# 游戏运营分析项目

# 课程目标

## 熟练使用Spark解决实现生产问题

## 熟悉游戏领域大数据分析相关概念

## 掌握游戏领域大数据分析的步骤和方法

# 项目背景

## 概述

目前大数据已经在各个领域得到广泛应用，在游戏领域玩家的各种行为会产生大量的数据，这些数据可以帮助游戏运维者了解玩家的行为和需求，通过玩家反馈的信息可以不断地矫正游戏中出现的问题，从而让游戏健康稳定的可持续运营。

## 游戏利用数据分析的特点

在游戏行业，大数据分析可能不会直接带来电商网站那样真金白银的收入，但其价值同样会体现在精准营销、客户体验优化等多个层面。全面的大数据分析能够有效提升玩家的留存率和转化指标，并且为游戏产品的研发提供指引。而个性化的精准营销同样与大数据分析紧密相关，像是针对不同性别、不同年龄、不同地域人群的广告精准投放，背后都要依靠基于360°用户视图的玩家特征分析。

一、性质：利用数据挖掘将产品中现有的数据转化为知识，帮助产品做出明智改进决策。

二、价值：用准确的数据结果来支持决策。

三、定位：在数据的基础上，让数据产生价值。

四、目标：发现问题、分析问题、提供建议、预测产品（深入数据分析并运用于业务）

# 游戏数据分析指标

## 新增玩家

* **新增玩家（DNU）** 当日新增加的玩家帐户数。
* **新玩家地区** 新增玩家帐号的地区分布情况。
* **新玩家渠道** 新增玩家帐号的渠道分布情况，每个玩家始终归入首次设备激活时的渠道。
* **新玩家性别** 新增玩家帐号的玩家性别情况，根据您的游戏中传入的性别确定玩家的性别。
* **新玩家年龄** 新增玩家帐号的年龄分布情况，根据您传入的年龄确定玩家的年龄段。
* **玩家帐户类型** 新增玩家帐号的帐户类型分布情况。

## 活跃玩家

* **DAU** 当日有开启过游戏的玩家数，其中当日新增玩家带来的活跃以区别的颜色显示。
* **WAU** 当日的最近一周（含当日的倒推7日）活跃玩家，将进行过游戏的玩家按照帐户进行排重。
* **MAU** 当日的最近一月（含当日的倒退30日）活跃玩家，将进行过游戏的玩家按照帐户进行排重。
* **DAU/MAU** DAU/MAU可体现玩家的总体粘度，衡量这30天内每日活跃玩家的交叉重合情况。此比例越趋近于1，代表月活跃中有更多玩家多日活跃；比例越趋近0 ，则代表大量玩家只在一日中活跃。当比例小于0.1时，游戏的自传播性将很差。DAU/MAU×30也用于大致衡量玩家平均每月活跃天数。
* **AVG|MDAVG** 指所选时期对应指标的平均值，MD则代表时期内指标的中位数。
* **活跃玩家游戏天数** 所选时期的活跃玩家的活跃日期减新增日期的日数差（即其已进行游戏的天数）的分布情况。
* **活跃玩家等级** 所选时期每日活跃玩家的等级分布情况，由每日数据组成的时期趋势可观测等级的总体走势。
* **活跃玩家地区** 所选时期排重活跃玩家的地区分布情况。
* **活跃玩家渠道** 所选时期排重活跃玩家的渠道分布情况，渠道按照其最初新增时的渠道确定。
* **活跃玩家性别** 所选时期排重活跃玩家的性别情况。
* **活跃玩家年龄** 所选时期排重活跃玩家的年龄分布情况。

## 玩家留存

* **次日留存率** 某日新增的玩家中，在下一日中还进行了游戏的玩家的比例。
* **7日留存率** 某日新增的玩家中，在该日后的7日中还进行了游戏的玩家比例。
* **30日留存率** 某日新增的玩家中，在该日后的第30日中还进行了游戏的玩家比例。
* **留存玩家分析** 圈定时期内符合次日/7日/30日留存条件的新玩家，根据其新增日的使用行为（等级、游戏次数、是否付费等）来进行分布，可有效的分析新增日行为与留存的相关性。

## 玩家流失

* **每日流失数** 到某日为止，玩家连续7/14/30天没有再玩过游戏，将被视为一个流失。
* **每日流失率** 每日的7日（14日/30日）不登陆流失玩家数占该日7日（14日/30日）前活跃玩家数的比例。
* **每日回流** 流失玩家在某日再次进行游戏，将视为当日的一个回流。回流玩家可用于沉默玩家唤醒活动后的效果评估。
* **流失用户分析圈** 定时期内符合7/14/30日流失条件的玩家，按照其流失前等级、已游戏天数、付费情况等条件来进行分布，可有效分析流失群体的特征，找出原因优化游戏。

## 游戏习惯

### 平均游戏时长与次数

* **每日** 当日玩家每次进行游戏的平均时长（当日总时长/游戏次数）和平均每玩家的游戏次数（游戏次数/日玩家数）。
* **自然周** 某自然周内，玩家每次进行游戏的平均时长（周时长/周游戏次数）和平均每玩家一周内的游戏次数（游戏次数/周排重玩家数）。
* **自然月** 某自然月内，玩家每次进行游戏的平均时长（月时长/月游戏次数）和平均每玩家一月内的游戏次数（游戏次数/月排重玩家数）。

### 游戏频次

* **游戏频次** 指玩家在一定期间中进行游戏的频繁度，将按照频度的不同将玩家分入不同的分布区间。报表中按照您所选日期自动扩充完整自然周/月来计算周和月的频次，跨多个周/月的数据查询会先计算单个周/月的分布数据，再组合各个区间的数据后集中进行展示。
* **月游戏天** 数玩家在一个自然月中活跃日数的分布情况。
* **周游戏天** 数玩家在一个自然周中活跃日数的分布情况。
* **周游戏次** 数玩家在一个自然周中游戏次数的分布情况。
* **日游戏次数** 玩家在一日中游戏次数的分布情况。

### 游戏时长

* **游戏时长** 指玩家在游戏内的游戏时长，按照不同时长将玩家分入不同的分布区间。对周使用时长将按照您所选日期自动扩充完整自然周来计算单用户时长，跨多周的查询会先计算单周的分布数据，再组合各个区间的数据后集中进行展示。
* **周游戏时长** 玩家在一个自然周中玩游戏总时长的分布情况。
* **单日游戏时长** 玩家在一日中玩游戏总时长的分布情况。
* **单次游戏时长** 玩家在一日中每次玩游戏的时长分布情况。
* **游戏间隔** 时期内，玩家每次玩游戏与上次游戏的时间间隔分布情况。

## 收入数据

* **收入金额** 所选时期的充值总金额根据玩家充值时的等级进行分布。
* **充值人次** 所选时期每次充值根据玩家充值时的等级进行分布。
* **各地区收入** 所选期间，各个地区的玩家充值金额的分布情况。
* **各渠道收入** 所选期间，各个渠道来源的玩家充值金额的分布情况。
* **分性别收入** 不同性别玩家的充值总金额情况。
* **各年龄段收入** 不同年龄段玩家的充值总金额情况。

# 数据处理

## 数据采集

### 离线数据采集

通过在游戏的服务端安装Logstash将数据定时采集到Elasticsearch中

### 实时数据采集

通过在游戏的服务端安装Logstash将数据定时采集到Kafka中

## 数据分析