



Расширяемость PostgreSQL и некоторые приложения

Олег Бартунов

oleg@sai.msu.su

Информационные технологии в Астрономии, ГАИШ-МГУ



Agenda

- Расширяемость БД
- Что такое PostgreSQL
- Расширяемость PostgreSQL
 - Btree
 - GiST — обобщенное поисковое дерево
 - GIN — обобщенный обратный индекс
- Приложения
 - Виртуальная Обсерватория ГАИШ
 - Полнотекстовый поиск в PostgreSQL



Расширяемость БД

- IT-цикл жизнедеятельности проекта:
 - Проектирование
 - Разработка
 - Поддержка
 - Следить за производительностью системы — правило нескольких секунд (конкуренты !)
 - Автоматические агенты
 - Увеличение информации, посещаемости
 - Развитие
 - Новые идеи, новые проекты, увеличение нагрузки



Расширяемость БД

- Расширяемость БД — важнейший фактор !
- **Новая функциональность:**
 - Новые типы данных (IP, ISSN...)
 - Новые запросы (похожие изображения)
 - Сохранение производительности и надежности
- **Масштабируемость**
 - Вертикальная, горизонтальная
- **Новая функциональность и масштабируемость связаны**



Расширяемость БД

- **Пример:**

{Фото}, {Пользователь}, {Тэги}

- **Соединения таблиц не масштабируются горизонтально**
- **Для этого требуются шарды (shards)**
- **Использование массивов помогает**
 - Требуется индексная поддержка для эффективной работы



Что такое PostgreSQL ?

PostgreSQL - это свободно распространяемая объектно-реляционная система управления базами данных (ORDBMS), наиболее развитая из открытых СУБД в мире и являющаяся реальной альтернативой коммерческим базам данных.

Произношение: **post-gress-Q-L, post-gres, пост-гресс, pgsql** (пэ-жэ-эс-ку-эль)

Web: **<http://www.postgresql.org>**

Лицензия: **BSD**



CODD 1969,1970
Relational model
1974-1975

**IBM
System R**

1973 **QUEL**

**UC Berkeley
INGRES**

Interactive Graphics REtrieval System

1979-1982
1994
2004
Ingres Co
**CA
Ingres r3**
\$\$\$

1985-1988
Postgres

1989 V1
1993 V4.2
1994 V5
**Postgres95
SQL**



1997-04-03
Agatha Christie

1996 V6
2005 V8
PostgreSQL

2008 8.3

1978
1979 **SQL**

SDL, RSI 1979

1981
SQL/DS

1983
DB2
\$\$\$

1983
ORACLE
1984 \$\$\$
RDb/VMS

2001
\$\$\$
INFORMIX
\$\$\$

1984
1995
1992
ILLUSTR
\$\$\$

1984-1987
NonStop SQL

1987
1992
1995
**Sybase
ASE**
\$\$\$

1993
MS SQL
\$\$\$

**TelegraphCQ
StreamBase**
\$\$\$

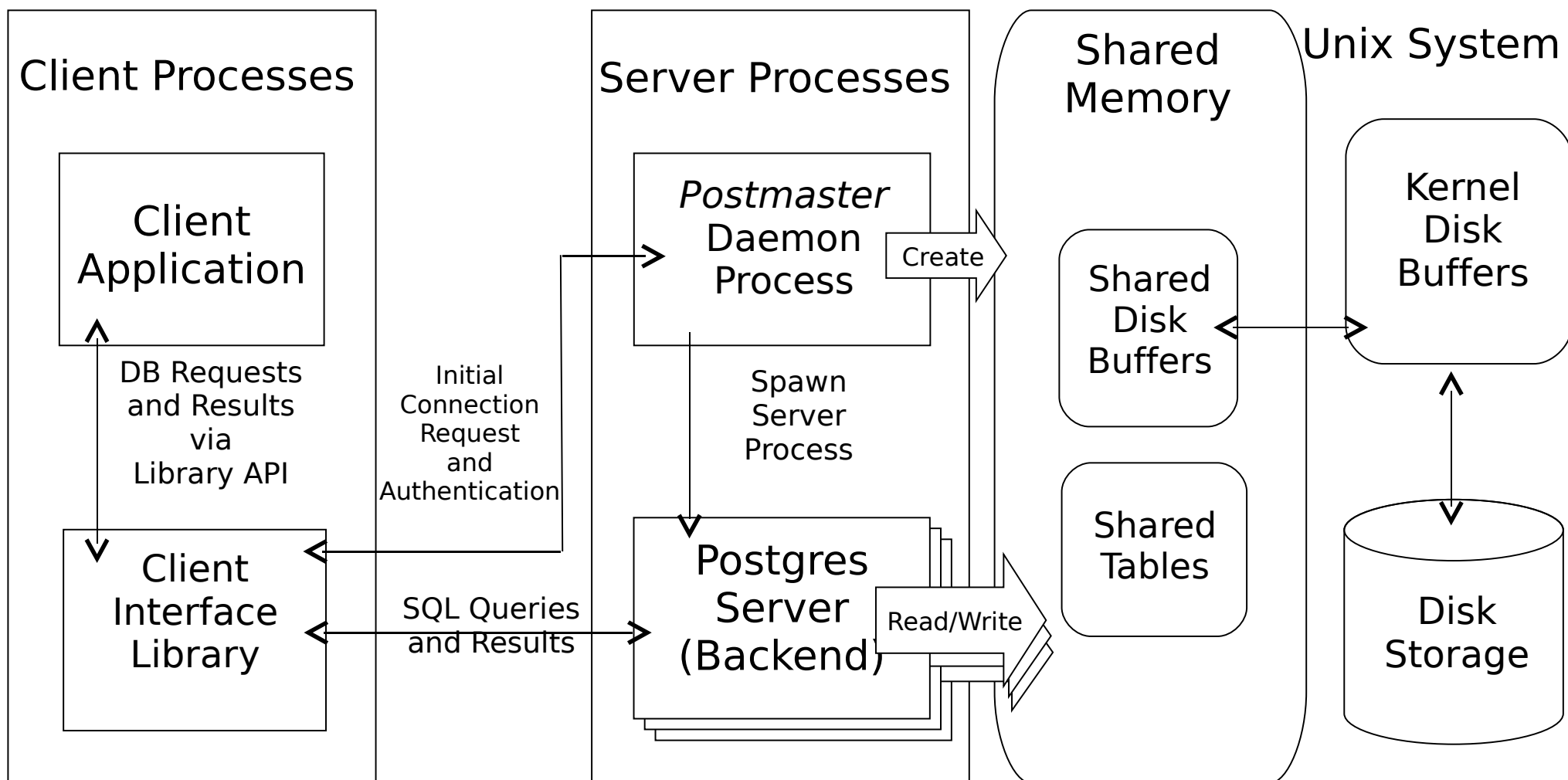
**C-Store
Vertica**
\$\$\$



PostgresPlus (EnterpriseDB)
Bisgres (GreenPlum)
Everest (Yahoo)



Что такое PostgreSQL: Architecture





Что такое PostgreSQL: Особенности

- Высокая степень параллелизма - MVCC
- Расширяемость
 - Типы данных, функции, агрегаты, операторы
 - Языки (sql, pl/pgsql, pl/perl, pl/tcl, pl/R, pl/java ...)
 - Индексы (Btree, GiST, GIN)
- Cost-based оптимизатор
- Хорошее соответствие ISO/ANSI SQL 92,99,2003
- Открытый код (BSD), открытая модель развития — нет владельца !



Название	ASE	DB2	FireBird	InterBase	MS SQL	MySQL	Oracle	PostgreSQL
Лицензия	\$\$\$	\$\$\$	IPL ²	\$\$\$	\$\$\$	GPL/\$\$\$	\$\$\$	BSD
ACID	Yes	Yes	Yes	Yes	Yes	Depends ¹	Yes	Yes
Referential integrity	Yes	Yes	Yes	Yes	Yes	Depends ¹	Yes	Yes
Transaction	Yes	Yes	Yes	Yes	Yes	Depends ¹	Yes	Yes
Unicode	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Schema	Yes	Yes	Yes	Yes	No ⁵	No	Yes	Yes
Temporary table	No	Yes	No	Yes	Yes	Yes	Yes	Yes
View	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Materialized view	No	Yes	No	No	No	No	Yes	No ³
Expression index	No	No	No	No	No	No	Yes	Yes
Partial index	No	No	No	No	No	No	Yes	Yes
Inverted index	No	No	No	No	No	Yes	Yes	Yes ⁶
Bitmap index	No	Yes	No	No	No	No	Yes	No
Domain	No	No	Yes	Yes	No	No	Yes	Yes
Cursor	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
User Defined Functions	Yes	Yes	Yes	Yes	Yes	No ⁴	Yes	Yes
Trigger	Yes	Yes	Yes	Yes	Yes	No ⁴	Yes	Yes
Stored procedure	Yes	Yes	Yes	Yes	Yes	No ⁴	Yes	Yes
Tablespace	Yes	Yes	No	?	No ⁵	No ¹	Yes	Yes
Название	ASE	DB2	FireBird	InterBase	MS SQL	MySQL	Oracle	PostgreSQL

чания:

- ◆ 1 - для поддержки транзакций и ссылочной целостности требуется InnoDB (не является типом таблицы по умолчанию)
- ◆ 2 - Interbase Public License
- ◆ 3 - Materialized view (обновляемые представления) могут быть эмулированы на PL/pgSQL
- ◆ 4 - только в MySQL 5.0, которая является экспериментальной версией
- ◆ 5 - только в MS SQL Server 2005 (Yukon)
- ◆ 6 - GIN (Generalized Inverted Index) с версии 8.2



Что такое PostgreSQL: Limitations

- Максимальный размер БД – unlimited
- Максимальный размер таблицы – 32Тб
- Максимальная длина записи – 400Gb
- Максимальная длина атрибута – 1 Gb
- Максимальное кол-во записей – unlimited
- Максимальное кол-во атрибутов – 250-1600
- Максимальное кол-во индексов - unlimited



Что такое PostgreSQL

- Поддержка:
 - Сообщество — мэйлинг лист
 - EnterpriseDB
 - SUN Microsystem (Solaris 10)
 - Много мелких компаний
- Более подробно о PostgreSQL можно прочитать в
http://www.sai.msu.su/~megera/postgres/talks/what_is_postgresql.html



Что такое PostgreSQL: Пользователи

- Skype - масштабируется до миллиарда польз.
- Hi5.com — 60 млн. пользователей, #8 Alexa traffic rank
- NyYearBook.com — 18,000 req/sec, 300 Gb database
- NASA — обработка спутниковых данных (MODIS)
- Yahoo — 2 Pb БД для работы с пользовательскими запросами

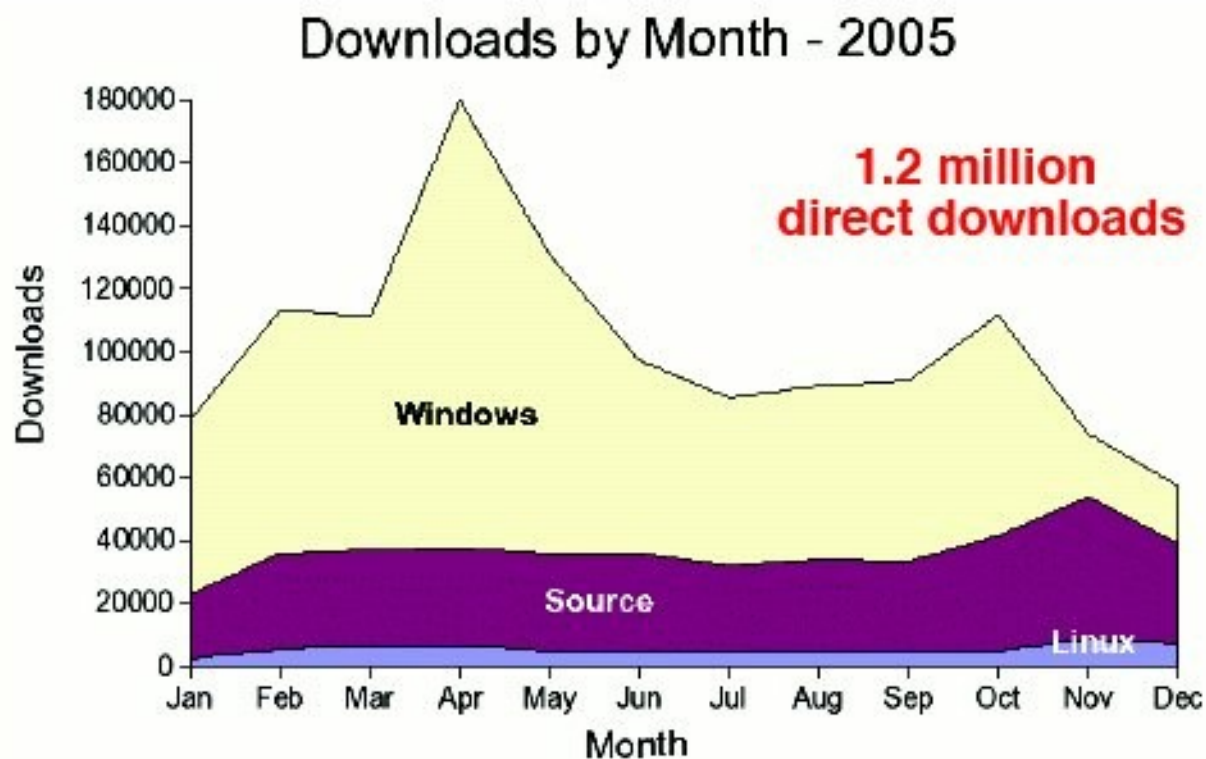


Что такое PostgreSQL: Пользователи

- Рамблер
- 1С:Предприятие
- MirTesen, MoiKrug.ru (Yandex)
- IRR.ru («Из рук в руки»)
- работа.ru, price.ru, РБК, МастерХост, Информ-мобил
- Военные — версия 7.X входит в МСВС



Что такое PostgreSQL





Что такое PostgreSQL: Бенчмарки

PostgreSQL 8.2 – 813.73 JOPS

- SPECjAppServer2004 2x Sun Fire X4200 appservers (8 cores, 4 chips) and 1 Sun Fire T2000 DB server (8 cores, 1 chips) with PostgreSQL 8.2.4
- HW: ~\$60,000; SW: \$0

Oracle 10g – 874.17 JOPS

- SPECjAppServer2004 1 HP rx2660 appserver (4 cores, 2 chips) and 1 rx2660 DB server (4 cores, 2 chips) with Oracle Database 10g Enterprise Edition Release 10.2.0.2
- HW: ~\$74,000; SW: ~\$110,000

Summer 2007: The 1st PostgreSQL Enterprise-level Benchmark (SPEC*)

*) SPEC – Standard Performance Evaluation Corporation, <http://spec.org>



Что такое PostgreSQL: Новости

- EnterpriseDB 2008
 - 10 млн. долларов инвестиций от IBM
 - Убран лозунг совместимости с Оракл
 - Rebranding EnterpriseDB
 - Postgres Plus — A true Enterprise Open Source DB
 - GridSQL — исполнение запросов в Grid среде
 - Postgres Plus Advanced Server — коммерческая версия
 - Cloud Edition (Elastra)
 - Совместимость с Оракл, миграция
 - DynaTune — автоматическая настройка



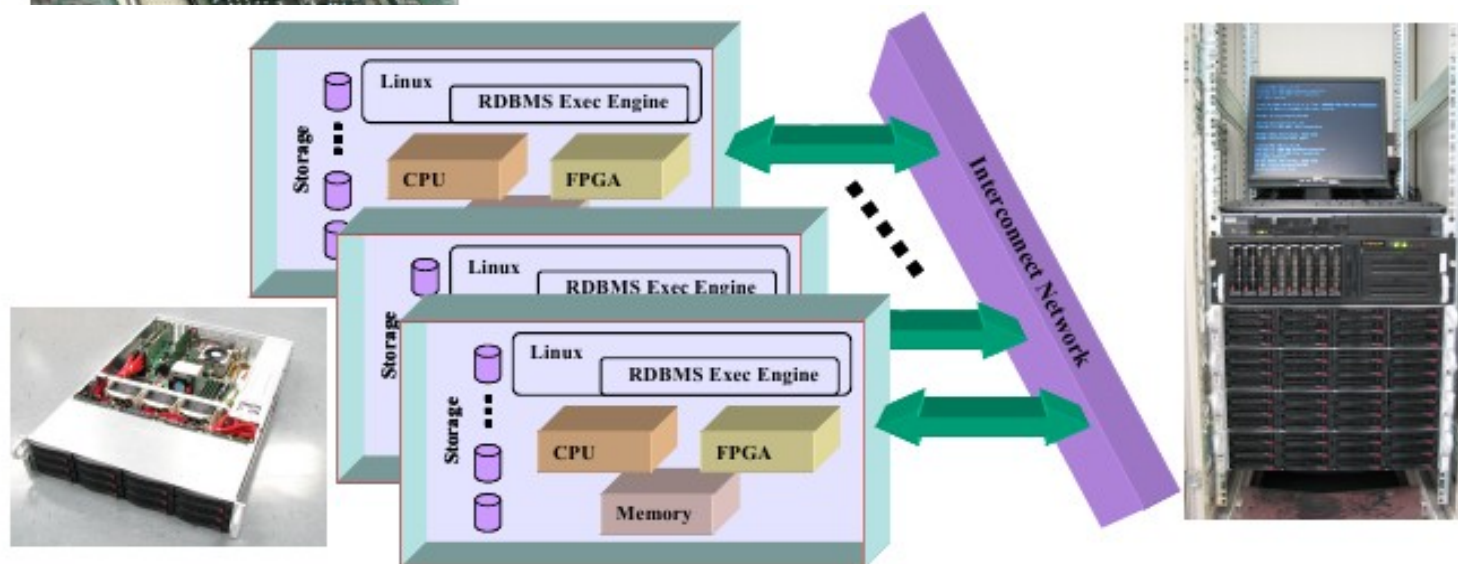
Что такое PostgreSQL: Новости

- XDI Appliance — PostgreSQL with support of shared-nothing parallel cluster environment with FPGA acceleration
 - 1GB/s SQL query per Node (TPC-H)
 - 1TB/min sustained SQL query 19 inch server rack
 - Decision Support Systems, Full table scan, Group By, Order By, Aggregation, multitable joins
- Web: <http://www.xtremedatainc.com/>



Что такое PostgreSQL: Новости

- XDI Appliance — 750,000 USD





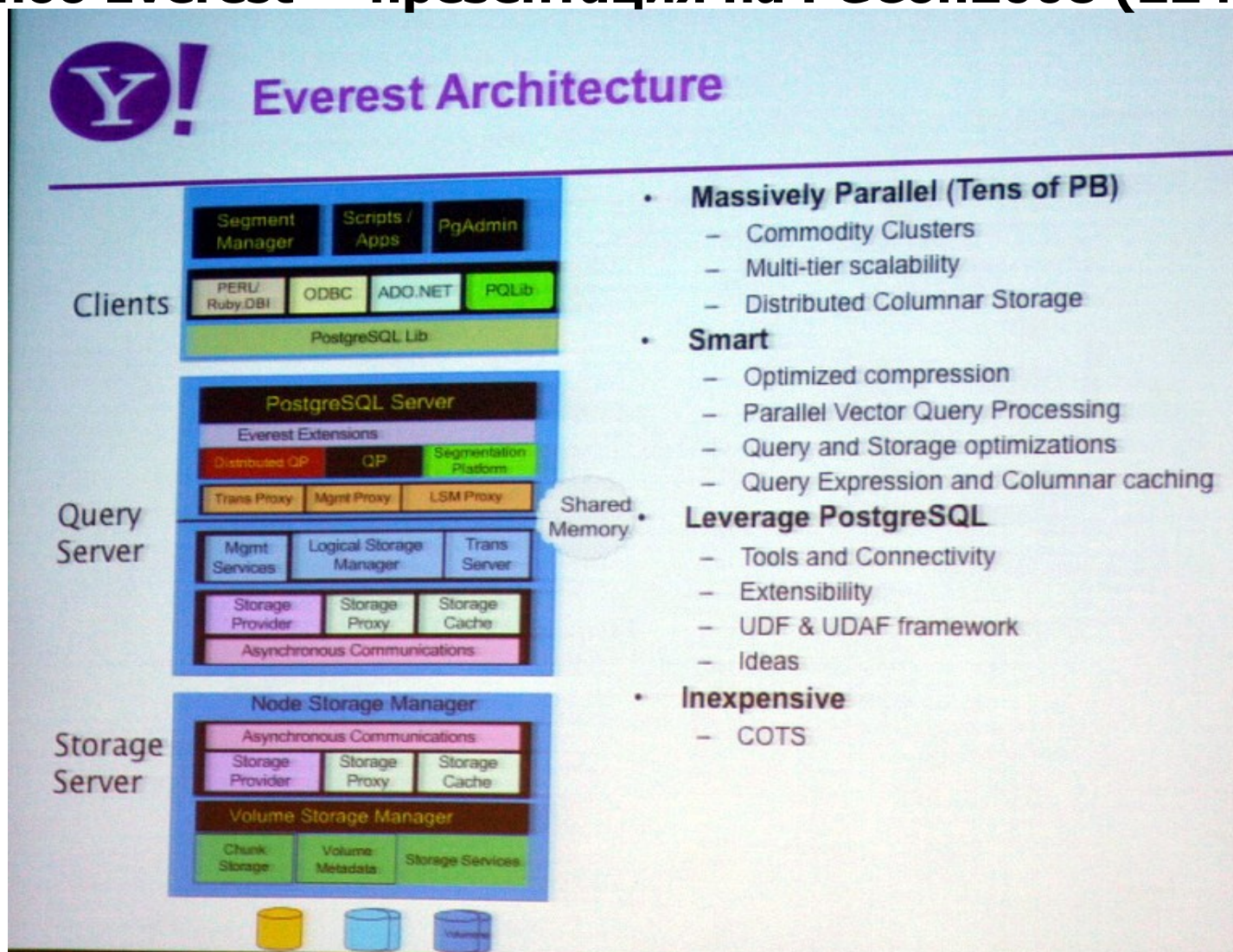
Что такое PostgreSQL: Новости

- Yahoo Inc. - 2 Pb database за год работы
 - Модифицированный PostgreSQL
 - Column-oriented хранилище
 - 24 млрд событий в день
 - Несколько триллионов записей в таблицах
 - 1/2 миллиарда пользователей в месяц
 - Рост до 5 Pb в след. году



Что такое PostgreSQL: Новости

Yahoo Everest — презентация на PGCon2008 (22 мая 2008)





Расширяемость PostgreSQL

- Новые типы данных требует индексные АМ (access methods). Разработка новых АМ, тестирование — трудно и утомительно
- Использовать Btree, Rtree вместо разработки новых АМ (access methods)
Michael Stonebraker, «Inclusion of new types in relational database systems», 1985
 - Повысить уровень абстракции процедур доступа и обновления записей
 - Только фиксированный набор операций (операции сравнения для Btree)



Расширяемость PostgreSQL

- Создание нового типа данных
 - Написать функции ввода/вывода
 - Определить тип (CREATE TYPE)
 - Создать операторы (CREATE OPERATOR)
 - Написать функции сравнения для **Btree** индекса
 - Оператор по-умолчанию для индекса по primary key (CREATE OPERATOR CLASS)
- для НОВЫХ ТИПОВ индекса использовать **GiST** или **GIN**)



Расширяемость PostgreSQL: GiST

- Generalized Search Tree (GiST)

J. M. Hellerstein, J. F. Naughton, and Avi Pfeffer. "Generalized search trees for database systems.", VLDB 21, 1995

- Многие AM можно представить как иерархию предикатов, в которой каждый предикат выполняется для всех подузлов этой иерархии
- Шаблон (template) для реализации новых AM

- GiST предоставляет

- Стандартные методы навигации по дереву
- Обновление дерева
- Конкурентность и восстановление после сбоя



Расширяемость PostgreSQL: GiST

- GiST позволяет реализовать новый АМ эксперту в области данных
- Поддерживает расширяемый набор запросов (в отличие от Btree)
- Новые типы данных обладают производительностью (индексный доступ, конкурентность) и надежностью (протокол логирования), как и встроенные типы



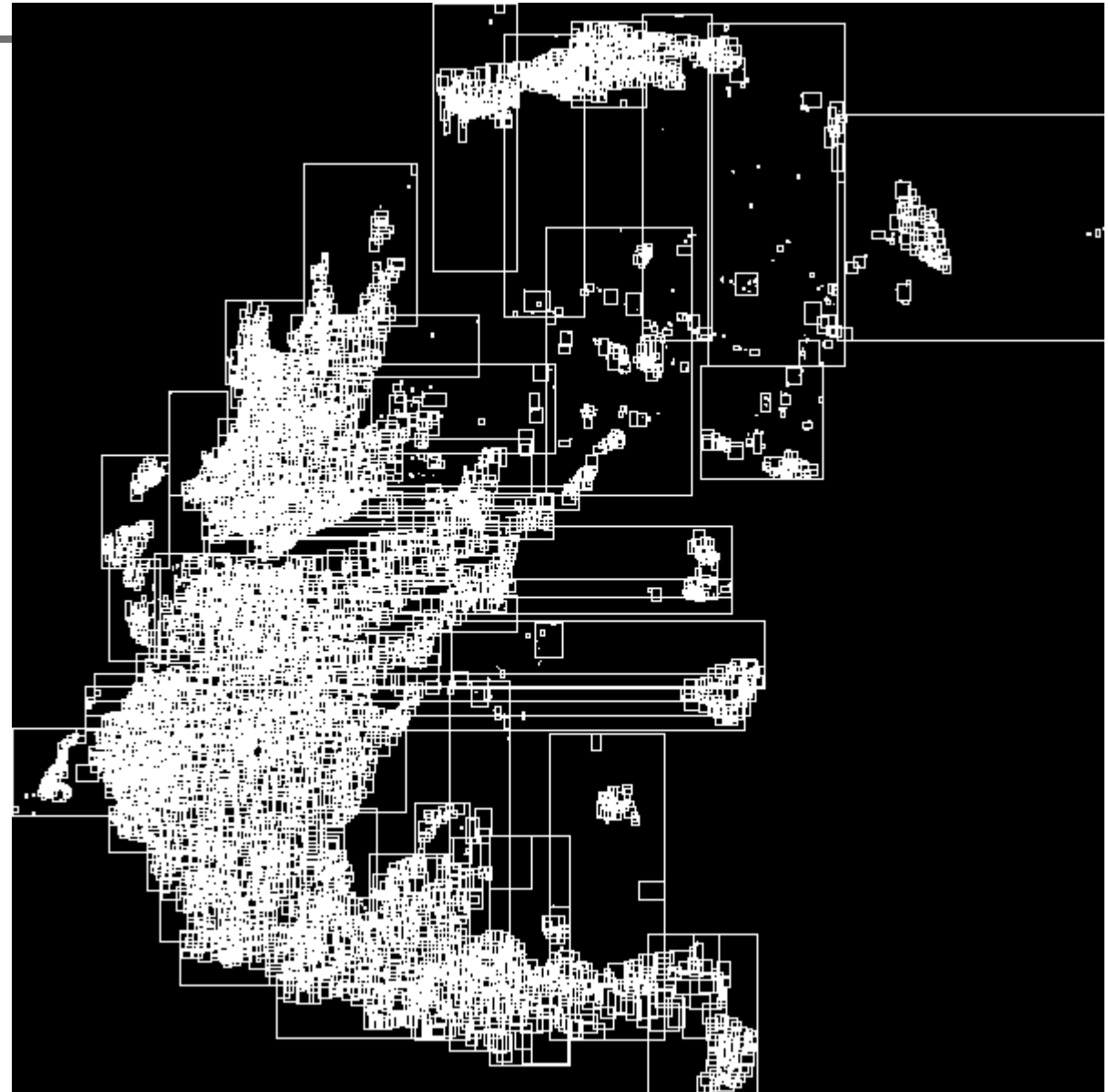
Расширяемость PostgreSQL: GiST

- Программный интерфейс GiST:
 - GISTENTRY * compress(GISTENTRY * in)
 - GISTENTRY * decompress(GISTENTRY * in)
 - bool equal(Datum a, Datum b)
 - float * penalty(GISTENTRY *origentry, GISTENTRY *newentry, float *result)
 - Datum union(GistEntryVector *entryvec, int *size)
 - bool consistent(GISTENTRY *entry, Datum query, StrategyNumber strategy)
 - GIST_SPLITVEC * split(GistEntryVector *entryvec, GIST_SPLITVEC *v)
- http://www.sai.msu.su/~megera/postgres/talks/gist_tutorial.html



Расширяемость PostgreSQL: GiST

- Пример — Rtree (GiST) для населенных пунктов Греции
 - Маленькие прямоугольники — исходные данные (MBR населенных пунктов)
 - Большие прямоугольники — 1-й уровень дерева
 - Подробности: http://www.sai.msu.su/~megera/wiki/Rtree_Index





Расширяемость PostgreSQL: GiST

- Intarray - АМ для целочисленных массивов
 - Операторы overlap, contains

$S1 = \{1, \mathbf{2}, 3, 5, 6, \mathbf{9}\}$

$S2 = \{1, \mathbf{2}, 5\}$

$S3 = \{0, 5, 6, \mathbf{9}\}$

$S4 = \{1, 4, 5, 8\}$

$S5 = \{0, 9\}$

$S6 = \{3, 5, 6, 7, 8\}$

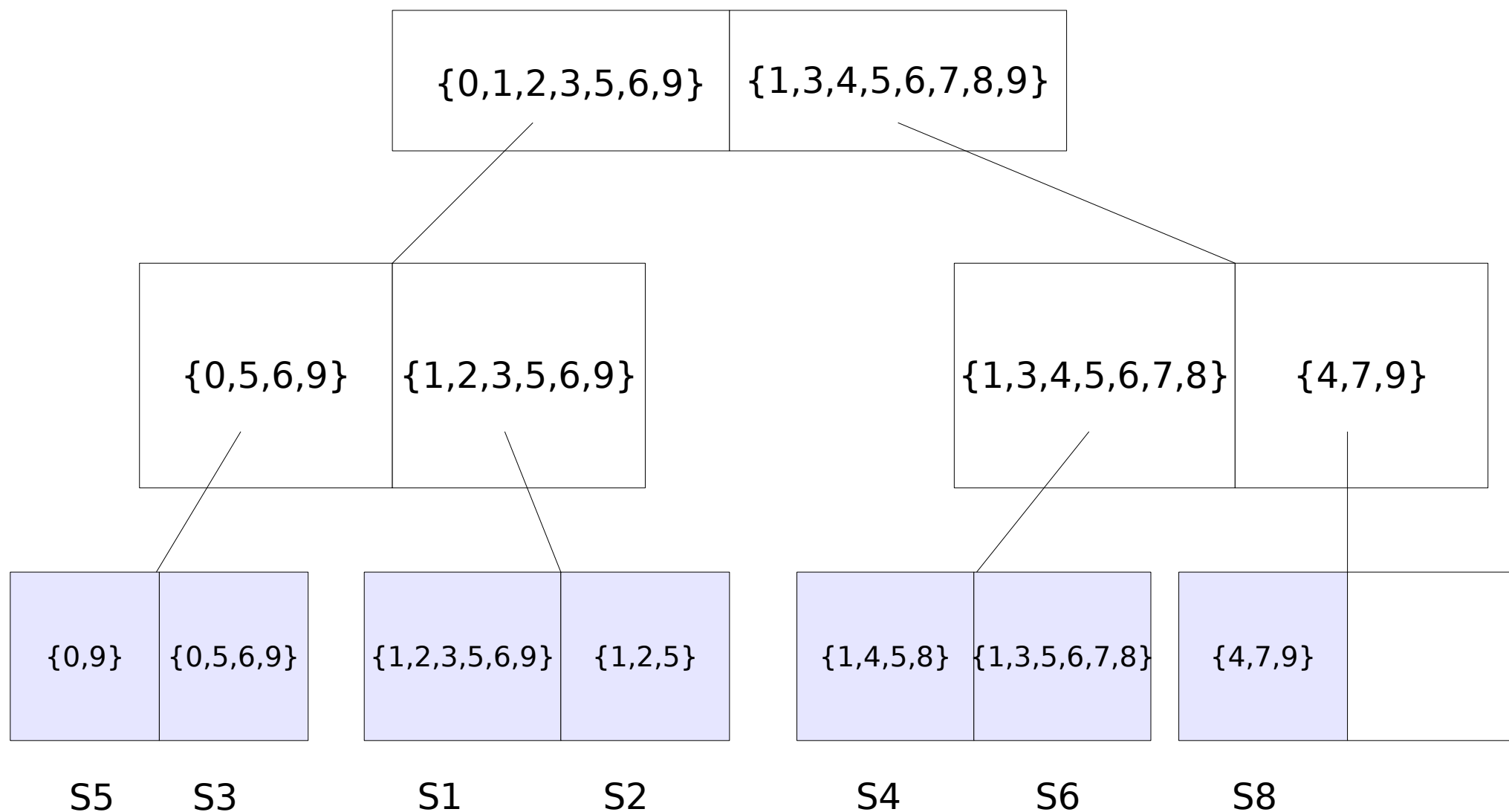
$S7 = \{4, 7, \mathbf{9}\}$

$Q = \{\mathbf{2}, \mathbf{9}\}$

"THE RD-TREE: AN INDEX STRUCTURE FOR SETS", Joseph M. Hellerstein



Расширяемость PostgreSQL: GiST





RD-Tree

QUERY

{2,9}

{0,1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8,9}

{0,5,6,9}

{1,2,3,5,6,9}

{1,3,4,5,6,7,8}

{4,7,9}

{0,9}

{0,5,6,9}

{1,2,3,5,6,9}

{1,2,5}

{1,4,5,8}

{1,3,5,6,7,8}

{4,7,9}

S5

S3

S1

S2

S4

S6

S8



RD-Tree

QUERY

{2,9}

{0,1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8,9}

{0,5,6,9}

{1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8}

{4,7,9}

{0,9}

{0,5,6,9}

{1,2,3,5,6,9}

{1,2,5}

{1,4,5,8}

{1,3,5,6,7,8}

{4,7,9}

S5

S3

S1

S2

S4

S6

S8



RD-Tree

QUERY

{2,9}

{0,1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8,9}

{0,5,6,9}

{1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8}

{4,7,9}

{0,9}

{0,5,6,9}

{1,**2**,3,5,6,**9**}

{1,2,5}

{1,4,5,8}

{1,3,5,6,7,8}

{4,7,9}

S5

S3

S1

S2

S4

S6

S8



RD-Tree

QUERY

{2,9}

{0,1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8,9}

{0,5,6,9}

{1,**2**,3,5,6,**9**}

{1,3,4,5,6,7,8}

{4,7,9}

{0,9}

{0,5,6,9}

{1,**2**,3,5,6,**9**}

{1,2,5}

{1,4,5,8}

{1,3,5,6,7,8}

{4,7,9}

S5

S3

S1

S2

S4

S6

S8



RD-Tree (GiST)

- Сигнатура слова — слово хэшируется в позицию '1'

w1 -> S1: 01000000 Document: w1 w2 w3

w2 -> S2: 00010000

w3 -> S3: 10000000

- Сигнатура документа (запроса) — суперпозиция (bit-wise OR) индивидуальных сигнатур

S: 11010000

- Фильтр Блума (Bloom filter)

Q1: 00000001 – exact not

Q2: 01010000 - may be contained in the document, **false drop**

- Сигнатура — неточное (lossy) представление док-та
 - + fixed length, compact, + fast bit operations
 - - lossy (false drops), - saturation with #words grows



RD-Tree (GiST)

- Пример — латинские поговорки

id		proverb
-----+-----		
1		Ars longa, vita brevis
2		Ars vitae
3		Jus vitae ac necis
4		Jus generis humani
5		Vita nostra brevis



RD-Tree (GiST)

word	signature
ac	00000011
ars	11000000
brevis	00001010
generis	01000100
humani	00110000
jus	00010001
longa	00100100
necis	01001000
nostra	10000001
vita	01000001
vitae	00011000

id	proverb	signature
1	Ars longa, vita brevis	11101111
2	Ars vitae	11011000
3	Jus vitae ac necis	01011011
4	Jus generis humani	01110101
5	Vita nostra brevis	11001011

False drop



RD-Tree (GiST)

Query

11011000

Root

11011011

11011001

10010011

← Internal nodes

1101000

11010001

11011000

10010010

10010001

← Leaf nodes



RD-Tree (GiST)

- Проблемы
 - Плохо шкалируется с ростом количества уникальных элементов (cardinality) и количеством записей
 - Индекс неточный (lossy), требует проверки false drops



Обратный индекс

Report Index

A

abrasives, 27
acceleration measurement, 58
accelerometers, 5, 10, 25, 28, 30, 36, 58, 59, 61, 73, 74
actuators, 4, 37, 46, 49
adaptive Kalman filters, 60, 61
adhesion, 63, 64
adhesive bonding, 15
adsorption, 44
aerodynamics, 29
aerospace instrumentation, 61
aerospace propulsion, 52
aerospace robotics, 68
aluminium, 17
amorphous state, 67
angular velocity measurement, 58
antenna phased arrays, 41, 46, 66
argon, 21
assembling, 22
atomic force microscopy, 13, 27, 35
atomic layer deposition, 15
attitude control, 60, 61
attitude measurement, 59, 61
automatic test equipment, 71
automatic testing, 24

B

backward wave oscillators, 45

compensation, 30, 68
compressive strength, 54
compressors, 29
computational fluid dynamics, 23, 29
computer games, 56
concurrent engineering, 14
contact resistance, 47, 66
convertors, 22
coplanar waveguide components, 40
Couette flow, 21
creep, 17
crystallisation, 64
current density, 13, 16

D

design for manufacture, 25
design for testability, 25
diamond, 3, 27, 43, 54, 67
dielectric losses, 31, 42
dielectric polarisation, 31
dielectric relaxation, 64
dielectric thin films, 16
differential amplifiers, 28
diffraction gratings, 68
discrete wavelet transforms, 72
displacement measurement, 11
display devices, 56
distributed feedback lasers, 38

E



Inverted Index

- Структура данных, которая для каждого ключа хранит список документов, содержащих его
- Тратим время на препроцессинг и экономим при поиске
- Синонимы: posting list, posting file, inverted file
- GIN (Generalized Inverted Index) - *Абстрагируемся от операции* — тип данных сам определяет какую операцию ускорять



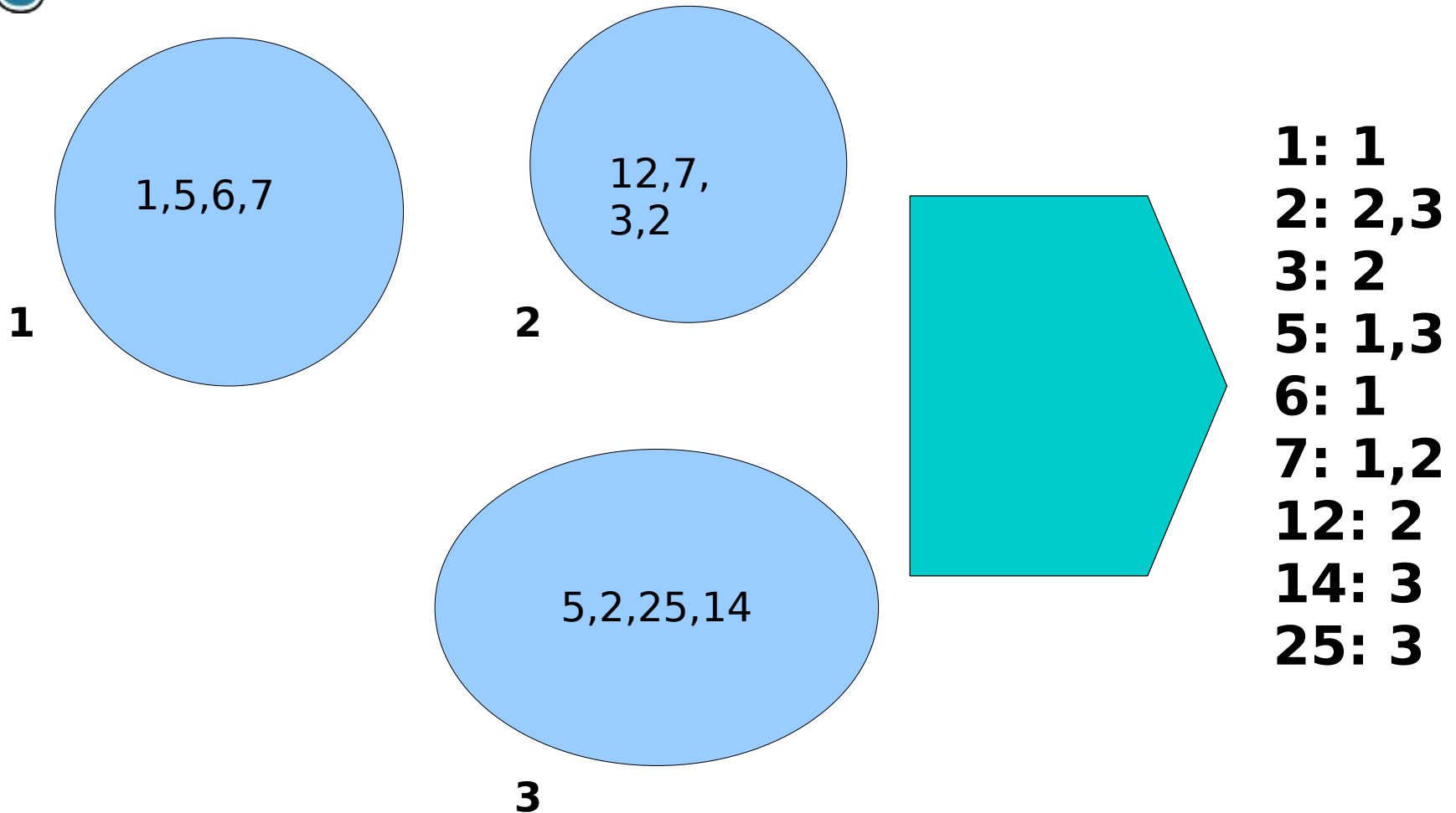
Generalized Inverted Index:API

Разработчик предоставляет 4 (5) функции:

- Datum* extractValue(Datum inputValue, uint32* nentries)
- int compareEntry(Datum a, Datum b)
- Datum* extractQuery(Datum query, uint32* nentries, StrategyNumber n, bool* pmatch[])
- bool consistent(bool check[], StrategyNumber n, Datum query, bool *needRecheck)
- int comparePartial(Datum query_key, Datum indexed_key, StrategyNumber n)



GIN



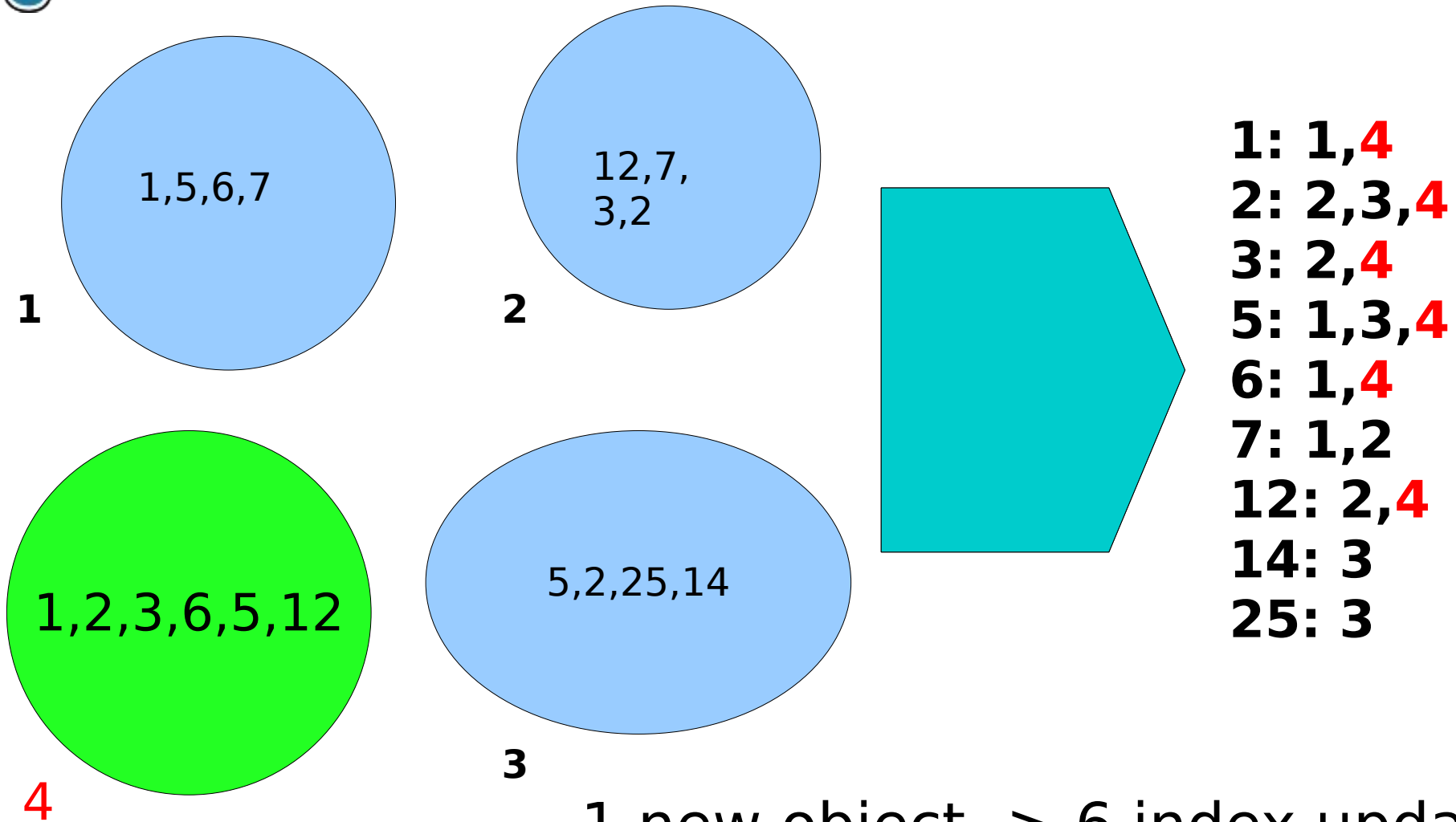


GIN

- Поддерживает разные типы данных
- Очень быстрый поиск по ключам — Btree
- Поддержка partial match
- Многоатрибутный индекс
- Хорошая масштабируемость (кол-во ключей, кол-во документов)
- Быстрое создание индекса
- Медленное обновление индекса :(
- Надежность и хороший параллелизм



GIN: Update problem



1 new object -> 6 index updates !



GIN: The Problem

```
CREATE TABLE  
INSERT 10,000 int[]  
CREATE INDEX
```

3.1 s + 11 s
13.1 s

```
CREATE TABLE  
CREATE INDEX  
INSERT 10,000 int[]
```

~0 s + 100 s
100 s

BULK index insert ~ 10 times faster !

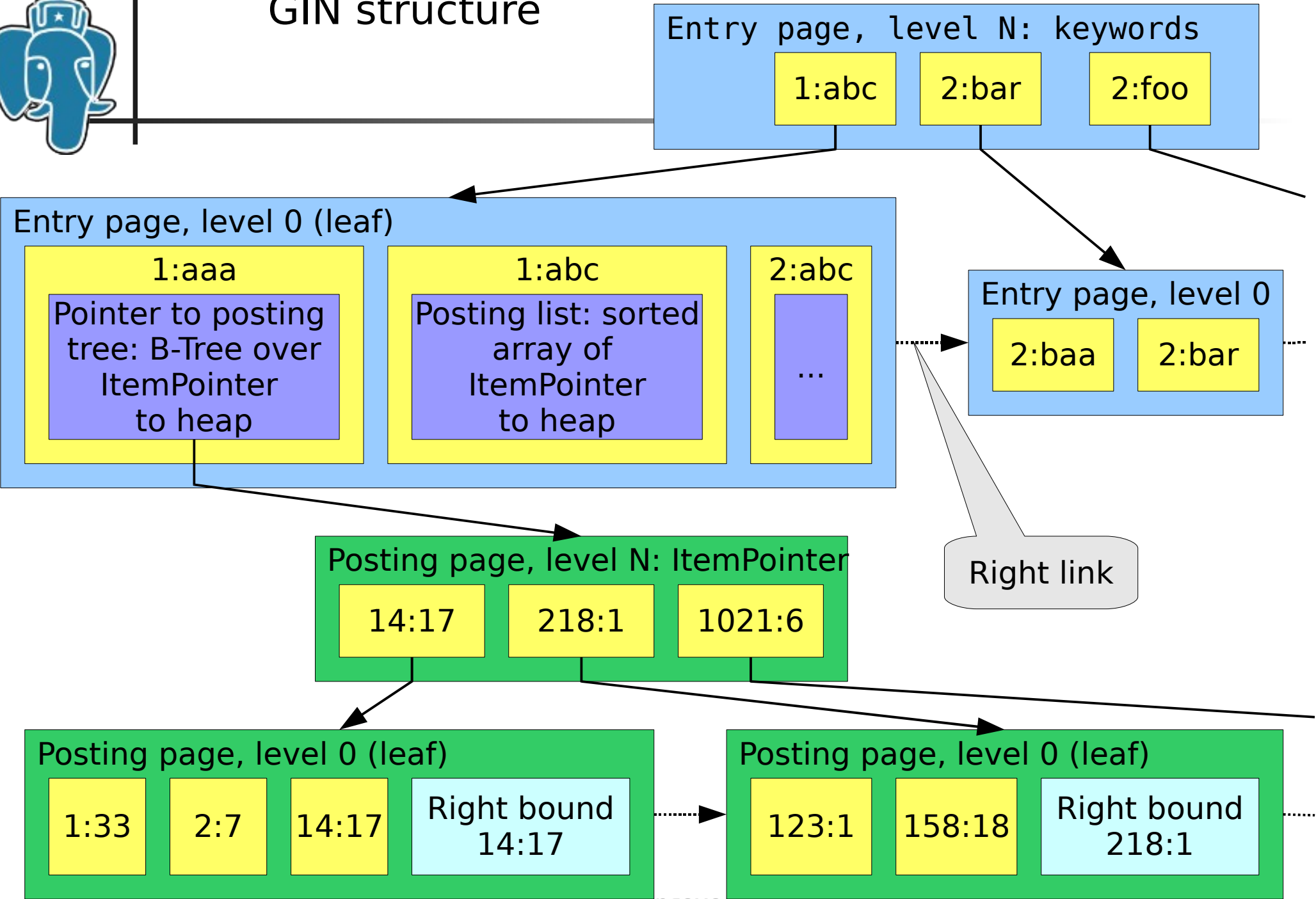


GIN: Быстрое обновление

- Постинг лист для больших значений заменяется на Btree — ускоряет поиск
- Обновления в индекс *откладываются* — используется техника bulk insert, как и при создании индекса



GIN structure





GIN: Быстрое обновление

```
CREATE TABLE  
CREATE INDEX  
INSERT 10,000 int[ ]
```

$\sim 0 \text{ s} + 100 \text{ s}$
100 s

```
CREATE TABLE  
CREATE INDEX  
INSERT 10,000 int[ ]  
VACUUM TABLE
```

$\sim 0 \text{ s} + 18 \text{ s} + 12 \text{ s}$
30 s

BULK INSERT	OLD_GIN	NEW_GIN
10 s	100 s	30 s



Приложения

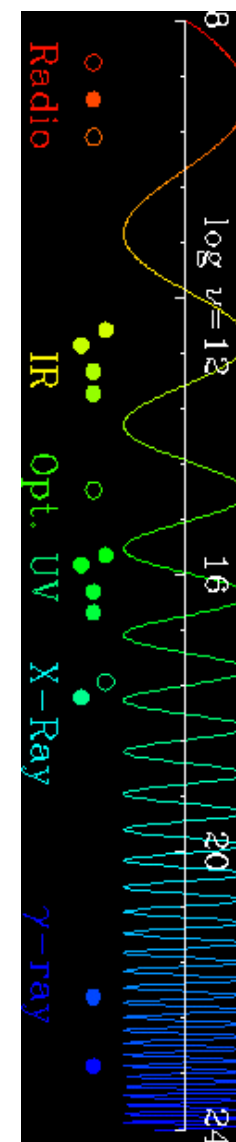
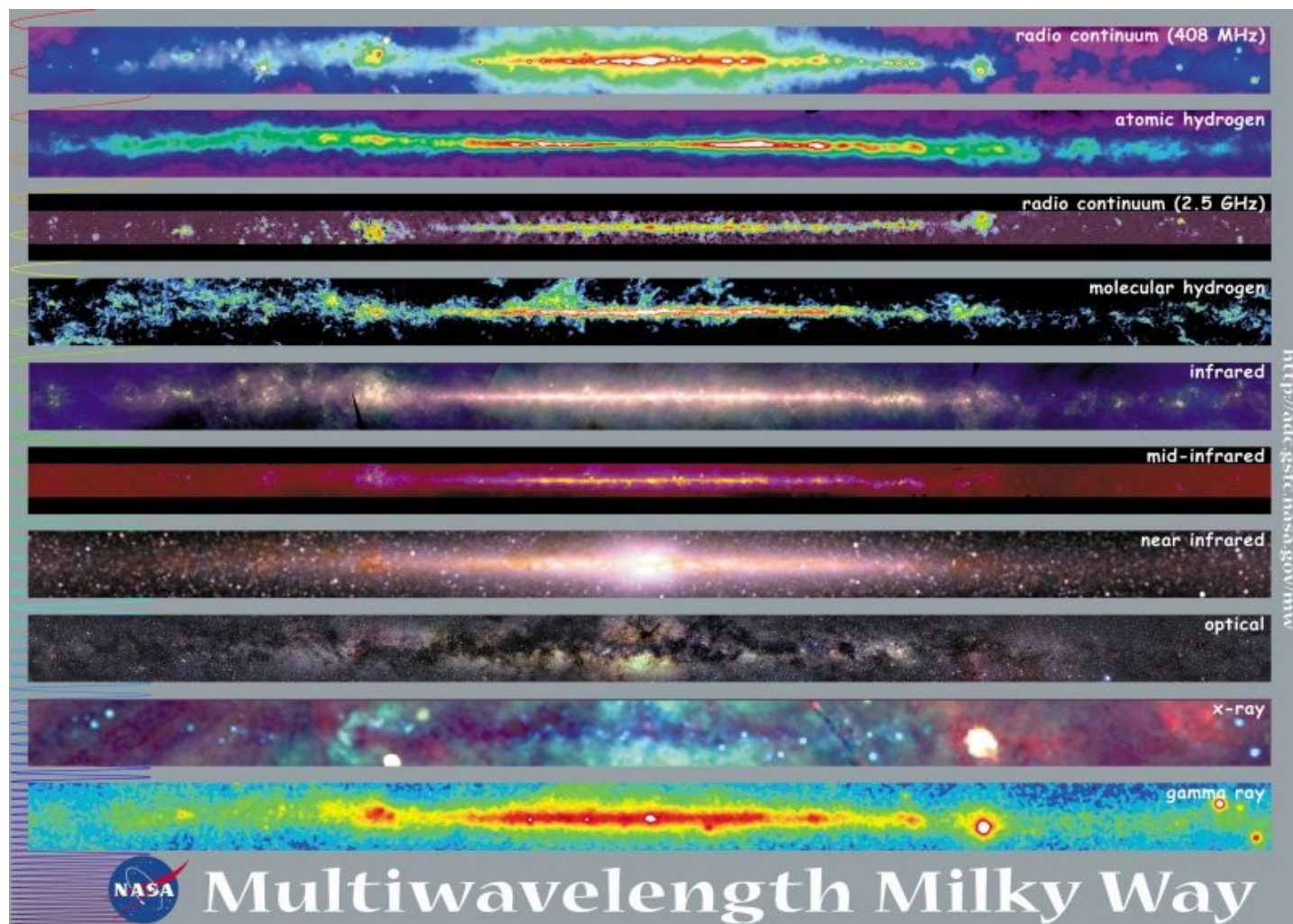
- Целочисленные массивы (GiST, GIN)
- Полнотекстовый поиск (GiST, GIN)
- Данные с древовидной структурой (GiST)
- Поиск похожих слов (GiST, GIN)
- Rtree (GiST)
- PostGIS (postgis.org) (GiST) — spatial index
- BLASTgres (GiST) — биоинформатика
- Многомерный куб (GiST)
-



Виртуальная Обсерватория



Многоволновая Астрономия





Data-intensive Astronomy

- Tycho Brahe's notebooks
 - (1570-1601) ~ 500Kb
- Palomar Observatory Sky Survey
 - 1950 – 10 Gb
- Sloan Digital Sky Survey
 - 2007 – 3 Tb
- Large Synoptic Survey Telescope
 - 2014 – 6.5 Pb per year



Что делать ?

- Монолитные программы,
Многопользовательские программы
- Клиент-Сервер
- Распределенные системы
 - RPC (Unix) — удаленный вызов процедур
 - RMI (Java, SUN) — удаленный вызов методов
 - COM/DCOM (Windows, Microsoft)
 - CORBA — (OMG)
- Проблемы со стандартизацией
интерфейсов !

Общение и обмен данными — 70-е годы прошлого века

Email:@address, text, smtp

Общение и обмен данными — 90-е годы прошлого века

WWW
URI, HTML, HTTP
TEXT

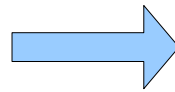
- **URI** - Universal Resource Identifier
- **HTML** - Hypertext Markup Language
- **HTTP** — Hypertext Transfer Protocol

Email:@address, text, smtp

Общение и обмен данными — начало 21 века

- **RDF** — Resource Description Framework
- **RDF(s)** — RDF Schema
- **OWL** — Web Ontology Language

WWW
URI, HTML, HTTP
TEXT



Semantic Web
RDF, RDF(s), OWL
ДААННЫЕ

Email:@address, text, smtp

Общение и обработка данных — начало 21 века

Web Services
UDDI, WSDL, SOAP
ПРОГРАММЫ

- **UDDI** - Universal Description, Discovery and Intergration
- **WSDL** — WS Description Language
- **SOAP** — Simple Object Access Protocol

WWW
URI, HTML, HTTP
TEXT

Semantic Web
RDF, RDF(s), OWL
ДАННЫЕ

Email:@address, text, smtp



Веб-сервисы

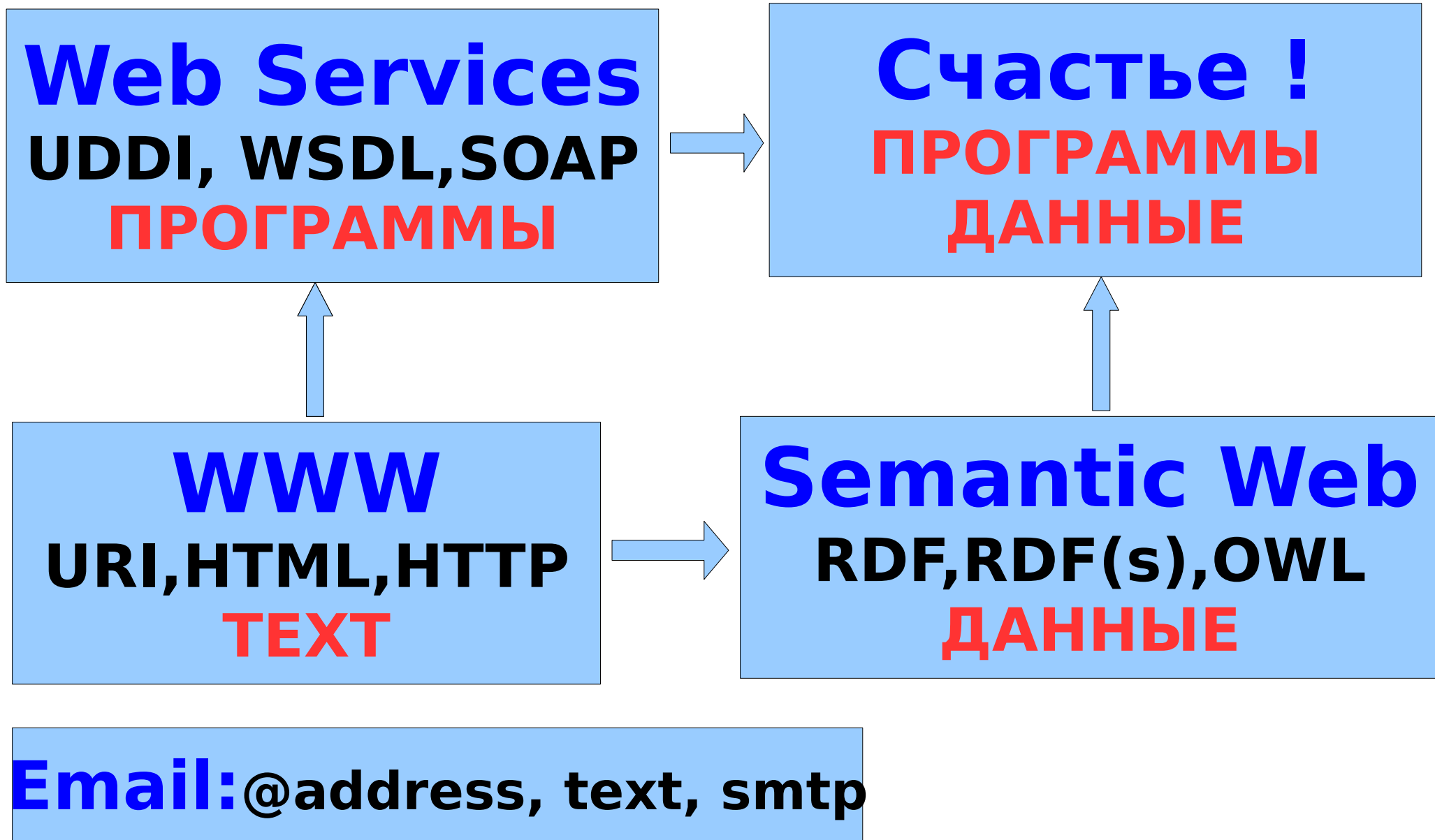
- Это программа, приложение
- Запускается через интернет
- Его описание доступно через интернет
- Как ее запускать описывается в WSDL файле, внутреннее устройство скрыто
- Результаты доступны как SOAP сообщение
- Обеспечивает межмашинное взаимодействие через сеть



Веб-сервисы

- Веб-сервисы можно рассматривать как подпрограммы. Отличия:
 - веб-сервис— **исполняемый** код
 - веб-сервис — **распределенный** код
 - **унифицированный** интерфейс
 - **унифицированные** протоколы
- Веб-сервисы - «кубики лего» для создания сложных веб-сервисов

WEB - universal medium for data, information, and knowledge exchange.



Virtual Observatory

Счастье для астронома

- Технология, которая призвана доставить астроному любые данные, независимо от их расположения и методов хранения
- Базируется на SOA + астрономические стандарты и соглашения
- Позволяет автоматизировать рутинные процедуры — от технологии «клик» к взаимодействию программных агентов
- WWT — Word Wide Telescope, телескоп, который не зависит от погоды !



SAI VO

- Узел Виртуальной Обсерватории ГАИШ
- SAI CAS — работа с очень большими каталогами
- Доступ к обзорам DSS
- Поисковая система по астрономическим абстрактам (arxiv.org)
- Астрономические сервисы
- Веб: vo.astronet.ru



SAI Catalog Access Services

- Хранилище для очень больших каталогов (~ 5 млрд. объектов, 6 Tb БД)
- Стандартизованный доступ к каталогам
 - Вебсервисы для программного доступа
 - Веб-интерфейсы для интерактивной работы
- Поддержка основных запросов (ConeSearch, CrossMatch)
- MySpace для пользовательских данных
- Plastic интерфейс для программного взаимодействия с другими сервисами



Sky Indexing

- Новая схема сегментации данных со сферическими атрибутами — Quad Tree Cube
- Разработана для поддержки VLDB Виртуальной Обсерватории ГАИШ (vo.astronet.ru)
- Реализован как расширение к PostgreSQL
- Лицензия GPL - q3c.sourceforge.net



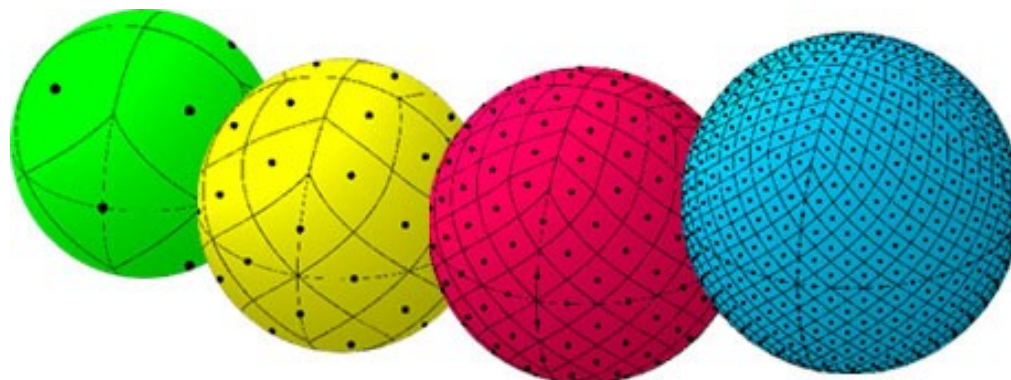
Basic Astronomical Queries

- ConeSearch (radial) query
 - find all stars in a circle on sky of a given radius and position of the center
 - spatial query – spherical coordinates !
- CrossMatch query
 - identify all stars in one set with another with given accuracy
 - fuzzy join of two tables

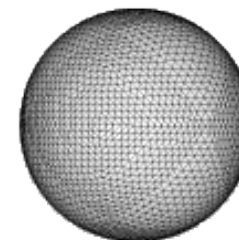
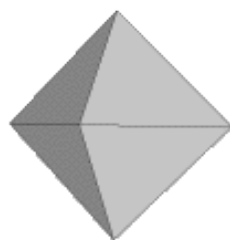
Sky segmentation



- HEALPix - Hierarchical Equal Area isoLatitude Pixelization of a sphere (CMB anisotropy)



- HTM - Hierarchical Triangular Mesh (SDSS project)





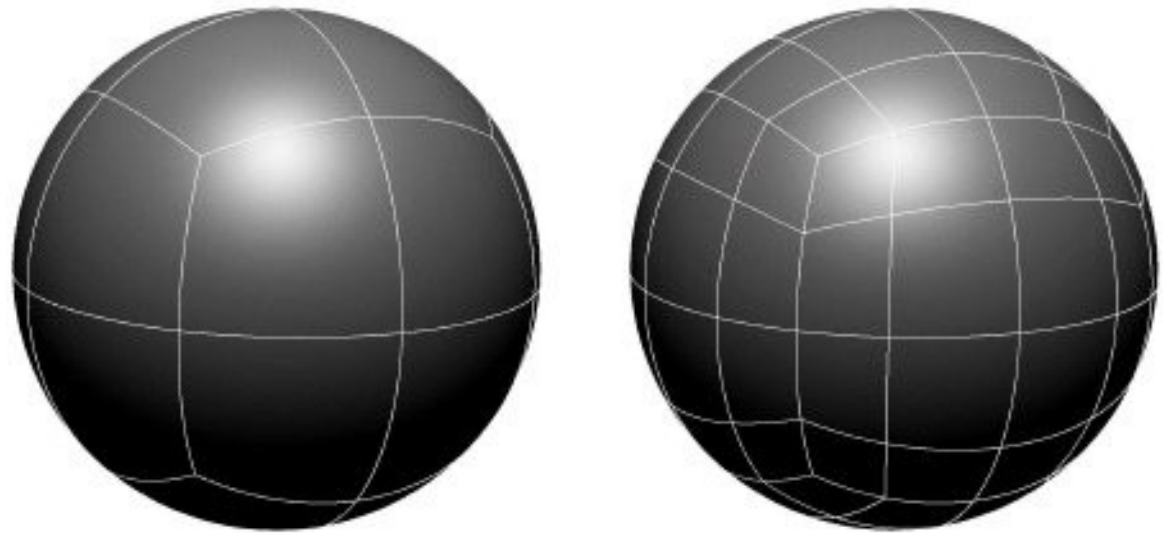
Sky indexing

- We need simple numbering scheme
 - for any point on sphere easy calculate pixel number
 - HEALPix provides
 - HTM needs complex recurrent procedure
- We need simple pixel shape
 - easy geometrical calculation
 - HEALPix pixel is a curvilinear quadrilaterals, very complex
 - HTM triangles are good



Sky Indexing: Q3C

- Cube inscribed into the sphere
- Quad-tree on a cube face maps 2D coordinates in the square (pixel) to the integer — IPIX values
- Central projection of the cube surface to the sphere



The sphere segmentation in Q3C



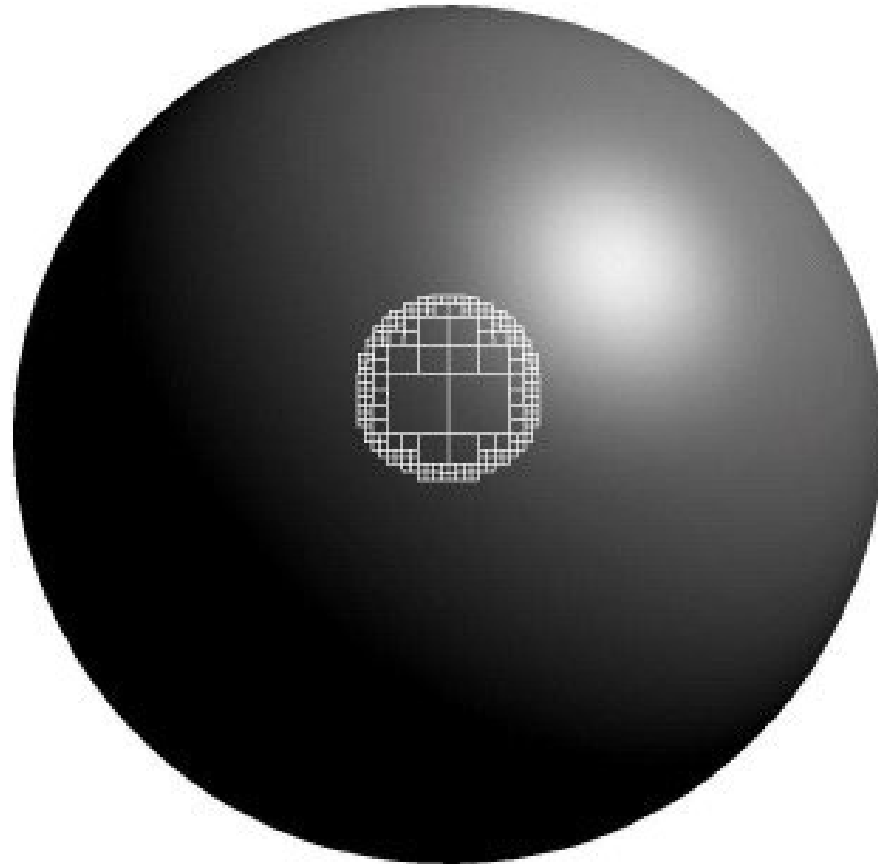
Sky Indexing: Q3C

- Q3C is a hybrid scheme
 - Simple mapping of point on sphere to the IPIX value as HEALPix
 - Simple pixel shape as in HTM
- Computations are much simpler
- Mapping property – nearby points on sphere have close IPIX values
- B-tree index on IPIX values speedup search



Q3C: Spatial Query

- Spatial query is segmented
- Each pixel represents continuous range of IPIX values





Q3C

- Q3C realized as a plugin to the PostgreSQL database (C-language)
 - download from q3c.sourceforge.net
 - SQL functions for
 - ConeSearch
 - CrossMatch (with variable error radii)
 - Supports also rectangular, polygonal queries



Q3C: Sample Queries

- Create index – no additional column !

```
CREATE INDEX usnob_idx on usnob (q3c_ang2ipix(ra,dec));  
CLUSTER usnob_idx on usnob;  
ANALYZE usnob;
```
- ConeSearch

```
SELECT * from usnob WHERE  
q3c_circle_query(ra,dec,10,30,1);
```
- CrossMatch (error radii 1arcsec)

```
SELECT * from 2mass,usnob WHERE q3c_join(2mass.ra,  
2mass.dec, usnob.ra, usno.dec, 0.00027)
```



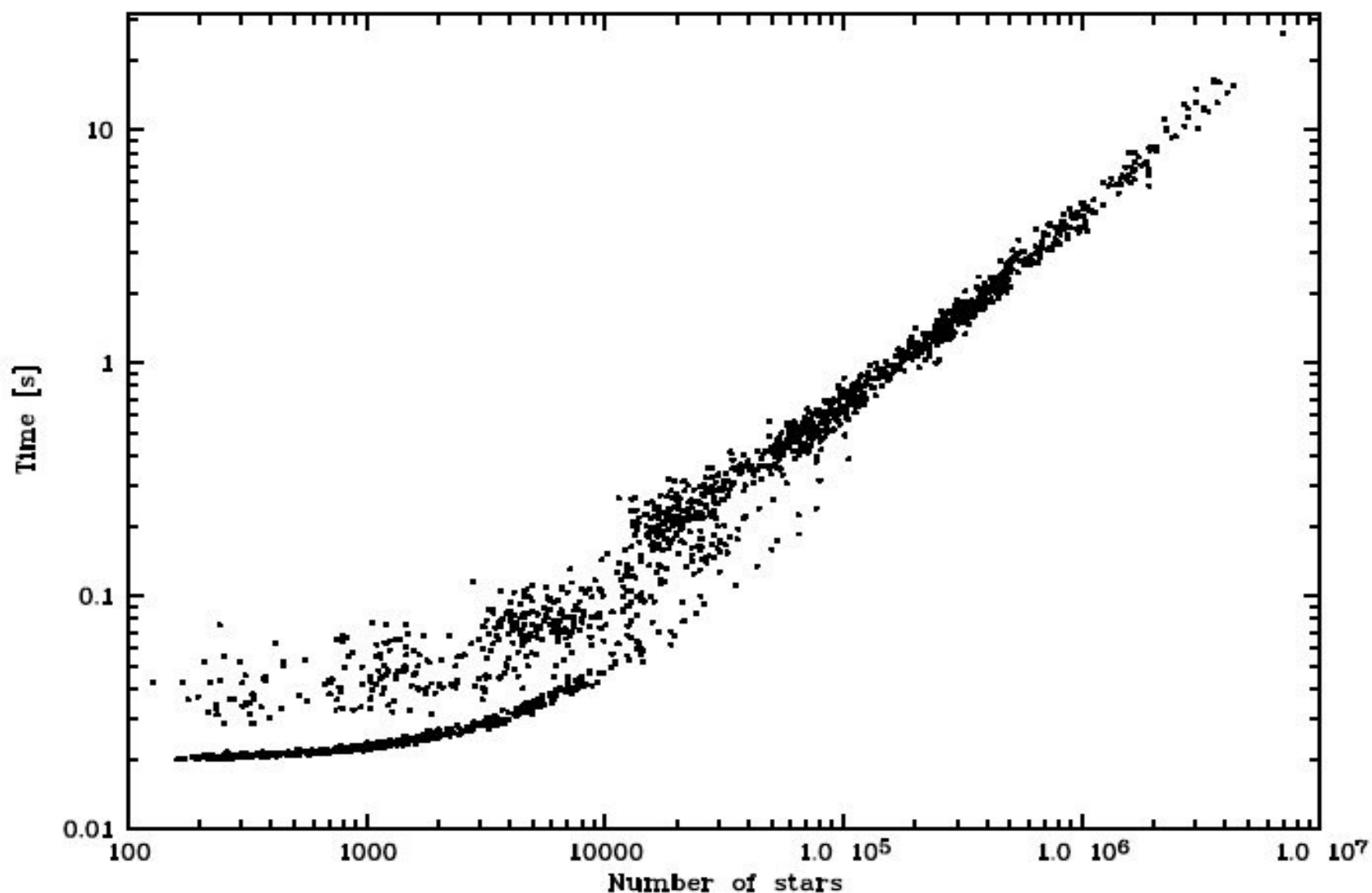

Q3C

- Q3C is extremely fast
 - ordering of the data induced by Q3C guarantee the optimal I/O performance
 - CPU doesn't limit the performance even on high depth of segmentation, since scheme is simple
 - PostgreSQL works well with complex queries produced by Q3C



Q3C: Performance

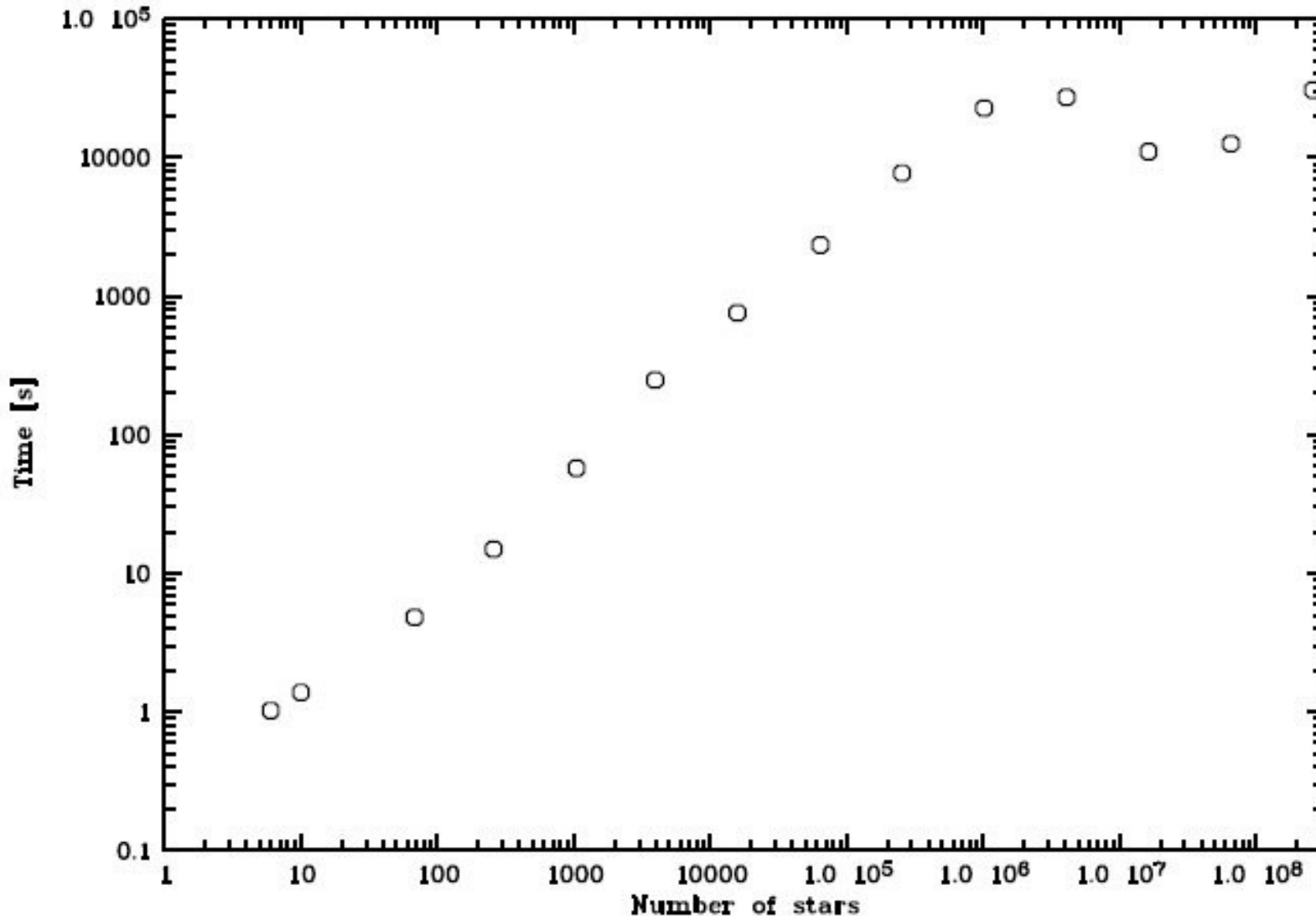
ConeSearch (USNO B1, 10^9 stars)





Q3C: Performance

CrossMatch USNO B1 x USNO B1, 10^9 stars)





Полнотекстовый поиск: GTSE

GTSE — Generalized Text Search Engine

- Полнотекстовые типы данных
 - **Tsvector** - полнотекстовое представление документа (прямой индекс)
 - лексемы с координатной информацией и важностью
 - **Tsquery** — полнотекстовый запрос
 - Лексемы с логическими операторами
- Полнотекстовый оператор `tsvector @@ tsquery`
'a fat cat sat on a mat and ate a fat rat':**tsvector** @@ 'cat & rat':**tsquery**;



Полнотекстовый поиск: GTSE

GTSE — Generalized Text Search Engine

- Полная поддержка ACID
- Online index — GiST, GIN
- Можно использовать для реализации ПОИСКОВЫХ ДВИЖКОВ:
 - OpenFTS (openfts.sourceforge.net) — внешний поиск, PostgreSQL используется как хранилище и исполнитель запросов
 - Встроенный поиск в PostgreSQL

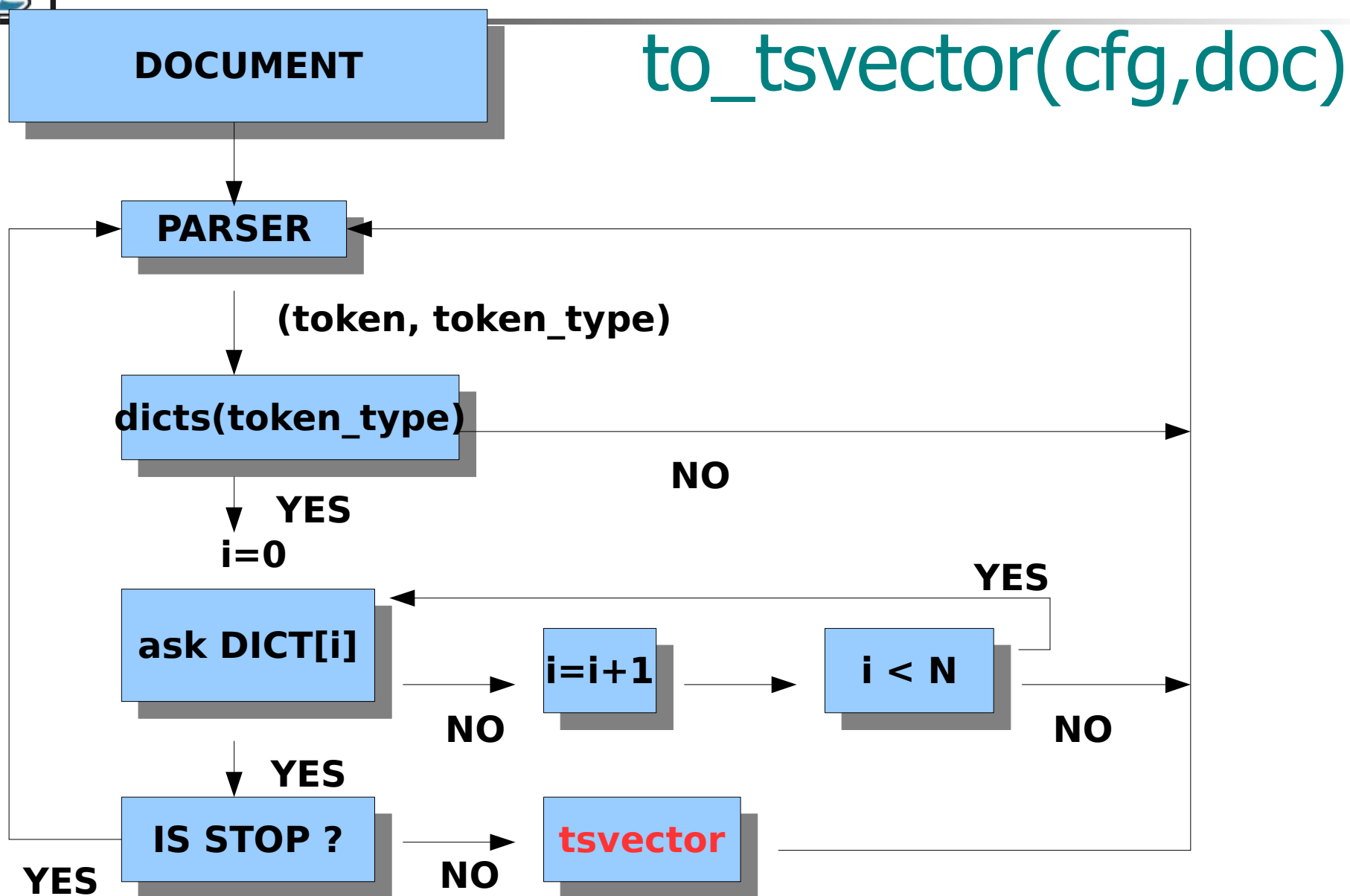


Полнотекстовый поиск PostgreSQL

- Программная «обвязка» предоставляет
 - API для словарей
 - API для парсеров
- SQL интерфейс для настройки поиска
- Встроенная поддержка для всех европейских языков
- Встроенные словари — simple, ispell, stemming, thesaurus, synonym
- Встроенный парсер — 23 типа лексем
- Ранжирование результатов поиска

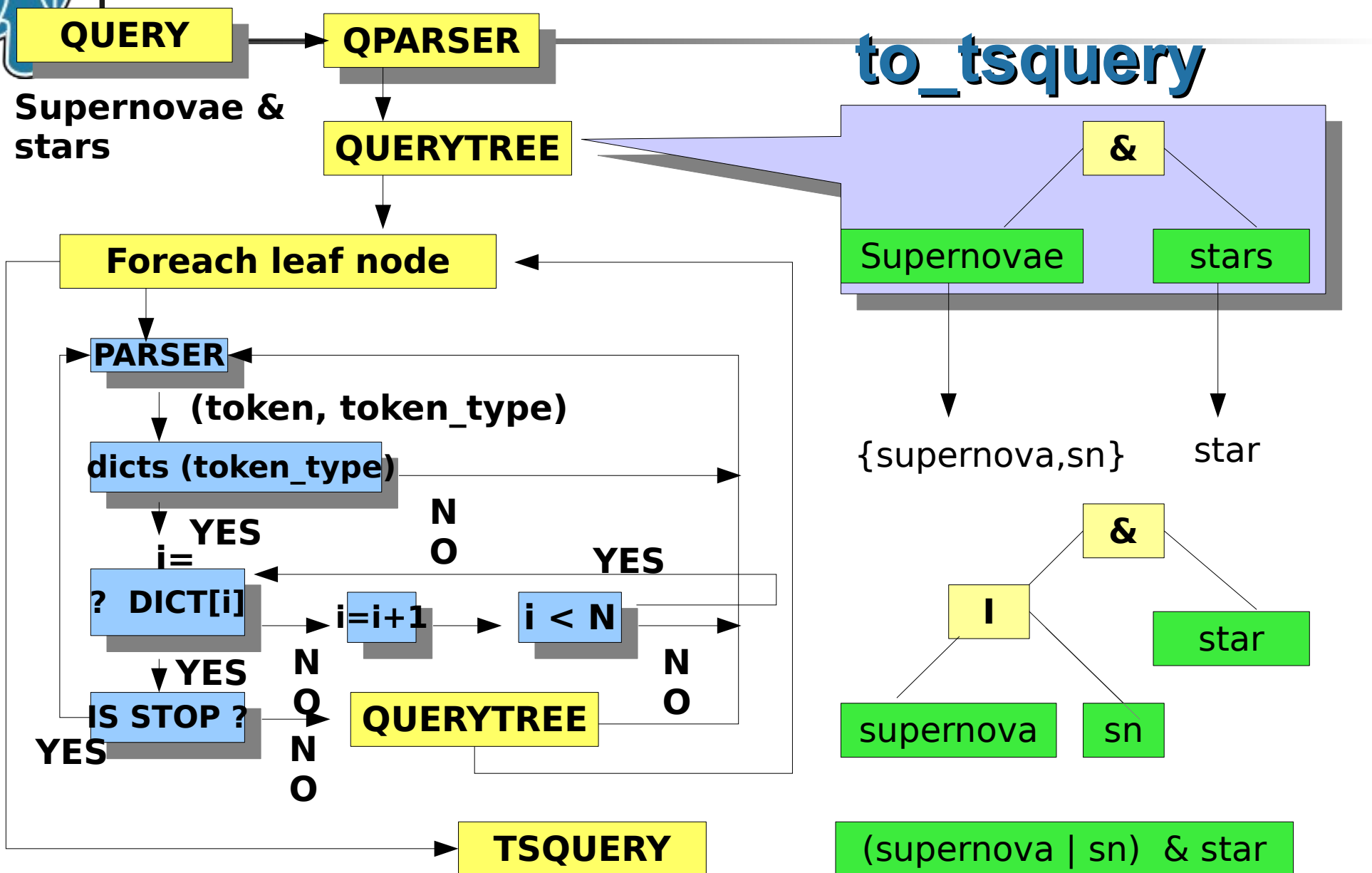


Полнотекстовый поиск PostgreSQL





Полнотекстовый поиск PostgreSQL





Словари

- Словарь — это программа !

```
=# select ts_lexize('intdict', 11234567890);  
ts_lexize
```

```
{112345}
```

```
=# select ts_lexize('roman', 'XIX');  
ts_lexize
```

```
{19}
```



Астрономический словарь (arxiv)

- Dictionary with regexp support (pcre library)

Messier objects

`(M|Messier)(\s|-)?((\d){1,3}) M$3`

catalogs

`(NGC|Abell|MKN|IC|H[DHR]|UGC|SAO|MWC)(\s|-)?((\d){1,6}[ABC]?) $1$3`

`(PSR|PKS)(\s|-)?([JB]?)(\d\d\d\d)\s?([+]\d\d)\d? $1$4$5`

Surveys

`OGLE(\s|-)?((\d){1,3}) ogle`

`2MASS twomass`

Spectral lines

`H(\s|-)?(alpha|beta|gamma) h$2`

`(Fe|Mg|Si|He|Ni)(\s|-)?((\d)|([IXV])+) $1$3`

GRBs

`gamma\s?ray\s?burst(s?) GRB`

`GRB\s?(\d\d\d\d\d\d)([abcd]?) GRB$1$2`



Спасибо за внимание !



Дополнительно



Что такое PostgreSQL: XML

- Новый тип данных: `xml`
- Стандартные (ISO/ANSI SQL) функции публикации SQL/XML
- DTD-валидация
- XPath: функция `xpath()` (с поддержкой Namespaces)
- Альтернативные функции экспорта объектов БД в XML
- B-tree и GIN-индексы для XML-данных, полнотекстовый поиск



Что такое PostgreSQL: XML

SQL:

```
SELECT XMLROOT (
  XMLELEMENT (
    NAME 'some',
    XMLATTRIBUTES (
      'val' AS 'name',
      1 + 1 AS 'num'
    ),
    XMLELEMENT (
      NAME 'more',
      'foo'
    )
  ),
  VERSION '1.0',
  STANDALONE YES
);
```

Результат:

```
<?xml version='1.0'
      standalone='yes' >
<some name='val'
      num='2'>
<more>foo</more>
</some>
```



Что такое PostgreSQL: XML

```
SELECT * FROM table1
WHERE ((xpath('//person/@smpl:pid', xdata,
ARRAY[ARRAY['smpl', 'http://example.com']]))[1])::text =
    '111';
```

```
CREATE INDEX i_table1_xdata ON table1 USING btree(
    (((xpath('//person/@smpl:pid', xdata,
ARRAY[ARRAY['smpl', 'http://example.com']]))[1])::text)
);
```

```
SELECT * FROM table1
WHERE ARRAY['111'] <@ xpath('//person/@smpl:pid', xdata,
    ARRAY[ARRAY['smpl', 'http://example.com']])::text[];
```

```
CREATE INDEX i_table1_xdata ON table
USING gin((xpath('//person/@smpl:pid', xdata,
ARRAY[ARRAY['smpl', 'http://example.com']])::text[]));
```