ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ СУБД И ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ПАРАЛЛЕЛЬНЫХ КОМПЬЮТЕРОВ В АСТРОНОМИИ НА ПРИМЕРАХ КЭ «ЛИРА» И «СВЕЧА»

М.Е. Прохоров, О.С. Бартунов, А.А. Белинский, А.И. Захаров, А.В. Миронов, Ф.Н. Николаев, М.С.Тучин

Развитие промышленных технологий привело к использованию в современных научных экспериментах новых приемников излучения, которые способны порождать очень большие объемы данных. Информационные системы разрабатываются для таких экспериментов с расчетом на размер хранилища в несколько петабайт в ближайшие 2-3 года (LSST, LHC). При этом хранилища в десятки терабайт уже находятся в использовании. Эту информацию требуется хранить, упорядочивать в соответствии с внутренней логикой эксперимента и обрабатывать. Для таких экспериментов нужны СУБД, способные работать с Очень Большими Базами Данных (VLDB) , имеющие специальную внутреннюю структуру и, возможно, распределенные. Для обработки экспериментальных данных необходимы параллельные суперкомпьютеры или кластеры. Мы проиллюстрируем все это на примерах разрабатываемых в ГАИШ МГУ космических экспериментов (КЭ) «Лира» и «Свеча».

Характеристики КЭ «Лира»

Цель КЭ - высокоточный многоцветный фотометрический обзор звезд всего неба до 16-17 звездной величины. В обзор войдут около 400 млн. звезд. Точность измерения блеска для звезд предельной величины составит 1%, а для ярких звезд (ярче 12 зв. величины) - 0.1%. Измерения будут вестись в 10 спектральных полосах от 0.2 до 1.0 мкм (т.е. в оптическом и близком УФ и ИК диапазонах) с борта Российского сегмента МКС [1].

Основная часть комплекса научной аппаратуры «Лира» - телескоп выполненный по схеме Риччи-Кретьена с линзовым корректором. Диаметр главного зеркала телескопа - 0.5 м, фокусное расстояние - 3 м, ширина поля зрения - 1.5 градуса.

В фокальной плоскости телескопа установлена мозаика из 11 пар ПЗС-матриц размером 2250x300 пкс каждая. На десять пар матриц будут нанесены интерференционные светофильтры, соответствующие полосам фотометрической системы КЭ, а на одну - широкополосное просветляющее покрытие, обеспечивающее максимальную чувствительность.

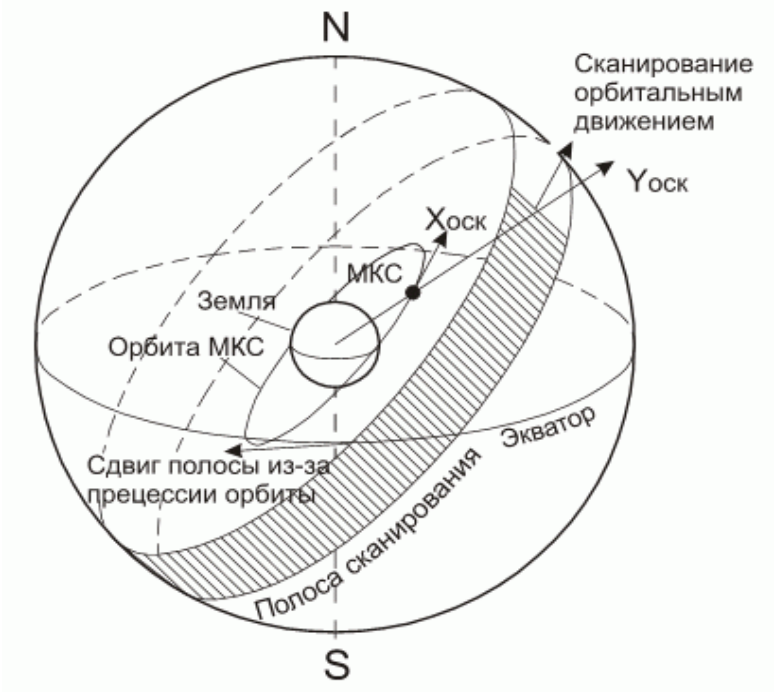

Рис. 1. Сканирование небесной сферы орбитальным движением МКС.

Наблюдения будут проводиться в сканирующем режиме. Это связано с тем, что МКС обращается вокруг Земли с сохранением так называемой орбитальной ориентации - одна сторона станции всегда обращена к Земле, а направление её оси примерно совпадает со скоростью МКС. Таким образом, за один орбитальный

оборот станция совершает один оборот вокруг оси перпендикулярной плоскости своей орбиты. Телескоп устанавливается на противоположной Земле стороне МКС и фиксируется в некотором положении относительно корпуса сегмента (ориентация телескопа изменяется приблизительно раз в месяц). Одна из типичных ориентаций, когда оптическая ось телескопа лежит в плоскости орбиты МКС. В этом режиме поле зрения инструмента перемещается по небесной сфере и заметает на ней полосу (см. Рис.1). Характеристики всех объектов в этой полосе измеряются при прохождении их через поле зрения. Ширина полосы сканирования составляет 1 градус. Наблюдения ведутся в среднем половину каждого орбитального оборота, когда на входное отверстие телескопа не попадают прямые лучи Солнца.

Для непрерывного ведения наблюдений ПЗС-матрицы работают в режиме с Временной Задержкой и Накоплением. Для этого ПЗС расположены в фокальной плоскости таким образом, что изображения звезд перемещаются вдоль их столбцов. В этом случае становится возможным перемещать пакеты накопленных в ячейках ПЗС фотоэлектронов в том же направлении и с той же скоростью, что и изображения звезд. В результате процессы наведения телескопа, накопления и считывания сигнала идут одновременно. На выходе мы получаем изображение полосы сканирования шириной 1 градус (4500 пкс) и длиной около 180 градусов (~90000 пкс).

Информация от телескопа по оптическому каналу связи передается в блок управления экспериментом «Лира», расположенный в обитаемой зоне внутри МКС. Там данные записываются на сменные внешние носители, которые раз в полгода доставляются на Землю в составе спускаемого груза. Такой способ передачи информации вызван отсутствием на МКС постоянного широкополосного канала передачи данных. Тип спускаемых носителей будет выбран на основе отношения емкости к массе и надежности хранения информации (в том числе радиационной стойкости).

Планируемая дата начала КЭ «Лира» - 2013 год, а длительность - 5 лет. За это время все объекты на небе будут измерены в среднем 100 раз (в каждой из спектральных полос). А в некоторых областях - вблизи полюсов мира - число измерений будет достигать 4-5 тысяч.

Потоки и объемы данных

Измерения в КЭ «Лира» идут в двух основных режимах.

- **Режим полной записи скана.**

МКС делает полный оборот вокруг Земли за 1.5 часа, следовательно, изображение звезды пересекает ПЗС по короткой стороне за 1 с, т.е. с матрицы считываются 300 строк в секунду. Тогда полный поток информации от фокальной плоскости (от 11 пар ПЗС) в этом режиме составит

F_full = 2 x 11 x 2250 x 300 x 16 = 240 Мбит/с

(11 - число пар ПЗС, 2250 - ширина ПЗС, 300 - частота считывания ПЗС, 16 - разрядность сигнала). За один оборот будет передано

N_1 = 1/2 x 1.5 x 3600 x F_full = 86 Гбит = 11 Гбайт

данных, а за сутки N_d = 180 Гбайт.

- **Режим выделения звезд**

Оценки показывают, что в КЭ «Лира» будет зарегистрировано до 400 млн. звезд, т.е. в среднем по 10^4 на кв.градус. Для получения характеристик звезды достаточно фрагмента скана размером 10x10 пикселей со звездой в центре (см. Рис. 2). В этом случае для сохранения информации о всех звездах из 1 кв.градуса достаточно 10^6 пкс, т.е. они занимают в среднем 5% площади небесной сферы (1 кв.градус для телескопа «Лира» содержит 4500x4500 пкс). Доля небесной сферы, занимаемой околозвездными фрагментами, зависит от плотности звезд. Она достигает 25% (50000 зв./кв.град.) вблизи направления на центр Галактики и падает ниже 1% (2000 зв./кв.град.) в галактических полюсах.

Отсчеты между звездами можно просто игнорировать или, лучше, осреднить по достаточно большим участкам для оценки фона неба (при этом необходимо использовать специальную технику считывания для снижения шумов).

Рис. 2. Считывание данных с ПЗС в режиме выделения звезд. Цифрами обозначены различные типы фрагментов: 1 - звезда, 10x10 пкс, 2 - протяженный объект, размер может отличаться от 10x10 пкс, 3 - полный фрагмент фона, размер 100x1 пкс, 4 - неполный фрагмент фона (примыкающий к фрагменту звезды или звезд), Nx1, (N<100).

В этом режиме полный поток информации от фокальной плоскости составит

F_* = n_* x 2 x 11 x 10x10 x 1/15 x 16 = 4 - 100 Мбит/с

(n_* - плотность звезд, 11 - число пар ПЗС, 1/15 - площадь проекции ПЗС в кв.градусах, 16 - разрядность сигнала). Поток информации о фоне при усреднении по области 100x1 пкс составит

F_bg = 2 x 11 x 2250 x 300 x 16 / 100 = 2.4 Мбит/с,

т.е. в большинстве случаем существенно меньше потока информации о звездах. Полный объем информации за один оборот составит N_2 = 2 - 34 Гбайт, а за сутки N_d2 = 30 - 500 Гбайт.

Полный объем информации

Основным режимом измерений является режим выделения звезд. Полная запись применяется только на небольшом числе участков небесной сферы, содержащих очень яркие протяженные объекты общей площадью около 100 кв.град. Однако объем полных сканов этих объектов сравним с объемом информации о звездах всего остального неба.

Помимо сканов неба научные данные КЭ включают в себя результаты калибровок фотоприемных устройств, которые проводятся вне сеансов наблюдений путем освещения матриц источником излучения с

известными характеристиками. Еще один источник данных - телеметрия аппаратуры и метки точного времени. Поток этих не превышает 30 Кбит/с.

Замечание. Применение к данным КЭ «Лира» процедур обратимого сжатия позволяет уменьшить в несколько раз объем информации, которую необходимо передать и сохранить на промежуточных носителях. Это снижает требования к пропускной способности каналов, емкости временных и внешних (спускаемых) носителей информации. В центре обработки данные будут содержаться в распакованном виде. Все оценки здесь приведены без учета упаковки.

Характеристики КЭ «Свеча»

КЭ «Свеча» - эксперимент следующего поколения. Его основная цель - построение высокоточного каталога координат и блеска звезд. Погрешность координат для ярких звезд будет составлять десятки микросекунд дуги. Фотометрическая погрешность будет такой же, как в КЭ «Лира», но для существенно более слабых звезд. КЭ «Свеча» будет проводиться на автономном спутнике, обращающимся вокруг Земли по геосинхронной орбите. Основой научной аппаратуры КЭ «Свеча» будут два длиннофокусных телескопа (F=30 м), строящих изображения на общей фокальной плоскости. Область наложения изображений используется для координатных измерений, а непересекающиеся области - для определения астрофизических характеристик объектов [2].

Также как в КЭ «Лира» измерения будут вестись в сканирующем режиме за счет вращения спутника. Типичная скорость вращения - один оборот в сутки. Ширина полосы сканирования - 1 градус, это определяет размер покрытой фотоприемникам области фокальной плоскости - 50x70 см. ПЗС матрицы расположены в 70 рядов, ряда содержит 17000 пкс. При сканировании со скоростью 1 оборот в сутки поток информации будет составлять 5 Гбит/с. Наблюдения будут вестись практически непрерывно.

Обработка информации на борту спутника не производится (за исключением упаковки), т.к. в КЭ «Свеча» измерения ведутся только в режиме полной записи скана. Передача на Землю осуществляется по широкополосному лазерному каналу связи (в близкой ИК области). ИК излучение в этом диапазоне поглощается плотной облачностью, поэтому для надежного приема данных предлагается создать несколько (3-4) пункта приема информации в местах с большим числом ясных дней и ночей. Передача будет вестись в те из приемных пунктов, где в данный момент отсутствует облачность. Для надежности передача может вестись одновременно на два пункта - всё время или во время перехода от одного приемного пункта к другому. К каждому из приемных пунктов должны быть подведен широкополосный канал передачи информации. Подобная система пунктов легко расширяется и позволяет обслуживать неограниченное число спутников или ретрансляторов на геостационарных и близких к ним орбитах - такая сеть будет крайне полезна для решения задач метеорологии, дистанционного зондирования Земли, контроля космического пространства и т.п.

Планируемая длительность КЭ «Свеча» составляет 7 лет. За это время на Землю будет передано 130 Пбайт данных.

Структура БД

БД, содержащая материалы КЭ и предназначенная для их обработки, должна состоять из нескольких частей различной структуры.

- Исходные данные

Структура исходных данных напрямую связана с основными режимами эксперимента, описанными выше. Среди исходных данных будут присутствовать структуры следующих типов:

1. изображения звезд (или небольших протяженных объектов);
2. осредненные записи фона;
3. полные фрагменты сканов;
4. данные калибровки;
5. телеметрические данные.

Характеристики записей приведены в следующей таблице.

Тип записи	Объем одной записи	Число записей	Полный объем
Звезда	250 B	$3–5 \cdot 10^{11}$	100 TB
Фон	10 B	10^{12}	10 TB
Полн. сканы	100 MB	$\sim 10^6$	60 TB
Калибровка	15 MB	10^4	150 GB
Телеметрия	10 B	$5 \cdot 10^9$	50 GB

Полный объем данных, который будет передан на Землю за 5 лет проведения КЭ «Лира», составит 150-200 Тбайт.

Эта часть БД функционирует в режиме WORM (Write Once Read Many), т.е. информация дописывается, но не удаляется и не изменяется. Для этой части БД лучше всего подходит структура ключ-значение.

- Метаданные, рабочие таблицы и связи.

Вторая часть БД содержит данные, с которыми ведется основная работа: предварительные значения блеска и координат звезд в сканах, ссылки на ту же звезду в другом скане или фильтре, привязка к фону, времени, текущим характеристикам аппаратуры и т.д. Для отдельных типов объектов, для которых используются собственные методы обработки (астероиды, переменные звезды и пр.), создаются отдельные таблицы. Эта часть БД должны иметь реляционную структуру. Полезным может оказаться использование колоночно-ориентированного внутреннего представления данных для всей или для части этой информации.

Традиционные СУБД хранят данные в виде записей, которые содержать все атрибуты (колонки). При чтении с диска поднимается вся запись, даже если запрашивается только один атрибут. Подобных накладных расходов можно избежать при хранении атрибутов отдельно - атрибутно-ориентированное (column-oriented) хранение. Из-за одинаковой природы данных они очень хорошо сжимаются, следовательно, занимают меньше места на диске и требуют меньшее количество очень медленных операций ввода-вывода. Соединение нескольких таблиц для такого хранения представляет сложную задачу, но, оказывается, что можно использовать алгоритмы поиска по сжатым данным и откладывать материализацию записей как можно дальше, что приводит к лучшей производительности, чем при традиционном хранении. Подобные СУБД имеют ряд других полезных качеств [3].

Для КЭ «Свеча» структура БД будет в целом подобна БД КЭ «Лира». Отличия будут состоять только в том, что первичными данными будут только полные записи сканов неба. Фрагменты с изображениями звезд и протяженных объектов, отчеты фона будут выделяться из сканов в ходе первичной обработки. Вторым отличием является в 1000 раз больший объем информации этого КЭ.

Типичные операции с данными

При поступлении новой партии наблюдательных данных их необходимо внести в базу исходных данных с распаковкой и соответствующим контролем. В ходе этой процедуры или сразу после нее необходимо обновить таблицы метаданных, для чего произвести следующие действия:

1. вычислить характеристики звезд в новых сканах и занести их в соответствующие таблицы;
2. отождествить звезды в разных спектральных полосах;
3. отождествить звезды в разных сканах;
4. отождествить звезды с независимым внешним каталогом (каталогами);
5. определить моменты наблюдения звезд;
6. сопоставить каждому моменту наблюдений звезды текущие характеристики аппаратуры;
7. обновить карту фона и сопоставить его значение каждой звезде;
8. выделить быстродвижущиеся объекты;
9. выделить ошибки и проблемные наблюдения;
10. Проверить попадание в скан известных малых тел Солнечной системы (астероидов).

Этот список не полон, возможно, потребуются еще какие-либо действия.

Для оперативного планирования хода КЭ необходимо знать покрытие неба сканами, т.е. иметь информацию о том сколько раз наблюдался тот или иной участок неба или объект.

По мере накопления данных необходимо выявлять и обрабатывать специальным образом переменные звезды, звезды со значимым изменением координат, астероиды, планеты и их спутники, и ряд других объектов, а также строить карты протяженных объектов и осредненного фона неба

Основной задачей для выпуска окончательного каталога является выявление и учет малых систематических ошибок. В математическом смысле это задача линейной регрессии большой размерности. Например, полное число наблюдений звезд (т.е. число регрессионных уравнений), для полного объема наблюдений составит порядка 10^11, а количество малых свободных параметров может достигать 10^6-10^9. Эта задача сводится к решению системы линейных уравнений с сильно разреженной матрицей с 10^11 неизвестными.

Для КЭ «Свеча» в типичных операциях с данными существует ряд важных отличий.

Во-первых, данные будут поступать только в виде полных сканов из которых необходимо будет выделять изображения звезд и других точечных источников, протяженных объектов и фон. Эта операция производится после записи данных в хранилище, но перед выполнением п.1 из приведенного выше списка стандартных действий.

Во-вторых, поскольку данные будут поступать непрерывно, стандартная обработка должна вестись в реальном времени.

И, наконец, учета систематических ошибок в итоговом каталоге придется проводить регрессию для 10^14-10^15 наблюдений с 10^8-10^11 малых свободных параметров.

Hardware & Software

Рассмотрим, какое аппаратное и программное обеспечение потребуется для хранения и обработки данных КЭ «Лира».

Для хранения исходных данных, основных таблиц и нескольких вариантов рабочих наборов данных потребуется дисковое пространство суммарным объемом 500-1000 Тбайт. Причем полный объем хранилища может наращиваться в ходе эксперимента по мере поступления данных. Полный объем будет необходим только к моменту его окончания, к 2018 г. Уже сегодня серийно выпускаются устройства хранения данных имеющие такой объем (Sun, HP и др.). Возможно также использование распределенных кластерных решений.

Для хранения исходных данных возможно использование СУБД Berkeley DB или ее аналог, рассчитанный на работу с большим числом записей (~10^11). В качестве реляционной СУБД предполагается использовать Oracle или PostgreSQL.

СУБД PostgreSQL выглядит наиболее привлекательной по нескольким причинам:

очень либеральная BSD-лицензия; устойчивое активное сообщество пользователей и, одновременно, надежное коммерческое сопровождение; расширяемость - возможность создания новых типов данных, запросов и индексов **предметными специалистами**; опробованные средства иерархического хранения астрономических данных (координаты на сфере). Существует ряд инновационных БД, производных от PostgreSQL. К ним относятся: PostgreSQL Plus Cloud Edition - поддержка расширяемости через Cloud Computing; Yahoo Everest - PostgreSQL с колоночно-ориентированным (см. ниже) хранилищем; GreenPlum и AsterData - параллельное исполнение запросов, колоночно-ориентированное хранилище, поддержка технологии MapReduce; TelegraphCQ и его коммерческая версия StreamBase - потоковая СУБД; кластерные и распределенные решения и т.д., что показывает крайнюю актуальность развития СУБД в этом направлении и важность работ начатых в ГАИШ.

Различные этапы обработки данных КЭ «Лира» существенно различаются по трудоемкости. Так для ввода и первичной обработки информации (действия под №№ 1-10 в разделе «Типичные операции») достаточно средств и вычислительной мощности СУБД. С учетом специфики перечисленных задач наиболее удобным сегодня выглядит многопроцессорное хранилище данных с единым дисковым пространством, например, Sun StorageTek 9990/9990V или HP StorageWorks XP24000. Для обработки объектов отдельных типов потребуется достаточно мощный компьютер или малый кластер, ориентированный на проведение вычислений, например, младшие модели СКИФ или Т-Платформы.

Отдельно стоит вопрос окончательной обработки каталога КЭ. Вычислительной мощности младшего суперкомпьютера хватит только на частичную обработку каталога - за 3-6 месячный период наблюдений или при существенно уменьшенном количестве параметров регрессии. С другой стороны, для остальной обработки такая высокая вычислительная мощность не требуется. Наиболее удобным выглядит использование для окончательной обработки каталога КЭ «Лира» внешнего суперкомпьютера высокого класса, например суперкомпьютера СКИФ МГУ «Чебышев».

Hardware & Software для КЭ «Свеча»

Несмотря на то, что начало КЭ «Свеча» пока планируется на 2020 г. Ожидать, что к тому времени появятся устройства хранения информации необходимого объема (несколько эксабайт) не представляется реальным. Поэтому в качестве основного варианта хранилища рассматриваются распределенные кластерные системы. Для текущей обработки информации потребуется параллельный суперкомпьютер среднего класса (по сегодняшним меркам). А для получения окончательного каталога - использование крупных вычислительных грид-систем и специально разработанных алгоритмов.

Другим вариантом может быть использование Cloud computing’a [4], т.е. аренда дискового пространства и компьютерной мощности в гигантских распределенных системах типа Amazon и Google. Для выбора этого

подхода необходимо решить ряд проблем, в частности, надежность хранения информации на больших временах (близка к решению), и проблема надежности недавно введенной информации (примерно за последние сутки).

VLDB сегодня в мире

Вот два факта:

– На сегодня официально анонсирована самая большая в мире база данных с активным доступом - Yahoo Everest, которая на май 2008 года имело хранилище размером более 2 Пбайт, несколько триллионов записей, с ежедневным поступлением около 24 млрд. событий и более 1/2 миллиарда пользователей в месяц. Ожидается в 2009 году рост базы данных до 5 Пбайт. Интересно отметить, что Yahoo Everest - это свободная СУБД PostgreSQL с распределенным вертикально-ориентированным хранилищем и поддержкой кластеризации.

– Большой Адронный Коллайдер (LHC), который ежегодно будет производить около 15 Пбайт данных, распределенное хранилище будет состоять из примерно 200 центров данных по всему миру.

Сообщества научное и разработчики СУБД решили взаимодействовать для создания новой SciDB (http://confluence.slac.stanford.edu/display/XLDB/SciDB) для XLDB (http://en.wikipedia.org/wiki/XLDB). Были выработаны основные требования к будущей базе данных для «Большой Науки»: открытая модель; отказ от строгого соблюдения ACID; хранилище, оптимизированное на чтение; загрузка данных только большими порциями (bulk load); масштабируемость на сотни Пбайт; интерфейсы к языкам и научным приложениям R, MATLAB, IDL, C++, Python; поддержка версионности данных и др.

Наработки группы ГАИШ

Практическое освоение VLDB – введено в стой хранилище данных объемом 36 Тбайт. На нем в течение 2 лет функционирует узел Виртуальной Обсерватории (vo.astronet.ru), предоставляющий интерактивный и программный доступ к терабайтным астрономическим каталогам.

Разработана первая версия модели БД для КЭ «Лира».

В течение многих лет ведется развитие СУБД PostgreSQL. Созданы и доступны пользователям системы расширяемости GiST и GIN, полнотекстовый поиск, эффективная работа с массивами, хранение и работа с очень большими астрономическими каталогами (данные со сферическими атрибутами).

Заключение

Для достижения целей современных научных проектов необходимы VLDB петабайтного объема – сегодня и эксобайтного – через несколько лет. Эти СУБД должны обладать специальной структурой и возможностями, как для хранения, так и для обработки информации. Кроме того это хранилище должно эффективно работать в связке в суперкомпьютерами разных классов: с «mini» - постоянно, «midi» - регулярно и с «high-end» гридами - для окончательной глобальной обработки результатов. ля последней задачи особую важность приобретают проблемы каналов связи, т.к. в силу ограниченности их скорости возникает необходимость работы сразу со всей БД, которая не помещается в оперативную память суперкомпьютера.

Работа была частично поддержана грантами РФФИ 09-07-00499, 09-07-00471.

ЛИТЕРАТУРА:

1. Прохоров М.Е. и др. Российский космический фотометрический эксперимент «Лира-Б». 37-й конф. «Физика космоса». Екатеринбург: Изд. Уральского ГУ, 2008, С. 141–163.
2. Прохоров М.Е., Захаров А.И., Миронов А.В. Космический астро- и фотометрический эксперимент «Свеча» // Конф. «Фундаментальное и прикладное координатно-временное и навигационное обеспечение (КВНО-2009)». СПб., ИПА РАН, 2009, С. 183–184.
3. Бартунов О.С., Прохоров М.Е. IT-АСТРОНОМИЯ. Что и почему нужно астрономии от информационных технологий. Конф. «Научный сервис в сети Интернет-2007», М., МГУ, 2007.
4. Прохоров М.Е., Бартунов О.С. "[За]облачные" базы данных для науки. Конф. «Научный сервис в сети Интернет-2008», М., МГУ, 2008, С. 431-432.