Greenplum DB 5.0 Roadmap

署名：姚延栋
职称：研发总监

Safe Harbor

"Any information regarding pre-release of Pivotal offerings, future updates or other planned modifications is subject to ongoing evaluation by Pivotal and therefore subject to change. The information is provided without warranty of any kind, express or implied. Pivotal has no obligation to update forward looking information in this presentation."

Greenplum is Growing Steadily

- Operating in 34 countries globally
- Pivotal engineering investment growing
- Open source code contribution growing
- Customer count and revenue growing
- 1417 commits to the github repo in 2016
- 8 Greenplum Database releases in 2016

Greenplum DB (GPDB) 5.0

- Greenplum DB: Production Ready Open Source MPP Database
- 5.0: First stable release after open source since 2015/10/27
- Greenplum 5.0 release planned early 2017

Agile Development Methodology

- Test cases
  - ICG inherited from PostgreSQL
  - Cdbfast/TINC
- Pair programming
- Continuous Integration/Continuous Delivery
- Transform ourselves on how to build distributed MPP database

PostgreSQL Based

- Horizontal Merge: 8.2 -> 8.3 -> 8.4 -> 9.x
- Vertical Merge
    - JSON/JSONB
    - Full Text Search
    - UUID Type
    - Raster PostGIS
    - Extension Framework
    - …

ORCA Open Source Optimizer

- First Cost Based Optimizer for BIG data
- Complex workloads for analytics produce large gains with ORCA
- Improved analyze performance on partition
- Parallelizing Union and Union All Queries
- Expanding index support to larger class of predicates
- Reduce optimization time

Query Execution

- LLVM dynamic code gen for faster query execution
- Dispatcher for performance and scalability
- Resource management
- Catalog data caching in the optimizer to speed short running queries

Storage & Backup/Restore

- Protection for data and performance improvement for backup/restore
- PostgreSQL WAL replication for segment mirroring
- Distributed Snapshot
- Fine grain incremental backup

External Tables

- S3
- Azure
- GPHDFS
- FDW

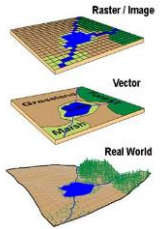
Greenplum Geospatial

Current Key Features:
- Geometry
- Geography
- Raster

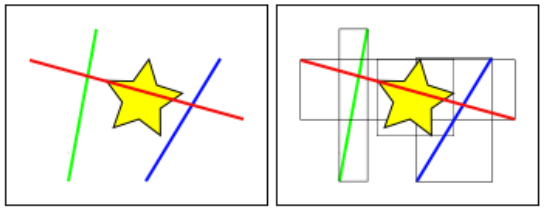
Ability to store geospatial data and query with with joins and operators

```
geodemo=# SELECT
nyc_subway_stations.long_name AS subway,
nyc_neighborhoods.name AS neighborhood
FROM nyc_neighborhoods
JOIN nyc_subway_stations
ON ST_Contains(nyc_neighborhoods.geom, nyc_subway_stations.geom)
WHERE nyc_neighborhoods.name = 'Greenwich Village';
                    subway                     |   neighborhood
-----------------------------------------------+------------------
 W 4th St (B,D,F,V) Manhattan                  | Greenwich Village
 14th St / Union Sq (4,5,6) Manhattan          | Greenwich Village
 14th St (1,2,3) Manhattan                     | Greenwich Village
 Bleecker St / Broadway-Lafayette St (6) Manhattan | Greenwich Village
 Christopher St / Sheridan Sq (1) Manhattan    | Greenwich Village
 Union Sq / 14th St (L,N,Q,R,W) Manhattan      | Greenwich Village
 6th Ave / 14th St (F,L,V) Manhattan           | Greenwich Village
 8th St / New York University (N,R,W) Manhattan | Greenwich Village
 Astor Pl (6) Manhattan                        | Greenwich Village
 W 4th St (A,C,E) Manhattan                    | Greenwich Village
(10 rows)
```

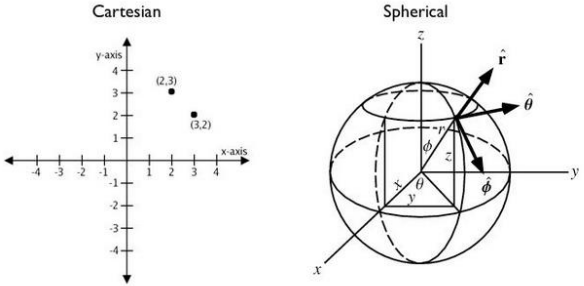
Raster Image Processing



Spatial Indexes & Bounding Boxes



Round earth calculations



Madlib: Scalable, in-Database Machine Learning

- Open source: http://madlib.incubator.apache.org/
- Supports PostgreSQL, Greenplum Database, Apache HAWQ
- Powerful machine learning, statistics and analytics for DS
  - Statistics
  - Supervised/Unsupervised Learning
  - Time Series Analysis
  - Text Analysis

Workload Management (WLM) (proprietary)

- Rule based query management to monitor/manager queries and resource queues
- More rules and actions

Command Center (GPCC) (proprietary)

- Redesign and reimplement of GPCC
- Integration with WLM

GPText: Full Text Search and Text Analysis (proprietary)

- Integrate GPDB with Solr Cloud
- Use SQL to perform full text search and text analysis
- Operation tools: recover, backup, restore, expand etc.

PL/Container (proprietary)

- Docker based containers
- Secure PL/Python and PL/R
- Flexible 3rd party libraries policy

G2C (Greenplum Gemfire Connector) (proprietary)

- Gemfire talk to GPDB
- GPDB talk to Gemfire
- Predicate pushdown and filtering
- Auto Synchronizer between GPDB and Gemfire

Greenplum DB Open Source Community

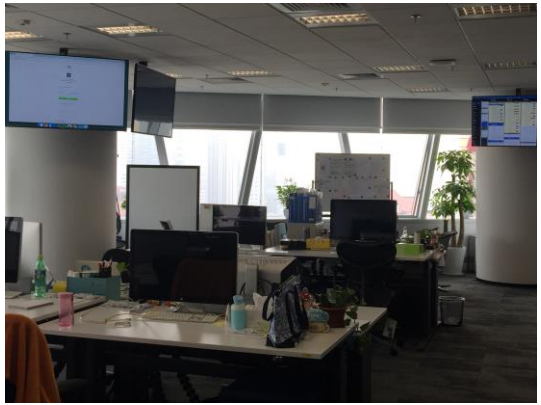
- Since open source from 2015/10/27
  - GPDB: 1630 stars, 458 fork, 300 watch
  - PostgreSQL: 2366 stars, 764 fork, 261 watch
- User groups
  - Wechat: 431
  - QQ: 131
- Greenplum DB Meetup at Beijing at 2016/11

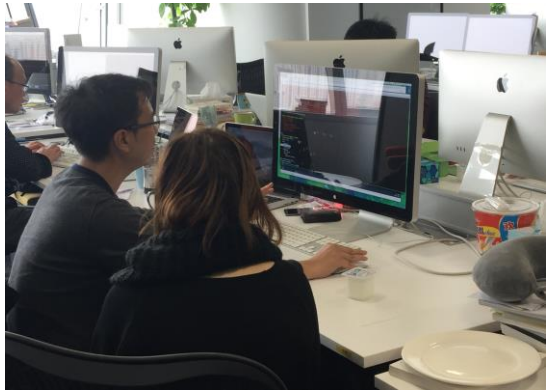
Contributions

- Pull Request (PR): 33 open, 942 closed within one year
- External Contributions from engineers of various companies: Aliyun, China Mobile, Inspur, Epam, …
  - More than 50% external contributions are from China

We are hiring

- MPP database kernel developers
- YOU will shape distributed MPP database technology trends





2016 The Computing Conference

THANKS

2016 杭州·云栖大会
THE COMPUTING CONFERENCE

扫码观看大会视频