

Considering the Data's Nature

I'm presently focusing on the initial analysis of the JSON data structure. My primary concern is understanding the user profile: the roles, experience, and the specific skill set of this FMCG marketing professional with some data analyst experience. This user profile will guide my further interpretation of the data.

Prioritizing Automation Opportunities

I've just finished a thorough needs analysis. My key focus right now is pinpointing the most impactful automation opportunities for this user. Manual data scraping and entity resolution are causing the most significant pain. I'm leaning heavily towards "ready-made solutions" and am exploring Power Automate Desktop as a primary tool due to the user's existing Power BI familiarity. I'm also mindful of potential obstacles like enterprise security and legacy systems during my product assessment.

Analyzing the Solution Space

I'm now revising my strategy. While "ready-made" is the goal, pure SaaS solutions might struggle with *internal system login* and *fuzzy entity matching*. Considering the user's Power BI and UiPath knowledge, and the hints in the JSON, the **Microsoft Power Platform** ecosystem emerges as the best path. It's a low-code "product" that fits the corporate landscape.

Developing the Technical Design

I'm currently crafting the technical architecture for the solution, keeping the user's non-technical background in mind. My design revolves around Power Automate Desktop for data ingestion from legacy web systems. This is ideal, given its ability to simulate keyboard/mouse interactions. I'm focusing on Excel on OneDrive for simple storage and using Power Query for fuzzy matching as a built-in AI method. Power BI will be the visualization tool. I'm addressing the login security and 2FA issue with an attended RPA approach.

Defining the Interface Design

My current focus is on the user interface and user experience, aiming for a "Control Center" metaphor. The flow is straightforward: launch from Power Automate Desktop (PAD), automated background processes, and then the user reviews results in Power BI. Key visualizations will include Before/After data comparisons, sales trends, and an exception list for manual review.

\n\nSchema and Data Structuring

I am now designing the data schema. The crucial entities are products from Source A, Source B, and a unified product record. The schema clearly maps raw data from each source to a central data point and also shows a mapping table with a confidence score. Data privacy is a prime concern, so all data stays on the local machine.

Documenting the Process

I'm presently documenting the entire process. I'm focusing on crafting a clear explanation of entity resolution that the non-technical user can readily understand. To that end, I'll be describing the "smart matching" concept. The document will also ensure I adhere to the "Buy over Build" preference, positioning the Power Platform as a tool kit. Furthermore, I will ensure local data processing is emphasized for data privacy.

基于您提供的 JSON State，特别是用户作为“快消行业市场营销”背景、痛点在于“内部系统割裂”与“实体匹配 (Entity Resolution)”，且明确倾向于“现成方案 (SaaS/No-Code)”的需求，我为您制定了以下基于 **Microsoft Power Platform (Power Automate + Power BI)** 的交付方案。

这套方案利用大多数快消外企/大厂已有的微软生态，无需额外采购昂贵的独立软件，同时满足 RPA（自动化操作）和 AI（智能匹配）的需求。

DOC_01_PRD.md

产品需求文档 (PRD): 营销数据自动化合并助手 (The Marketing Data Unifier)

1. 核心价值主张 (Value Proposition)

打造一个***“零代码数字员工”***，它能像真人一样自动登录公司旧系统下载报表，并利用模糊匹配算法自动合并不同来源的商品数据，将用户从每周 4 小时的“复制粘贴”和“肉眼比对”中解放出来，专注于高价值的市场策略分析。

2. 用户故事 (User Stories)

- **As a** 品牌市场经理 (Marketing Manager),
I want to 点击一个桌面按钮就能自动完成三个内部系统的登录和数据下载,
So that 我不需要每周重复输入密码和点击繁琐的菜单，避免机械性劳动。

- **As a** 数据分析兼职人员,
I want 系统能自动识别“A系统里的可口可乐 330ml”和“B系统里的 Coca-Cola Can 330”是同一个商品,
So that 我不需要在 Excel 里手动一行行做 VLOOKUP 或人工勾选, 解决由于命名不规范导致的数据孤岛。
- **As a** 策略制定者,
I want to 在 Power BI 中直接查看合并后的全渠道销售分析,
So that 我能快速发现渠道间的价格冲突或库存机会。

3. 功能清单 (Feature List)

P0: 核心闭环 (Must-have)

1. **桌面端自动抓取 (RPA)**: 基于 Power Automate Desktop, 模拟鼠标点击, 完成登录、导航、导出 Excel 的全过程。
2. **智能模糊匹配 (Fuzzy Matching)**: 利用 Power Query 的模糊合并功能 (Fuzzy Merge), 设定相似度阈值 (如 0.85), 自动关联不同系统名称不一致的同一商品。
3. **异常数据标记**: 对于匹配度低于阈值的数据, 生成“待人工确认清单”, 确保数据准确性。
4. **本地数据存储**: 所有数据仅在本地 Excel 或公司 OneDrive 流转, 确保绝对合规与安全。

P1: 体验升级 (Nice-to-have)

1. **邮件自动触发**: 数据更新完成后, 自动发送 Teams 通知或邮件告知“日报已就绪”。
2. **AI 增强清洗**: 接入 AI Builder (如需更高级的非结构化数据提取), 从 PDF 扫描件中提取数据。

4. 验收标准 (Success Metrics)

- **时间节省**: 数据准备时间从 **4小时/周** 缩减至 **15分钟/周** (仅需人工处理少量未匹配项)。
- **准确率**: 商品自动匹配准确率达到 **90%** 以上。
- **稳定性**: 在连续 4 周的运行中, 无需人工干预流程 (除 2FA 验证外)。

DOC_02_Tech_Architecture.md

🚧 技术落地架构 (Technical Implementation Guide)

1. 架构设计图



Syntax error in text
mermaid version 11.12.1

2. 技术栈选型 (Tech Stack)

鉴于用户的“非技术背景”和“现成方案”偏好，我们采用 **Microsoft Power Platform** 生态，这是最符合快消外企环境的低代码方案。

- **抓取/操作层 (RPA): Microsoft Power Automate Desktop (PAD)**
 - *理由*: 这是 Windows 自带或易于安装的工具，专门用于对付那些没有 API、界面老旧的内部网页。它通过“录制”用户操作来生成脚本，零代码。
- **逻辑/数据清洗层: Excel Power Query 或 Power BI Dataflows**
 - *理由*: 用户熟悉 Excel。Power Query 内置了 "Fuzzy Merge" (模糊合并) 功能，无需写 Python 代码即可解决“实体匹配”痛点。
- **存储层: SharePoint / OneDrive for Business**
 - *理由*: 企业级安全，自动同步，方便团队共享。
- **展示层: Power BI**
 - *理由*: 用户已认知该工具，且能直接连接 Power Query 处理好的数据。

3. 关键难点与攻克 (Key Challenges & Solutions)

- **难点 A: 企业系统的登录限制与验证码 (2FA)**
 - *预判*: 很多内部系统有动态验证码或 SSO 跳转，全自动脚本难以通过。
 - *解决方案*: 采用 **Attended RPA (有人值守自动化)** 模式。
 - 脚本启动后，自动输入账号密码。
 - 遇到验证码时，脚本设计一个“等待窗口”，由用户手动输入验证码或在手机上确认。
 - 确认后，脚本继续接管剩下的下载工作。
- **难点 B: 不同系统商品名称差异巨大 (Entity Resolution)**

- 预判: 系统 A 叫 "Coke Zero", 系统 B 叫 "CocaCola Zero Sugar"。
- 解决方案: 使用 Power Query 的 **Fuzzy Merge**。
 - 设置 Similarity Threshold (相似度阈值) 为 0.8。
 - 建立一个 Transformation Table (转换表), 将常用的别名 (Mapping表) 维护在 Excel 中, 辅助算法匹配。

DOC_03_UX_Concept.md

交互与体验设计 (UX/UI Concept)

1. 界面形态 (Interface Metaphor)

产品不是一个独立的 App, 而是一套**“桌面小助手 + 数据驾驶舱”**的组合。

- 控制端: Power Automate Desktop 的悬浮窗 (类似一个简单的“播放器”)。
- 结果端: Power BI 的交互式报表。

2. 用户交互流程 (User Flow)

1. 启动 (Trigger):

- 用户周一早上上班, 打开电脑, 点击桌面上的 Run Weekly Data Fetch 按钮。

2. 协作 (Collaboration):

- 屏幕自动打开浏览器, 填入密码。
- (弹出提示框): “检测到需要手机验证码, 请在手机确认后点击‘继续’。”
- 用户点击继续, 随后可以去泡咖啡, 屏幕上鼠标自动快速点击下载文件。

3. 反馈 (Feedback):

- 5分钟后, 电脑右下角弹出通知: “数据下载完毕, 正在进行智能合并...”

4. 消费 (Consumption):

- 用户打开 Power BI 仪表盘。
- 首页显示: “本周数据已更新。共处理 500 条记录, 自动匹配 480 条, 有 20 条需人工确认。”

5. 修正 (Correction):

- 用户点击“待确认列表”, 在 Excel 维护表中简单修正那 20 条无法匹配的新品, 刷新即可。

3. 可视化建议 (Visualization)

针对市场营销场景，建议 Dashboard 包含：

- 漏斗图 (Funnel Chart)**：展示从“全渠道流量”到“实际销量”的转化。
- 桑基图 (Sankey Diagram)**：极佳地展示“数据匹配流向”—— 左边是来源 A/B/C 的杂乱商品名，汇聚到右边的统一标准商品名，直观展示 AI 的归因效果。
- 红绿灯指标 (KPI Cards)**：对于价格异常（如某渠道破价）的商品，用红色高亮显示。

DOC_04_Data_Schema.md

数据流与 Schema 设计 (Data Flow & Schema Design)

1. 核心实体定义 (Entity Definition)

为了解决数据孤岛，我们需要构建一个***“主数据标准 (Master Data)”***。

- Entity A: Raw_System_Export (原始凭证)**
 - 来源：各个割裂的内部系统导出的 CSV/Excel。
- Entity B: Product_Master_Map (商品主映射表)**
 - 作用：这是核心资产，记录了 AI 匹配的结果和人工修正的规则。
- Entity C: Unified_Sales_Fact (统一销售事实表)**
 - 作用：清洗后的最终数据，用于分析。

2. 数据结构 (Schema Example)

推荐使用扁平化的表格结构（适合 Power BI/Excel 处理）。

表 1: Unified_Product_Master (标准商品表)

Field Name	Type	Description
Master_SKU_ID	String	公司统一的标准货号 (PK)
Standard_Name	String	标准商品名称 (如: Coca-Cola 330ml Can)

Field Name	Type	Description
Category	String	品类
Brand	String	品牌

表 2: Source_Mapping_Rules (映射规则表 - 核心痛点解决)

Field Name	Type	Description
Source_System	Enum	来源系统 (e.g., 'ERP', 'Tmall', 'Distributor')
Raw_Product_Name	String	原始系统中的名称 (e.g., '可乐 330 拉罐')
Mapped_Master_ID	String	对应的标准货号 (FK -> Master_SKU_ID)
Match_Confidence	Float	匹配置信度 (0.0 - 1.0), 低于 0.8 标记为需审核
Is_Manual_Override	Boolean	是否经过人工修正

表 3: Daily_Sales_Fact (销售事实表)

Field Name	Type	Description
Date	Date	销售日期
Master_SKU_ID	String	标准货号
Channel	String	销售渠道
Sales_Volume	Integer	销量
Revenue	Decimal	销售额

3. 数据流转与合规 (Privacy & Compliance)

- 数据流向:
Internal Systems (Web) -> Local Temp Folder (Downloads) -> Excel Power Query (ETL) -> Power BI Dataset
- 合规性设计:
 - 不上传公有云 AI: 鉴于您在企业环境, 我们**不推荐**将销售数据发送到 ChatGPT/OpenAI API 进行匹配。
 - 本地计算: Power Query 的模糊匹配算法完全在本地 CPU 运行, 数据不出内网, 符合最严格的企业数据安全标准 (DLP)。

iii. **凭证管理**：Power Automate Desktop 支持加密存储密码，避免明文写在脚本中。