

# 大数据技术在产品测试中的实践分享

---

马航 广联达

# 个人简介

马航

2007年加入广联达，曾负责计价产品测试，现在工程信息产品部负责测试工作。

2007年，本科毕业于北京航空航天大学，软件工程专业。

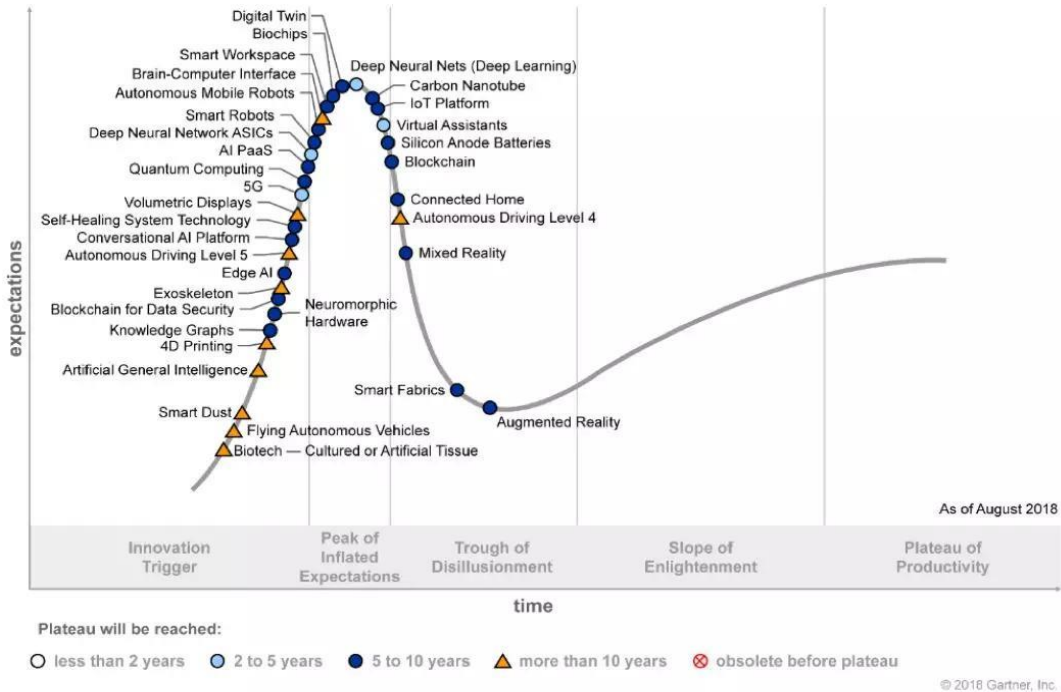
# 目录

1. 背景介绍
2. 大数据测试的思考
3. 几个例子
4. 总结和展望



# 大数据技术与建筑行业的结合

- 1. Gartner2018发布的技术成熟度曲线大数据不再是“新兴”技术，到了真正使用大数据解决问题的时代
- 2. 建筑行业数字化水平有很大提升空间



## 建筑业是数字化水平最差的行业之一



麦肯锡全球机构  
行业数字化指数:

数字化程度相对低 数字化程度相对高

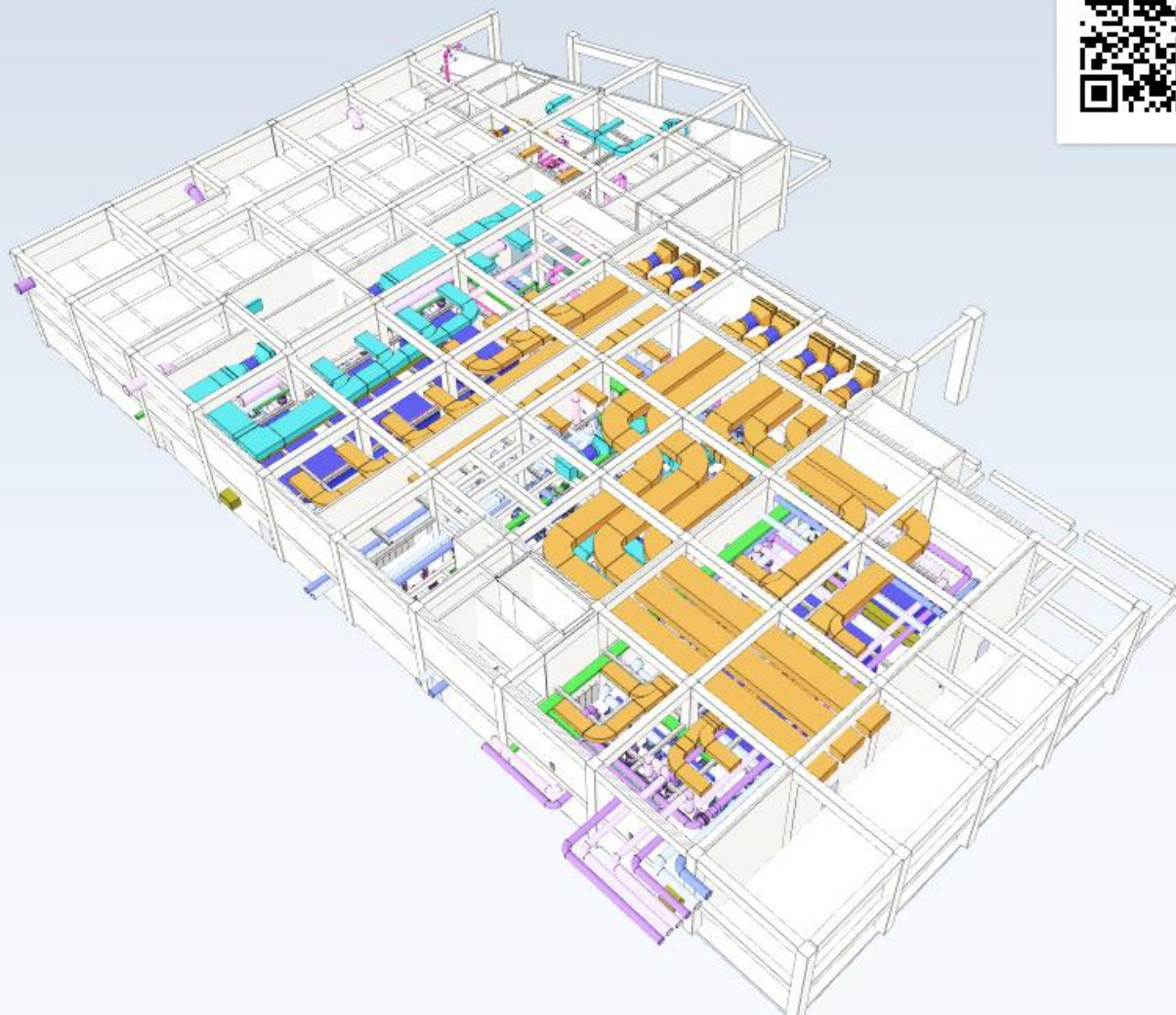
(来自2015或最新的数据)

# 广联达科技股份有限公司 (股票代码: 002410)

Glodon 广联达

广联达是一家立足建筑产业，围绕建设工程项目的全生命周期，以提供建设工程领域专业应用为核心支撑，以产业大数据、产业征信、产业金融等为增值服务的平台服务商。





扫一扫，在手机上查看



<http://bimface.com/sample>

扫一扫，在手机上查看





# 建筑行业产生的数据

BIM(Building Information Modeling)

数字建筑，智能硬件，智能家装，VR



建筑实体-信息模型-数字孪生

招投标数据

工程造价指标信息数据

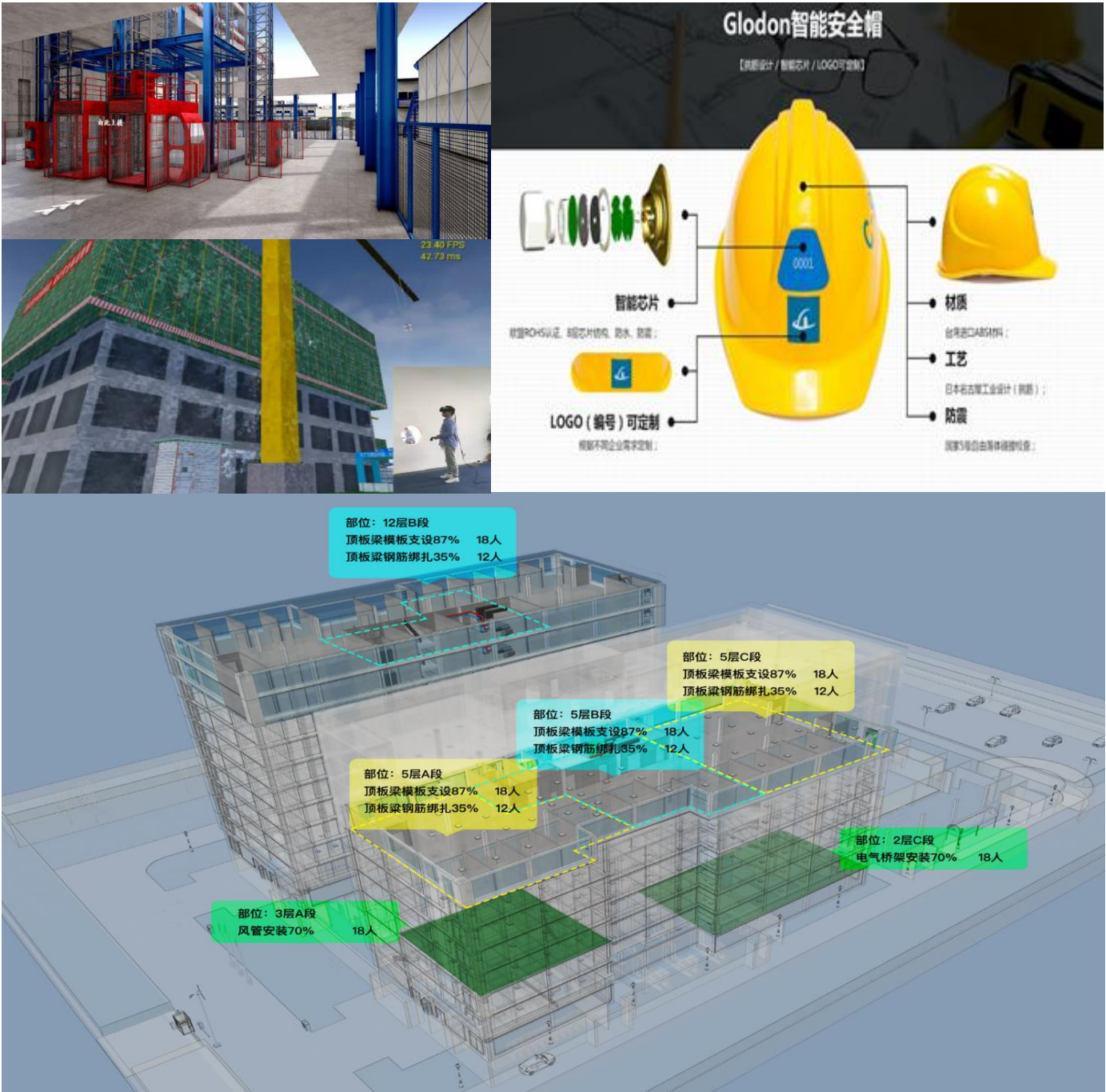
施工过程数据

建筑材料信息数据

从业者数据

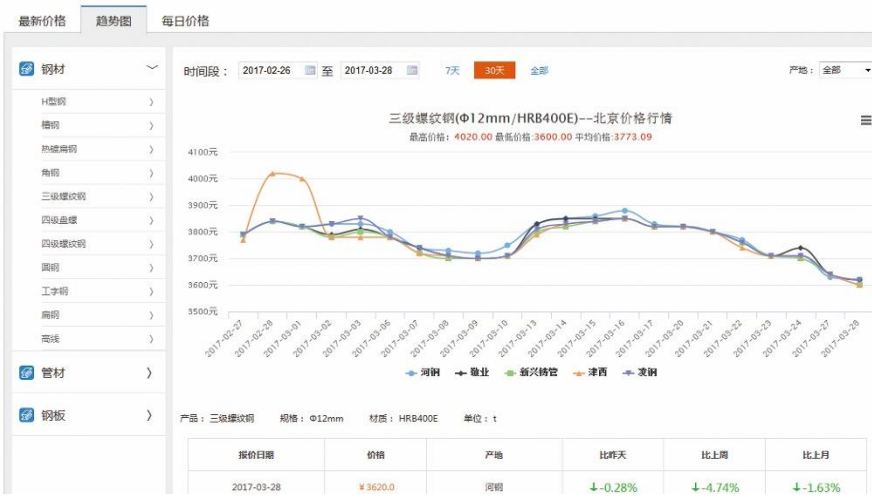
行业标准行业规范数据

.....



# 建筑成本大数据

SPU商品  
材料报价  
标讯  
注册账号  
信息价日均载价



展开全部科目 显示含量指标 显示单价指标 显示计算基数 查看已删除项目

序号	项目	工程内容	广联达大数据		备注 双主材需填	操作科目	
			造价指标参考范围 (元/m²)	单方造价 (元/m²) 双主材数据调整			
	主体工程		4318.45~4816.01	4551.48	91029.64		
▶ 1	基础工程	钻孔灌注桩、地下连续墙、土方开挖及外运、筏板...	982.05~1114.49	1012.46	20249.22	×	+
▶ 2	地下结构	传统现浇钢筋混凝土、钢结构、加气混凝土砌块...	1395.52~1557.95	1471.29	8827.74	×	+
▶ 3	地下建筑与装修	防水、保温、粗粉刷...	484.77~539.79	517.07	3102.42	×	+
▶ 4	地下室内装饰	电梯厅、走廊、卫生间、楼梯间、车库等精装修...	216.70~271.37	241.45	1448.70	×	+
▶ 5	上部结构-裙房	现浇钢筋混凝土承重墙、柱及梁框架及叠合板、钢...	2008.33~2514.94	2237.69	3132.77	×	+



广材网 www.gcliao.com 北京 地区切换

建筑材料价格网

市场 信息价 供应商

询价 搜索 发布人工询价

地区： 全国 北京

同材料不同品牌对比

品牌	单位	含税市场价	品牌	单位	含税市场价	含税单价	税率	报价时间	供应商	纳税人识别号
新斯得	台	收 1184.65	1247.0	17%						
新斯得	台	收 1184.65								
普雷	台	收 1196.25	1329.0	17%						

共2个品牌3条报价

同材料不同供应商对比

品牌	单位	含税市场价	含税单价	税率	报价时间	供应商	纳税人识别号
新斯得	台	收 1184.65	1247.0	17%	2017/02/25	【北京市】上海沪盛建设工程有限公司	【一星】
普雷	台	收 1196.25	1329.0	17%	2016/11/10	【北京市】上海冠志阀门机械有限公司	【一星】
中	台	收 1061.2	5306.0	17%	2016/11/07	【北京市】北京上瑞中阀门销售中心	【一星】

共3个品牌3条报价，查看全部详情

不同材料，详细参数和价格对比

品牌	单位	含税市场价	含税单价	税率	报价时间	供应商	纳税人识别号
新斯得	台	收 1184.65	1247.0	17%	2017/02/25	【北京市】上海沪盛建设工程有限公司	【一星】
普雷	台	收 1196.25	1329.0	17%	2016/11/10	【北京市】上海冠志阀门机械有限公司	【一星】
中	台	收 1061.2	5306.0	17%	2016/11/07	【北京市】北京上瑞中阀门销售中心	【一星】



# 互联网测试和垂直行业测试对比

对行业背景和业务知识的要求

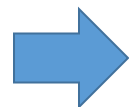
测试范围和测试外延的不同

# 建筑行业大数据带给测试的挑战

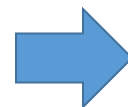
1. 数据分散在各上下游企业里，没有真正实现开放、关联、融合、共享
2. 大量的实时数据，包括设计参数、造价信息、工程信息等基础性数据和现场采集和收集的数据
3. 各个产品之间数据关系复杂，测试难度大
4. 大数据并不是精确的计算，没有简单的判定标准，没有可视化的界面呈现，无法直观判断测试结果

# 大数据处理流程和数据价值闭环

1. 业务问题理解
2. 数据采集
3. 数据预处理
4. 数据存储与管理
5. 建模和数据分析
6. 数据可视化展示
7. 解决方案



1. 结合业务不断迭代
2. 问题分析
3. 持续优化



1. 大数据资源
2. 大数据技术
3. 商业应用



# 在思考大数据该怎么测试之前

1. 怎么才算测试的好？：要测什么？为什么这么测？准备怎么测？测试分析-设计-执行
2. 评价好坏的标准是什么？：覆盖度？测试深度？测试效率提升？没有泄露bug？
3. 如何让客户满意？
4. 如何让领导满意？
5. 如何让自己满意？KPI->OKR
6. 成为主管，怎么在时间，成本，质量之间找到平衡点？什么时间开始？什么时候停止？
7. 规划未来，业界有哪些优秀做法？能不能引入进来？适不适合我们的现状？性价比？
8. ....

更多时候，常常是寻找一种现实和理想的平衡，向自己内心深处寻求一种和解  
努力做一名正直、有底线、有想法、有能力、有眼光、有勇气的测试工程师

# 迭代优化

1. 先把事情完成，解决0到1的问题
2. 再想办法做深入，解决1到100的问题
3. 随着经验和能力的提升，把时间减下来
4. 把成本砍下去
5. ....

前提条件是测试活动本身的数据积累和数据度量

数据是客观的，分析数据可以再现过程，复盘过程可以定位问题，聚焦主要问题可以逼近理想的解决方案。

# 一点感悟

1. 单机软件，云，云+端
2. Web，app，小程序
3. Delphi，C++，Java，Ruby，Python，Go
4. 大数据，人工智能，机器学习，物联网
5. ....

技术驱动业务？业务驱动技术？

你的自动化测试到底有多自动化？

测试会一直存在吗？

英雄造势，智者顺势



# 几个例子

1. 以大数据技术的处理流程为主线，以每个阶段技术和产品的结合作为切入点。
2. 主要列举数据获取、数据清洗，数据分析与数据挖掘，几个典型场景下测试实践，以及过程优化改进。
3. 最后总结和梳理大数据测试方法，从质量度量、质量监控、质量标准维度做的思考和工作。

# 数据获取

数据的格式:

1. txt
2. Excel
3. Json
4. XML
5. 纸质表格
6. 电子版图片
7. 视频流数据
8. .....

- [illegible]

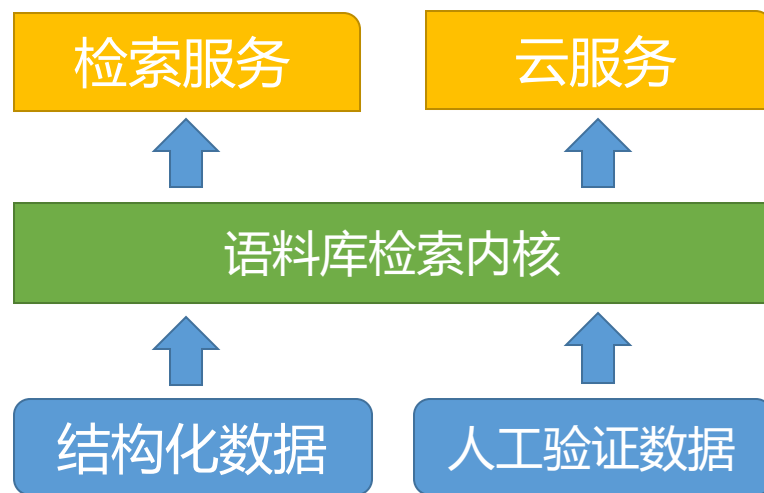




# 数据清洗和数据加工

建立语料库：

参考语言学科成熟的语料库知识体系，尝试搭建建筑行业招投标领域专业的语料库



标讯数据经过清洗和人工审核之后，要经过分类和提取字段值，为下一步处理做数据准备：

1. 通过数据分类，识别出招标标讯，中标标讯，采购标讯
2. 从分类后的数据中提取招标单位、招标金额、招标代理机构、中标单位、中标金额等字段

【新系统】【中标公示】龙广镇中心幼儿园设备采购项目中标公示				
【发布时间：2018/10/12】【阅读次数：82】【我要打印】【关闭】				
<hr/>				
<ol style="list-style-type: none"> <li>项目名称: 龙广镇中心幼儿园设备采购项目</li> <li>项目编号: GZACKYG02018-08-06-2</li> <li>信息序列号: 州公勘采【201809】-227号</li> <li>信息联系人: 朱晓星</li> <li>信息联系人电话: 0859-3918233</li> <li>信息用途: 采购技术要求及合同履行日期; 详见招标文件</li> <li>采购方式: 公开招标</li> <li>采购日期: 2018-10-12</li> <li>公告网址: 贵州省政府采购网</li> <li>开标时间: 2018-10-12</li> <li>评标地点: 黔西南州公共资源交易中心</li> <li>评标委员会成员名单:</li> <li>王学军、陈忠安、黄应强、王安全、赵久强</li> <li>开标时间: 2018-10-12</li> <li>中标(成交)信息</li> </ol>				
序号	中标供应商	中标标的名称	主要中标内容	中标金额(元)
1	贵州普路数字设备有限公司	贵州普路数字设备有限公司龙广镇中心幼儿园(新建)特色教室产品区)	中心幼儿园设备	1590749

```
Testing...
Test Loss: 0.23, Test Acc: 93.01%
Precision, Recall and F1-Score...
precision    recall  f1-score   support

 建设工程      0.93      0.98      0.96     1207
 土地招拍挂      0.99      0.97      0.98      231
 物资采购      0.60      0.50      0.61      165

avg / total          0.93      0.93      0.92     1603

Confusion Matrix...
[[1185  2 20]
 [  6 224  1]
 [  83  0 82]]
Time usage: 0:00:04
PyDev console: starting.
```

[illegible]

# 数据清洗和数据加工

如何把供应商报价单和线上已有SPU数据建立关联：

- 1. 报价单数据来源多样，格式也不统一
- 2. 供应商报价的材料，品种、规格、名称、分类，与已有材料库中的材料没有对应关系
- 3. 收集的数据去重、合并、归类、建立与库里材料的关联，本身工作量巨大，但是是绕不开的一步工作
- 4. 采用机器学习的方法自动分类，提高效率

ROCKWOOL  
专业防火岩棉保温系统  
ROCKWOOL FireSafe Insulation  
(Guangzhou) Co., Ltd.  
www.rockwool.com.cn

报价函

收信人单位: 中雨华联建筑公司  
收信人: -  
收信人电话: -  
工程項目: 海口美兰机场二期

我方参考号: Q20170118JX  
报价日期: 2017-01-18  
收信人传真: -

尊敬的客户:  
首先万分感谢阁下及贵公司的垂询, 我司就所需产品的报价和条款如下:

序号	产品描述	数量	单位	单价(元/m2)	备注
1	建筑用岩棉 TR-S60.FF				



广材网  
建筑材料网

报价函

收信人单位: 中雨华联建筑公司  
收信人: -  
收信人电话: -  
工程項目: 海口美兰机场二期

我方参考号: Q20170118JX  
报价日期: 2017-01-18  
收信人传真: -

尊敬的客户:  
首先万分感谢阁下及贵公司的垂询, 我司就所需产品的报价和条款如下:

序号	产品描述	数量	单位	单价(元/m2)	备注
1	建筑用岩棉 TR-S60.FF				



广材网  
建筑材料网

报价函

收信人单位: 中雨华联建筑公司  
收信人: -  
收信人电话: -  
工程項目: 海口美兰机场二期

我方参考号: Q20170118JX  
报价日期: 2017-01-18  
收信人传真: -

尊敬的客户:  
首先万分感谢阁下及贵公司的垂询, 我司就所需产品的报价和条款如下:

序号	产品描述	数量	单位	单价(元/m2)	备注
1	建筑用岩棉 TR-S60.FF				



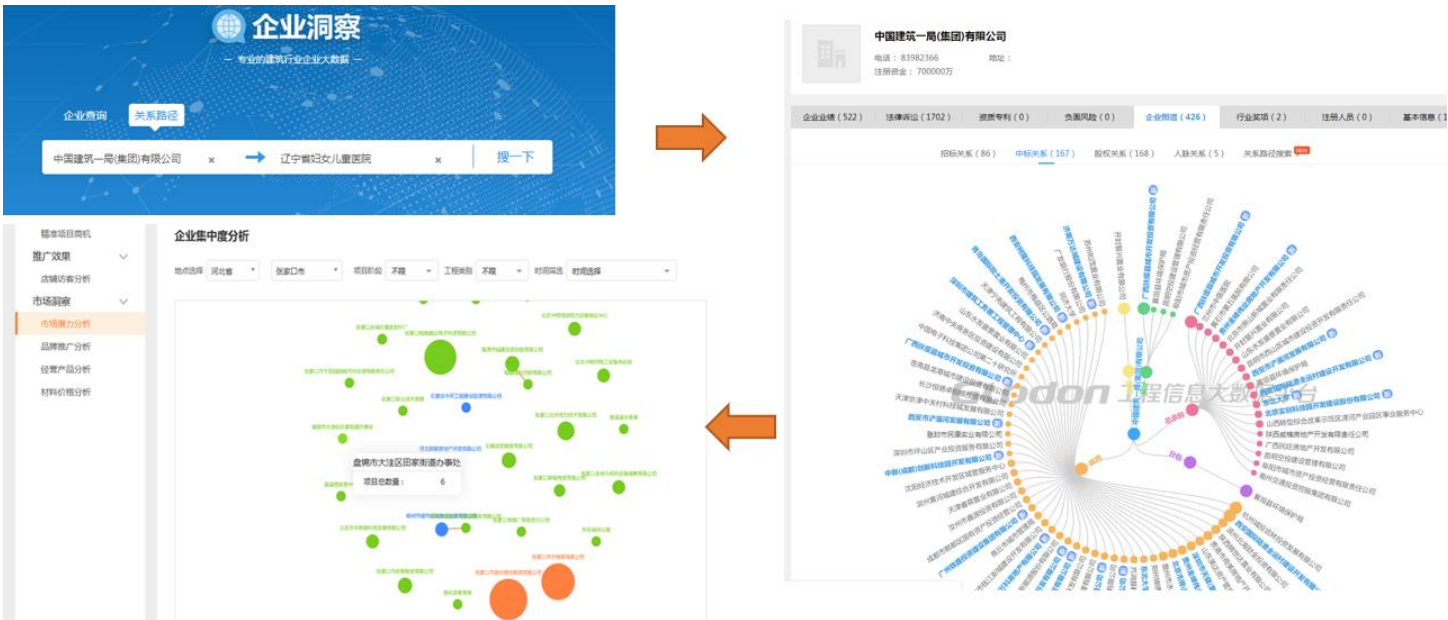
# 数据分析和可视化

## 基于大数据分析界定清单综合单价区间

清单标签数据																
清单状态： <div>请选择状态</div> <div>审核通过</div> <div>审核不通过</div> <div>删除</div>																
			ID	清单名称	清单特征	清单量	综合单价	线上价格区间	拟发布价格区间	单位	状态	定额编号				
1	+			回填土方	1.密实度要求:投标人根据招标图纸等招标资料自行考虑 2.填方材料品种:发包方认可之外购土方 3.填方来源、运距:根据施工组织设计或方案,由投标人统筹安排,自主决定 4.其他情况:此项目未尽内容但为完成该项目所需的请投标单位在综合单价内考虑	1896	6.26	5.450-7.920	5.450-7.920	m3	审核通过	1-2-24				
2	-			回填砂	1.密实度要求:投标人根据招标图纸等招标资料自行考虑 2.填方材料品种:外进及分层回填经认可之中砂或粗砂 3.填方来源、运距:根据施工组织设计或方案,由投标人统筹安排,自主决定 4.其他情况:此项目未尽内容但为完成该项目所需的请投标单位在综合单价内考虑	126.4	11.17	11.170-11.170	11.170-11.170	m3	审核通过	1-2-24				
清单定额信息						清单人料机信息										
	编码	名称			单位	含量	数量	单价		编码	类别	名称	规格及型号	单位	数量	预算价
1	1-2-24	装载机装运土方 运距150m以内			m3	1	126.4	11.17		1	人工费	其他工		工日	6.32	
										2	机械费	轮胎式装载机	2m3	台班	0.7584	
										3	材料费	黄砂中砂		kg	126.4	

# 数据分析和可视化

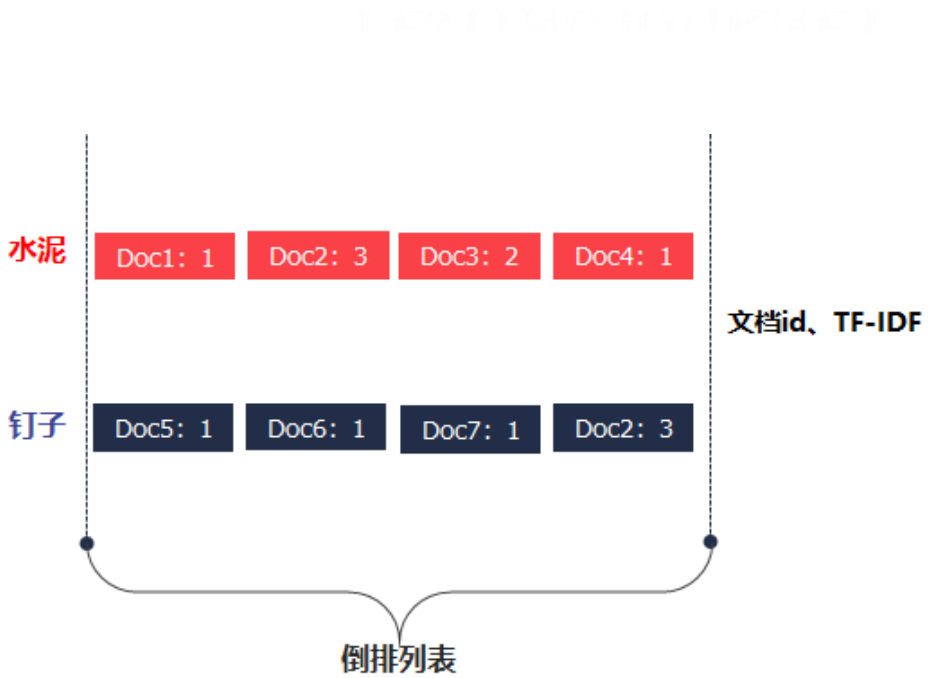
- 1. 企业之间关系查询
- 2. 政务大数据的应用
- 3. 数据大屏



# 搜索测试

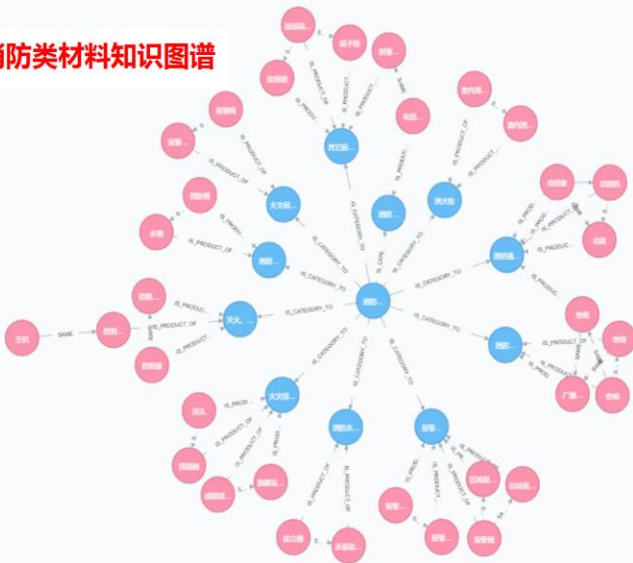
搜索测试关注几个重点：

- 1. 测试结果正确：召回率/准确率、分词、知识图谱
- 2. 搜索结果排序正确：打分、定制规则、推广效果
- 3. 搜索效率
- 4. 搜索稳定性：自动化、监控体系建立



搜索监控信息列表						
	用户信息	搜索条件	搜索结果	跟进状态	跟进人	完成时间
客户编码: vip643072 账号: vip643072 所在省: vip有效用户	地区: 内蒙古自治区 内蒙古自治区 搜索词: 软管 其它: 品种: 地埋式喷头* 材料类别: 洒水喷头	结果地区: 匹配方式: 精确匹配 页面结果: 无结果	分 站: 待处理 搜索规则: 待处理 同 义 词: 待处理 询价交付: 未交付 询价跟进: 未跟进 最终交付: 未交付 最终跟进: 未跟进	分站: 内蒙分站-	搜 索: 2018-11-28 19:25 分 站: 搜索规则: 同 义 词: 询价交付: 询价跟进: 最终交付: 最终跟进: 分站结项时间: 0 小时	
客户编码: vip643072 账号: vip643072 所在省: vip有效用户	地区: 内蒙古自治区 内蒙古自治区 搜索词: 鸭嘴 其它:	结果地区: 河北省 石家庄市 匹配方式: 精确匹配 页面结果: 推荐全国	分 站: 待处理 搜索规则: 待处理 同 义 词: 待处理 询价交付: 未交付 询价跟进: 未跟进 最终交付: 未交付 最终跟进: 未跟进	分站: 内蒙分站-	搜 索: 2018-11-28 19:13 分 站: 搜索规则: 同 义 词: 询价交付: 询价跟进: 最终交付: 最终跟进: 分站结项时间: 0 小时	
客户编码: vip643072 账号: vip643072 所在省: vip有效用户	地区: 内蒙古自治区 内蒙古自治区 搜索词: 鸭嘴 其它:	结果地区: 河北省 石家庄市 匹配方式: 精确匹配 页面结果: 推荐全国	分 站: 待处理 搜索规则: 待处理 同 义 词: 待处理 询价交付: 未交付 询价跟进: 未跟进 最终交付: 未交付 最终跟进: 未跟进	分站: 内蒙分站-	搜 索: 2018-11-28 19:13 分 站: 搜索规则: 同 义 词: 询价交付: 询价跟进: 最终交付: 最终跟进: 分站结项时间: 0 小时	

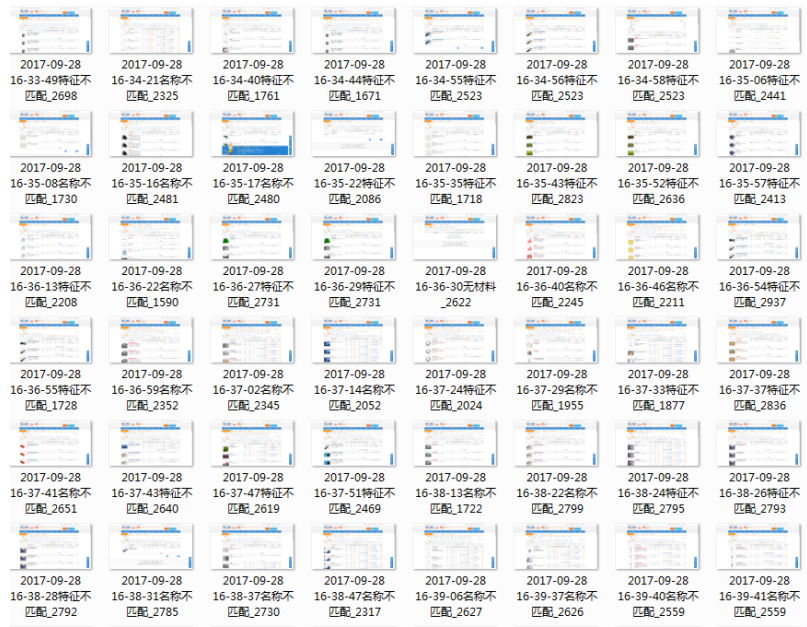
消防类材料知识图谱



# 搜索测试

搜索测试关注几个重点：

- 1. 测试结果正确：召回率/准确率、分词、知识图谱
- 2. 搜索结果排序正确：打分、定制规则、推广效果
- 3. 搜索效率
- 4. 搜索稳定性：自动化、监控体系建立

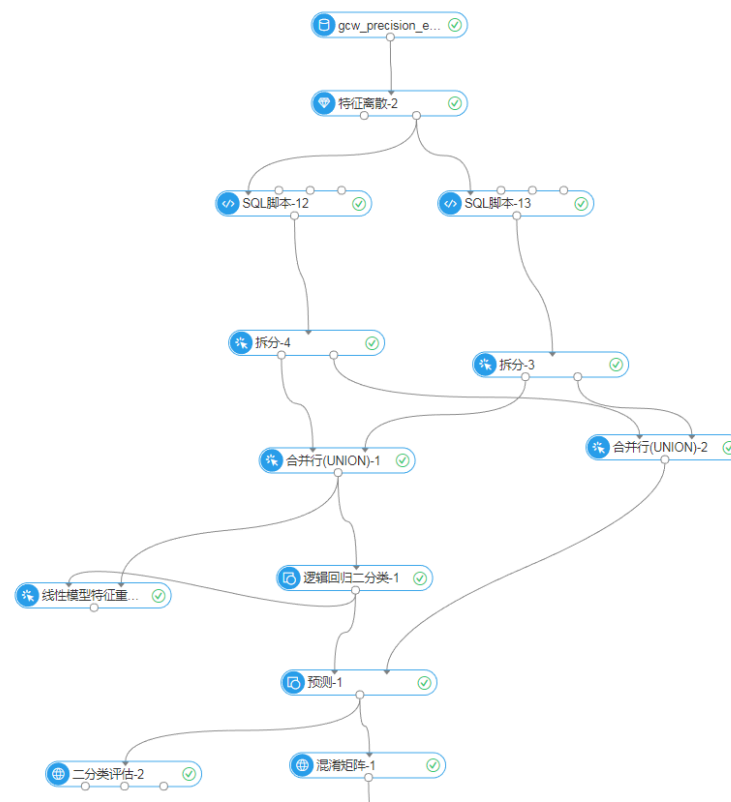
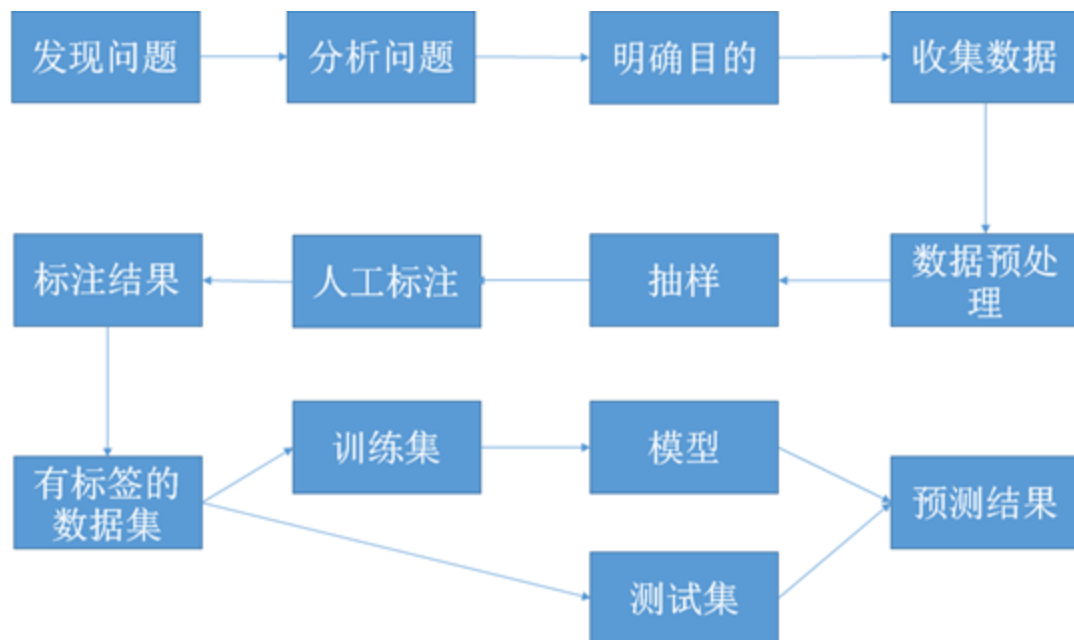


搜索成功数	搜索总数	覆盖率	准确率
9003414	10371270	86.81%	90.83%
1016498	1124243	90.42%	95.99%
959945	1090120	88.06%	92.43%
126337	151627	83.32%	91.09%
649914	743859	87.37%	90.27%
149603	173145	86.4%	91.5%
138323	159714	86.61%	89.8%
290093	329169	88.13%	93.24%
146805	169988	86.36%	90.15%
133266	151023	88.24%	90.72%
500405	555984	90%	93.01%
182166	209099	87.12%	90.33%
143554	168094	85.4%	89.95%
83584	100014	83.57%	89.58%
2579	17267	14.94%	74.59%
126880	147797	85.85%	89.22%
284614	332214	85.67%	90.78%
409285	477628	85.69%	89.61%
527765	599451	88.04%	89.28%
150883	179224	84.19%	90.94%
1033218	1186794	87.06%	90.61%
206940	240606	86.01%	88.28%
107212	129808	82.59%	89.56%
416560	469090	88.8%	89.85%
259058	304010	85.21%	92.04%
288070	332928	86.53%	90.32%
197835	226566	87.32%	92.64%
0	4059	0%	0%
321821	376476	85.48%	91.28%
90491	106933	84.62%	87.33%
1	15969	0.01%	81.37%
56648	69591	81.4%	88.38%
3061	28470	10.75%	43.08%
0	0	0%	0%
0	123	0%	0%
0	187	0%	0%

# 搜索测试

基于机器学习的搜索测试尝试：

1. 数据预处理
2. 使用有监督的学习，训练模型





# 智能推荐测试

1. 标签数据
2. 用户画像
3. 个性化推荐

选择你感兴趣的内容

---

 人工价格

免费开通微信端查询入口

 信息价更新提醒

☐ 订阅地区一： 请选择订阅地区 ☐ 订阅地区二： 请选择订阅地区

 一周价格动态

☐ 全选

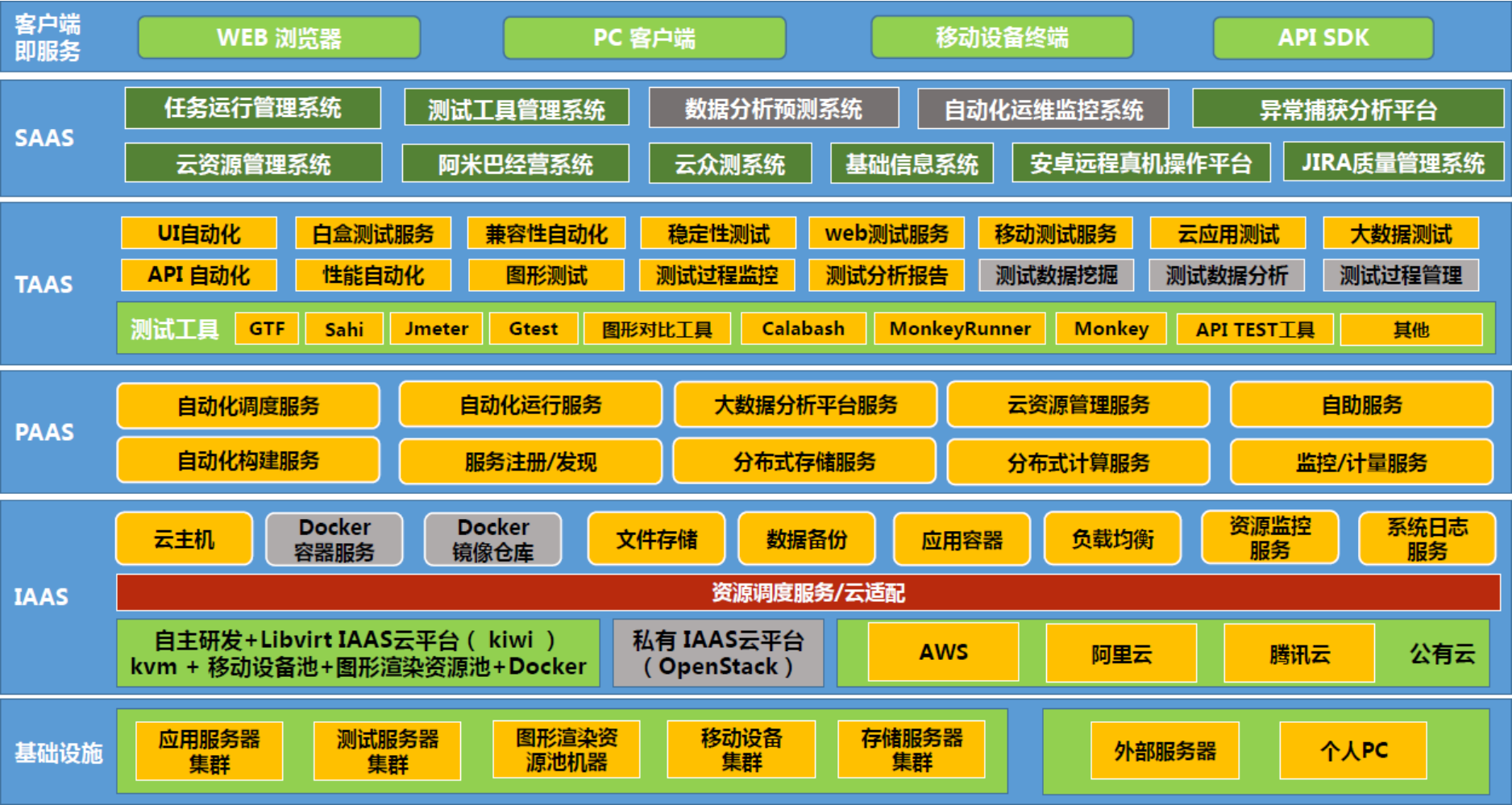
 数据包价格优化

☐ 全选



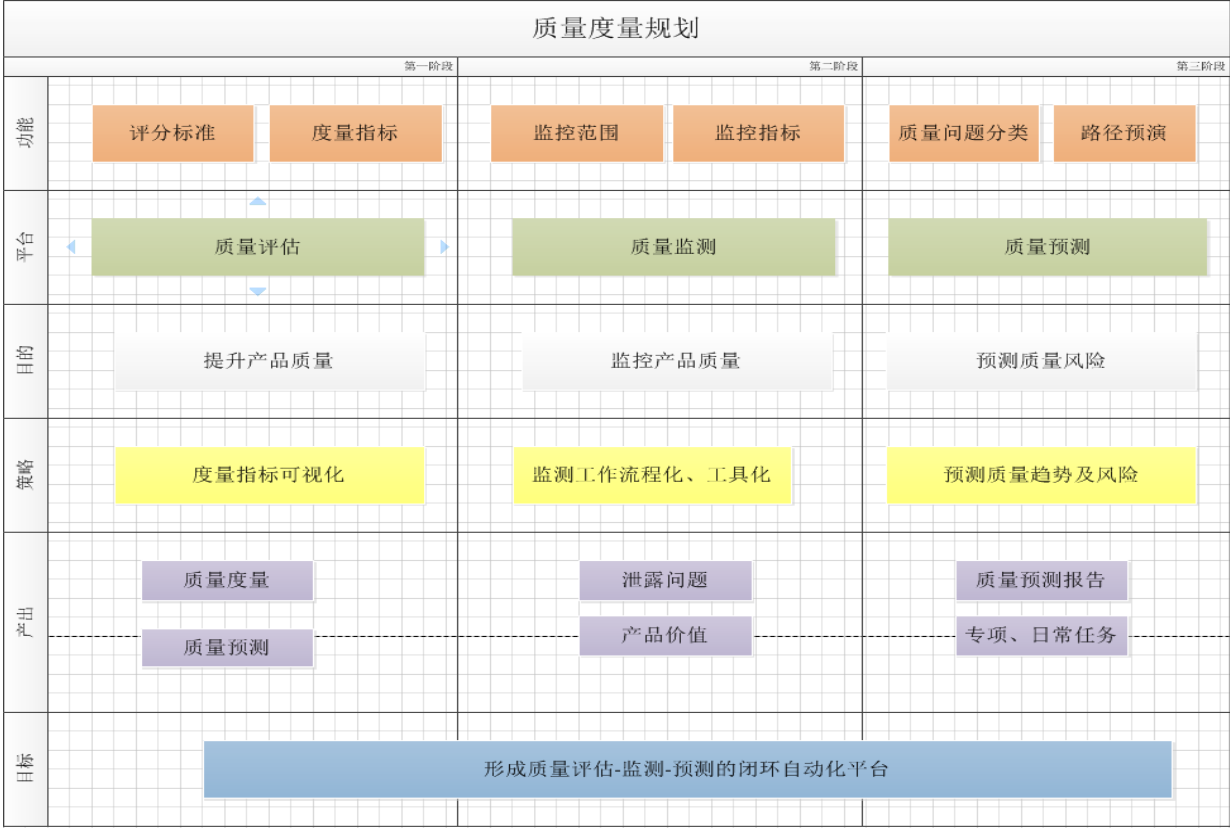
# 测试技术架构

## 集团云质量平台



# 总结

- 1. 对质量评价和质量度量的思考
- 2. 形成跨产品的质量评价体系
- 3. 动态监控平台



# 展望

1. 大数据能够在哪些业务单点发挥作用？
2. 如果从单点连接成质量闭环？
3. 如何打造和完善质量保证的整个体系？

# 展望

1. AI技术应用于测试
2. 增强式分析
3. 智能空间
4. 量子计算



# 感谢您的参与!



关注公众号获得更多软件测试知识