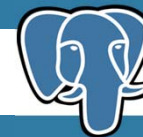


VITESSE DATA

DeepGreen DB: 性能优化、开发方向

CK Tan

Vitesse Data, Inc.













创始人

- CK Tan
 - 田丰
-
- 美国威斯康星大学硕士 / 博士、数据库系
 - 原 Greenplum Database 团队 2006-2009
 - gpfdist, external table, column store, executor opt, storage opt, gpmon, hashagg opt, etc.



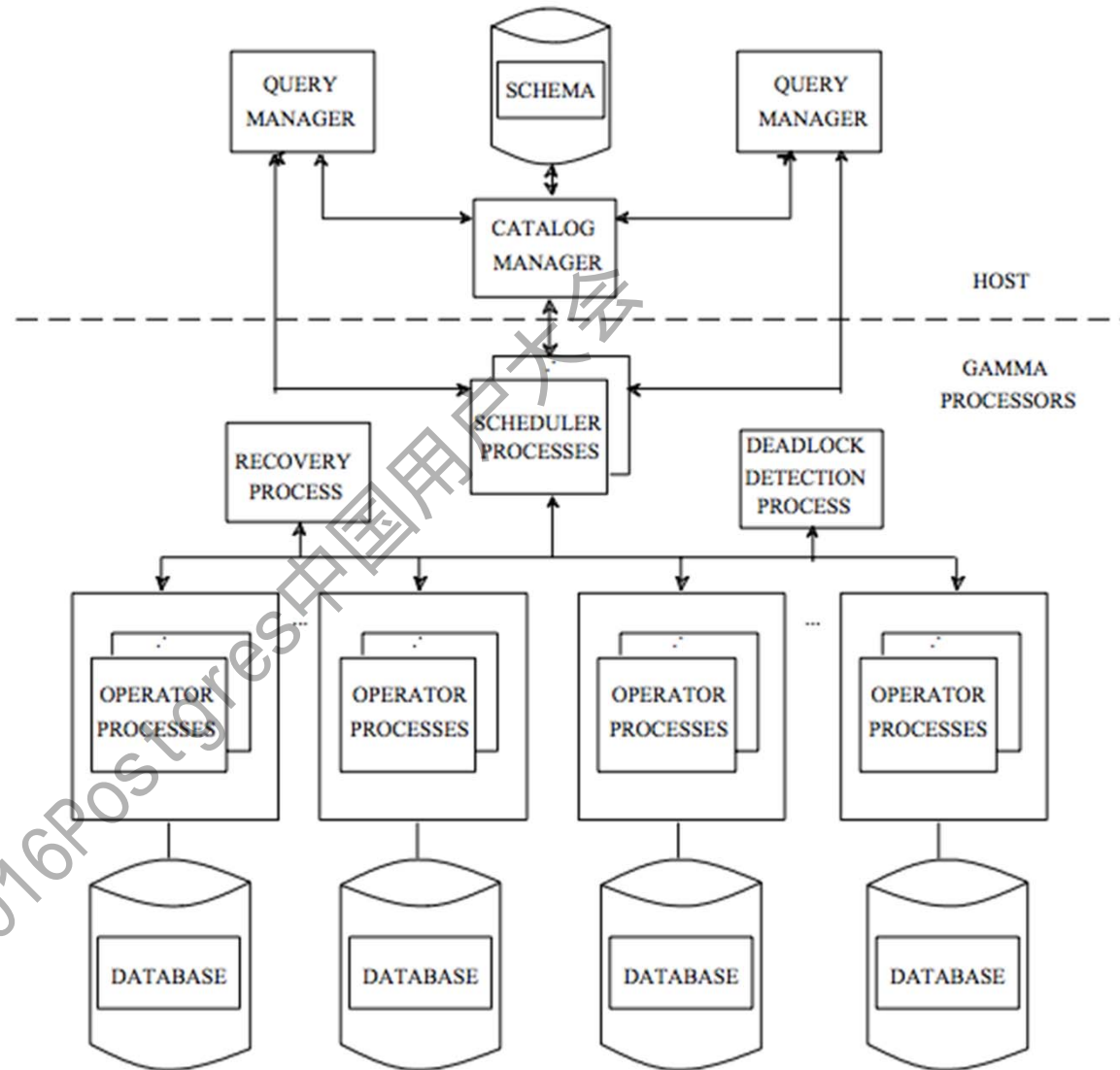
MPP 起源：威斯康星大学

- Gamma Database Machine, 1985-1990
- 20 VAX 11/750
- 32 intel iPSC/2



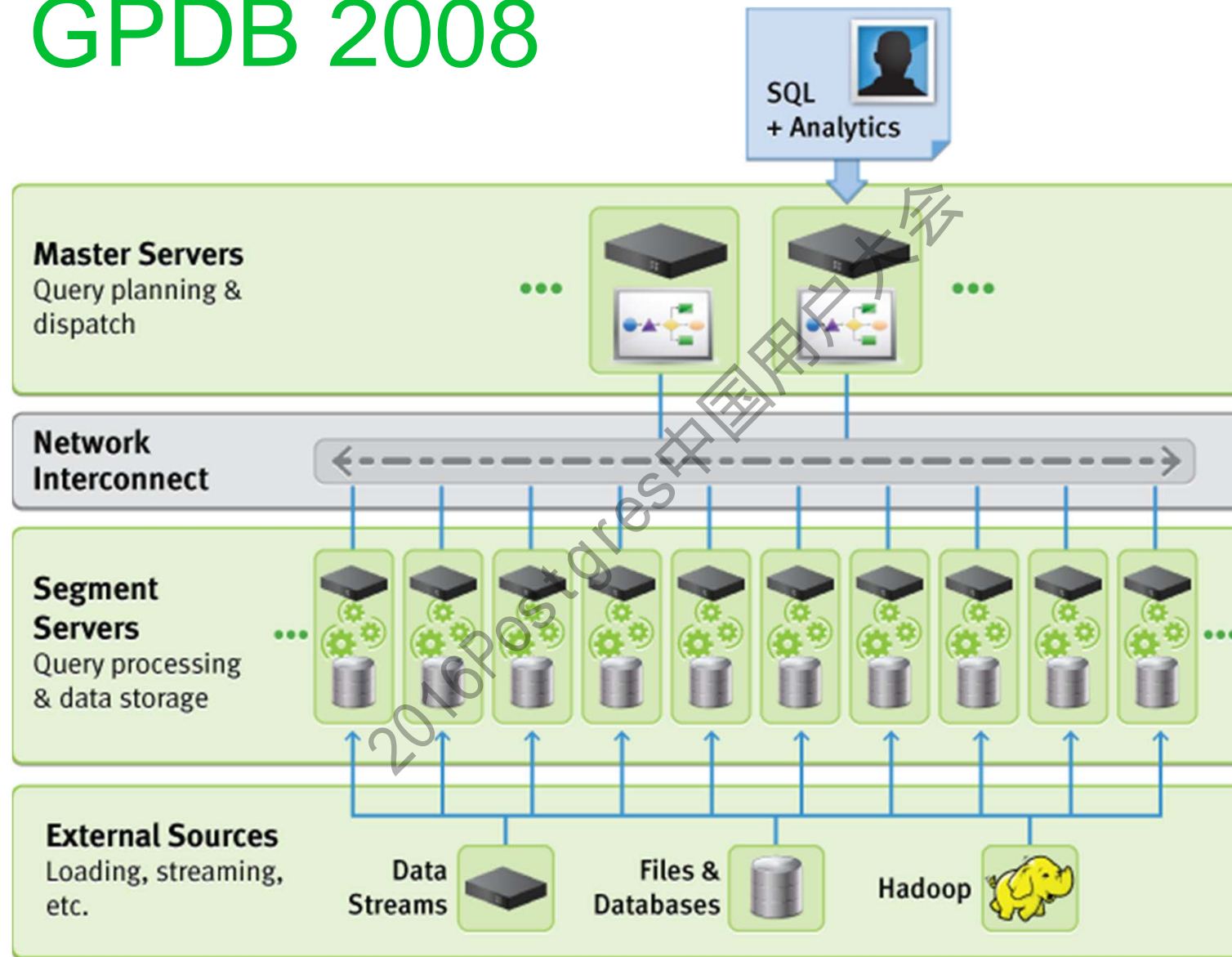


Gamma 1990

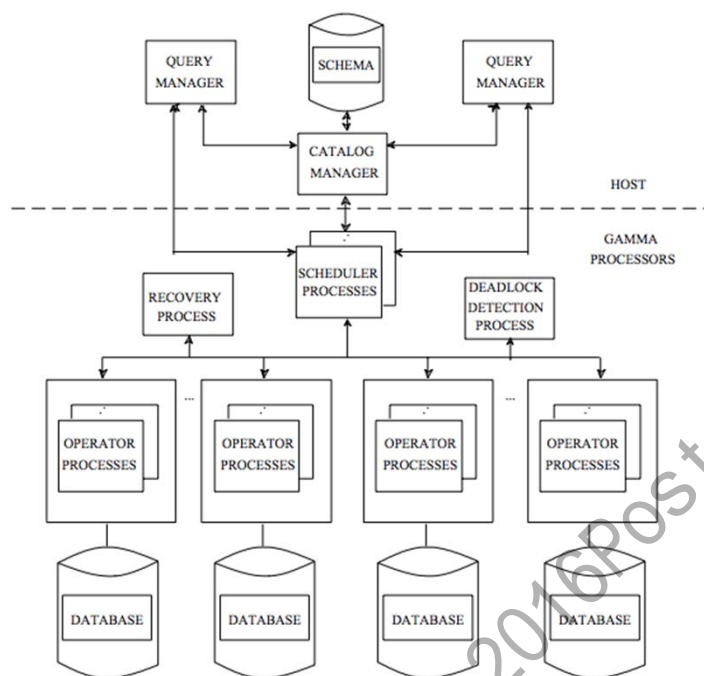


Gamma Process Structure
Figure 2

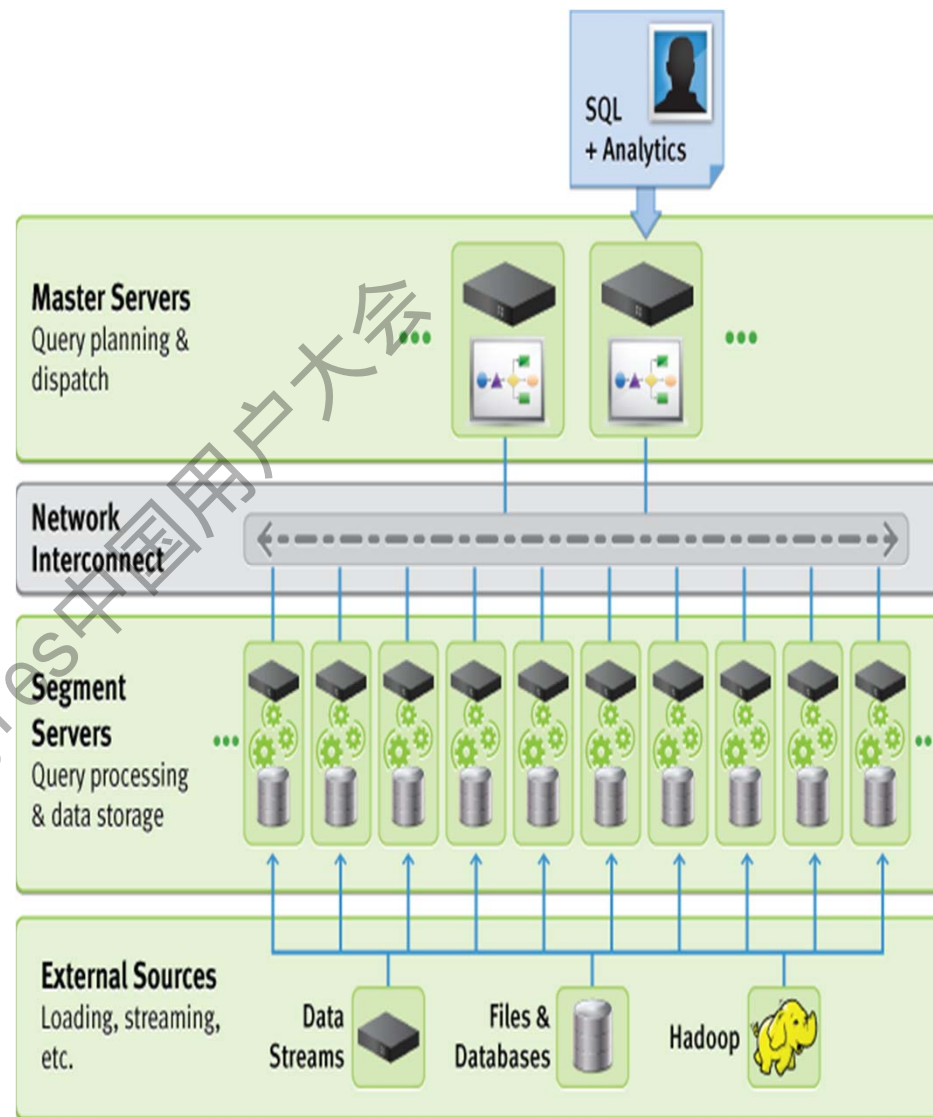
GPDB 2008



万变不离其宗



Gamma Process Structure
Figure 2



Vitesse Data 简介

- 成立于 9月2014
- 产品
 - vitesse db 9.3, 9.4, 9.5
 - deepgreen db, loft, xdrive
- 产品发布: 30+ 次



性能优化

- LLVM JIT
- Hash table
- Spill Framework
- Planner
- CSV Parser - SIMD
- lz4, zstd 压缩
- approx count distinct

LLVM JIT 黑技术



此 JIT 非彼 JIT

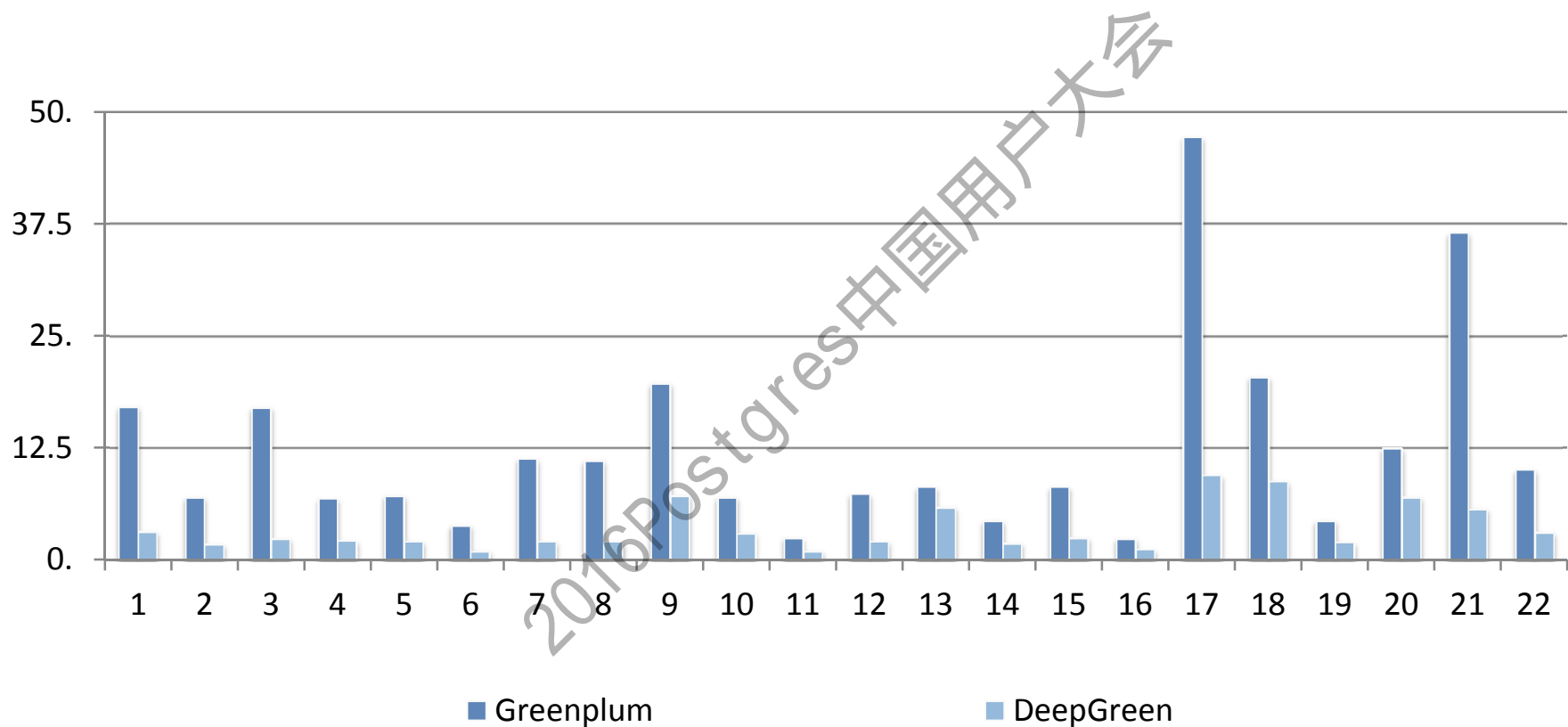
Just-in-time compilation of query plan

- 将整个查询转换成一个汇编语言程式
- 有效去除 x86 执行器与内存的摩擦

LLVM 只是工具。用法不同，效果各异。

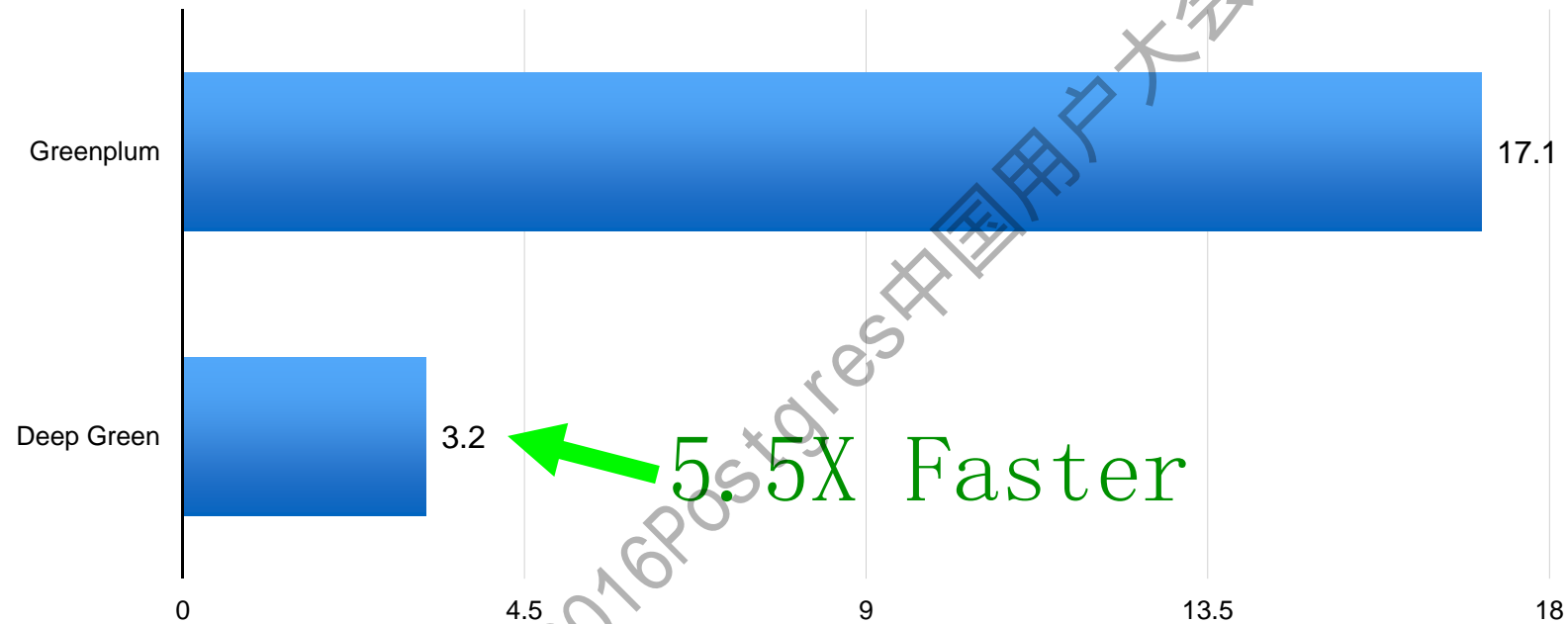
shipping since ... 3/2015

TPCH 10g



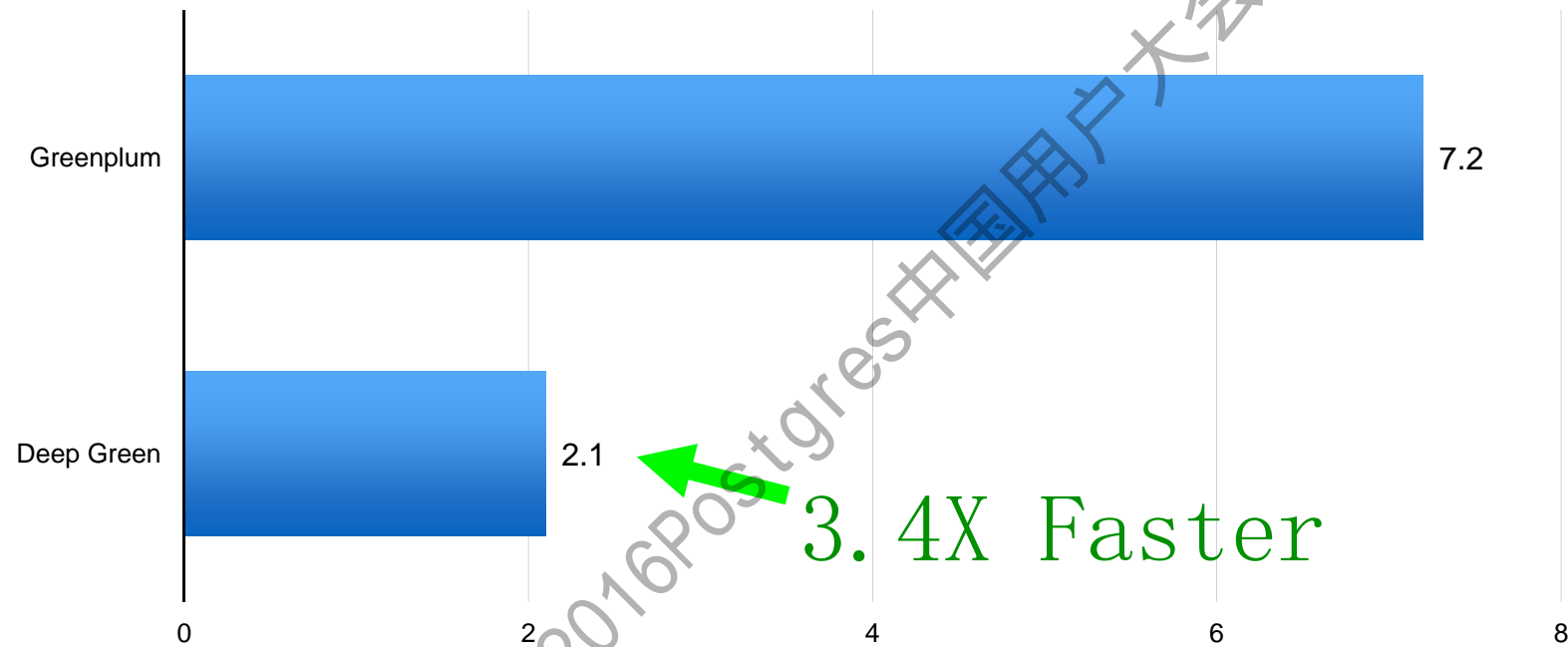
TPCH Q1

Q1 runtime in seconds (lower is better)

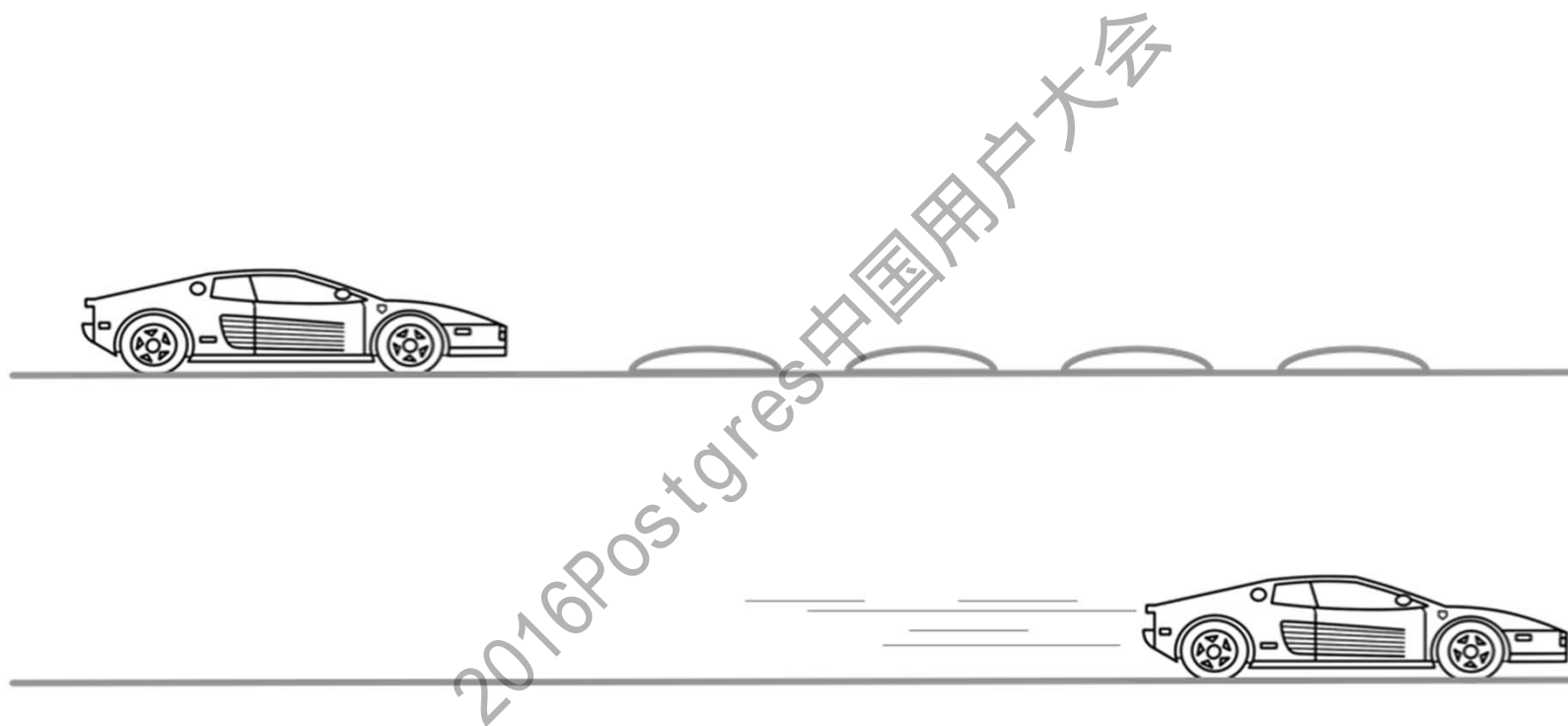


TPCH Q5

Q5 runtime in seconds (lower is better)



请别如此对待您的x86



CSV 解析器

- 已完成 SIMD: parse 8 bytes at a time.
 - 比 GPDB CSV parser 大约快 2 倍。
- 开发中 AVX-512: parse 64 bytes at a time
 - 估计可比现在加速至少 2 倍。

压缩器：lz4, zstd, zlib

	压缩率	压缩时间	解压缩时间
memcpy	1.00	4200 MB/s	4200 MB/s
lz4	1.61	690 MB/s	2220 MB/s
zstd	2.88	240 MB/s	620 MB/s
zlib -1	2.73	59 MB/s	250 MB/s

2016Postgres中国用户大会

10倍 (lz4 vs memcpy)

10倍 (lz4 vs zstd)

4倍 (zstd vs memcpy)

2倍 (zstd vs zlib)

Approximate count distinct

```
select count(distinct url) from page_view;
```

- 必须记住每一个URL – 非常耗内存。

```
select approximate_count_distinct(url) from  
page_view;
```

- hyper log log algorithm
- 1% 错误，但快 3 倍。

XDRIVE

2016Postgres中国用户大会

xdrive

理解

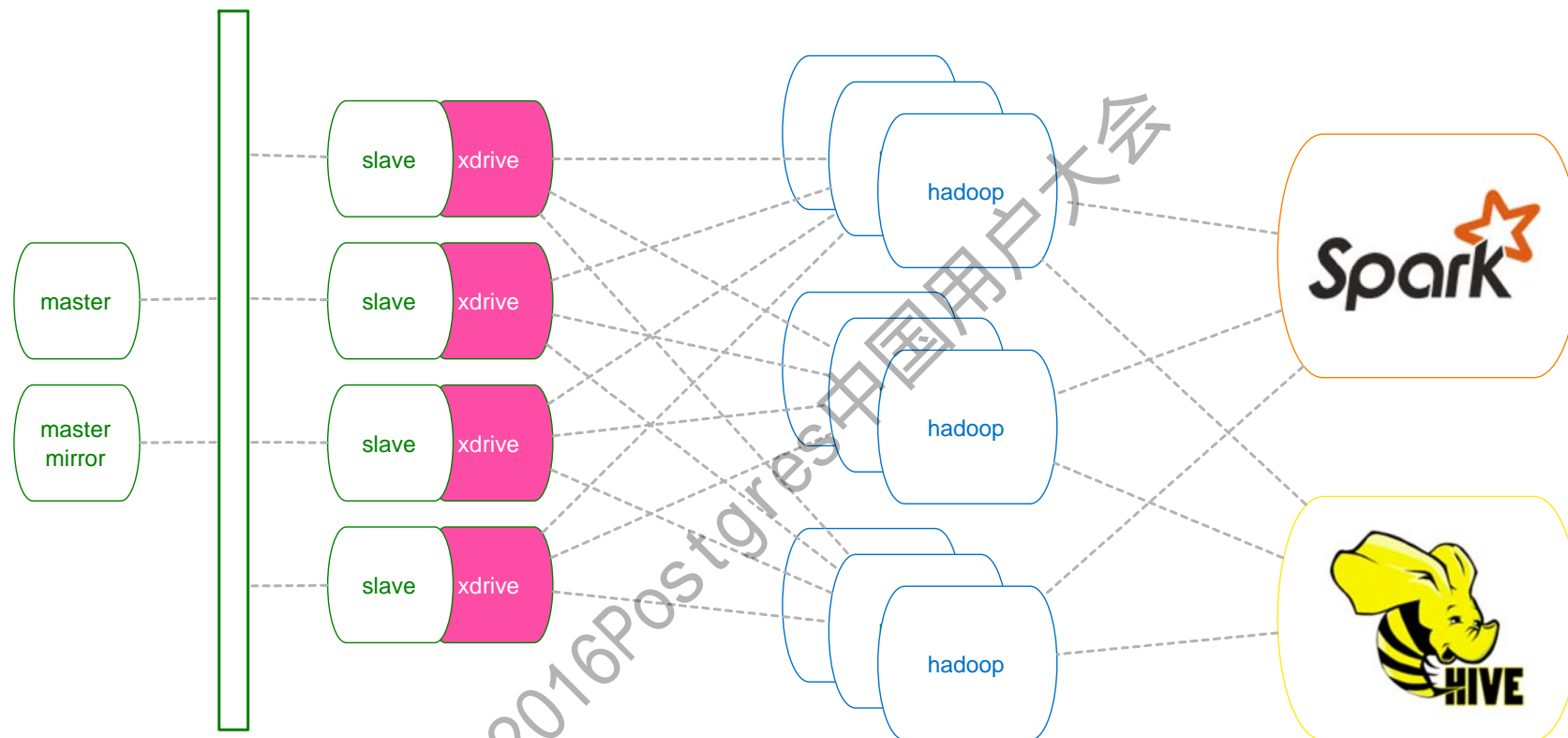
- gpfdist for Hadoop
- DeepGreen DB 上加了 HAWQ 功能

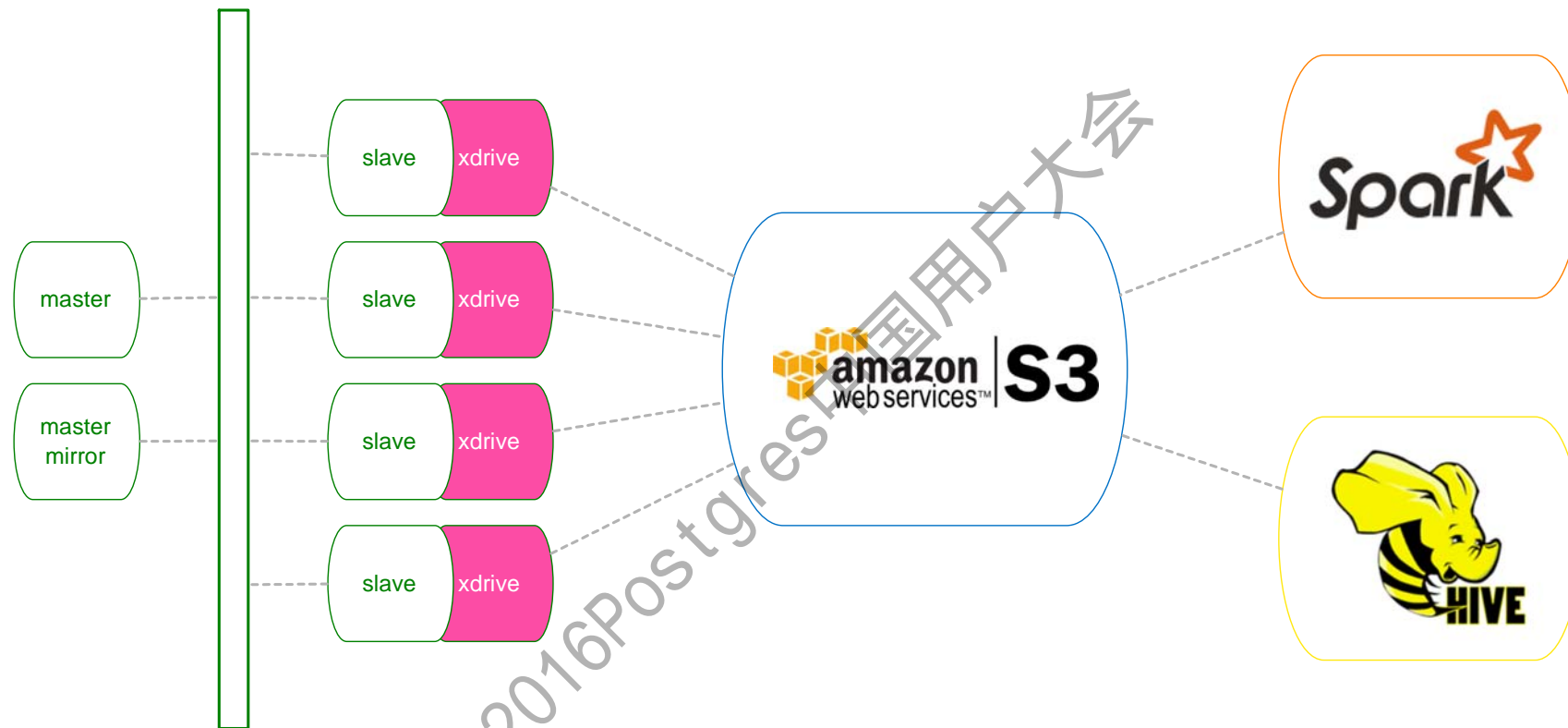
高扩展性

- HDFS, S3, Hive
- NFS, Ceph FS
- Local FS

xdrive: 灵活、便捷

- 只需一个配置文件
- 可以在任何地方灵活运行
- 可以连接多个 NFS, Hadoop, S3 等系统
- 可以转换多种文件 csv, parquet, spq, orc
- 嵌入 DeepGreen. 服务 Spark.

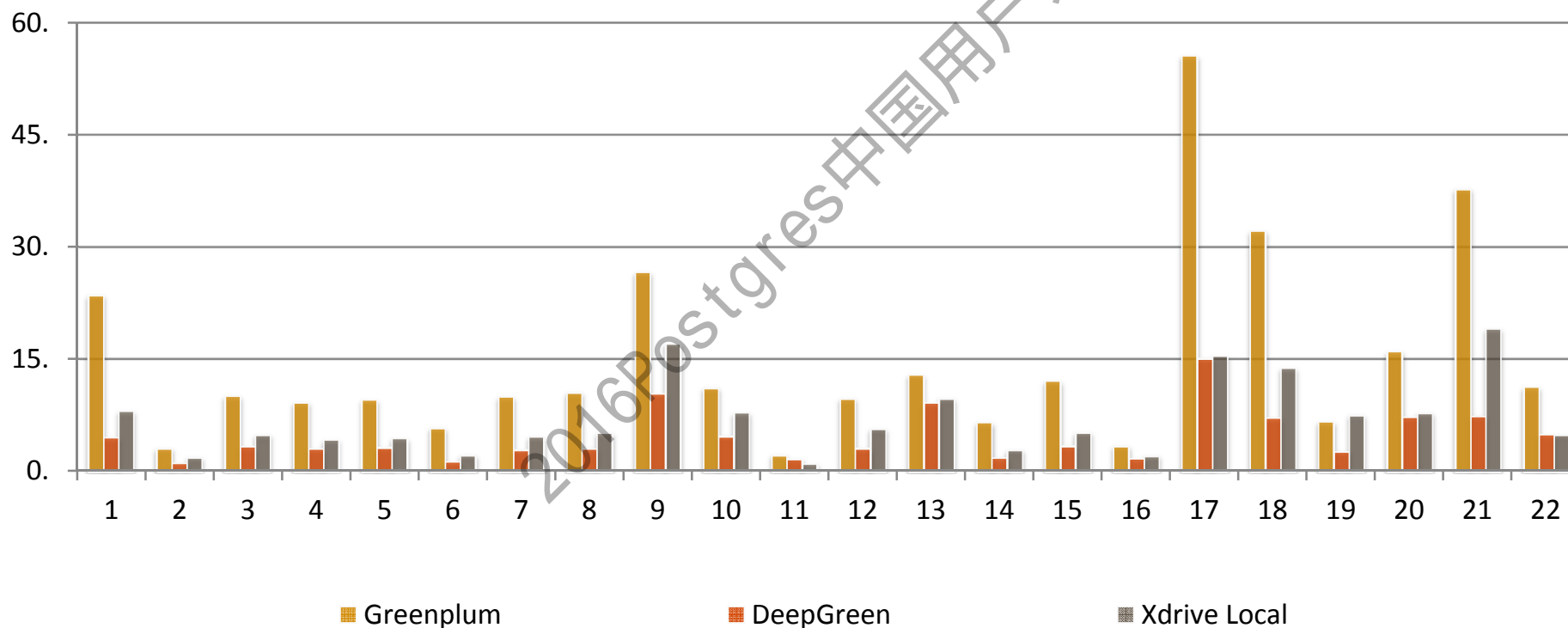




2016Postgres中国用户大会

xdrive: 性能

Greenplum vs DeepGreen vs DeepGreen + XDrive Local



xdrive: 性能 [2]

- xdrive 外表性能远高于 GPDB 外表
- xdrive 外表性能高于 GPDB 内建的 Heap 表
- xdrive 外表性能略低于 DeepGreen 内建的 Heap 表
 - 主要是数据从 HDFS 读出

xdrive: 双活

企业里的两派人马

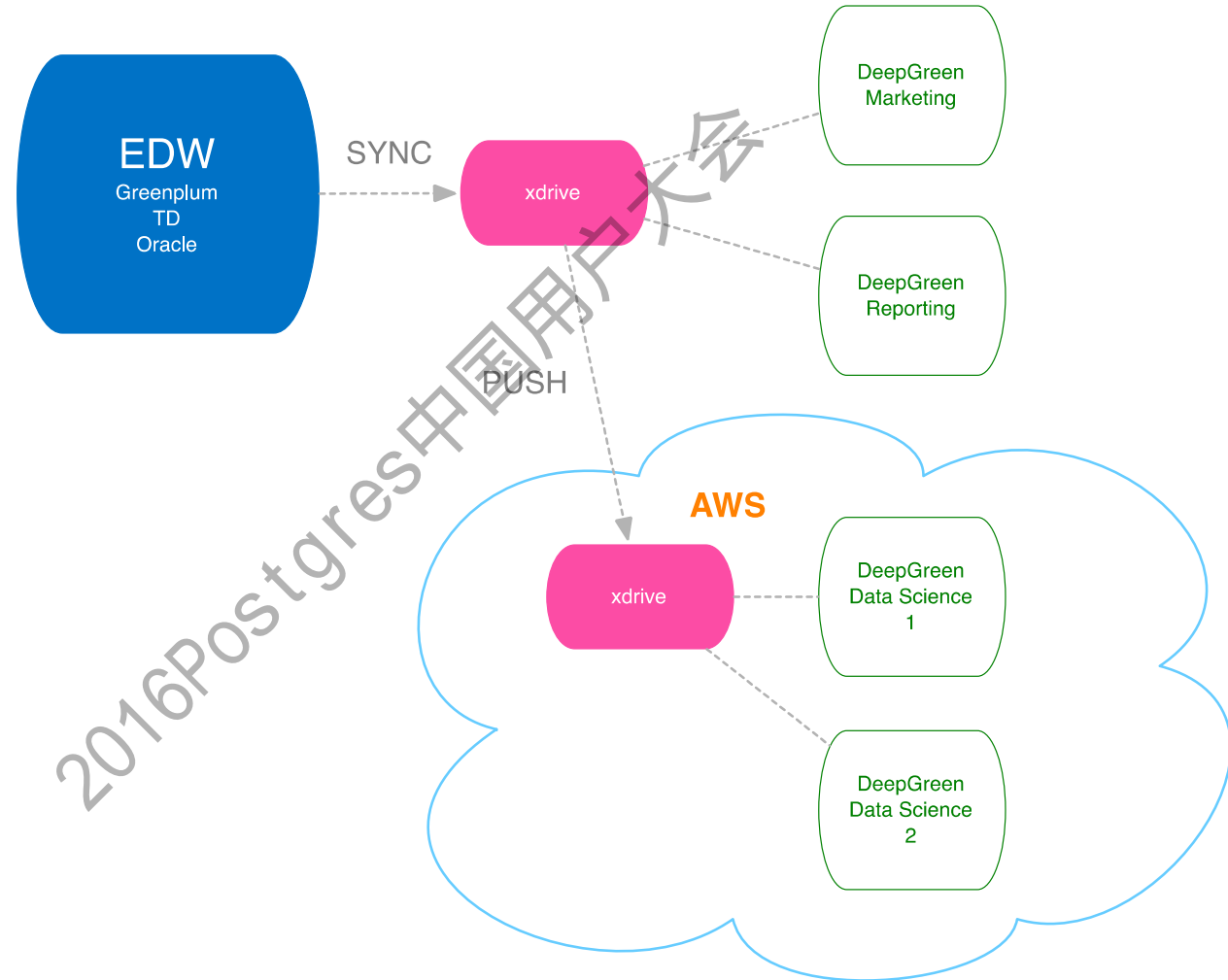
- SQL 组
- spark / hive
- 通过 xdrive
 - 共享原数据
 - 共享分析结果

消除对立

xdrive: Data Mart

- EDW 太贵或太忙
- 各个部门有不同的需求
 - 不同数据
 - 不同新鲜度
 - 不同用户群
- 复制数个 data mart 可大量减轻 EDW 负荷

xdrive: Data Mart

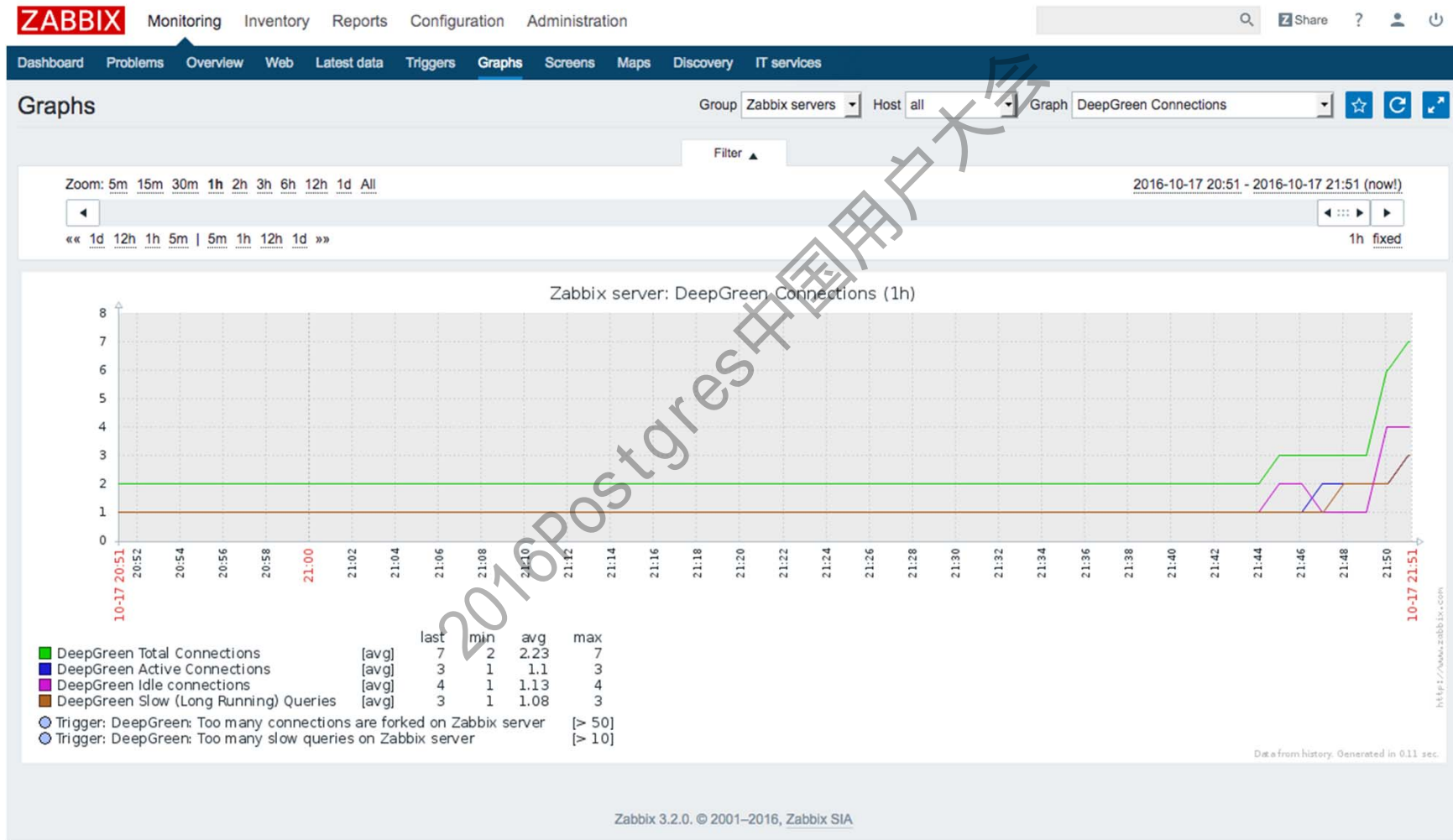


ZABBIX

2016Postgres中国用户大会



DeepGreen DB + Zabbix 监控



DeepGreen 2017 开发方向

- xdrive plugin
- session query monitor
- new utilities with local agent (in GO)
- new interconnect with local hub
- GPU

颠覆性的 PG 9.6

- 一年一版 = GPDB 永远落后 PG 十年
 - 必须考虑非常手段
- multiple backend
 - 足以完全改变 GPDB 的进程架构。
 - 针对 GPDB 的 SLICE，改成 co-backend.
 - 针对 GPDB 的镜子，改成 pg standby.
- 其他

Thanks!

Q & A

2016Postgres中国用户大会