# 概述

## 软件概述

白鲸开源是Apache DolphinScheduler和Apache SeaTunnel原班人马成立的开源商业公司，提供功能更强多、稳定性更强的商业版本解决用户调度，数据开发、数据同步和ETL的问题，目前支持192种数据库的ETL与数据开发，全面替换Informatica与Talend等工具相应功能，在中信证券、中信建投、中国银行、中国人保、中国人寿、旺旺集团等多个行业头部企业都有成功商业版和实施替换案例。



WhaleStudio是一款通过拖拽式的实现调度和ETL功能，支持快速融合在企业/合作伙伴的系统当中，感兴趣直接发邮件给service@whaleops.com，或者扫描以下二维码：



项目测试

* 工作流编排： 允许用户以编程方式定义工作流，将一系列任务组织成有向无环图（DAG），并指定任务之间的依赖关系和顺序。
* 调度和执行： 它可以按照预定的时间表和依赖关系调度任务，并执行这些任务。
* 监控和告警： 提供了用户友好的用户界面，用于监控工作流的运行状态、任务的执行情况和日志信息。同时，它还支持告警功能，可以在任务失败或达到特定状态时发送通知。
* 任务编写和定制：系统提供了丰富的内置操作符，涵盖了常见的数据处理、连接器和工作流程操作。
* 数据源支持MySQL、Oracle等结构化和半非结构化数据源，支持信创数据源等。
* 数据转换和清洗： 对数据进行实时转换、清洗和处理，确保数据质量和一致性。
* 可视化和监控： 提供用户友好的界面和监控工具，帮助用户监视数据流和任务的运行情况，以及识别潜在的问题或瓶颈。

## 产品需求理解

# 系统功能

## 系统架构

### 系统架构设计

本系统架构为分布式多集群部署，支持进行分组管理，如下图：



* 水平扩展性：本系统支持水平扩展，允许通过增加调度器、执行器和工作节点来提高处理能力。此特性使用户能够根据业务需求动态扩展资源，从而满足不同规模和负载的要求。
* 高可用性配置：为了实现高可用性，系统提供了多种配置选项，如通过负载均衡器（例如 NGINX 或 HAProxy）进行流量分发，配置多个元数据库实例以避免单点故障，并支持多节点部署的调度器和执行器，保障系统的稳定运行。

支持集群部署与故障转移，从而实现不停机扩展和缩减系统服务：



1. Edge Node： Edge Node 是集群的管理节点，负责管理整个集群。它提供 Web UI 和 API 来管理、配置和监控集群中的各个 Collector 节点。
2. Collector Node： Collector Node 是实际执行数据处理任务的节点。它们在集群中分布，执行数据流水线并处理数据。Collector 节点通过与 Edge Node 通信，获取配置信息和任务调度。

。

### 系统稳定性

本系统通过多管理节点部署，实现了工作流和数据同步功能，并具备节点故障转移能力，从而支持不停机的系统扩展与缩减。多管理节点架构确保在任意节点发生故障时，其他节点能够无缝接管任务，保障任务调度的连续性和稳定性。多节点并行调度不仅显著提高了系统的处理效率，还增强了系统整体的高可用性和可靠性：

* 先进的架构：工具将采用现代化的架构设计，确保高效稳定的数据处理和调度能力。
* 原生分布式部署：支持在分布式环境中进行部署，以适应大规模数据处理需求。



本项目预计采购的数据调度工具将满足以下高稳定性和成熟度要求：

1.  **系统架构稳定性**：工具将保持系统架构的稳定性，确保在长期运行中的高可靠性。
2.  **成熟技术采用**：采用经过市场验证的成熟技术，以满足不同层次用户的业务需求。
3.  **高压力场景考验**：经过高压力场景的严格测试，验证其在复杂和高负载环境下的表现。

|  |  |
| --- | --- |
| 指标项 | 实际测试指标 |
| 故障发现 | 1s |
| 自动切换 | 1s |
| 可用率（年） | >=99.9999% |