

数据集成流程及管理制度

商业研究部 | Data Engineering Group

一、数据库管理制度

- a) 务必保持信息同步渠道，指派固定对接人，提前告知变动计划，包括：新的地址和端口及启用的时间、后续是否有再次变动的可能性等
- b) 需要同步的信息：数据库地址、端口等绝对必要的连接信息

二、数据源管理制度

- a) 务必提供数据源（数据库）相关文档：ER 图、表结构说明/注释等
- b) 数据表或表结构变动时（删除/修改），由 DB 对接人提前发出通知，通知内容需要囊括：变动的具体时间、即将受影响的表名、计划改动/新增/删除的字段等

三、数据接入流程

1. 数据接入前

- a) 明确对接人：需要一名 DB 对接人和一名业务对接人，共同负责统筹管理数据接入项目中出现的信息同步以及可能存在的协作问题及潜在风险
 - DB 对接人
 - 业务对接人
- b) 梳理业务情况：业务对接人和 DB 对接人协助数仓开发工程师熟悉业务流及对应的数据流，需提供文档：
 - 产品文档：产品流程图等

- 数据库文档：ER 图、表结构文档等

c) 明确数据库信息：

- 数据库地址
- 数据库端口
- 其他会影响数据接入效果的信息

2. 数据接入中

a) 数据库连通测试：需要 DB 对接人协助

- 添加白名单
- 提供账号密码
- 联合调试

b) 数据入库：入库过程中出现的问题，会及时反馈给 DB 对接人，沟通、排查

3. 数据接入后

a) 同步任务监控：数仓负责数据异常监控，出现问题需定位问题并及时反馈给分析师及相关业务对接人

b) 异常任务反馈：数仓负责数据问题的初步定位并反馈给分析师和 DB 对接人，如果是数据源问题，需要 DB 对接人和业务对接人参与排查及解决