

Р.Ю. Машинистов  
М.А. Григорьева

# Семинары лаборатории

# Организация

- Организация регулярных научных семинары лаборатории (каждые 2 недели)
  - Руководитель семинара Машинистов Р.Ю.
  - Заместитель руководителя семинара Григорьева М.А.
- Места проведения:
  - НБИКС (зд. 190) комн.207, переговорная
  - территория МГУ в Ломоносовском корпусе

# Области интересов

- темы связанные с работами Лаборатории
  - SQL/noSQL, PanDA@NRC-KI, webFTS,...
- темы связанные с экспериментами и проектами, в которых участвуют сотрудники
  - LHC, WLCG
- темы "общеобразовательного" характера
  - BigData, HEP SW, ....

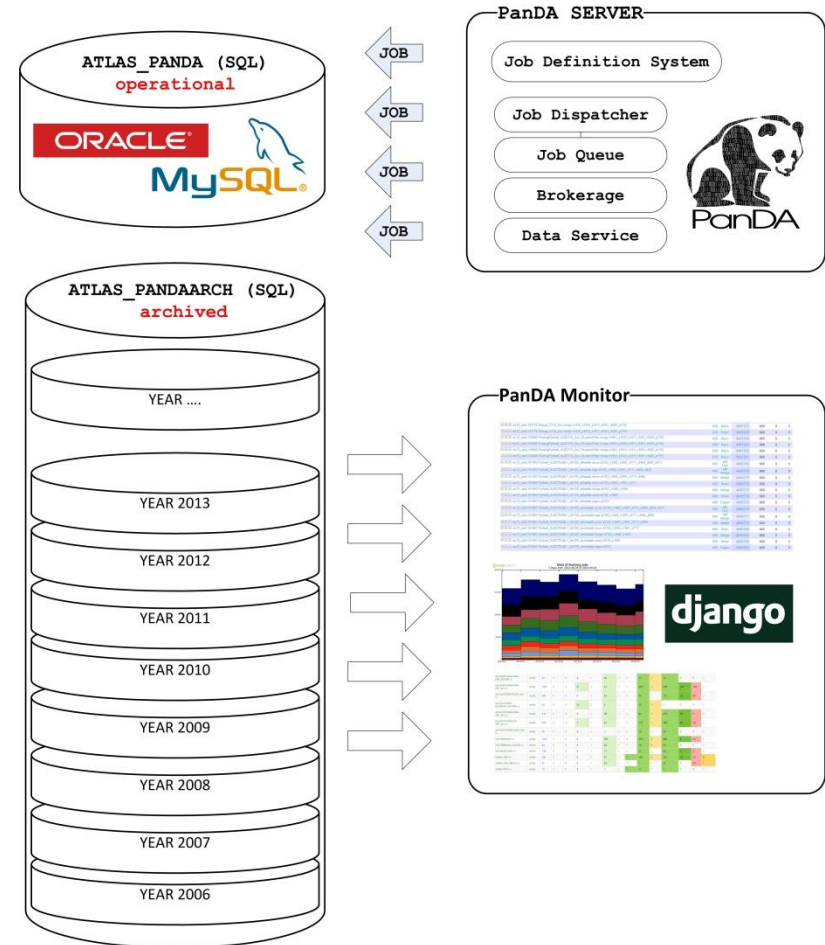
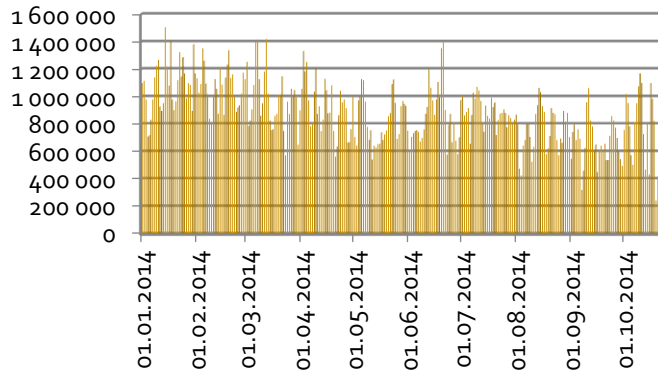
# Использование NoSQL для хранения «больших данных»

- ✓ PanDA обрабатывает **МИЛЛИОНЫ** задач ежедневно
- ✓ Любые изменения статуса задач, и других важных параметров, отображаются в базе данных
- ✓ Оперативные (за последние несколько дней) данные хранятся в отдельной схеме. Задачи, которые имеют статус **'finished'** или **'failed'** перемещаются в архивную схему PANDA

**ATLAS\_PANDA => ATLAS\_PANDAARCH**

Полный архив PanDA сейчас содержит более 900 миллионов задач, выполненных начиная с 2006 года.

**PANDAARCH.JOBSARCHIVED**  
Records



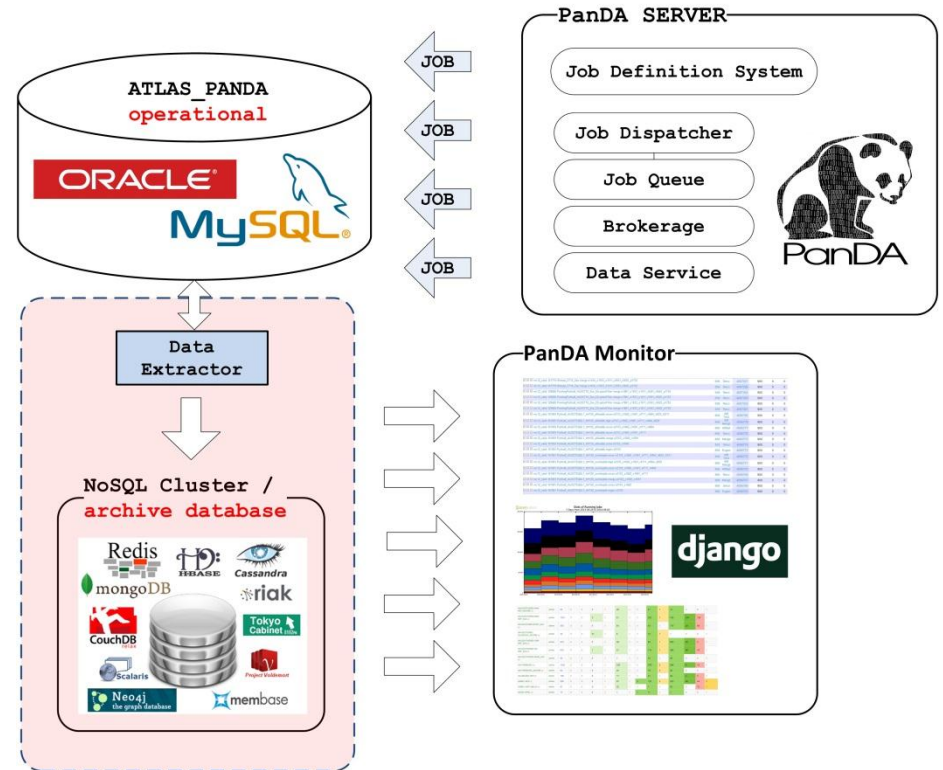
# Использование NoSQL для хранения «больших данных»

Мы столкнулись со следующими проблемами:

- ✓ пользовательская нагрузка на БД сильно возросла, начиная в 2010 года
- ✓ оперативная база стала сильно фрагментированной из-за постоянного переноса данных в архив и их удаления из основной базы

## Методы решения:

- ✓ Разработка NoSQL хранилища архивных данных
  - ✓ Анализ и тестирование скорости доступа к архивным данным в следующих NoSQL системах: Cassandra, MongoDB, HBase
  - ✓ Разработка моделей данных для NoSQL хранилищ
  - ✓ Применение аналитических инструментов в NoSQL системах



# Планируемые семинары по базам данных:

1. Обзор NoSQL базы данных Cassandra
2. Обзор NoSQL базы данных MongoDB
3. Обзор NoSQL базы данных HBase
4. Анализ использования MapReduce как аналитического инструмента для NoSQL хранилищ