

# 概述

## 个人信息

- 姓名：张宏标
- 所属团队：计算平台事业部-MaxCompute-Data Plus AI
- 实习时间：2025.05.29-2025.08.15

## 教育经历

- 硕士：北京邮电大学-计算机学院-软件工程专业
- 本科：山东大学-软件学院-软件工程专业

## 过往实习与项目经历

参与的项目	负责的工作	个人收获
AUTOSAR汽车工具栈	AUTOSAR架构下的RTE层级模板代码开发工作	对于新领域以及标准规范文档的快速接受能力
美团到店研发平台MDO系统	MDO系统的功能设计与开发以及系统维护工作，包括一个体测系统的重构工作	对于大型项目的开发经验、在原有代码的基础上二次开发的能力，企业内的正式开发规范与流程

## 实习情况

在2025.06-2025.08期间，我以基础平台研发实习生身份加入Data Plus AI团队，参与MaxFrame计算框架的开发工作。期间主要参与引擎侧的开发工作，完成了在MCSQL引擎下的extract\_kv、collect\_kv算子的从设计到测试的全链路开发工作，同时也完成了PivotTable等多个算子的测试工作；在管控链路层，参与了GU售卖链路的开发工作。

## 主要产出

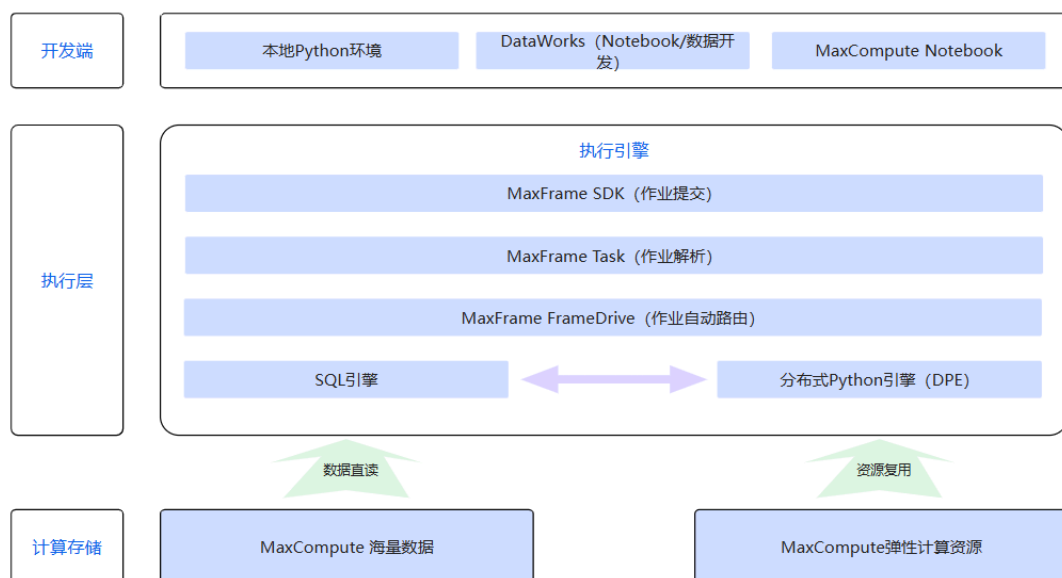
- Extract\_kv、Collect\_kv算子开发**

**业务背景：**MaxFrame是一个提供基于Python编程接口的分布式计算框架，提供MaxFrame SDK用于Python生态中的大数据处理和AI开发。其优势之一在于提供与Pandas兼容的DataFrame API，减少用户的学习与上手难度。所以项目的核心开发工作之一就在于尽快提高兼容Pandas的API数量，也就是算子覆盖率。本工作基于此完成了Extract\_kv、Collect\_kv算子的开发，提高了算子覆盖率。

**核心目标：**参照PyOdps的规范完整的实现MCSQL引擎中的Extract\_kv、Collect\_kv两个算子的功能。

  - Extract\_kv：**将指定列中的Key-Value对按照特定的kv\_delim和item\_delim拆分展开为多列。
  - Collect\_kv：**将多个普通列转换为Key-Value列。

## 实现逻辑：



一个算子主要包括两部分，第一部分是表示层，负责在client端构建计算节点，另一部分是执行层，在MCSQL中的执行方式为将算子的功能逻辑转化为符合MaxCompute语法的SQLcode并交由SQLtask执行。实现时也主要实现这两部分，在表示层对输入参数进行预校验等操作，然后进行算子的init并设置inputs等属性，再然后进行返回结果ouputs的schema构建，最终返回一个OP对象。实现层则负责将OP对象转换为SQLcode并提交给SQLtask。

### 项目难点及解决方式：

- extract\_kv算子会展开生成多列，MCSQL引擎对于SubDAG生成的SQLcode是连续执行且SQL执行过程中不支持动态新增列。

**解决方案：**使用flatmap自定义函数获取指定列中所有的key并将其执行设置为算子的inputs属性，同时将算子的acceptence属性设置为BREAK\_BEFORE，将算子的acceptence属性设置为BREAK\_BEFORE当计算图执行到该算子时，会中断并执行inputs以获取具体数据。

- MaxFrame是惰性计算框架，需要在执行具体计算前指定output schema以供后续算子的执行，但是extract\_kv算子中新增加的列数和列名不确定，无法指定schema。

**解决方案：**在算子中断执行inputs后，拿到具体的执行结果时根据结果使用dynamic update机制动态的更新output schema。

- 对于多个Key-Value对之间以及Key和Value之间如何进行拆分

**解决方案：**前期尝试自己实现拆分，先将一行内的多个Key-Value对拆分到多行，然后将每行的单个Key-Value对拆分为Key和Value两列，然后依据原DataFrame的index将最初属于同一行的多个Key再转换到同一行当中，但始终无法解决在分布式环境下的index唯一性问题。后查询文档使用STRTOMAP函数将字符串直接转换为Key Value Map。并使用Key获取对应Value。

## 成长收获

### ✂ 实习收获

在阿里云的这段实习，对于我的技术能力和职业素养提升都非常大。首先是对于业内前沿技术的学习。分布式对于学生的我来说一直都仅限于理论知识的学习，很少有这样的机会能够将过去所学的理论和具体的技术实现结合起来。其次是对企业级项目的开发的有了更加深刻的体会，过去参与的项目大多将代码开发作为重点，只需要完成功能的实现即可，其余部分大多都是模糊的。而这次在组内参与公司成熟系统的开发，全程体验了工业级项目的严谨性，体验了从需求评审到最终上线的全过程，让我真正明白功能实现只是一小部分，更要考虑可维护性、扩展性和质量保障，这种“大局观”和严谨态度将使我终身受益。最后是对AI的使用，阿里云非常鼓励使用AI工具来提升工作效率，我们团队还开过有关如何使用AI辅助工作的交流会并形成文档。现在，AI已经成为我日常工作的一部分，这极大的提高了我的工作效率。

## 🔒不足与反思

---

在实习的过程中也认识到了很多自己身上的不足之处，知识面不够以及代码能力不够强这些就不提了，对于代码的阅读能力不足是我认为我最大的缺陷，过去习惯于直接上网搜索文档和别人提供的解决方案，现在猛然遇到了需要直接去阅读代码来学习和解决问题的场景，很不适应，每次都要付出很多时间和精力去阅读代码，这个应该是我后续首要解决的问题。其次是开发流程的不规范，很多时候还没有定义清楚需求和解决方案，就急于进行代码的开发工作，导致代码要多次修改，最终事倍功半。最典型的例子就在于对Extract\_kv算子的开发过程中，对于字符串的拆分过程我一直尝试自己实现，废了很大功夫，后来发现MaxCompute中就有个STRTO\_MAP函数可以直接实现我需要的功能。现在已经学会了在定义清楚需求并形成完整且完善的解决方案之后再进行代码的开发。

## 未来规划与设想

### 📅个人发展工作规划

---

- **短期规划：**完成从学生到工程师的转变，包括思想、心态、思考问题的方式与角度等。将所学知识应用到真正的企业级开发，依托组内业务项目，快速提高自己的代码能力和技术水平，拓宽技术视野，积累开发经验；同时补充自己的非技术向能力，包括但不限于团队协作能力、表达沟通与文档写作能力等。
- **长期规划：**在特定领域或方向上积累更多经验，研究技术实现细节，深究底层原理，希冀成为一个在该领域有所深耕的技术专业人员。尝试以一个更高的视角去理解整体项目的设计与规划等，从一个单纯的技术人员逐步向系统分析或项目架构师转变。