



DTCC

2016中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2016

数据定义未来



SequeMedia
威拓传媒

IT168

ChinaUnix

ITPUB

推荐系统架构演进 典型套路

当当网 推荐系统开发经理

张相於

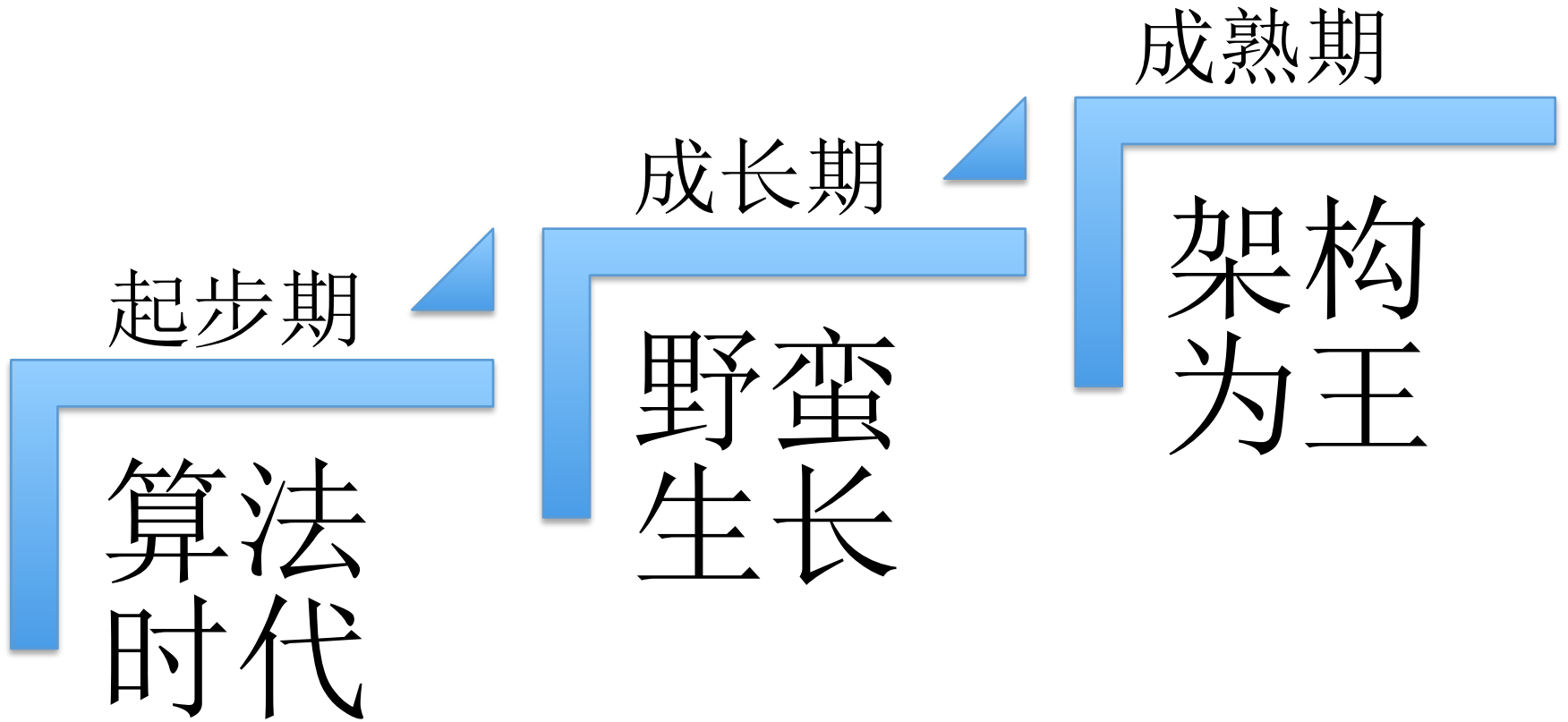
zhangxy@live.com

套路，都是套路。

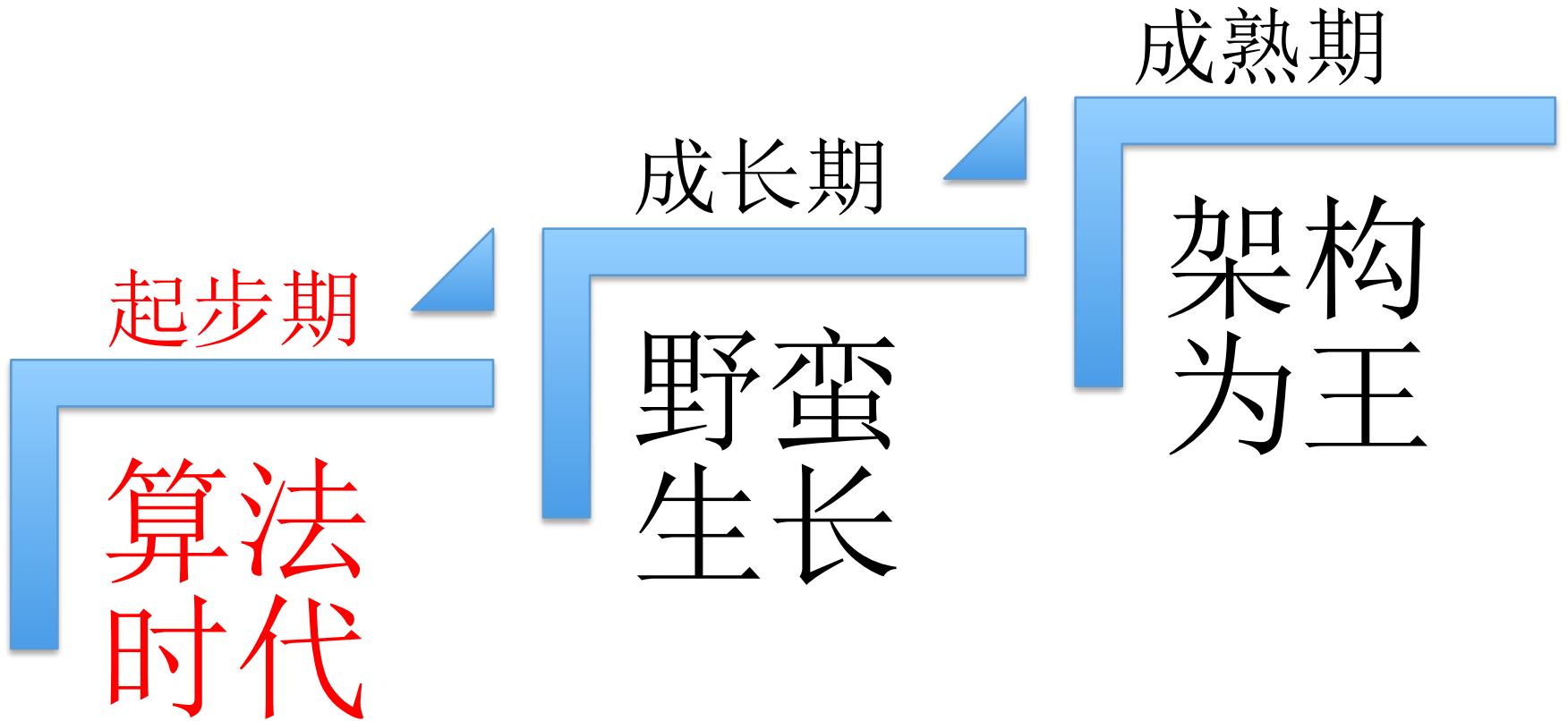
大纲



发展历程



发展历程



算法时代——动机和特点

动机

数据积累

业务需要

业务特点

从0到1

关注效果

意图单纯

特殊逻辑少

人工干预少

算法时代——算法

算法设计

选择

- 相似度：协同过滤|基于内容的算法

召回

- 全部商品。

排序

- 相似度

调优

- 参数调优

算法特点

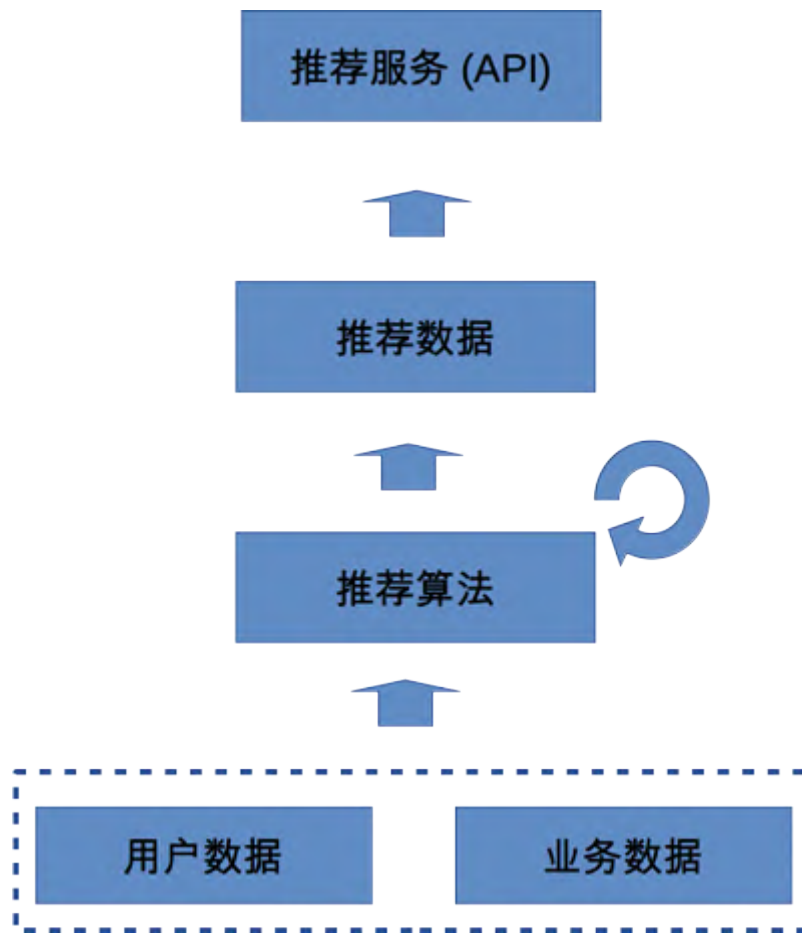
单一性

- 单一算法
- 单一模块

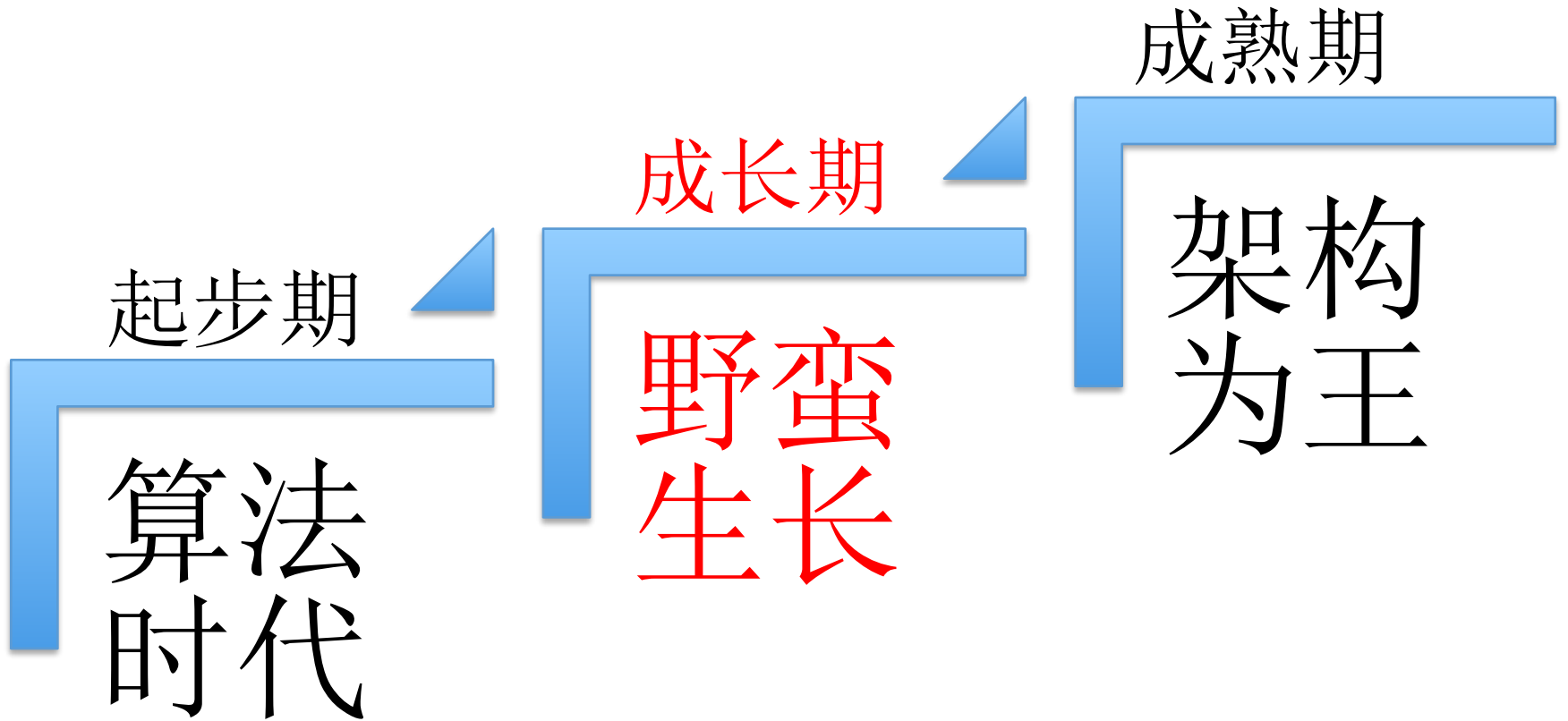
干扰少

- 数据干扰少
- 业务干扰少

算法时代——架构示例图



发展历程



野蛮生长——动机和诉求

动机



正向

- 效果好=>推广



负向

- 更高要求=>改进
- 数据复杂化
 - 作弊、刷单、促销

诉求

模块

- 模块逻辑清晰、独立
- 模块多样化

算法

- 增加算法
- 细致调优
- 数据预处理

野蛮生长——特点和问题

特点

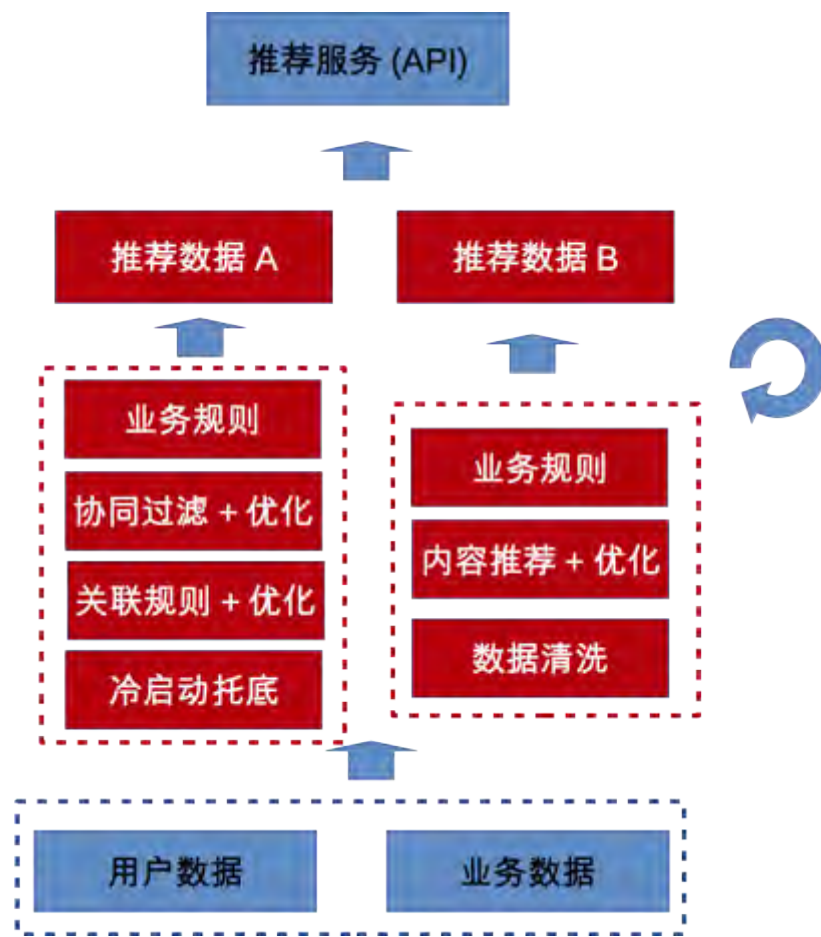
- 业务诉求快速得到满足
- 系统开始变得定制化
- 改进多发于算法程序本身，而非独立处理。

问题

- 代码开始膨胀
- 算法逻辑开始混乱
- 业务逻辑和算法开始混杂
- 不同算法各自为政

逐步发展、潜移默化、长期存在

野蛮生长——架构示例图



发展历程



**“架构是本无所谓有，
无所谓无的，踩得坑多
了，也便有了架构。”**

——尼古拉斯·迅哥

“An expert is a person who has made all the mistakes that can be made in a very narrow field.”

——Niels Bohr

野蛮生长的代价

运行链条长

结构不统一

血统难追溯

运行速度变慢

僵尸逻辑

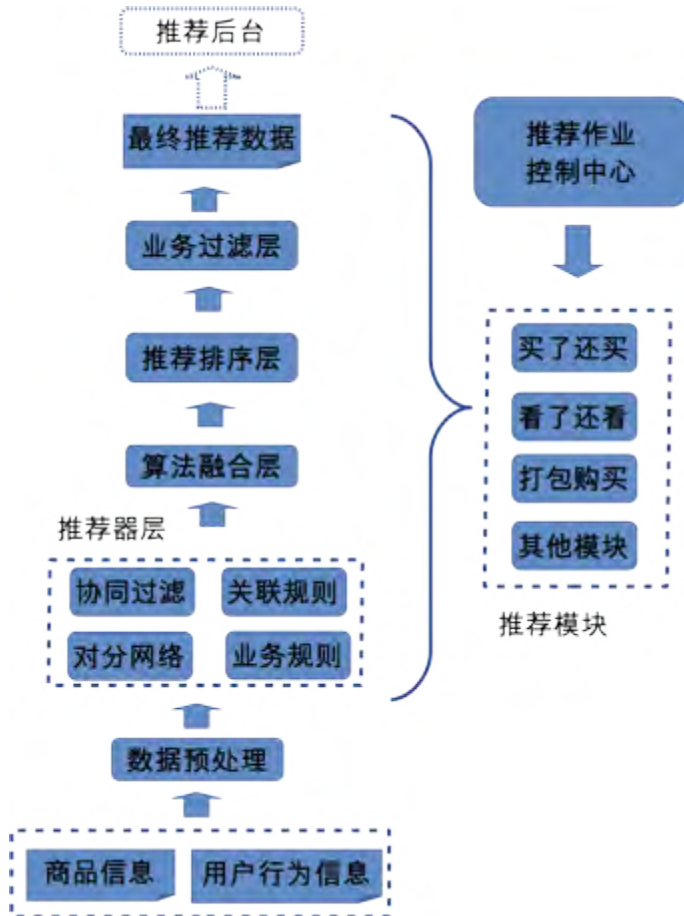
算法实验支持不足

**Fail fast. Fail often.
Fail forward.**

架构设计的原则—哈希表原则



架构为王——架构示例图



- 优化方向
 - 总结=>抽象
 - 扩展=>泛化
- 优化收益：
 - 运行链条缩短+统一
 - 内部排序+融合排序+统一排序
 - 融合层标记血统
 - 针对性性能优化

经验&教训

- 推荐系统：算法 -> 调优 -> 多算法 -> 架构
- 架构是水到渠成，不是杞人忧天。
- 避免拿来主义，探索自身特点。
- 不要怕混乱，混乱代表空间。
- 快速生长，逐步重构。



TIANKO

SequeMedia
威客传媒

IT168

ChinaUnix

ITPUB