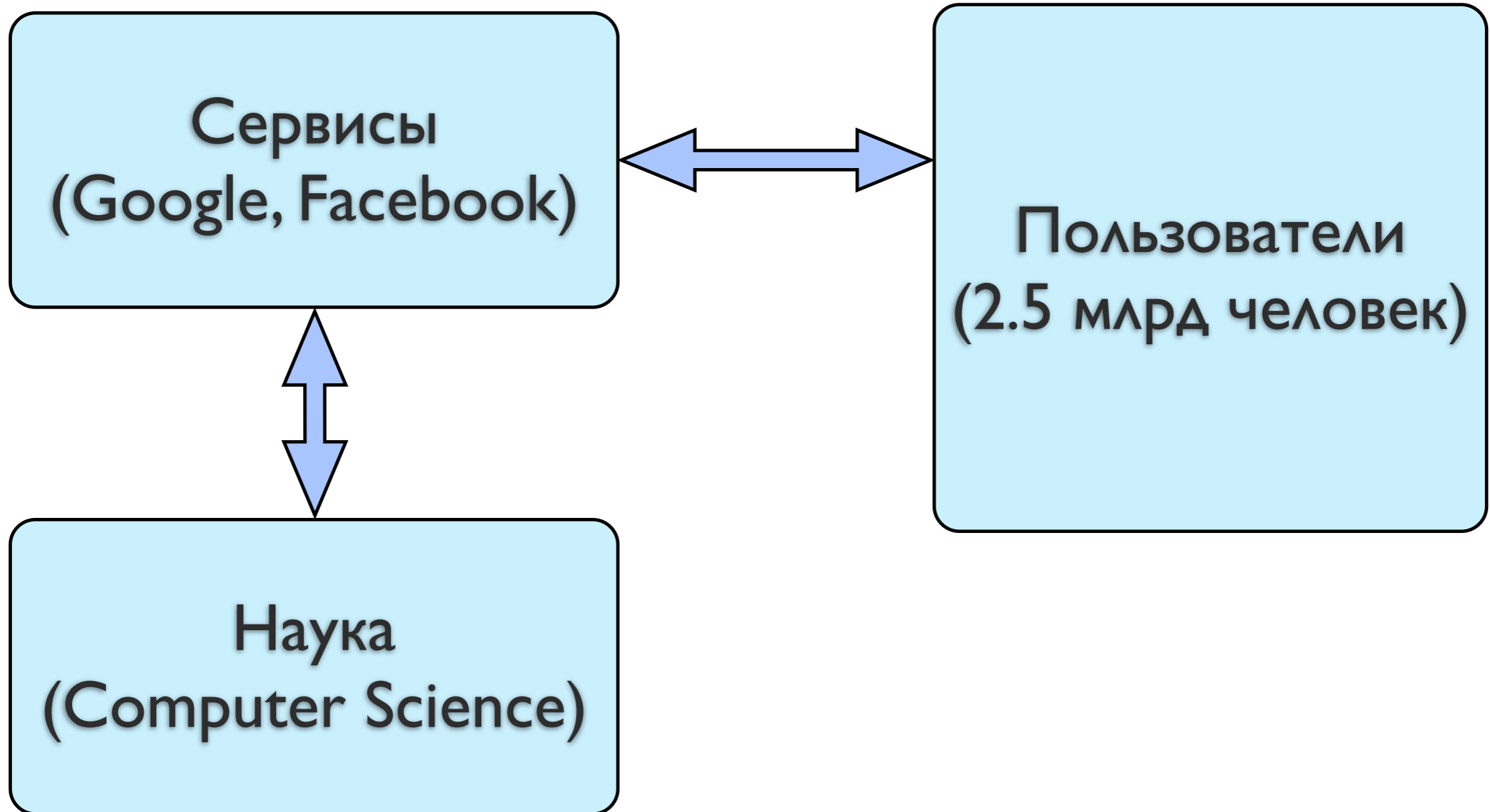
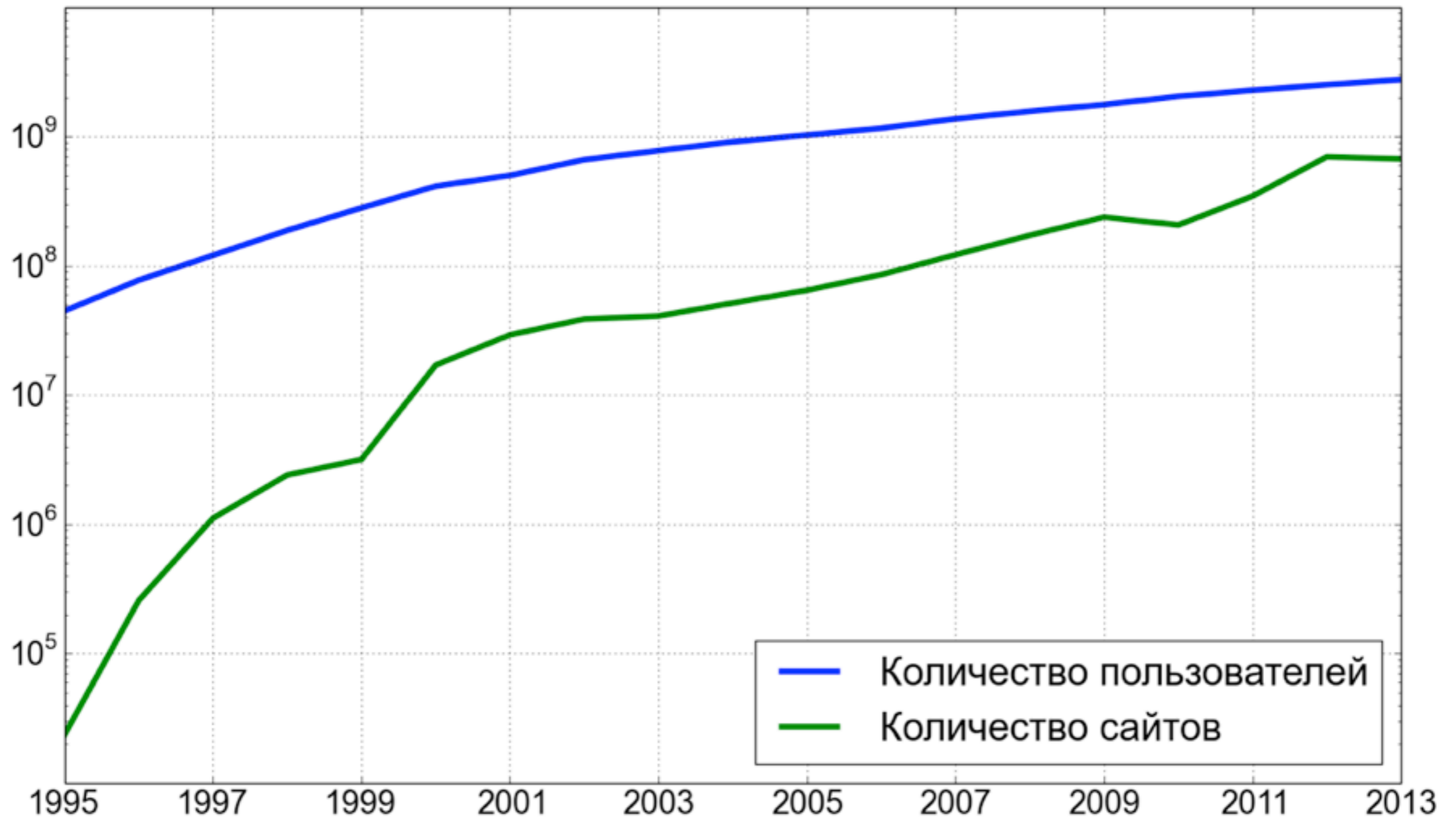


Мир IT



История



История

	1995	2014
RAM	64Mb	128Gb
CPU	200Mhz	100Ghz
HDD	64Gb	10Tb

История

Объем, которые хранят крупнейшие компании

1995	2014
1-10 Тб	1000-10000 Пб

Причины роста данных

- Рост активность пользователей
- Персонализация
- Новые типы данных

Big Data и индустрия

Скорость обработки



Количество данных

Требования индустрии

- Надежность хранения данных
- Простота обработки
- Высокая доступность

ОТВЕТ ИНДУСТРИИ

- MapReduce, Hadoop
- BigTable, HBase
- MealWheel, Storm
- Omega, Mesos
- Spanner

Идея MapReduce

- Не надо думать о распределенность
- Данные должны быть доступны всегда
- Над данными может одновременно работать много людей

Google Spanner

Google's scary powerful Spanner database gets us one step closer to Skynet

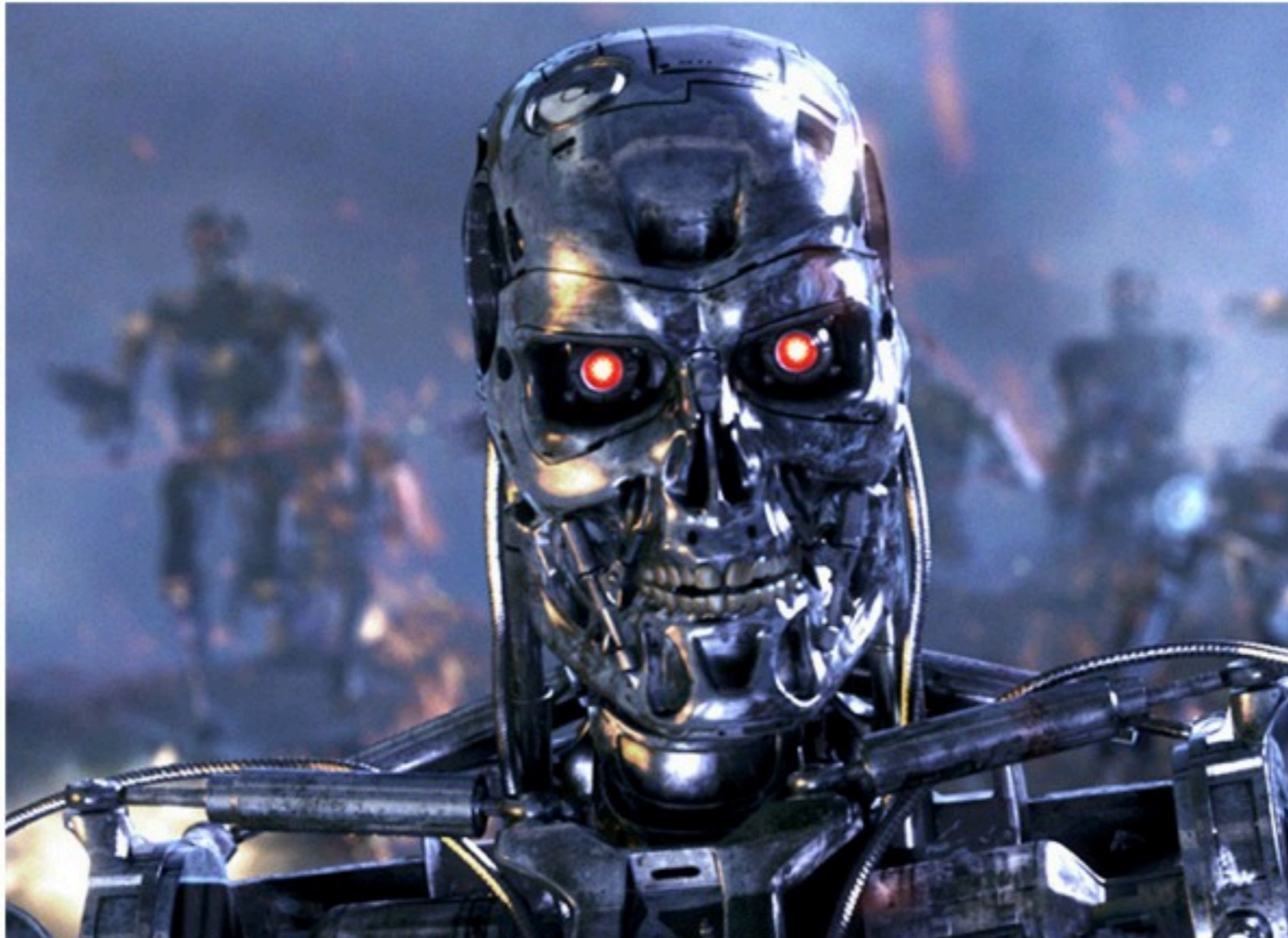
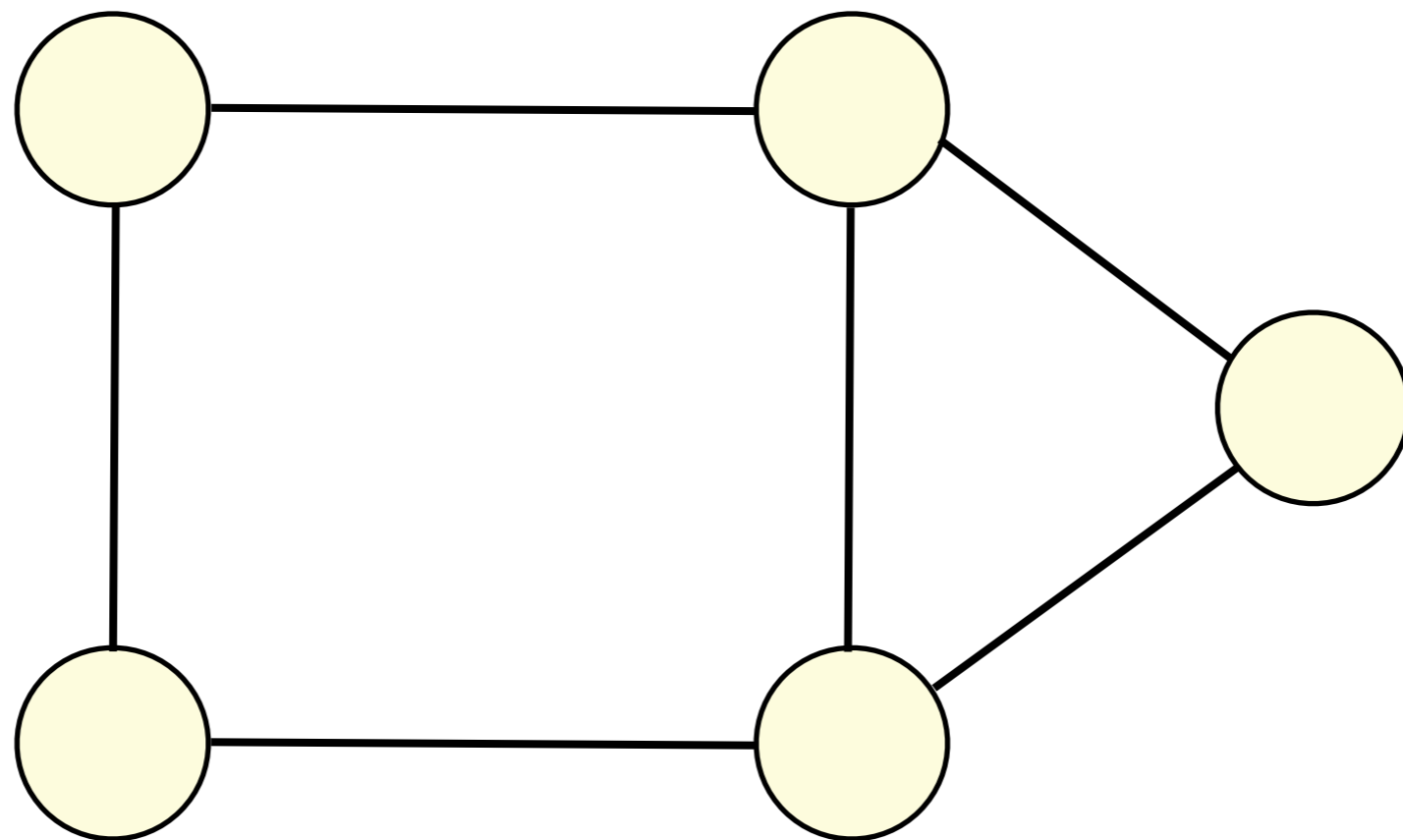


Image Credit: Terminator 3: Rise of the Machines/Warner Brothers

Big Data и наука

Компьютеры соединенные по сети

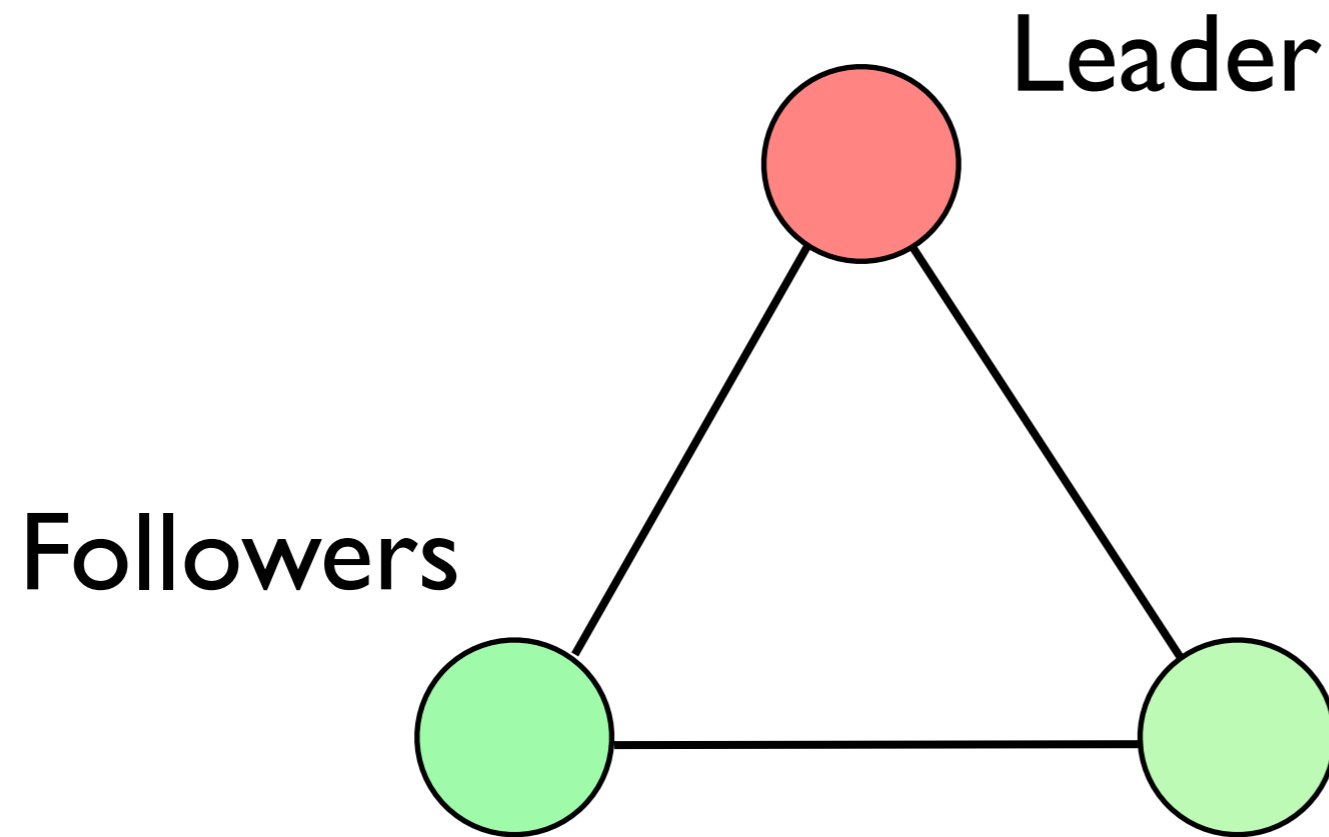


Big Data и наука

Модель распределенных вычислений

- ◆ Компьютеры соединенные по сети
- ◆ Могут ломаться
- ◆ Может пропадать сеть
- ◆ Может отключаться ДЦ
- ◆ Безопасность

Поиск кворума

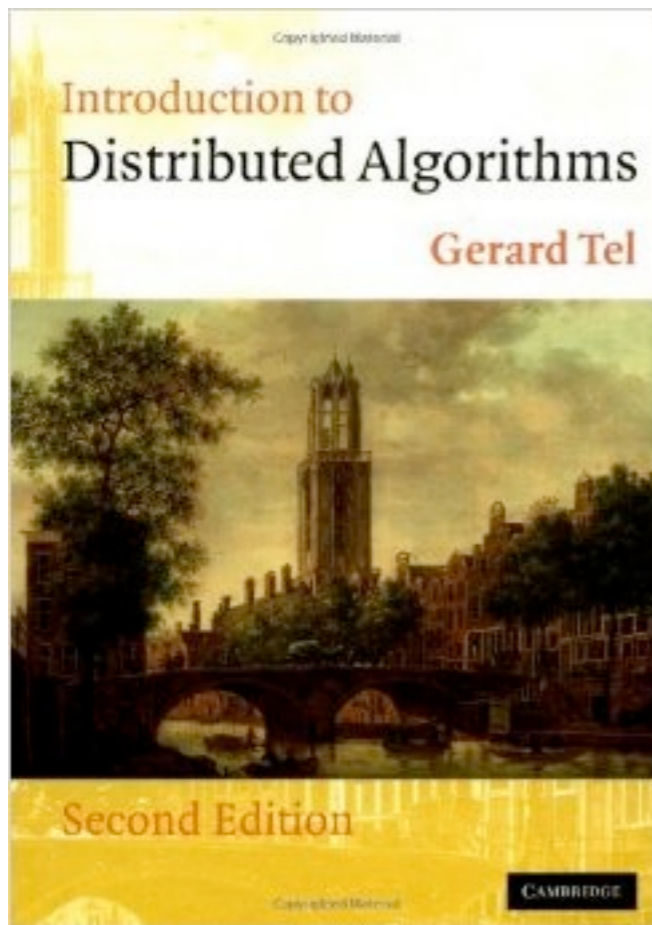


Paxos

L. Lamport – The Part-Time Parliament

“My attempt at inserting some humor into the subject was a dismal failure. People who attended my lecture remembered Indiana Jones, but not the algorithm.”

Распределенные вычисления



Теория:

- Очень сложно
- Непрактичные результаты

Практика:

- Дорого тестировать
- Требуется быстроедействие

Big Data и пользователи

- Важная часть интернета в целом
- Amazon AWS
 - Dropbox
 - HTC
 - AirBnb

Будущее Big Data

- Простые и удобные инструменты
- Рост разнообразного контента и способов его обработки
- Больше взаимодействия между индустрией и наукой