

二，使用数据集市时的常见问题：

参考文档：[永洪 Z-Data Mart 手册.pdf](#)

1. 数据集市安装和部署时要注意哪些？

数据集市分为本地集市和 MPP 数据集市。如果用户数据量在 GB 级别，并且单机服务器配置还可以，采用单机版数据集市系统可以达到数据加速的功能。如果用户数据在 TB 级别，就可以采用一个 MPP 多机版数据集市系统来支撑秒级的数据分析响应速度。

MPP 数据集市系统中，参与通讯的节点有以下几种类型：

- Naming Node : n, 命名节点，存储数据集市系统的元数据。
- Map Node : m, Map 节点，存储一部分 Map 数据的元数据和物理数据，并执行 Map 任务。
- Reduce Node: r, Reduce 节点，存储一部分 Reduce 数据的元数据和物理数据，并执行 Reduce 任务。
- Client Node : c, 客户端节点，发起对数据集市系统的访问。

安装时，可以选择多个节点一起安装，如下图：



- . 如果资源足够，命名节点分开，单独一台做。
- . Reduce 节点和客户端节点最好在一起，Reduce 的结果会给客户端，返回结果集很大，如果不和 Client 在一起的话，会有延迟。
- . Map 节点和 Reduce 节点最好分开，它俩都会占用大量资源。
- . 如果资源足够，可以在多台机器上装 Map 节点，可以并行工作，均摊工作量。

. 如果一台机器的性能足够好，可以安装多个应用，如多个 Map 节点，只需要配置端口偏移量 (dc.port.offset=1) 即可。

2. 数据导入集市时，如何写入 GMeta 值到云文件上？

✓ 如何写入 Meta 值：

在云任务的中，可以使用脚本通过 setMeta(String name,) 方法给导入到入库的云文件加入标签。如：

```
setMeta( "date" , new Date(2015,3,31)) , setMeta( "boolean" ,true), setMeta  
("region", "beijing");
```

另外，系统会自动加上文件名作为其中的一项摘要信息：_FILE_NAME_。

✓ 有何作用：

当用户创建一个云查询时，他/ 她可能不想基于云文件夹下面所有的云文件来运行查询。因为很多时候这样去运行查询没有必要，却非常消耗资源。这时，云查询可以定义文件过滤器来限制需要访问的文件。这个文件过滤器将基于命名节点的文件上的 GMeta 来运行，直接找出云查询需要访问的那些云文件，这样能极大地提升云查询的运行性能，并减少资源消耗。

3. 在云任务中，如何按时间动态生成云文件？

在创建任务计划时，文件名不填，通过脚本实现文件名动态生成，file=文件名。

例如文件名按日期命名的方法为

```
var fmt = new java.text.SimpleDateFormat('yyyyMMdd');  
var date = fmt.format(new java.util.Date());  
file = "test" + date;
```

这样如果今天为 2015 年 1 月 5 日，则生成的云文件为 test20150105 开始的 zb 文件，按天增量更新时不会冲突。

另外，文件夹和是否追加也可以通过脚本指定，folder=文件夹名，append=true/false.