

Опыт построения корпоративных хранилищ данных на платформе Teradata

Андрей Прохоров

Консультант

Компания NCR подразделение Teradata

E-mail: ap185022@ncr.com

You've never seen your business like this before.



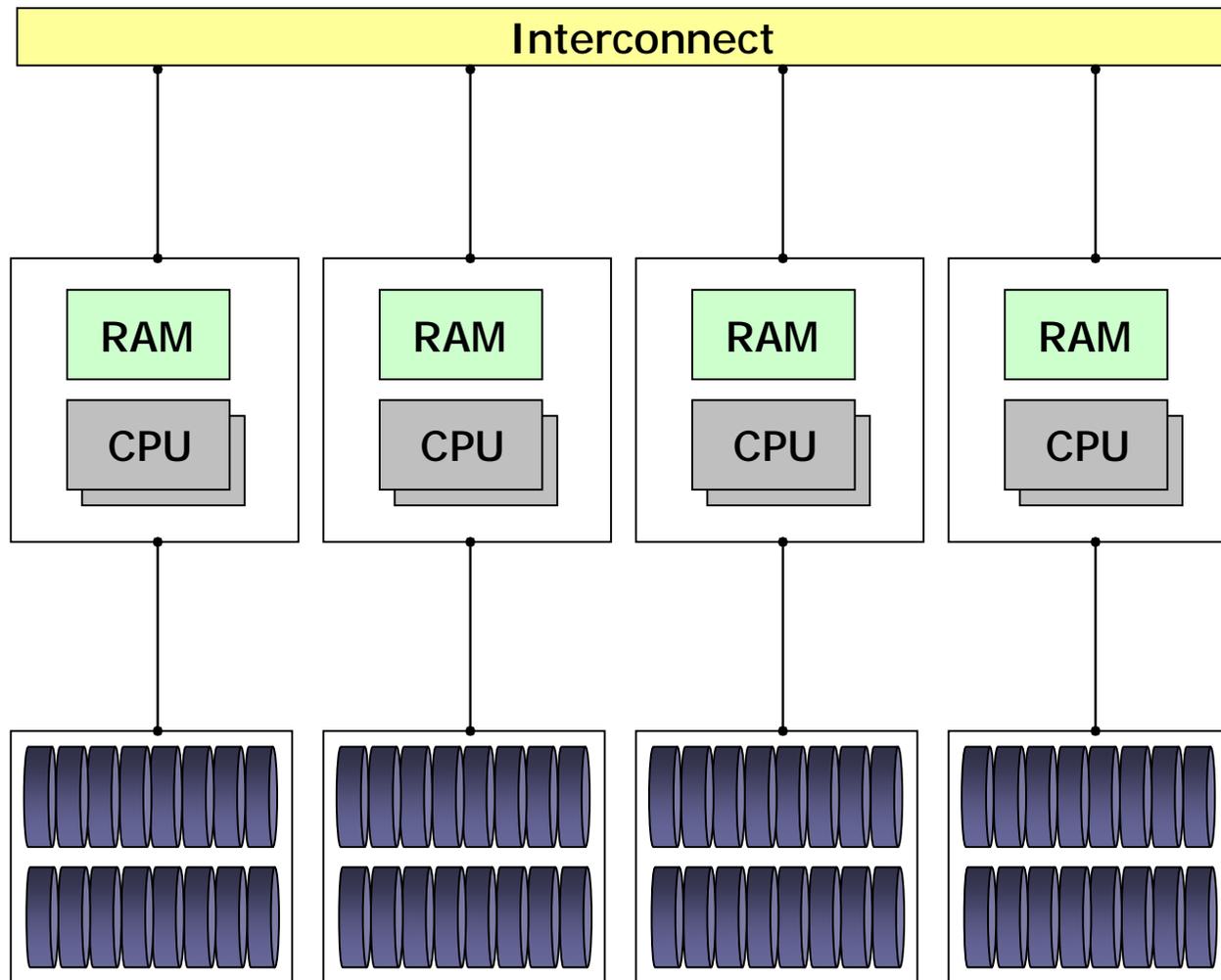
Программа

- История Teradata
- Архитектура СУБД
- Централизованная архитектура корпоративного ХД
- Модель данных
- Организация ETL процессов
- Заключение

История Teradata

- 1979 – Основание Teradata Corp в Лос-Анжелесе, Калифорния
Начало разработки массивно параллельной системы
- 1984 – Teradata выпускает первую систему DBC/1012
Первым заказчиком становится Wells Fargo Bank of Cal.
- 1989 – Начало партнерства Teradata и NCR для выпуска следующей версии DBC
- 1991 – NCR Corporation поглощается AT&T
- 1992 – Teradata поглощается NCR
- 1996 – AT&T выделяет NCR Corp. вместе с подразделением Teradata
- 1997 – Teradata становится признанным лидером в технологии DW
- 2000 – в промышленную эксплуатацию введена система объемом более 100 ТБ
- 2004 – В России открыто представительство подразделения Teradata
- 2005 – Первый заказчик Teradata в России

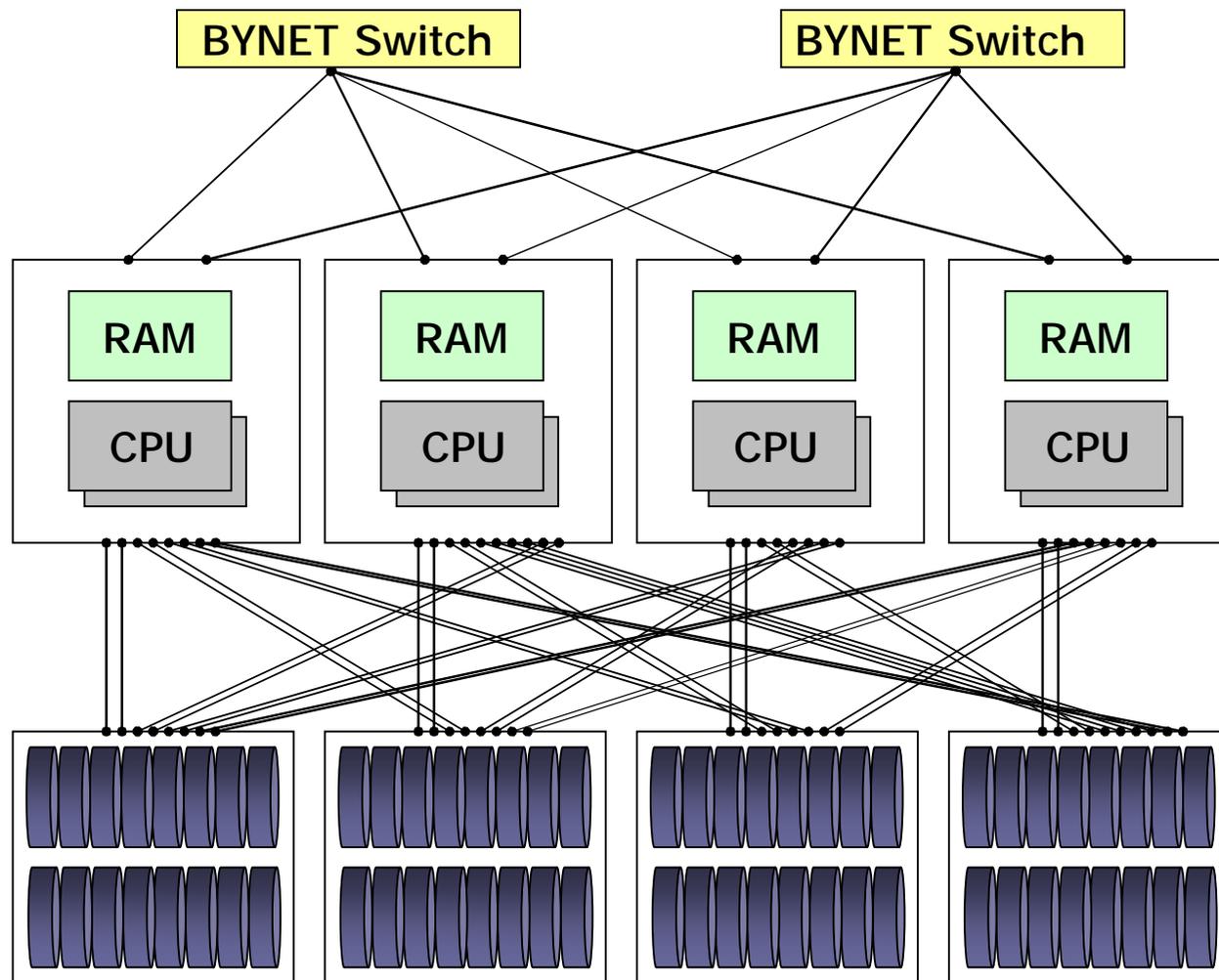
Популярное представление об архитектура shared nothing



Типичные сомнения в архитектуре shared nothing

- Масштабируемость межузлового соединения
- Аппаратная надежность
- Возможность равномерного распределения нагрузки

Пример аппаратной архитектуры Teradata для 4-х узлов



Единственная патентованная аппаратная компонента

Fiber Channel

2-х процессорные Intel-серверы на базе:

- MP-RAS
- Linux
- Windows 2003

Fiber Channel

Системы хранения:

- Enginio
- EMC

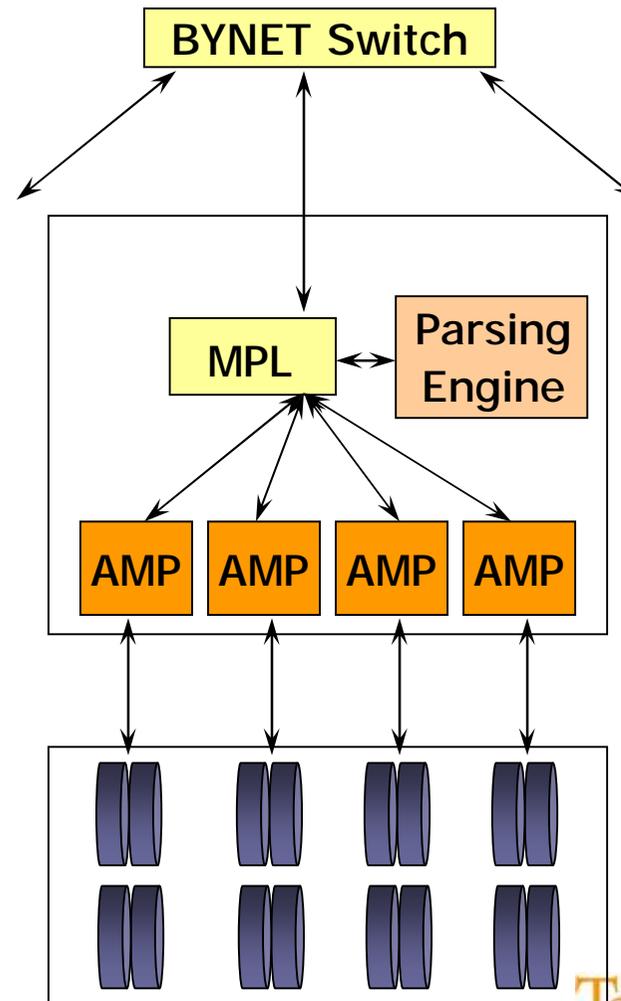
RAID 1

Возможности аппаратной архитектуры

- Линейная масштабируемость до 1024-х узлов
- Отсутствие **single point of failure**
- Равномерное распределение данных при помощи хэширования

Программная архитектура Teradata

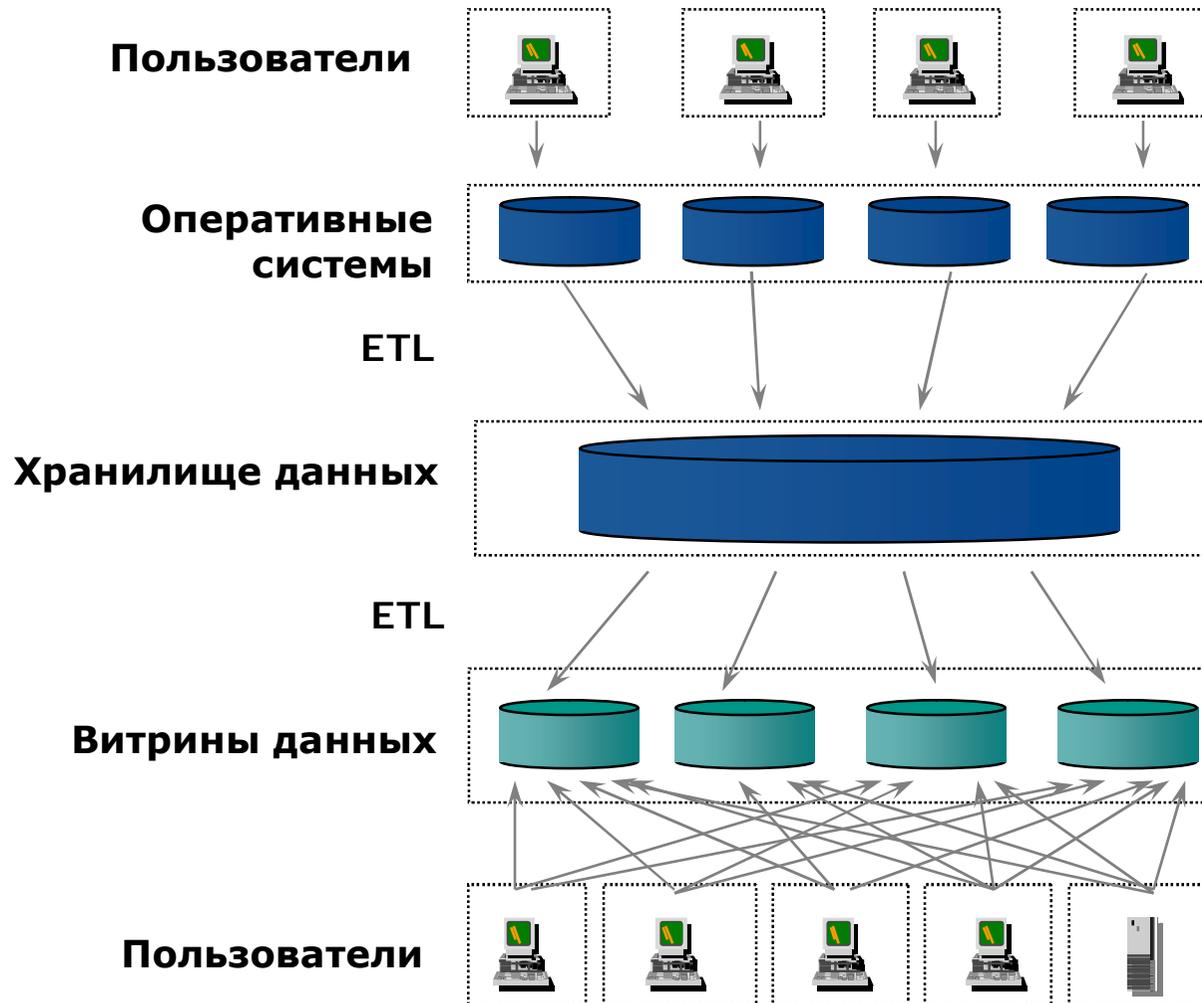
- **Message Parsing Layer**
 - > Коммуникационная компонента
- **Parsing Engine**
 - > Управление сессиями
 - > Построение плана запроса
 - > Создание задач для каждого AMP
 - > Сбор результатов от AMP-ов и их передачу клиенту
- **Access Module Processor**
 - > Выполнение инструкций PE
 - > Управление доступом к выделенной области дискового пространства



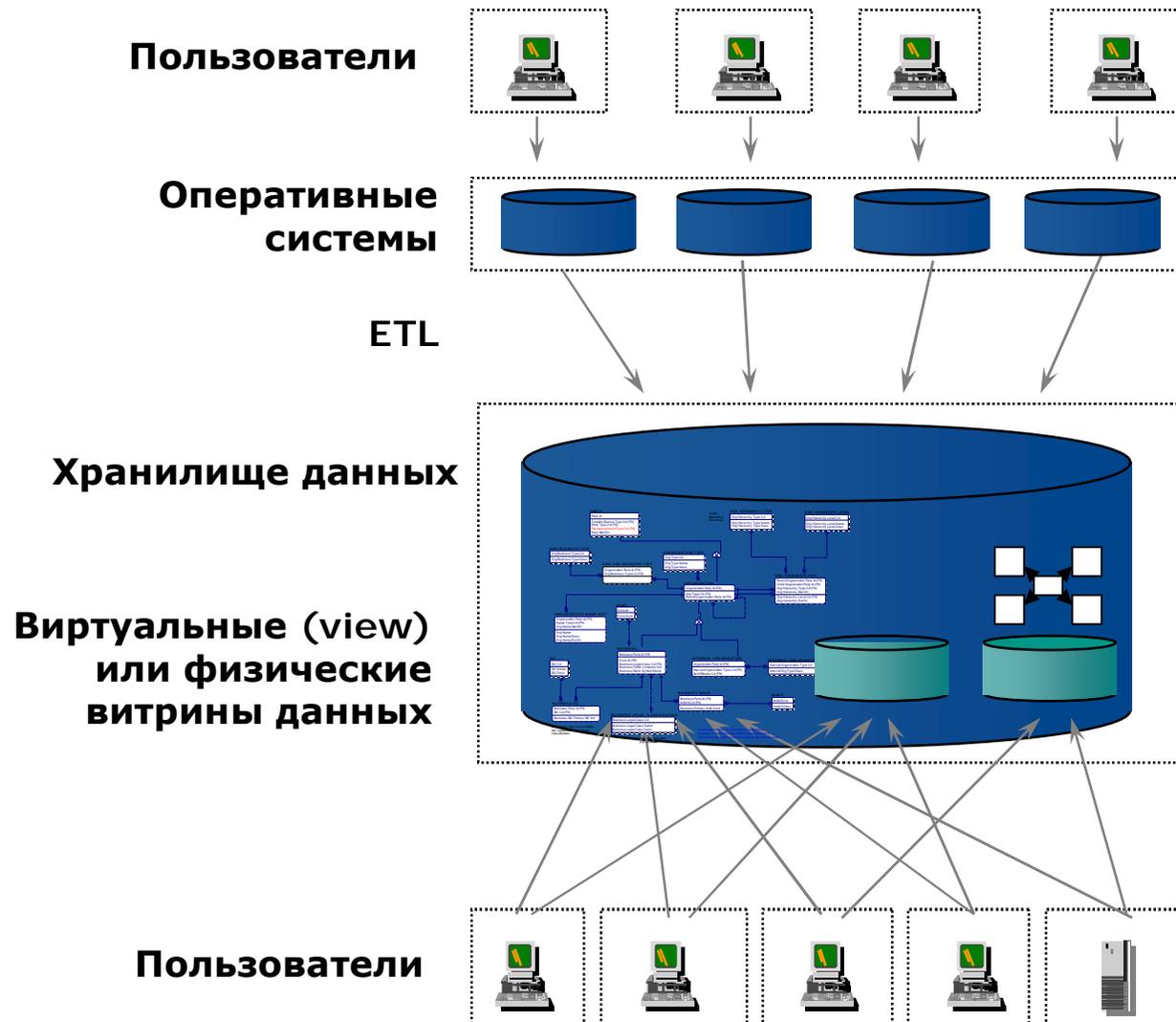
Главные особенности СУБД Teradata

- Специализация на системах поддержки принятия решения
- Высокая производительность
- Линейная масштабируемость
- Простота администрирования
- Поддержка основных стандартов SQL, ODBC, OLE DB, JDBC, .Net Provider

Популярная архитектура корпоративного хранилища данных



Рекомендуемая архитектура корпоративного хранилища данных на платформе Teradata

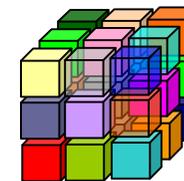
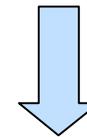
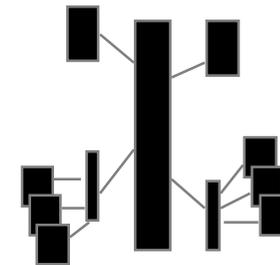


Причины использования централизованной архитектуры

- Преимущества
 - > Возможность доступа бизнес-пользователей к полной детальной информации
 - > Возможность перераспределения ресурсов
 - > Ограниченное количество администрируемых систем
 - > Минимальный маршрут данных в ETL процессе от источника до бизнес-пользователя
- Требования
 - > Линейная масштабируемость платформы
 - > Эффективные алгоритмы обработки запросов
 - > Средства управления ресурсами и нагрузкой

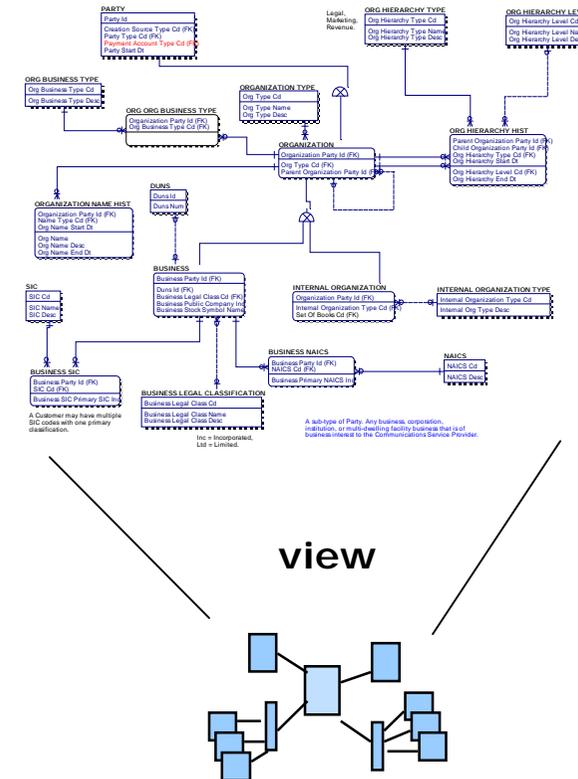
Многомерная модель данных

- Классические подходы
 - > Star schema
 - > Snowflake schema
- Преимущества
 - > Признание классиков
 - > Доступность для понимания
 - > Легкость интеграции с OLAP-средствами
- Недостатки
 - > Недостаточная гибкость
 - > Избыточность

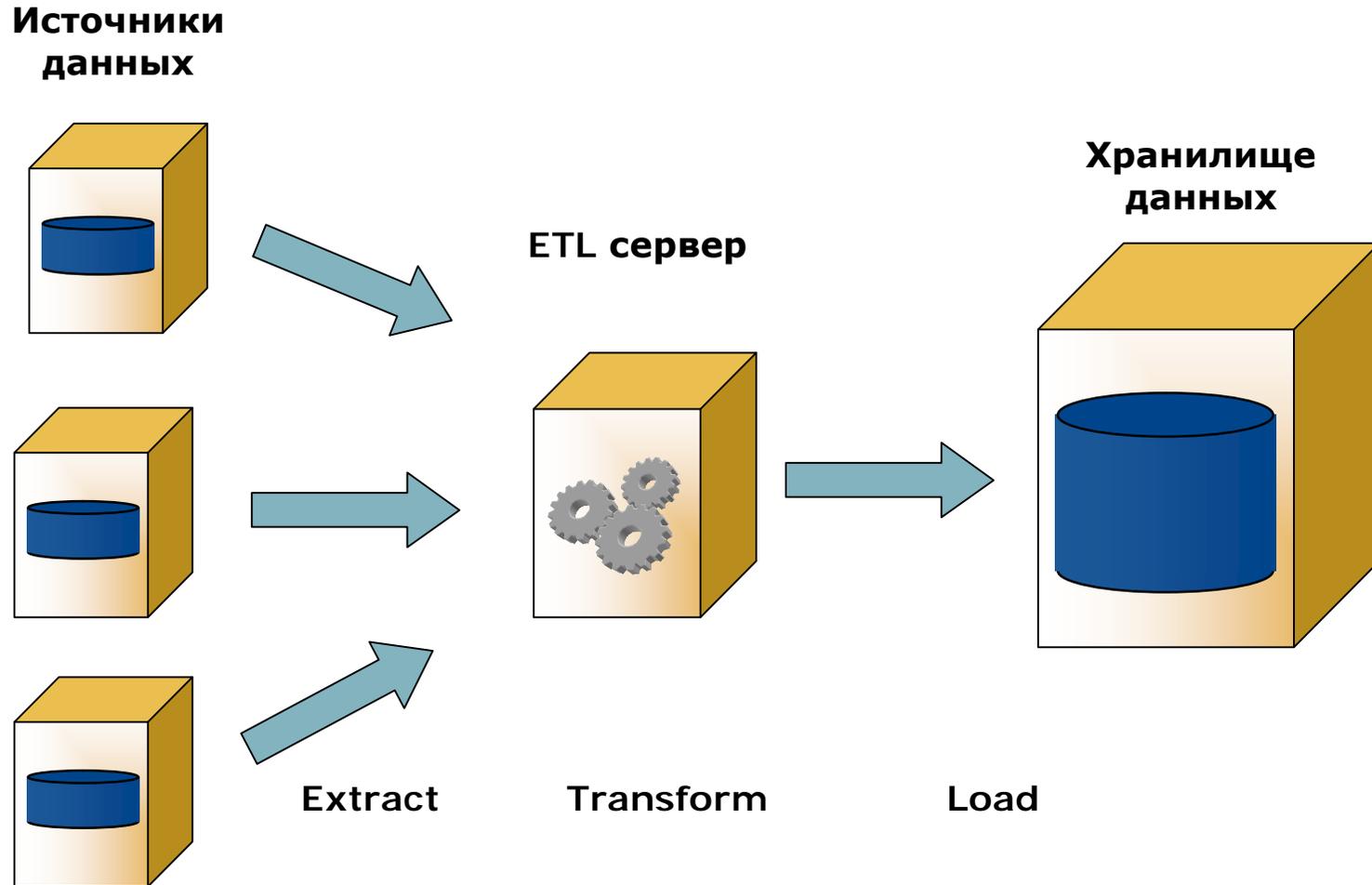


3NF - рекомендация Teradata

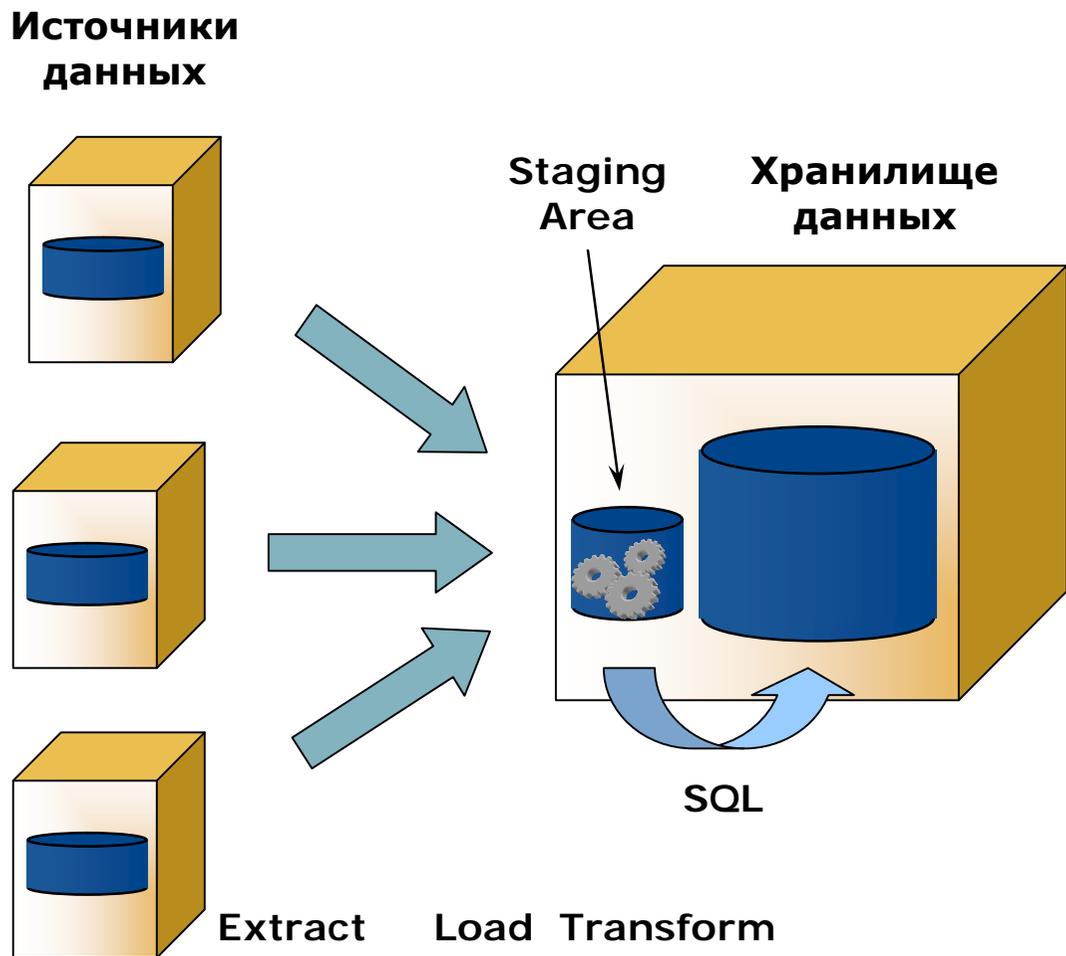
- Гибкость модели
 - > минимальные изменения при расширении
 - > Отсутствие искусственных ограничений
 - SCD
 - Петли
- Неизбыточность
- Возможно построить многомерную модель через логические витрины данных



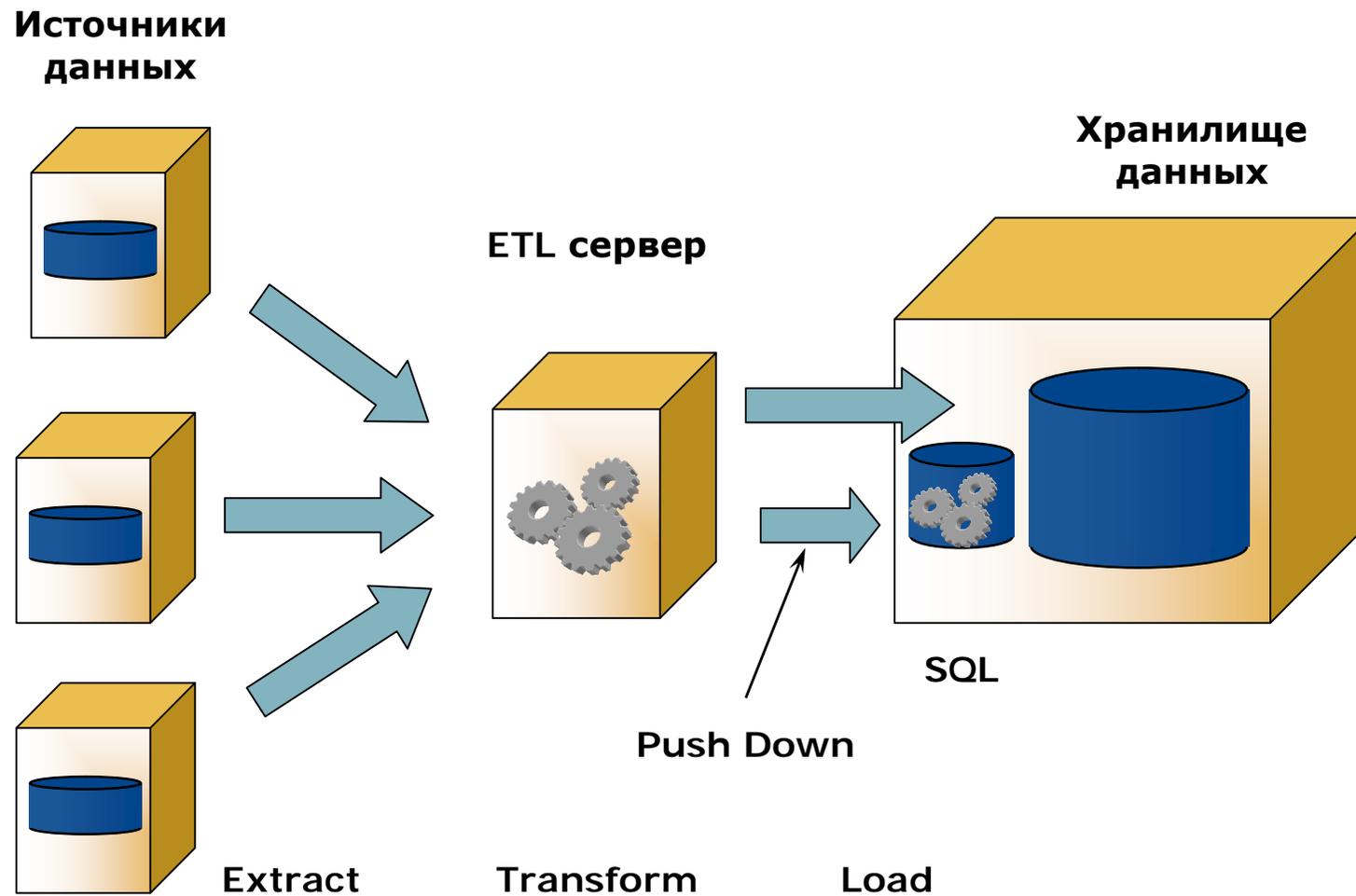
Архитектура Extract Transform Load



Архитектура Extract Load Transform



Архитектура Extract Load Transform + Push Down



Достоинства и недостатки

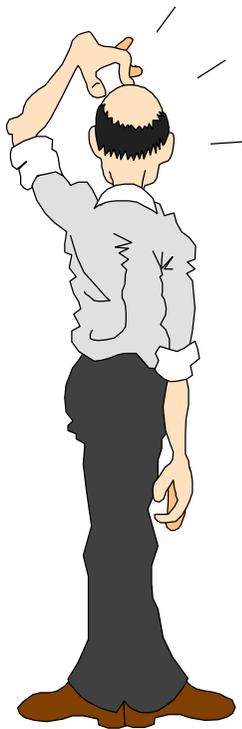
ETL

- (+)
 - > Большой выбор средств
 - > Множество готовых интерфейсов к источникам
- (-)
 - > Немногие средств обладают высокой производительностью
 - > Взрывная нагрузка на выделенное оборудование

ELT

- (+)
 - > Высокая производительность
 - > Эффективное использование оборудования
- (-)
 - > Ограниченный выбор средств
 - > Не все преобразования эффективно реализуются при помощи SQL

Опыт построения корпоративных хранилищ данных на платформе Teradata



Вопросы?

Teradata
a division of  NCR

Андрей Прохоров

Консультант

NCR

115054, Москва

Космодамианская наб. 52,
стр. 5, этаж 7

Тел.: +7 495 961 3027

Факс: +7 495 961 3031

E-mail: ap185022@ncr.com

www.teradata.com

Teradata
a division of  NCR