险净风险保额
长险护理险净风险保额
长险客运交通工具意外伤残净风险保额
长险航空意外伤残净风险保额
长险重大自然灾害意外伤残净风险保额

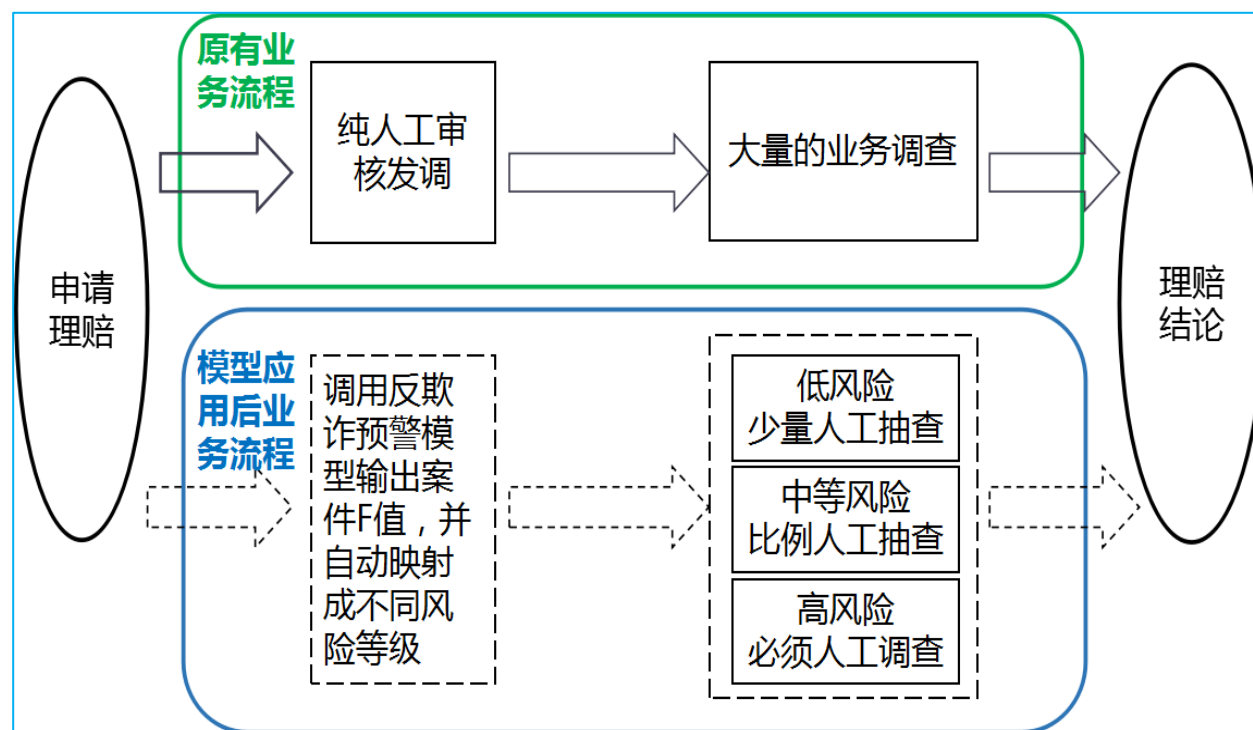
.....

整合多系统全国数据 **提升**核保智能化作业

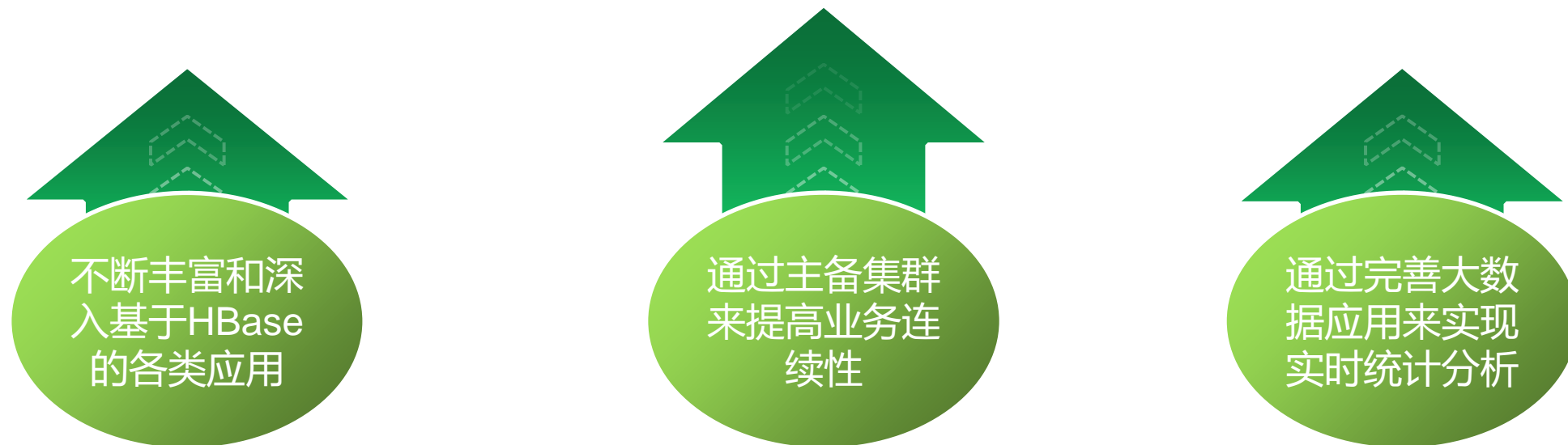


智能理赔反欺诈

智能理赔反欺诈：实现毫秒级反馈欺诈风险评估结果



下一步工作





敬请指正