1. **会议信息**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **会议日期** | **会议时间** | **会议地点** | **发起部门** | **记录人** | **联系电话** |
| 2016-05-17 | 10:00~12:00 | 创牌中302A | 690PSI | 胡艺友 |  |
| **出席人**：卢瑾、毛振兴、宁鹏涛、李鑫、胡艺友 | | | | | |
| **发送**：卢瑾、李鑫、胡艺友  **抄送**：许学平、黄裔、邹晓慧 | | | | | |

1. **会议主题**

|  |
| --- |
| **主题**：方正爬取数据对接实施讨论 |
| **目的**（拟解决的问题或产出的内容）：方正爬取数据对接方案确认 |
| **议程：** |

1. **会议纪要**

| No. |  |
| --- | --- |
|  | 方正数据存储在HBase当中，hadoop版本为1.2.1，数据通过方正自己开发的检索系统查询  可以直接开放数据接口，也可以开放Web Service接口  数据每半小时更新一次  数据内容：   1. 海尔制定关键词，见《方正HBase家电关键词》（来源于徐美霖）。 2. 除家电相关关键词，也有房地产这类与家电行业相关的数据。房地产数据主要来源于搜房网 3. 数据既有海尔相关的家电数据，也有竞品数据 4. 原始数据获得之后会做清洗，也会对标题、内容的聚类分析及情感分析，   方正数据存储服务器在北京，系统本身压力不大，但是，宽带带宽压力会对数据传输压力影响较大，建议非工作时间进行抽取。  前台展现、业务结果数据约60M/日 |
|  | Hbase系统对接人为 宁鹏涛 |

**四、本次会议调度及决议**

1、决议状态：完成（按计划完成）；进行中（尚未到达计划完成日期任务执行中或任务的计划延期且经过同意）；未完成（未按计划完成）。

2、各参会人员按照决议要求及时落实解决，并请按决议到位时间反馈闭环结果。

| No | 调度及决议项 | 描述 | 责任人 | 到位时间 | 闭环结果 |
| --- | --- | --- | --- | --- | --- |
|  | HBase介绍 | 发送HBase相关介绍文档 | 宁鹏涛、卢瑾 | 6月21日 |  |
|  | 数据对接 | 分析HBase数据并使用DataFlow、Vector等进行连接、抽取测试 | 黄裔 | 待定 |  |

**滚动待关闭决议**

| No | 调度及决议项 | 描述 | 责任人 | 到位时间 | 闭环结果 |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

**五、评价方式(根据具体使用范围,输出可执行和落地的机制)**

1.决议拖期反馈, …… 2.决议未按进度要求完成,………