

Исполнитель запросов в системе обработки данных из разнородных источников

Самарин Алексей Владимирович, 544 гр.

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра системного программирования

Научный руководитель:

проф., д.ф.-м.н. Новиков Б.А.

Рецензент:

ассистент кафедры информатики Чернышев Г. А.

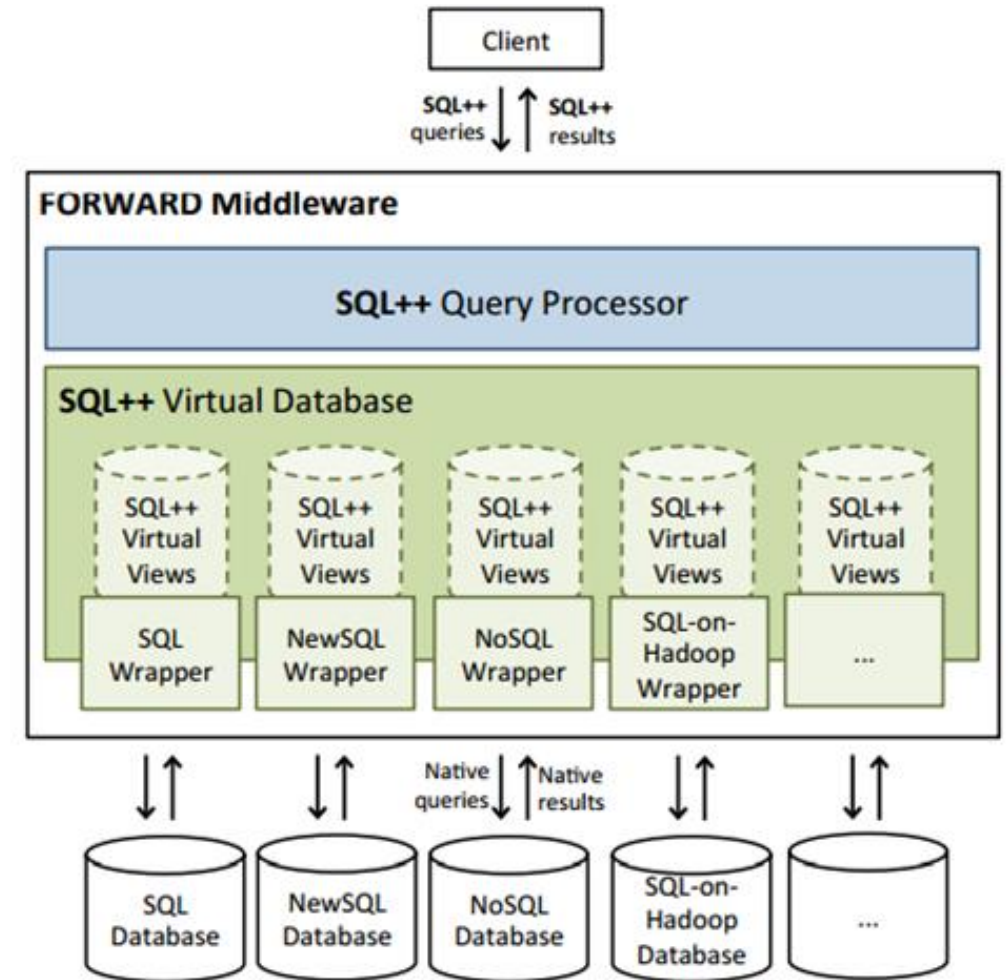
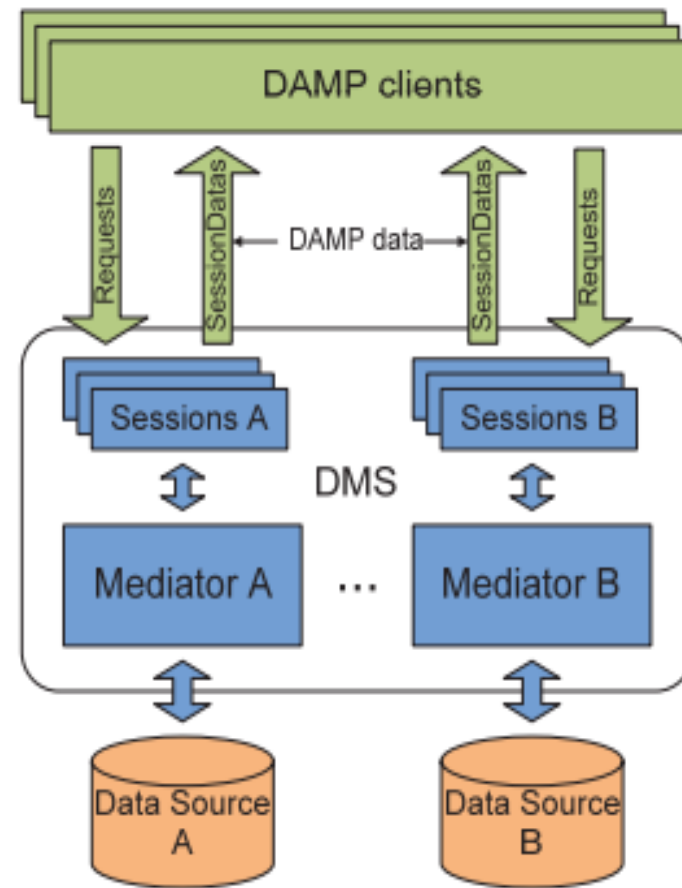
Введение

- Рост объемов хранимой информации
- Гетерогенные источники
- Поиск по релевантности
- Передача и декомпозиция запросов
- Выполнение нечетких запросов
- Приближенное исполнение



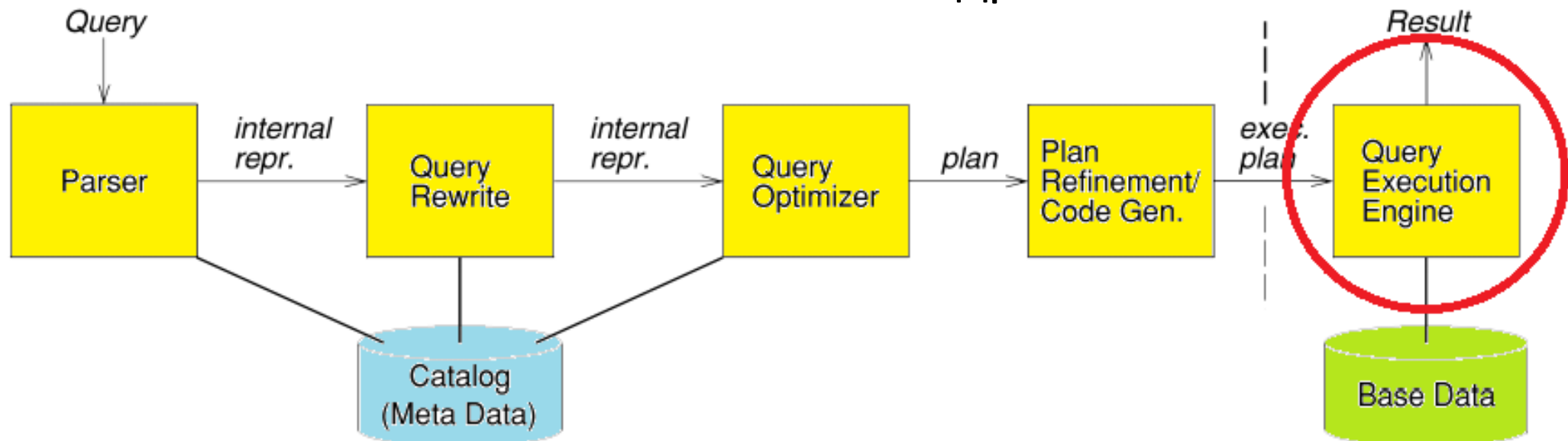
Технологии обработки информации из различных источников

- ODBC, JDBC, ADO...
- DAMP
- FORWARD (SQL++)
- Pentaho



Организация системы обработки информации из гетерогенных ресурсов

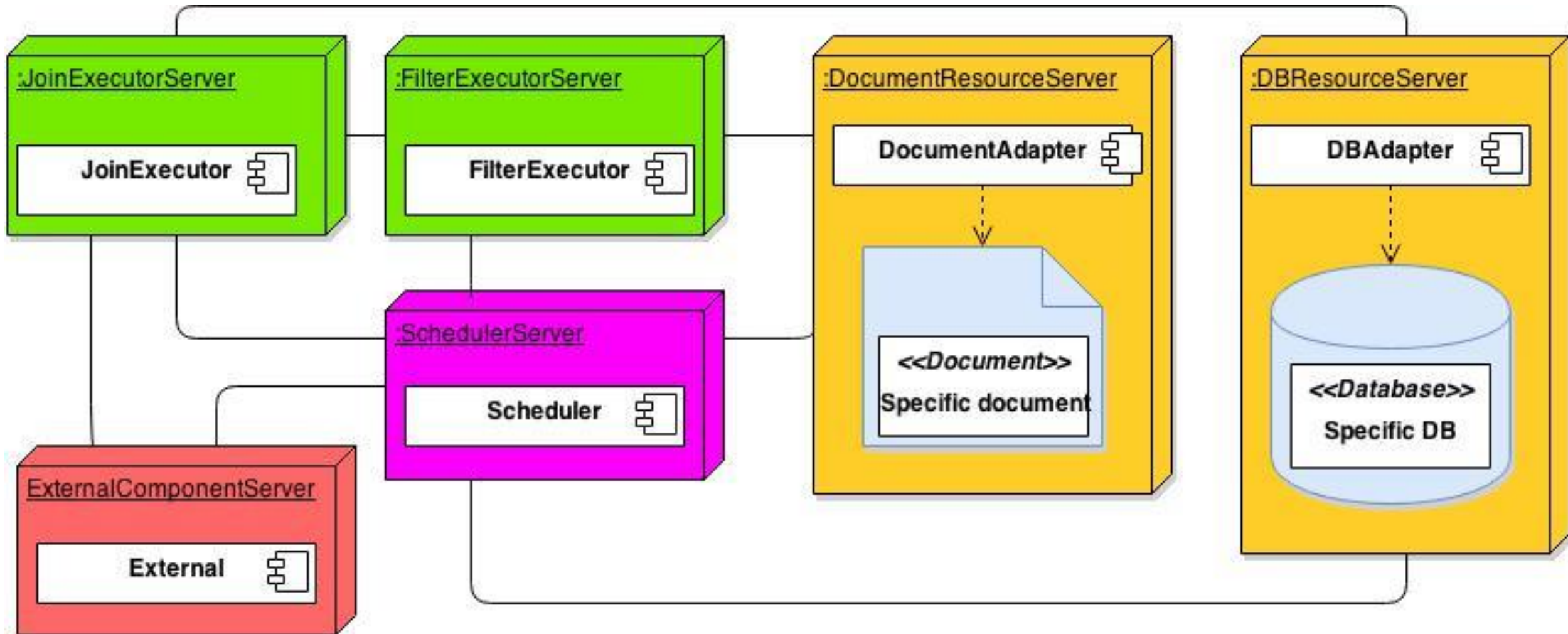
- Q-set
- Расширенная реляционная алгебра
- Средства описания запросов
- Оптимизатор
- Ядро исполнения



Постановка задачи

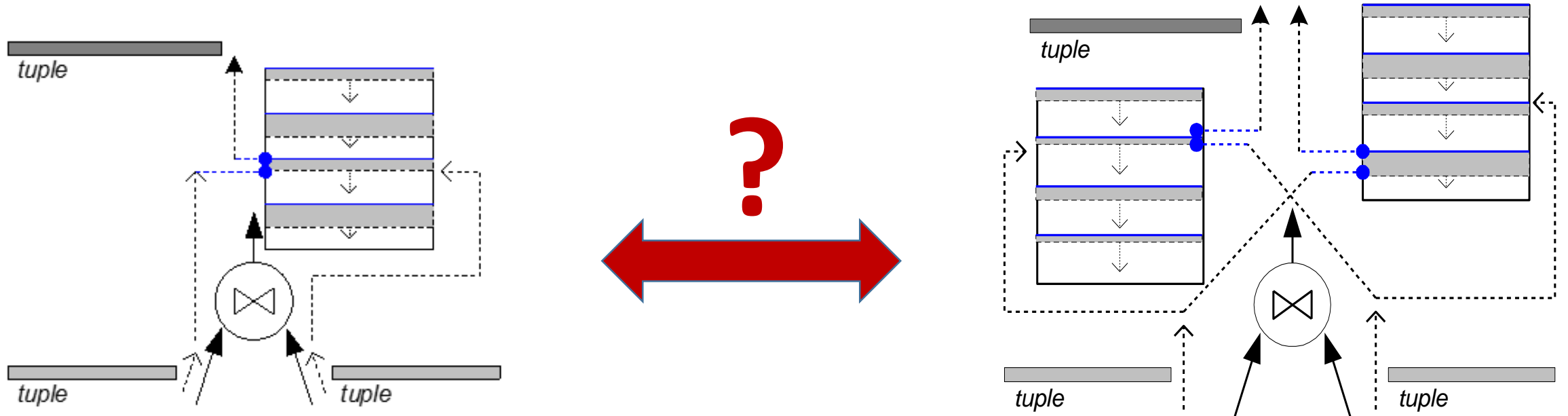
- Разработать архитектуру ядра исполнения запросов
- Обеспечить поддержку конвейерной обработки данных в рамках разрабатываемой архитектуры
- Обеспечить возможность интеграции с языковыми компонентами и оптимизатором
- Обеспечить возможность интеграции новых исполнителей и адаптеров источников
- Произвести апробацию разработанной архитектуры при взаимодействии с реализованным ранее исполнителем/источником данных на основе AsterixDB и высокоуровневыми языковыми средствами для описания запросов

Архитектура ядра исполнения запросов

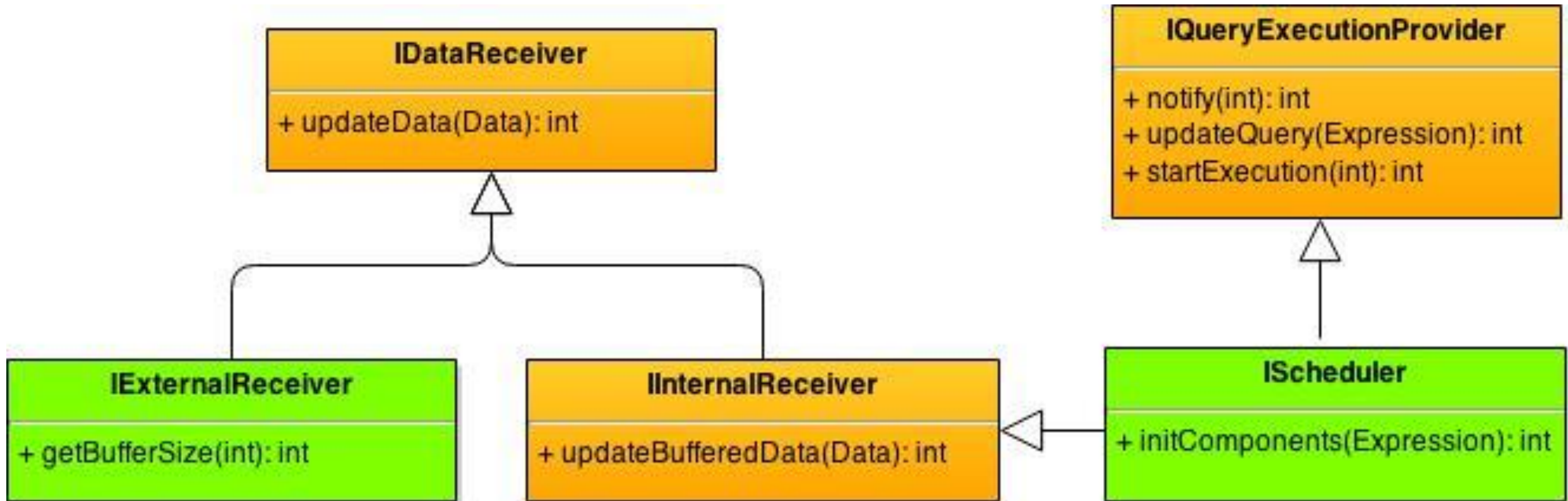


Поддержка конвейерной обработки данных

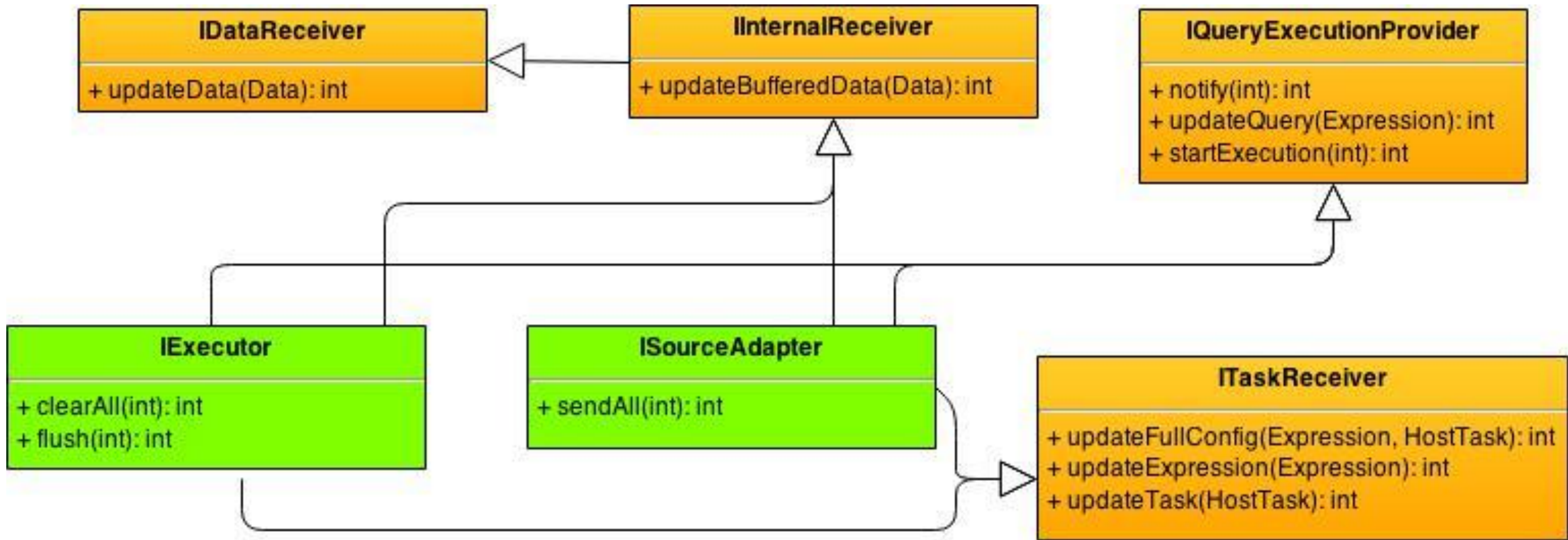
- Уменьшение времени отклика
- Балансировка загрузки хостов системы
- Возможное увеличение необходимого объема ресурсов



Взаимодействие с внешними компонентами системы

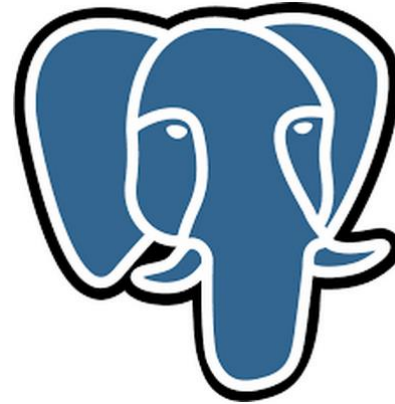
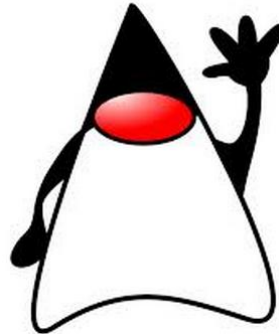


Исполнители и адаптеры источников

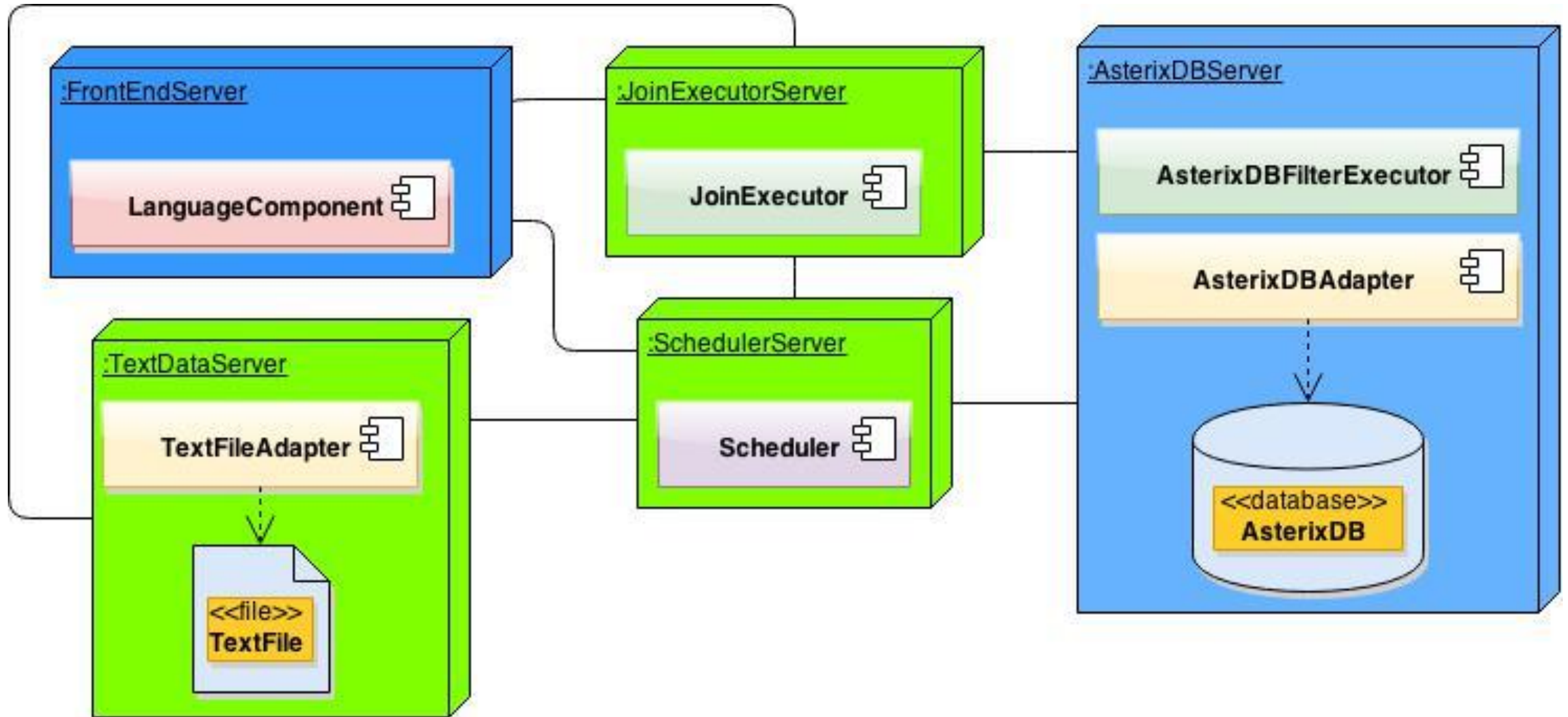


Технологии реализации

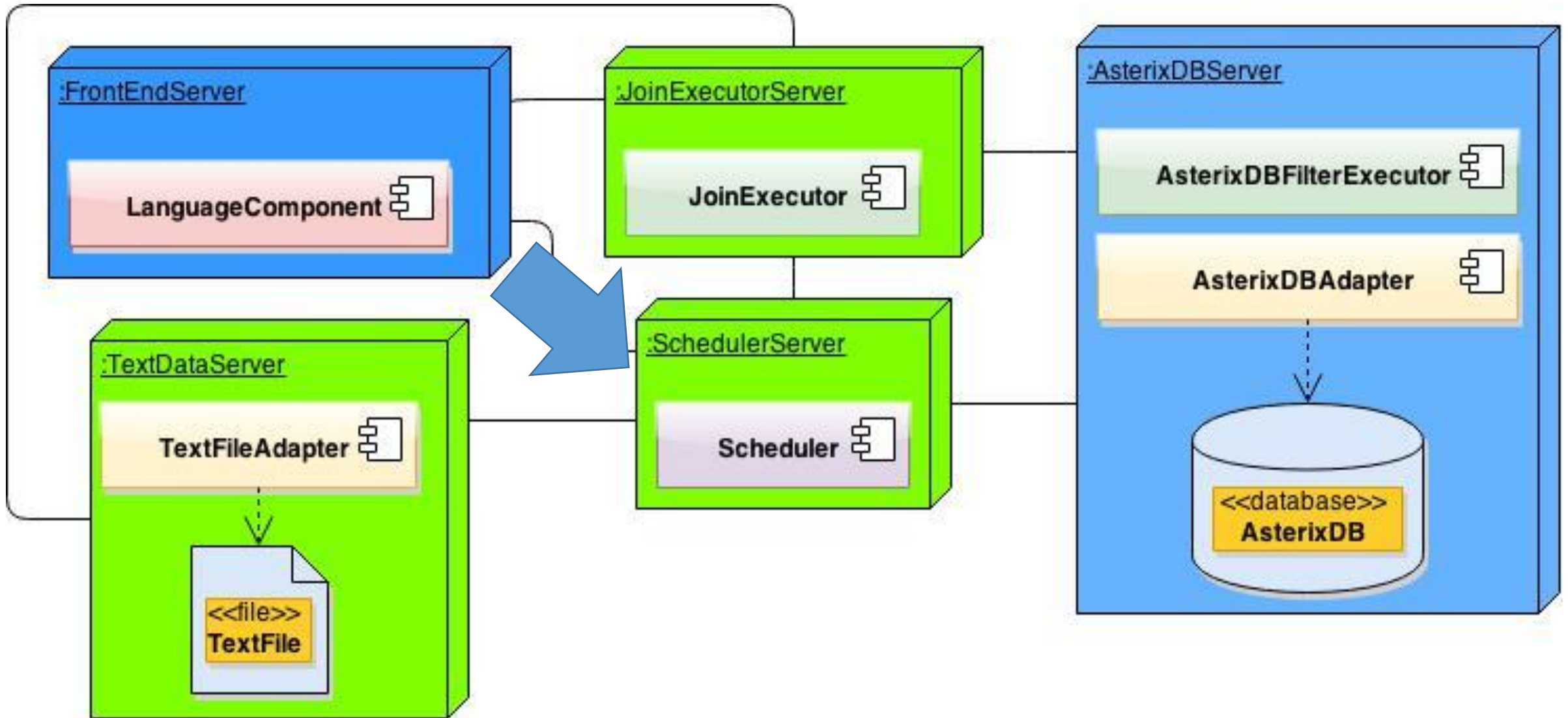
- Apache Thrift
- C++
- Java
- AsterixDB
- PostgreSQL



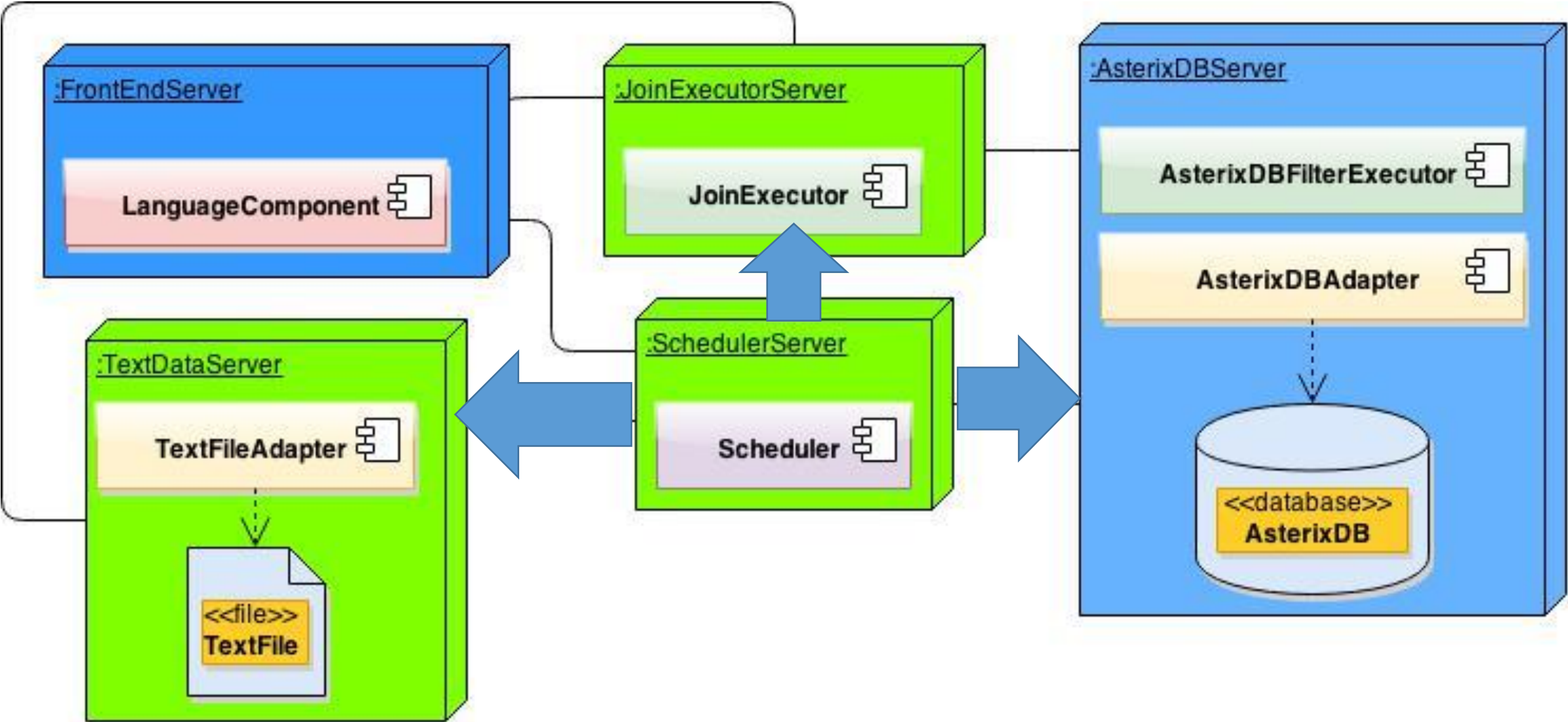
Апробация #1



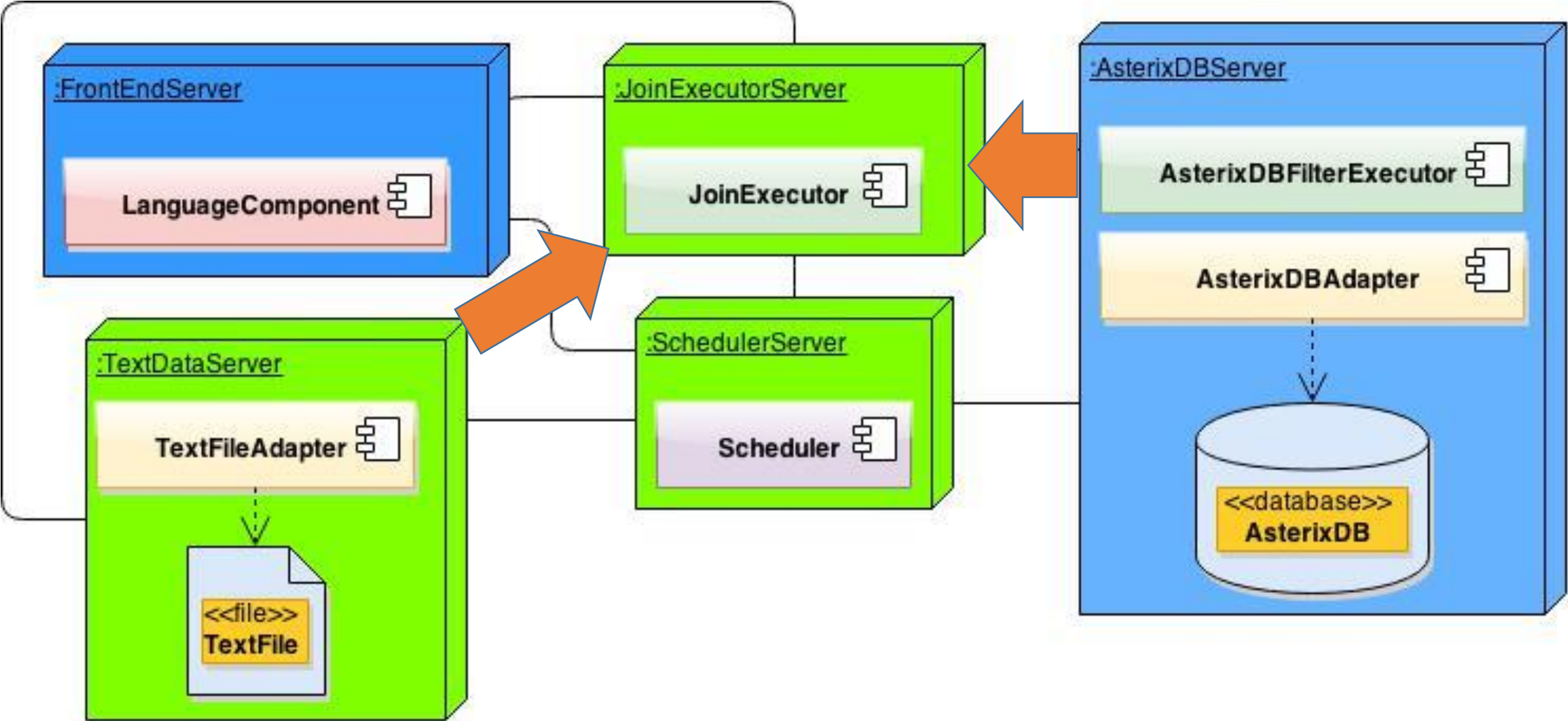
Апробация #2



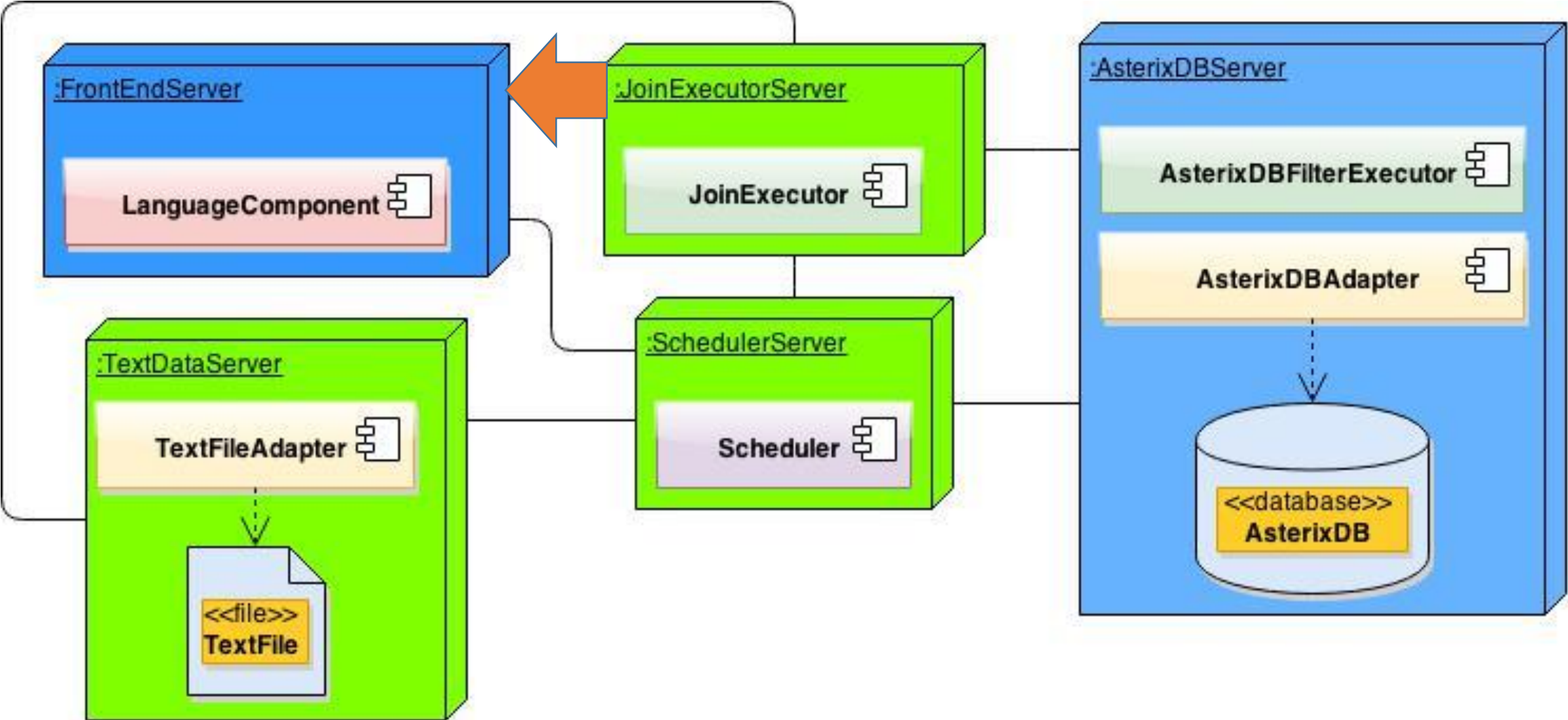
Апробация #3



Апробация #4



Апробация #5



Результаты

- Разработана архитектура ядра исполнения запросов
- Обеспечена поддержка конвейерной обработки данных в рамках разрабатываемой архитектуры
- Обеспечена возможность интеграции с языковыми компонентами и оптимизатором
- Обеспечена возможность интеграции новых исполнителей и адаптеров источников
- Произведена апробация разработанной архитектуры при взаимодействии с реализованным ранее исполнителем/источником данных на основе AsterixDB и высокоуровневыми языковыми средствами для описания запросов
- Результаты данной работы представлены на ISSAT2015