arXiv:2303.03313v1 [astro-ph.IM] 6 Mar 2023

Rubin天文台科学处理流水线的数据管理与执行系统

Nate B. Lust，[1] Tim Jenness，[2] James F. Bosch，[1] Andrei Salnikov，[3]Nathan M. Pease，[3] Michelle Gower，[4] Mikolaj Kowalik，[4]Gregory P. Dubois-Felsmann，[5] Fritz Mueller，[3] 以及 Pim Schellart[1] 天体物理科学系，普林斯顿大学，普林斯顿，新泽西州 *08544*，美国[2]*Rubin*天文台项目办公室[2]，*950 N. Cherry Ave.*，图森，亚利桑那州 *85719*，美国

[3]*SLAC* 国家加速器实验室，*2575 Sand Hill Rd.*，门洛帕克，加利福尼亚州 *94025*，美国

[4]*NCSA*，伊利诺伊大学厄巴纳-香槟分校，*1205 W. Clark St.*，厄巴纳，伊利诺伊州 *61801*，美国

[5]*IPAC*，加州理工学院，*MS 100-22*，帕萨迪纳，加利福尼亚州 *91125*，美国

摘要。 我们介绍了Rubin天文台用于数据存储/检索和流水线代码执行的系统。用于数据存储和检索的层称为 Butler。它由一个关系数据库和一个用于管理数据位置的系统组成；前者称为注册表，用于跟踪元数据及其关系，后者称为数据存储。这些系统共同构成了一个抽象层，科学算法可以基于该抽象层编写。该抽象层管理了预期的大规模数据量所带来的复杂性，使算法能够独立编写，同时又能自动连接成连贯的处理流水线。该系统包含用于执行这些流水线的工具，这些工具将流水线转换为执行图，而执行图中包含存储在 Butler 中的具体数据。流水线基础设施从设计上具有可扩展性，能够在从笔记本电脑到多设施数据中心的各种环境中执行。本报告将重点介绍数据管理方面，并概述流水线及相应执行图的创建过程。

1. 引言

O'Mullane等（2023）介绍了 Rubin 科学处理流水线（Ivezi´c 等，2019）背后的核心设计原则。秉持这一理念，我们创建了名为 Butler 和中间件的系统。这些系统负责管理大量文件的存储、检索和组织所涉及的复杂性。它们还会存储有关数据的信息，例如各个数据单元之间的关系。在 Butler 和中间件中，单个数据单元称为数据集。这些系统起到抽象层的作用，算法代码构建于其上，因此算法的作者无需操心数据管理以及与其他算法编排的复杂性。

算法无需操心数据管理以及与其他算法编排的复杂性。

2. Butler 与中间件

Butler 是一个了解所有数据的系统，因此用户或代码作者无需操心这些事。它知道数据存储在哪里、以何种格式存储，以及各数据集之间的关系。该系统的用户只需处理内存中的 Python 对象，Butler 会负责管理适当的磁盘存储格式，并使用格式化器在两者之间进行转换。此外，这种抽象还使我们能够透明地管理同一种底层磁盘格式与各种内存数据表示之间的转换，例如 PyArrow 与 Pandas DataFrame。

Butler 是大多数用户和软件所使用的高级接口。它由两个较低层级的组件组成：注册表和数据存储。注册表是一个兼容多种不同实现的 SQL 数据库。它的职责是维护所有与数据相关的信息，包括唯一标识符以及数据集之间的关系。数据存储负责实际存储数据集。它能够识别用于存储数据的文件格式及其对应的 Python 内存类型。文件本身可以存储在一个或多个后端：POSIX、S3、Google 云服务、WebDAV；也可以作为中间对象保留在内存中，而不写入磁盘。

我们将与 Butler 交互的软件统称为中间件。这些软件负责处理从组织用户查询到协调 Butler 中存储数据的处理等各项工作。有关 Butler 和中间件的更深入介绍，请参见 Jenness 等人（2022）。

3. 维度

如前所述，维度是一种在大量可能性中组织、定位和识别数据的方式。在开始追踪任何数据之前，针对特定项目的维度集合便已固定。这可以视为该项目的“数据模型”，它完整描述了可能与任意给定数据集相关联的所有标识符。

维度之间还可以相互关联。例如，Rubin 维度系统包含曝光和探测器这两个维度。Rubin 的仪器本身也是一个维度，其中有 189 个独立探测器，因此给定一次曝光，就意味着探测器维度上存在 189 个相应的数据点。反过来，也可以针对给定的探测器，考察数据如何沿曝光维度变化（因此可将其作为时间的替代指标）。维度之间既可以像前例那样以逻辑方式关联，也可以在空间或时间上相互关联。

4. 数据集

为了完成管理数据集的工作，Butler 对数据集的理解远不止是存储在某处磁盘上的一堆比特。Butler 跟踪的每个数据集都由三部分组成：数据集类型、数据 ID 和集合。

数据集类型描述了数据集“是什么”。它有一个用作标识符的名称，以及一个存储类，用于将磁盘上的格式映射为内存中的 Python 对象，

以及一个维度子集。该子集用作标识符键，用于描述如何寻址上述数据集类型。

数据 ID 类似于给定数据集类型的地址。它是将数据集类型的维度映射到相应具体值的一种映射。这些值在所有维度的集合中构成一个“数据坐标”，就像 x、y、z 在笛卡尔空间中构成的坐标一样。

集合是一种将数据归组的方式。集合本身只是一个名称，但它允许 Butler 中两个或更多可能位于同一“数据坐标”的数据集共存，而不会发生冲突或歧义。不过，在给定集合中，每个坐标只能有一个数据集。

5. 利用 Butler 中的数据

Rubin 处理流水线中的所有算法都是称为任务的特殊对象。任务是单一的逻辑作业，可以组合到其他任务中，并具有一个单独定义且可持久化的配置架构。所有任务都对内存中的数据产品进行操作，这些数据产品通常由另一个任务提供。

任务有一个特殊子类，称为 PipelineTasks。它们通过向中间件提供信息，让中间件代表任务与 Butler 执行输入/输出，从而获取内存中的数据。PipelineTasks 会声明一个由“连接”组成的接口，这些连接是任务所需的输入/输出数据集类型，以及连接对应的维度。它们还会声明任务将要处理的一组维度。将这些维度与连接的维度结合起来，中间件便能决定如何向任务提供输入。例如，扩展上面的例子，如果任务声明其维度为曝光，输入为曝光、探测器，那么单次任务调用将获得 189 个输入（每个探测器一个）。相反，如果任务的维度也包括曝光和探测器，那么任务将被分别调用 189 次，每次接收一个输入。

6. Pipelines 与图

以 YAML 语法编写的 Pipeline 文件提供了一种将多个 PipelineTask 组合在一起的方法。在一个 Pipeline 中，每个 PipelineTask 都通过唯一标签声明，并且可以包含对任务默认值的配置覆盖。Pipeline 的设计具有特定用途，例如准备校准数据，或处理来自特定仪器的数据。因此，同一个 PipelineTask 可能会以不同的配置出现在多个 Pipeline 中。通过使用唯一标签，还可以使同一个 PipelineTask 在一个 Pipeline 中以不同的唯一配置出现多次。

中间件会将这些流水线转换为按执行依赖顺序排列的有向无环图（DAG）。具体来说，它会结合配置检查 Pipeline 中每个任务声明的连接，而配置可能会修改这些连接。由于排序由中间件完成，因此在流水线中声明 PipelineTasks 的顺序并不重要。该系统使任务作者无需面对整体复杂性，并提供了最大的可复用性，因为它能够在各种不同的上下文中使用同一任务。它还让作者能够在更熟悉的软件开发体系中工作，只需关注接口，而不必处理整个应用程序。

中间件构建好流水线图后，该图便可用作后续处理步骤的输入。流水线图会与 Butler 位置、用户指定的数据集合以及用户指定的数据约束一同传入。随后，中间件会检查流水线图中的输入节点，确定 Butler 中符合该节点的数据集类型和维度的数据集合。利用这些输入以及 Butler 所掌握的维度间关系信息，中间件便能根据预测将要创建、使用和输出的数据集，补全所有其余节点。这些新节点均称为 Quantum，分别代表对某个任务的一次调用。新有向无环图中的所有节点合称为 QuantumGraph。该图会在任何数据处理开始之前创建，以便分配或安排适当的资源。同一段代码可用于在各种计算平台上处理数据，从笔记本电脑上运行数百个 Quanta，到如 Gower 等人（2023）所述，在物理硬件或云端的计算集群上执行生产级任务。

7. 结论

Butler 和中间件使 Rubin 能够将所有数据访问操作从任务作者或最终用户处抽象出来。抽象数据访问也使得可以通过 PipelineTasks 和 Pipelines 来抽象数据处理。这些抽象使开发能够更快速、更独立地进行，同时也提供了一种机制，使同一代码库既能满足个人开发需求，也能扩展到运行完整的巡天。这些系统正用于处理昴星团望远镜的 Hyper Suprime-Cam PDR4，取代 Rubin 先前版本使用的旧中间件，例如 Bosch 等人（2018）所述的版本。它也用于美国国家航空航天局的 SPHEREx 任务，该任务拥有自己的数据模型。所有代码均为开源代码，可从 Python 包索引获取。

致谢。本材料或工作部分由国家科学基金会通过合作协议 AST-1258333 和合作支持协议 AST1836783 提供支持，这些协议由大学天文研究协会（AURA）管理；另有美国能源部根据与 SLAC 国家加速器实验室签订的合同 No. DE-AC02-76SF00515 提供支持，该实验室由斯坦福大学管理。

参考文献

Bosch, J., 等. 2018, 《日本天文学会刊》, 70, S5. arXiv:1705.06766 Gower, M., 等. 2023, 载于 ADASS XXXII，由 S. Gaudet、S. Gwyn、P. Dowler、D. Bohlender 和 A. Hincks 编（旧金山：ASP），ASP 会议系列第 TBD 卷，TBD. arXiv:2211.15795 Ivezi´c, Ž., 等. 2019, 《天体物理学杂志》, 873, 111. arXiv:0805.2366 Jenness, T., 等. 2022, 载于《天文学软件与网络基础设施 VII》，SPIE 会议录第 12189 卷，1218911. arXiv:2206.14941 O'Mullane, W., 等. 2023, 载于 ADASS XXXII，由 S. Gaudet、S. Gwyn、P. Dowler、D. Bohlender 和 A. Hincks 编（旧金山：ASP），ASP 会议系列第 TBD 卷，TBD. arXiv:2211.13611