

数据与算法的故事: 4

吕正华

2019.12

自我介绍

- ▶ Pivotal 资深软件工程师，开发 Greenplum 内核
- ▶ kainwen@gmail.com
- ▶ 个人主页: <https://kainwen.com>
- ▶ 2014 年毕业于电子系智能感知实验室, 工学硕士
- ▶ 2012 年第一次担任数据与算法助教

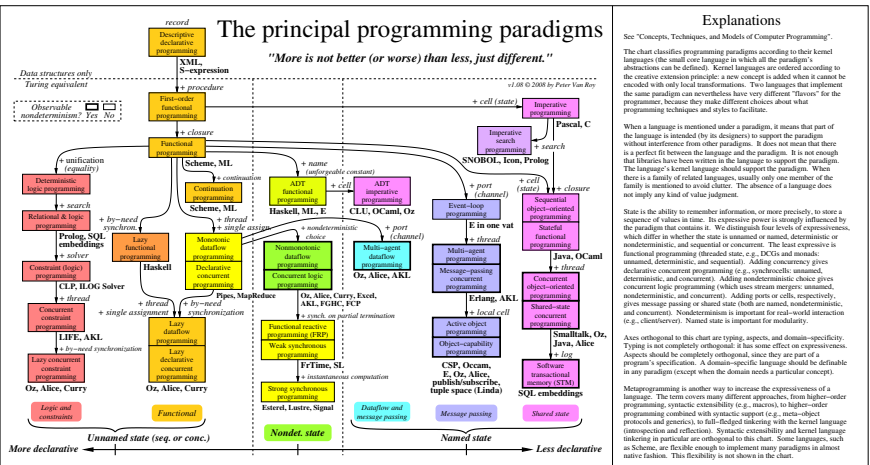


Figure: 编程语言范式

声明式 VS 命令式

- ▶ $\text{sqrt}(x) :: \text{real} \rightarrow \text{real}$
- ▶ 数学定义: $r^2 = x \wedge r \geq 0 \implies \text{sqrt}(x) = r$
- ▶ 但数学定义仅仅是定义，并没有告诉我们如何去得到 r
- ▶ 我们可以编写牛顿法等程序去求解
- ▶ 但如果只需要有了定义，就能有结果呢？这就是声明式编程

Programming in Logic

Prolog (Programming in Logic 的缩写) 是一种逻辑编程语言。它创建在逻辑学的理论基础之上，最初被运用于自然语言等研究领域。

一个求连通性的 Prolog 程序

```
1 link(a, b).  
2 link(b, c).  
3 link(a, c).  
4 link(c, d).  
5  
6 path(A, B) :-  
7     link(A, X), path(X, B).
```

- ▶ 小写开头标识符的是 atom, 可以认为是常量
- ▶ 大写开头标识符是逻辑变量, 可以被实例化
- ▶ 被 :- 连接起来的语句是规则 (rule)
- ▶ 上述过程的求解搜索过程是深度优先搜索

使用 Prolog 搜索锁模式

- ▶ 数据库里非常重要的领域和技术是并发控制
- ▶ Postgres 是用锁来调度事务，有八种锁模式
- ▶ Greenplum 基于 Postgres，但是一个分布式数据库
- ▶ 分布式数据库 DML 语句之间的特殊问题
 - ▶ EvalPlanQual 问题
 - ▶ 更新分布键问题

调度分布式系统的 DML

- ▶ 添加一种锁的类型
- ▶ 如何设计不同 DML 的持锁模式?
- ▶ Prolog 求解约束方程: [lockmode.pl](#)
- ▶ 讨论的邮件列表: [Serialization of split-update on QD](#)
- ▶ Pull Requests: [PR 6765](#)

关于搜索算法的理解

- ▶ 有定制的数据结构存储访问过的元素
- ▶ 这个数据接口决定了搜索的过程
- ▶ 数据结构是栈，则为深度优先搜索
- ▶ 数据结构是队列，则为广度优先搜索

用 SQL 编写 Dijkstra 算法

- ▶ 参考博文: [Dijkstra in SQL](#)
- ▶ SQL 也是一种声明式语言, 关系代数为基础
- ▶ 现代的 SQL 功能非常强大了
- ▶ 递归 CTE 使得 SQL 图灵完备

参考和推荐的资料

- ▶ Concepts, Techniques, and Models of Computer Programming

结束

谢谢！
Q&A