# 旅游住宿分析平台说明

## 背景

随着网络和信息技术的不断普及，人类产生的数据量正在呈指数级增长，而云计算的诞生，更是直接把我们送进了大数据时代。“大数据”作为时下最时髦的词汇，开始向各行业渗透辐射，颠覆着很多特别是传统行业的管理和运营思维。在这一大背景下，大数据也触动着酒店行业管理者的神经，搅动着酒店行业管理者的思维；大数据在酒店行业释放出的巨大价值吸引着诸多酒店行业人士的兴趣和关注。探讨和学习如何借助大数据为酒店行业经营管理服务也是当今该行业管理者面临的挑战。

**一、大数据有助于精确酒店行业市场定位**

成功的品牌离不开精准的市场定位，可以这样说，一个成功的市场定位，能够使一个企业的品牌加倍快速成长，而基于大数据的市场数据分析和调研是企业进行品牌定位的第一步。酒店行业企业要想在无硝烟的市场中分得一杯羹，需要架构大数据战略，拓宽酒店行业调研数据的广度和深度，从大数据中了解酒店行业市场构成、细分市场特征、消费者需求和竞争者状况等众多因素，在科学系统的信息数据收集、管理、分析的基础上，提出更好的解决问题的方案和建议，保证企业品牌市场定位独具个性化，提高企业品牌市场定位的行业接受度。

企业想进入或开拓某一区域酒店行业市场，首先要进行项目评估和可行性分析，只有通过项目评估和可行性分析才能最终决定是否适合进入或者开拓这块市场。如果适合，那么这个区域人口是多少？消费水平怎么样？客户的消费习惯是什么？市场对产品的认知度怎么样？当前的市场供需情况怎么样？公众的消费喜好是什么等等，这些问题背后包含的海量信息构成了酒店行业市场调研的大数据，对这些大数据的分析就是我们的市场定位过程。

企业开拓新市场，需要动用巨大的人力、物力和精力，如果市场定位不精准或者出现偏差，其给投资商和企业自身带来后期损失是巨大甚至有时是毁灭性的，由此看出市场定位对酒店行业市场开拓的重要性。只有定位准确乃至精确，企业才能构建出满足市场需求地产品，使自己在竞争中立于不败之地。但是，要想做到这一点，就必须有足够量的信息数据来供酒店行业研究人员分析和判断。在传统情况下，分析数据的收集主要来自于统计年鉴、行业管理部门数据、相关行业报告、行业专家意见及属地市场调查等，这些数据多存在样本量不足，时间滞后和准确度低等缺陷，研究人员能够获得的信息量非常有限，使准确的市场定位存在着数据瓶颈。随着大数据时代的来临，借助数据挖掘和信息采集技术不仅能给研究人员提供足够的样本量和数据信息，还能够建立基于大数据数学模型对未来市场进行预测。当然，依靠传统的人工数据收集和统计显然难以满足大数据环境下的数据需求，这就需要依靠相关数据公司(如深圳乐思软件)自动化数据采集工具的帮助。

**二、大数据成为酒店行业市场营销的利器**

今天，从搜索引擎、社交网络的普及到人手一机的智能移动设备，互联网上的信息总量正以极快的速度不断暴涨。每天在Facebook、Twitter、微博、微信、论坛、新闻评论、电商平台上分享各种文本、照片、视频、音频、数据等信息高达的几百亿甚至几千亿条，这些信息涵盖着、商家信息、个人信息、行业资讯、产品使用体验、商品浏览记录、商品成交记录、产品价格动态等等海量信息。这些数据通过聚类可以形成酒店行业大数据，其背后隐藏的是酒店行业的市场需求、竞争情报，闪现着巨大的财富价值。

在酒店行业市场营销工作中，无论是产品、渠道、价格还是顾客，可以说每一项工作都与大数据的采集和分析息息相关，而以下两个方面又是酒店行业市场营销工作中的重中之重。一是通过获取数据并加以统计分析来充分了解市场信息，掌握竞争者的商情和动态，知晓产品在竞争群中所处的市场地位，来达到“知彼知己，百战不殆”的目的；二是企业通过积累和挖掘酒店行业消费者档案数据，有助于分析顾客的消费行为和价值趣向，便于更好地为消费者服务和发展忠诚顾客。

以酒店行业在对顾客的消费行为和趣向分析方面为例，如果企业平时善于积累、收集和整理消费者的消费行为方面的信息数据，如：消费者购买产品的花费、选择的产品渠道、偏好产品的类型、产品使用周期、购买产品的目的、消费者家庭背景、工作和生活环境、个人消费观和价值观等。如果企业收集到了这些数据，建立消费者大数据库，便可通过统计和分析来掌握消费者的消费行为、兴趣偏好和产品的市场口碑现状，再根据这些总结出来的行为、兴趣爱好和产品口碑现状制定有针对性的营销方案和营销战略，投消费者所好，那么其带来的营销效应是可想而知的。因此，可以说大数据中蕴含着出奇制胜的力量，如果企业管理者善于在市场营销加以运用，将成为酒店行业市场竞争中立于不败之地的利器。

**三、大数据支撑酒店行业收益管理**

收益管理作为实现收益最大化的一门理论学科，近年来受到酒店行业人士的普遍关注和推广运用。收益管理意在把合适的产品或服务，在合适的时间，以合适的价格，通过合适的销售渠道，出售给合适的顾客，最终实现企业收益最大化目标。要达到收益管理的目标，需求预测、细分市场和敏感度分析是此项工作的三个重要环节，而这三个的环节推进的基础就是大数据。

需求预测是通过对建构的大数据统计与分析，采取科学的预测方法，通过建立数学模型，使企业管理者掌握和了解酒店行业潜在的市场需求，未来一段时间每个细分市场的产品销售量和产品价格走势等，从而使企业能够通过价格的杠杆来调节市场的供需平衡，并针对不同的细分市场来实行动态定价和差别定价。需求预测的好处在于可提高企业管理者对酒店行业市场判断的前瞻性，并在不同的市场波动周期以合适的产品和价格投放市场，获得潜在的收益。细分市场为企业预测销售量和实行差别定价提供了条件，其科学性体现在通过酒店行业市场需求预测来制定和更新价格，最大化各个细分市场的收益。敏感度分析是通过需求价格弹性分析技术，对不同细分市场的价格进行优化，最大限度地挖掘市场潜在的收入。

大数据时代的来临，为企业收益管理工作的开展提供了更加广阔的空间。需求预测、细分市场和敏感度分析对数据需求量很大，而传统的数据分析大多是采集的是企业自身的历史数据来进行预测和分析，容易忽视整个酒店行业信息数据，因此难免使预测结果存在偏差。企业在实施收益管理过程中如果能在自有数据的基础上，依靠一些自动化[信息采集](http://www.knowlesys.cn/product_webdataminer_index.html)软件来收集更多的酒店行业数据，了解更多的酒店行业市场信息，这将会对制订准确的收益策略，盈得更高的收益起到推进作用。

**四、大数据创新酒店行业需求开发**

随着论坛、博客、微博、微信、电商平台、点评网等媒介在PC端和移动端的创新和发展，公众分享信息变得更加便捷自由，而公众分享信息的主动性促使了“网络评论”这一新型舆论形式的发展。微博、微信、点评网、评论版上成千上亿的网络评论形成了交互性大数据，其中蕴藏了巨大的酒店行业需求开发价值，值得企业管理者重视。

网络评论，最早源自于互联网论坛，是供网友闲暇之余相互交流的网络社交平台。在微博、微信、论坛、评论版等平台随处可见网友使用某款产品优点点评、缺点的吐槽、功能需求点评、质量好坏与否点评、外形美观度点评、款式样式点评等信息，这些都构成了产品需求大数据。同时，消费者对企业服务及产品简单表扬与评批演变得更加的客观真实，消费者的评价内容也更趋于专业化和理性化，发布的渠道也更加广泛。作为酒店行业企业，如果能对网上酒店行业的评论数据进行收集，建立网评大数据库，然后再利用分词、聚类、情感分析了解消费者的消费行为、价值趣向、评论中体现的新消费需求和企业产品质量问题，以此来改进和创新产品，量化产品价值，制订合理的价格及提高服务质量，从中获取更大的收益。

## **项目起源**

贯石科技是一家以云酒店管理系统为核心产品的软件互联网公司，公司的核心产品就是酒店管理平台，为中小型酒店、客栈、旅店等提供酒店管理服务，并且维护着所有客户的业务经营的数据中心。随着客户量的拓展，有每天都有庞大酒店经营业务数据，为了更好的为客户提供增值服务，公司决定开发大数据平台，为酒店提供数据咨询服务。

## **项目主题**

旅游住宿大数据平台对酒店行业数据的分析分为

质量分析----提升产品

用户画像----了解客户

收益分析----优化销售

舆情分析----品牌构建

竞争分析----企业定位

五大主题，数据来源总体分为两类：

1. 云酒店平台
2. 网络爬虫

## **需求**

### 质量分析（产品）主题

客房分析

1. 计算每个月，每个酒店，每个客房的入住率及收入

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 日期（月） | 酒店id | 房间ID | 预定间夜数 | 入住间夜数 | 入住率 | 价格 | 总金额 | 实际总金额 |  |
|  |  |  | 在booked\_room表中 | checkin\_room | 入住间夜数/本月总天数 |  | 定价\*间夜数 | 实际收费金额 |  |
|  |  |  |  |  |  |  |  |  |  |
| 日期（月） | 酒店id | 房型 | 房间数 | 预定间夜数 | 入住间夜数 | 入住率=入住间夜数/(总房间数\*当月天数) | 价格 | 总金额 | 实际总金额 |
|  |  |  |  |  |  |  |  |  |  |
| 日期（月） | 酒店id | 价格区间（100以下的房间，100~200的房间，200~300的房间，300以上的房间） | 房间数 | 预定间夜数 | 入住间夜数 | 入住率 | 价格 | 总金额 | 实际总金额 |
|  |  |  |  |  |  |  |  |  |  |
| 日期（月） | 酒店id | 房型数量 | 房间数 | 预定间夜数 | 入住间夜数 | 入住率 | 价格 | 总金额 | 实际总金额 |

评价得分分析

平台排名分析

市场平均分析

### **用户画像（客户）主题**



为企业的客户分析人群画像，深度了解自己客户的基本属性、消费属性、行为属性等:

客户画像分析的客户属性能知道客户的消费能力层次、常去旅游景区、公寓还是酒店，是从外地来的还是本地的。从而对自己的客户做到了如指掌，做到数据驱动运营与服务。

基于自己的客户画像分析，进行精准的营销，帮助企业找到真正需要产品的客户:

分析自己的客户常在某景区活跃，从而可以对该精准客户制定媒介投放策略。

分析某商圈的客流属性，帮助企业找到最佳开店位置

分析某地周边客户人群画像、周边商圈的客户画像能够知道该地区客户的消费能力、消费喜好、受众群体是年轻人还是商务精英，从而知道该地是否适合自己的品牌入住及哪个层次的品牌入住。

帮客户分析其竞争对手的客群画像:

分析系统可以获取竞品的客流，每日有多少流入多少流出，是男顾客多还是女顾客多，消费层次如何。从而与自己的品牌的一个数据对比，寻找差距点，提升品牌影响力与服务质量。

大数据分析出消费者的喜好，根据消费者的喜好研发新产品，满足客户需求打造爆款:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 日期（月） | 酒店id | 区域（省） | 入住用户数量 | 入住总间夜数 | 人均间夜数 | 总消费金额 | 人均消费金额 | 人均间夜消费金额 |  |
|  |  |  |  |  |  |  |  |  |  |
| 日期（月） | 区域（省） | 入住用户数量 | 入住总间夜数 | 人均间夜数 | 总消费金额 | 人均消费金额 | 人均间夜消费金额 |  |  |
|  |  |  |  |  |  |  |  |  |  |
| 日期（月） | 客户id（手机） | 订单数 | 订单总金额 | 总间夜数 | 总消费金额 | 平均间夜消费金额 | 入住酒店数 |  |  |

客户画像的分析得出很多年轻人爱好玩游戏，那么我们是否可以在娱乐中心新增最新的潮流游戏。如：VR游戏等提升客户的满意度。分析的场景和维度很多，以此类推。

用户区域分析

### **收益分析（渠道）主题**

入住率分析

订单量分析

年、季度、月、同比环比分析

蓝色：同比 红色：环比

客户渠道来源分析

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 日期 | 酒店id | 地域（省市） | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |
| 日期 | 地域（省市） | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |  |
| 日期 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |  |  |
| 日期 | 渠道 | 酒店id | 地域（省市） | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |
| 日期 | 渠道 | 地域（省市） | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |  |
| 日期 | 渠道 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 | 房间数 | 总间夜数 | 预定间夜数 | 入住间夜数 | 预订率 | 入住率 | 月收入 | 房间均收入 |  |  |

### **舆情（品牌）主题**

事情舆情监控

覆盖全网媒体资讯平台，含主流微博，微信公众号，保证监测信息的全面

对信息进行实时采集，最短采集频率1次/分钟

利用先进的人工智能技术，实现信息内容的精准挖掘，针对不同的行业领域特点进行情感正负面判断，提高价值。

从不同维度对舆情信息的分布、趋势进行分析并从中提取热点事件、重要舆情信息。

通过邮件将资讯、点评、指数信息在第一时间传递，让企业不遗漏任何重要 信息。

可自定义预警的规则，通过个性化的设置系统自动预警，更加人性化的预警





爬虫爬取数据内容：

1. 对酒店的实时总评分，各维度的评分，进行监控和更新，对每一天的评论数进行统计，每天总共评论数，好评数，中评数，差评数，环比上升或下降百分比

维度：时间（天），酒店，平台

平均总评分=每条评论的综合评分和/评论条数

各维度的平均评分=各维度评分和/评论条数

评论总记录数

好评数

中频数

差评数

环比上升下降百分比

（评论综合评分总和，评论各维度评分总和）

1. 差评预警，差评来源，差评各维度得分，差评评价内容，差评链接地址

点评信息采集：

1. 点评平台：携程，艺龙，去哪儿，途牛，大众点评网，美团。。。。。
2. 点评人的账号信息：平台id：202132689 \_\_ 去哪儿耍呀

点评师级别，点评次数，遍布城市数

1. 标题
2. 评分：总评分 各维度评分：卫生、服务、隔音、安全
3. 入住时间信息 入住类型 是否有经平台下订单入住
4. 评价内容

### **竞争分析主题**

酒店标签分析

同等竞争对手收益比对

不同标签类型数据分析

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 日期 | 酒店id | 地区 | 房间数 | 房型数 | 最高房价 | 最低房价 | 房间均价 | 全月间夜金额 | 酒店标签：中高端，根据口径计算中高低端 |
| 日期 | 酒店标签 | 地区 | 酒店数 | 房间数 | 房型数 | 最高房价 | 最低房价 | 房间均价 | 全月间夜金额 |
| 日期 | 地区 | 酒店数 | 房间数 | 房型数 | 最高房价 | 最低房价 | 房间均价 | 全月间夜金额 |  |
|  |  |  |  |  |  |  |  |  |  |
| 日期 | 酒店id | 标签 | 预订率 | 入住率 | 同地区预订率 | 同地区入住率 | 总预订率 | 总刚入住率 |  |

## **大数据平台技术架构**

### **数据来源**

### **数据抽取（ETL）**

关系型数据库采集方式：

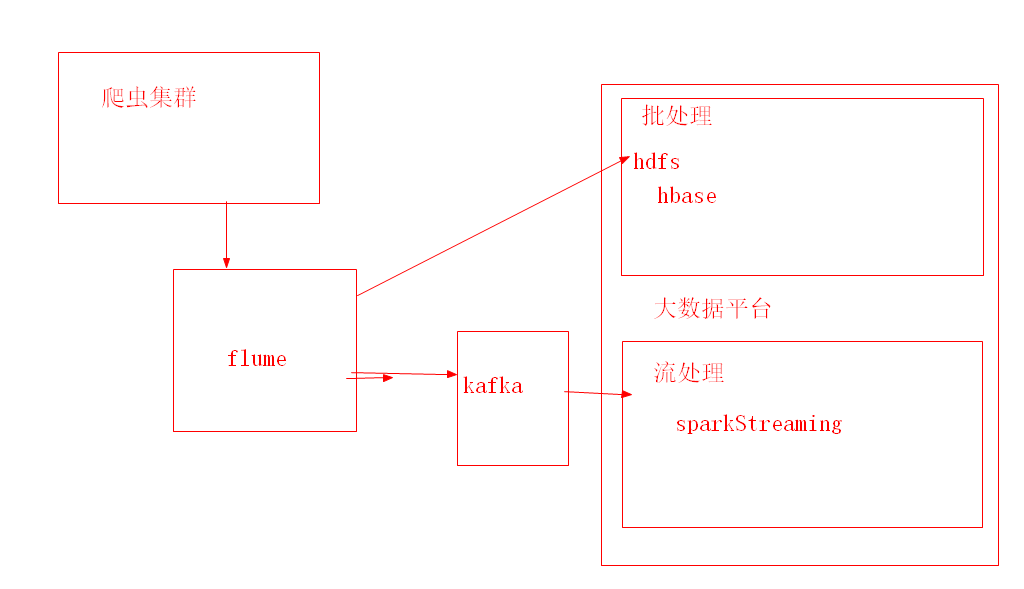
1.sqoop直连业务库，把数据导入进来

2.业务平台导出分析平台需要的数据，定期（每日）用ftp方式传递到大数据平台，传递结束以后启动mr或者hive脚本进行清洗入库

3.spark或sparksql直接从业务库抽取数据进入大数据平台

4.其他etl工具kettle

网络爬虫采集数据入库：



### **数据仓库**

hive作为数仓的基础

对数据的处理可以选择：

Spark sql

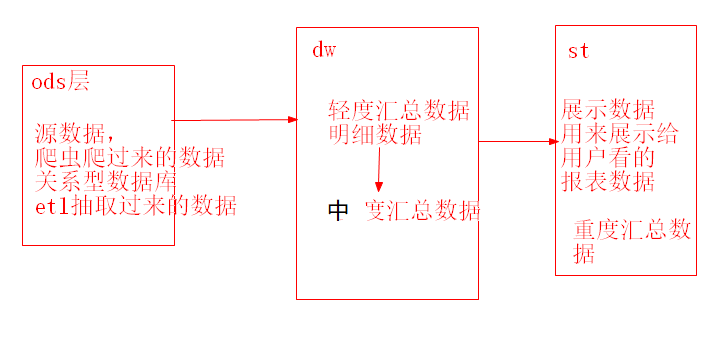
Hive脚本

Impala

Spark

Pig

MR



调度系统

Oozie

Azkaban

自己研发的调度系统

### **数据展现**

web项目

SSM

Jquery + bootstrap + ECharts + mysql(postgresql)

## **系统架构**

### **数据存储**

Hdfs --数据分析过程中的所有数据保存在hdfs上，对接hive

Hbase--明细数据，历史数据保存在hbase对接查询应用

Mysql(postgresql)--报表数据(重度汇总后的数据)，适合图表展现的数据，监控数据，实时数据

### **数据处理**

spark 或 spark sql

### **数据查询**

web service(ssm + hbase) 提供接口 供用户客户端调用

## **硬件要求**

10个节点：（主机节点内存256G，硬盘做RAID）

单节点服务器

处理器：2cpu 32核 2.6GHZ 20M缓存

内存：128G （DDR3 8G\*16）

硬盘：1T\*4 SATA 7200转

## **软件要求**

Linux centos 6.5

-- 订单数据表

tb\_booking

-- 客户表

tb\_customer

-- 预定房间表

tb\_bookedroom

-- 入住

tb\_check\_in

-- 入住房间

checkin\_room

1. 数据导入，批量导入，还要做定期导入，给定时间参数，把特定时间参数的数据导入到hive，并其支持程序的重复运行。
2. Hive中的表建分区，按照时间，dateday年月日作为分区依据
3. 维度数据导入，一次性批量导入

房间（tb\_roominfo）---》房型（tb\_roomtype）---》酒店（tb\_hotel）---》公司（tb\_company）

4.时间维度表