



Российские
интернет-технологии
2010

Астрономические вызовы технологиям баз данных

Иван Золотухин, Олег Бартунов



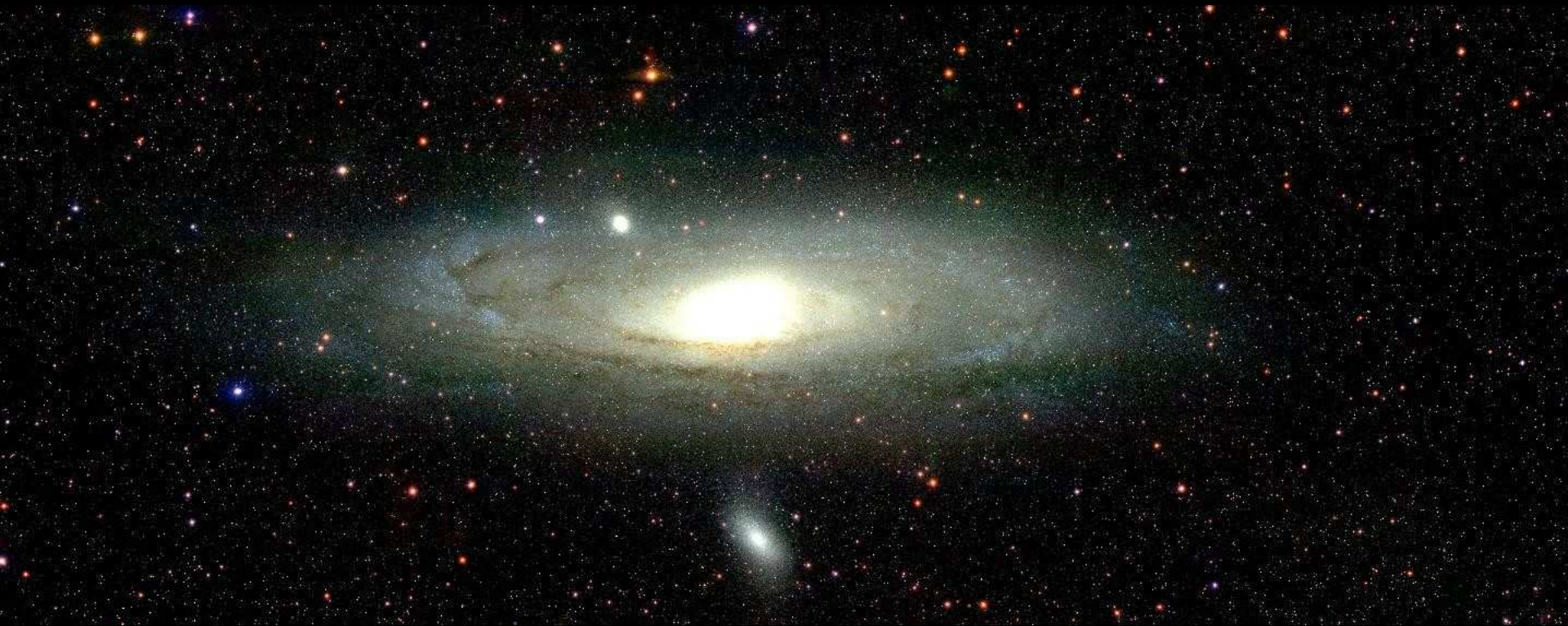
Земля

- 6.8 млрд человек
- 4.6 млрд SIM-карт
- 40 млрд фоток, Facebook
- 1 млн покупок в час, WallMart



Небо

- 100 млрд звезд в галактике
- 200 млрд галактик
- Видимое вещество – 4% Вселенной





Наука

- Слышим голос Бога, хотим поймать его за руку
- Глубокие обзоры неба помогают решать фундаментальные проблемы



SDSS

- **Sloan Digital Sky Survey**
- 2.5-метровый телескоп
- Первый свет в 2000 г. в Нью-Мексико
- Спектры и фото 600 млн объектов





SDSS

- 140 ТБ изображений
- 10 ТБ каталог
- 30 млрд рядов в РСУБД



SDSS

- Публичный SQL-интерфейс!
- 50 млн SQL запросов в год
- Огромный успех в научном сообщ-ве
- \$10 000 за ссылку



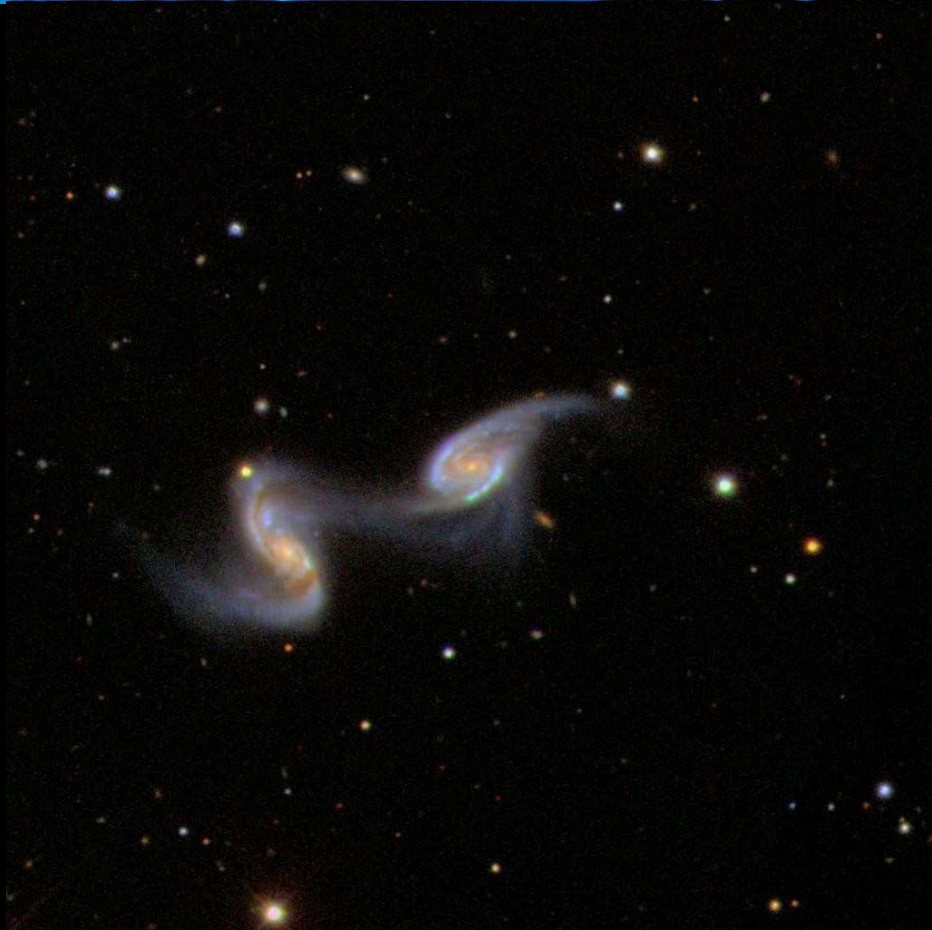
SDSS

- Jim Gray, Microsoft Research
- DBMS: GrayWulf (MS \$QL)
- Не у всех есть Turing Award Winner
- Что делать остальным?..



Galaxy Zoo

- Citizen Science
- 270 000 участников
- 60 млн. классификаций за 2 года







Gaia

- Европейский спутник
- Запуск: «Союз» из Фр. Гвианы, 2012
- Вращается вокруг своей оси
- 1 000 измерений 1 млрд звезд



Gaia

- Как скачать 100 ТБ телеметрии с 1.5 млн километров?
- Сложный cooking данных



Gaia

- 10^{21} Flops для обработки
- 2 года вычислений на 200 TFlops/сек
- Публичный доступ к 10 ПБ продукту
- Какая СУБД? Вопрос открыт. SciDB?



LSST

- **L**arge **S**ynoptic **S**urvey **T**elescope
- Начало работы в 2014 г. в Чили
- Огромный public-private фонд
- Bill Gates и Google вложились в кассу



LSST

- Объектив Canon, 8400 мм, f/1.25
- Поле в 1 000 раз больше Хаббла
- Глубокий обзор всего неба за 3 дня...





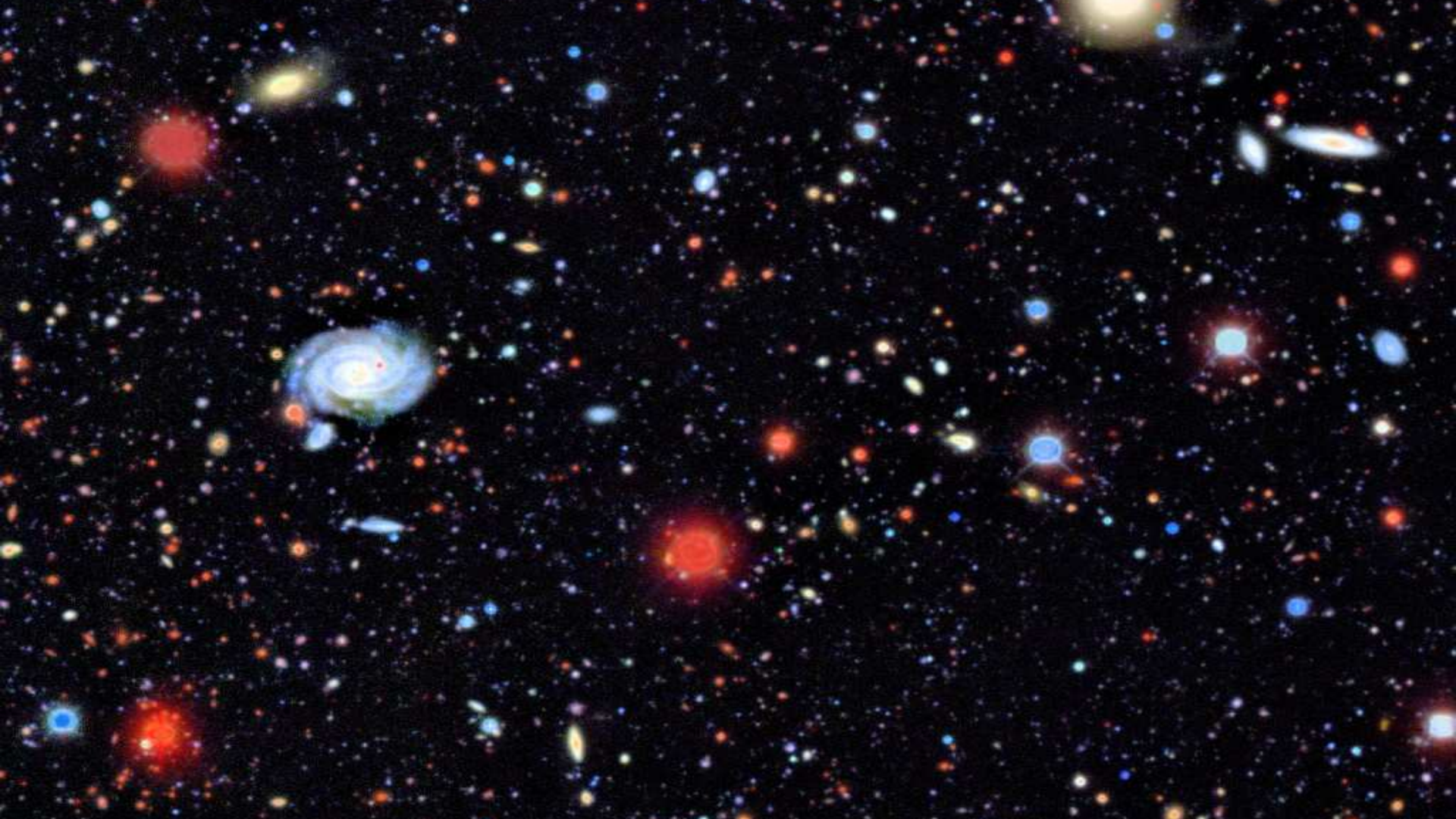
LSST

- 3200 Mpix матрица
- 1 пиксель = 2 байта (raw)
- 15 сек экспозиция, 2 сек считывание
- 3 ГБ/сек запись



LSST

- Real-time обработка: < 10 сек / кадр
- 5 000 операций на пиксель
- 9 TFlops/сек в пике
- 30 ТБ за ночь





LSST

- 30 ПБ каталог
- 100 ПБ изображений
- Google Sky и Microsoft WWT
- Публичный доступ, citizen science!



Мысли

- Вычисления двигаются к данным
- MapReduce: решает только embarrassingly parallel problems
- Выигрывают те, кто знает «SQL»



Будущее

- Это цунами данных!
- Большой вызов человечеству
- Data intensive science
- CIO: Chief Information Officer



Российские
интернет-технологии

2010

**THE SMARTEST PEOPLE ON
THE PLANET ARE PUTTING
THEIR HEADS TOGETHER.**

SciDB Open Source DBMS for Scientific Research



Очень много данных !

VLDB -> XLDB

Very Large → Extremely Large
XXX Tb XXX Pb



Стало очень много данных !

- Успехи в технологии сенсоров
 - Большие размеры
 - Качественные
 - Доступные
 - Все диапазоны спектра
- «Sensor-centric» science !



Что такое научные данные ?

- «Сырые» (raw) данные — данные непосредственно из сенсора **Хранить вечно !**
 - Nikon's raw 12/14 bit
- **Мы все больше абстрагируемся от объекта изучения**



Что такое научные данные ?

- Обработка (cooking) «сырых» данных — сложная процедура.
 - Изучение звезд: **глаз**-фото-ЭОП-ccd
 - Открытие частиц: треки в камере Вильсона, сейчас CCD в LHC
 - Raw Converter: Capture NX, NX2



Что такое научные данные ?

- Результат обработки - Научные данные
 - Публикуются каталоги
 - используются в исследованиях
 - jpg 8-bit → flickr, photosight,...



Принцип науки

Воспроизводимость

Научных Результатов

под угрозой !



Воспроизводимость результатов

- Как ссылаться на данные (purl) ?
 - Раньше: Название журнала, том, страница, год
 - [Oops! This link appears to be broken.](#)
 - Данные меняются в архивах
 - «Бегущая ссылка»?page=237



Воспроизводимость результатов

- Как обеспечить доступность данных ?
 - Свободный обмен данными
 - Независимость от одного вендора
 - Вопросы лицензии



Воспроизводимость результатов

- Как обеспечить сохранность данных ?
 - «Сырые» данные хранить вечно !
- Как обеспечить целостность данных ?
 - Большинство проектов хранят метаданные в БД, а объекты - вне.



Воспроизводимость результатов

- Происхождение данных (data provenance)
 - Качество данных, Источники данных
 - Какие операции привели к появлению этих данных
 - Какие изменения в данных привели к изменению в результатах



Специфика научных данных

- Данные только добавляются, WORM. Изменение данных приводит к появлению новой версии.
- Научные данные - это результаты экспериментов, вычислений



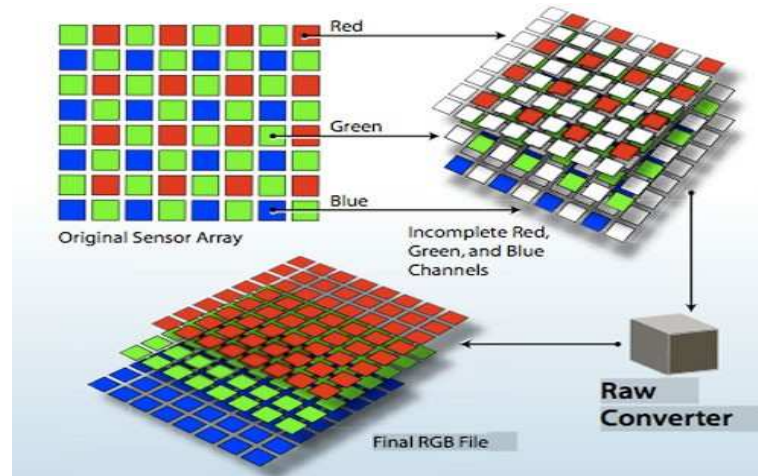
Специфика научных данных

- Данные с погрешностью измерений (error bar)
- Цензурированные данные - астрономия, медицина, биология



Специфика научных данных

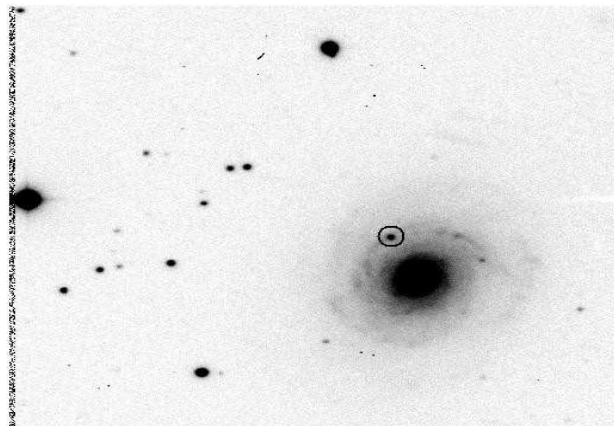
Современная наука построена вокруг сенсоров - «sensor-centric science»



Специфика научных данных

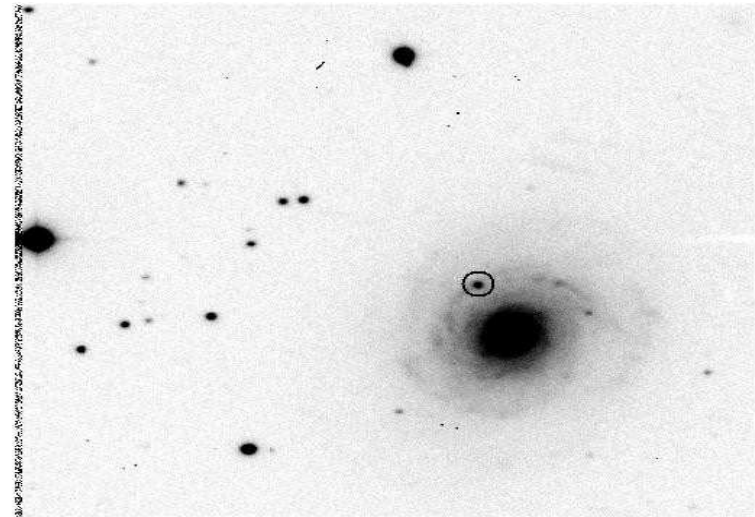
Сенсорные данные естественно хранить в массивах

- SN 2008fv в галактике NGC 3147 в Драконе
(Дмитрий Цветков, ГАИШ) - 2-d массив пикселей



Специфика научных данных

RDBMS неэффективны,
так как не имеют
внутренней
упорядоченности как у
массивов !





I	J	V
0	0	0.02
0	1	0.01
0	2	0.002
...	...	...
1	2	0.5
...	...	...
3	3	0.02

```
SELECT A1.I, A1.J, AVG(A2.V)
```

```
FROM Observation A1, Observation A2
```

```
WHERE A2.I BETWEEN A1.I - 1 AND A1.I + 1
```

```
      AND A2.J BETWEEN A1.J - 1 AND A1.J + 1
```

```
GROUP BY A1.I, A1.J;
```

**Декларативный язык запросов и
логическая модель хороши !**

**Но на практике проблема состоит в
необходимости для каждой ячейки
массива пересчитать массив :(**



Компьютерные технологии изменились

- Классические RDBMS <- system R
 - 1 большой \$\$\$\$ компьютер (mainframe)
 - БД не помещалась в оперативную память - дисковые структуры, индексы
 - Треды как квазипараллельность



Компьютерные технологии изменились

- Компьютерный мир изменился
 - «Железо» дешевое, много RAM, много ядер.
35K\$ - 1Tb RAM, 100 миллионов запросов в сутки
 - Распределенность: кластеры-датацентры
 - Быстрые и широкополосные сети



Используемые решения

- Метаданные в СУБД, куча данных в файлах (xPb)
 - BaBar, LHC, LCLS, PanSTARRS
- Все в СУБД
 - Ebay(xPb, Terradata, GreenPlum), WalMart (xPb, Oracle), SDSS (xTb MS SQL), Genome



Используемые решения

- Home-grown системы (xPb)
 - АТТ, Google, Yahoo, Amazon, Facebook
- Все это нетиражируемые решения !



Что делается

- Текущие тренды
 - Shared-nothing parallel database
 - Дешевые сервера
 - Агрессивная компрессия данных



Что делается

- Попытки продвинуть технологии
 - Aster Data, Vertica, ParAccel, Greenplum, Neteza
- Нерешенные проблемы
 - Эффективная поддержка массивов
 - Data provenance
 - Данные с ошибками



XLDB

- XLDB — сверхбольшие базы данных
- Большая наука - астрономия, биология, науки о земле, физика высоких энергий + разработчики СУБД (MIT, Yahoo, Microsoft, IBM, BEA) + eBay
 - 3 встречи: 2007, 2008, 2009



XLDB

- Традиционные СУБД бесполезно «штопать» - они разрабатывались для других машин, других данных и других требований



- SciDB - новая СУБД с нуля !
- LSST — тестовая площадка
- Майк Стоунбрейкер — лидер проекта



- Ingres, Postgres,
- Illustra, StreamDB,
- Vertica. H-store

Первая фотография разработчиков: Лион, 2009





Полный цикл анализа данных в БД !

- **Data cooking**
- **Feature extraction**
- **Data mining**
- **Data sharing**



Scidb

- Открытая модель развития, open-source
- Ориентирована на науку
- оптимизация для аналитики, не OLTP, ACID → BASE
- Поддержка данных с ошибками



Scidb

- многомерные вложенные массивы — основная структура данных, декларативный язык. Разреженные массивы с неровными краями
- Масштабируемость на сотни петабайт, от ноутбука до кластера, минимум административных затрат



Scidb

- Интерфейсы к научным приложениям (R, Matlab, IDL), к языкам (C++, Python)
- No overwrite storage, поддержка версионности данных (named versions), дельта-компрессия
- Происхождение данных, аннотирования данных



Scidb: Статус

- Основная команда разработчиков (в основном из России) почти сформирована
- Работают первые запросы в односерверном режиме, в течение месяца — запросы на нескольких серверах
- Работаем над тестовым кластером (100 нод)



Scidb: Статус

- Летом ожидается выпуск версии V1
- Первый заказчик — геном человека

Пожелайте успеха !