

Проектирование программных систем

Хританков А.С.

Системы обработки массивов данных

- Медленная сеть
 - Изменение соотношения мощности процессора к пропускной способности сети и объему дисков
 - Возврат к централизованной обработке, перенос вычислений ближе к данным
- Разнородность вычислений
 - CPU, GPGPU, ASIC, FPGA, Mobile
 - Масштабируемые распределенные вычисления
 - Повсеместные вычисления
- Большие данные
 - Нечисловые и неструктурированные данные
 - Скорость поступления превышает возможности по размещению (медленный диск)
 - Достоверность и корректность данных неизвестна
 - Объем данных требует горизонтального масштабирования
 - Не понятна ценность собираемых данных

Архитектура систем обработки массивов данных

- Не программные а информационные системы
 - Архитектура приложений
 - Логическая структура: потоки данных и модули
 - Платформа реализации
 - Алгоритмическое обеспечение
 - Архитектура данных
 - Организационная архитектура
 - Физическая архитектура

Методы проектирования

- Data-Flow Diagrams
 - Пояснения по нотации
 - Пример на открытую и закрытую модели
- Structured Systems Analysis and Design Method
 - Логическая структура
 - Контекстная диаграмма
 - Транзакционный анализ
- Сопоставление платформе реализации
- Прототипирование, «выращивание» систем, fitness-for-purpose

Программная платформа реализации

- Эталонные архитектуры (по классам систем)
 - Вычислительный кластер (HPC)
 - Кластер обработки данных (Batch processing, M/R, Lambda)
 - Система обработки событий (Time-Series, Event Sourcing, Stream)
 - Business Intelligence (Reporting, Predictive analytics)
 - Хранилище данных (Data lake, DWH / EDW / Data Fed, Polyglot Persistence)
- Атрибуты качества для систем обработки данных
 - Scalability, Reliability, Security, Efficiency
- Применение Attribute-Driven Design (дополнительно)
 - ADD v2.0 -> v3.0
 - Big-Data Architecture Design (BDD)

Организационная структура (опционально)

- Общие
 - Контроль и обеспечение качества
 - Безопасность и защита информации
 - Управление инфраструктурой
- BI часть
 - Управление информацией и знаниями
 - Базовые и производные данные
 - Жизненный цикл данных
 - Апробация и исследование гипотез
- BigData часть
 - Получение и хранение данных
 - Промышленное применение
 - Проверка и переход к применению результатов

Физическая архитектура (опционально)

- Проектирование ЦОД
- Проектирование НРС и Grid систем
- Облачные вычисления
 - Совместное использование вычислительных ресурсов
 - Промежуточное программное обеспечение
 - XaaS: IaaS, PaaS, SaaS

Документирование

- Зачем нужна документация:
 - обучение, коммуникации со stakeholders, сопровождение и развитие, анализ и моделирование (vs. мысленные модели)
- Документирование архитектуры
 - Перспектива (viewpoint) и представление (view)
 - Пример с проекциями детали
 - Виды перспектив IEEE: структуры, поведение, размещения
 - Система перспектив 4+1 (Крютчен, RUP)

Литература

- Modern Structured Analysis, E. Yourdon
- A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)
- Documenting Software Architectures: Views and Beyond, Second Edition.
- Статья крютчена про 4+1