

Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М. А. Бонч-Бруевича»
Кафедра Безопасности информационных систем

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Сетевые информационные технологии

Лекция 12

Компьютерные сети

Кривцов Александр Николаевич

an.krivtsov@gmail.com

План лекции

- **Компьютерные сети. Основные понятия и технологии**
- **Облачные вычисления**
- **Проблемы и перспективы поиска и обработки данных**

1. Эволюция компьютерных сетей

Первые компьютеры **1950-х гг.** основным предназначением являлось небольшое число избранных операций. Данные компьютеры не применялись для интерактивной работы пользователя, а использовались в режиме пакетной обработки. Системы пакетной обработки обычно строились на базе **мейнфрейма**, который является мощным и надежным компьютером универсального назначения. (перфокатор)

Пакетный режим являлся самым эффективным режимом использования вычислительной мощности, так как он позволял выполнить в единицу времени больше пользовательских задач, чем любые другие режимы.

Во главе угла находилась **эффективность работы самого дорогого устройства вычислительной машины**, которым являлся процессор, в ущерб эффективности работы использующих его специалистов.

1. Эволюция компьютерных сетей

В начале **1960-х гг.** появились новые способы организации вычислительного процесса, позволяющие учесть **интересы пользователей**. Началось **развитие интерактивных многотерминальных систем разделения времени**. В данных системах на компьютере работали сразу несколько пользователей. Каждый из них получал в распоряжение терминал, который помогал ему производить общение с компьютером. Несмотря на то что вычислительная мощность оставалась полностью

В данный период был справедлив так называемый **закон Гроша**, эмпирически отражающий уровень технологии того времени. По этому закону **производительность компьютера** была **пропорциональна квадрату его стоимости**, следовательно, за одну и ту же сумму было выгоднее купить одну мощную машину, чем две менее мощных, так как их суммарная мощность оказывалась намного ниже мощности дорогой машины.

1. Эволюция компьютерных сетей

В **1970-х гг.** наблюдался **технологический прорыв в сфере производства компьютерных компонентов**, что выразилось в **появлении БИС**. Их небольшая стоимость и огромные функциональные возможности позволили создать **мини-компьютеры**, которые стали реальными конкурентами мейнфреймов. Со временем потребности пользователей вычислительной техники увеличивались, появлялась необходимость получения возможности обмена данными с другими близко расположенными компьютерами.

Изначально для соединения компьютеров друг с другом применялись самые разнообразные **нестандартные устройства с собственными способами представления данных** на линиях связи, своими типами кабелей и т. п. Такие устройства были способны **соединять только те типы компьютеров, для которых были разработаны**

1. Эволюция компьютерных сетей

В **1980-х гг.** Появились **стандартные технологии объединения компьютеров в сеть** – **Ethernet, Arcnet, Token Ring**. Сильный импульс для их развития дали ПК. Персональные компьютеры стали преобладать в локальных сетях, при этом не только как клиентские компьютеры, но и как центры хранения и обработки данных, т. е. сетевых серверов, потеснив при этом с привычных ролей мини-компьютеры и мейнфреймы .

Для того чтобы создать сети, **достаточно было приобрести сетевые адаптеры соответствующего стандарта**, например Ethernet, **стандартный кабель**, соединить адаптеры и кабель **стандартными разъемами** и установить на компьютер **какую-либо из имеющихся сетевых операционных систем**, например NetWare.

1. Эволюция компьютерных сетей

В настоящий момент вычислительные сети непрерывно развиваются, и достаточно быстро. Соединяющий компьютеры пассивный кабель в них сменили разнообразные типы коммуникационного оборудования – коммутаторы, маршрутизаторы, шлюзы. Из-за использования такого оборудования **появилась возможность построения больших корпоративных сетей**, которые насчитывают тысячи компьютеров и имеют сложную структуру. Вновь **появился интерес к крупным компьютерам**. Системы, которые состоят из сотен серверов, обслуживать сложнее, чем несколько больших компьютеров. Поэтому на новом этапе эволюции **мейнфреймы возвращаются в корпоративные вычислительные системы**. При этом они являются полноправными сетевыми узлами, поддерживающими **стандартные сетевые протоколы** глобальных сетей.

1.1. Основные программные и аппаратные компоненты сети

Сетью называют группу компьютеров, соединенных между собой при помощи специального оборудования, которое, при установленных специальных программных средствах, обеспечивает обмен информацией в реальном режиме времени.

вычислительная сеть является сложным комплексом взаимосвязанных и согласованно функционирующих программных и аппаратных компонентов. Это **многослойная модель**, состоящая из:

- 1) компьютеров;
- 2) коммуникационного оборудования;
- 3) операционных систем;
- 4) сетевых приложений.

1.2. Классификация сетей

Критерием для **классификации** сетей является их **масштаб**.

LAN - *Local Area Network* - локальная сеть, компьютеры расположены на близком расстоянии, как правило в пределах одного здания - ЛКС.

MAN - *Metropolitan Area Network* - районная сеть, компьютеры расположены в пределах городского района или небольшого города.

WAN - *Wide Area Network* - глобальная сеть, охватывающая страну, континент.

HAN - *Home Area Network* - домашняя сеть.

1.3. Локальные компьютерные сети (ЛКС)

Локальная компьютерная сеть объединяет небольшое количество компьютеров и позволяет пользователям совместно использовать ресурсы компьютеров, а также периферийных устройств (принтеров, плоттеров, дисков, модемов и др.), подключённых к сети.

Подключение компьютера к локальной сети осуществляется с помощью специальной платы - **сетевого адаптера**. Основной функцией сетевого адаптера является приём и передача информации в сети.

Некоторые сетевые адаптеры (например **EtherNet**) позволяют объединять в сеть компьютеры различных аппаратных и программных платформ (IBM-совместимые, Macintosh, Unix-компьютеры).

Скорость передачи данных по локальной сети обычно находится в диапазоне от 10 до 100 Мбит/с.

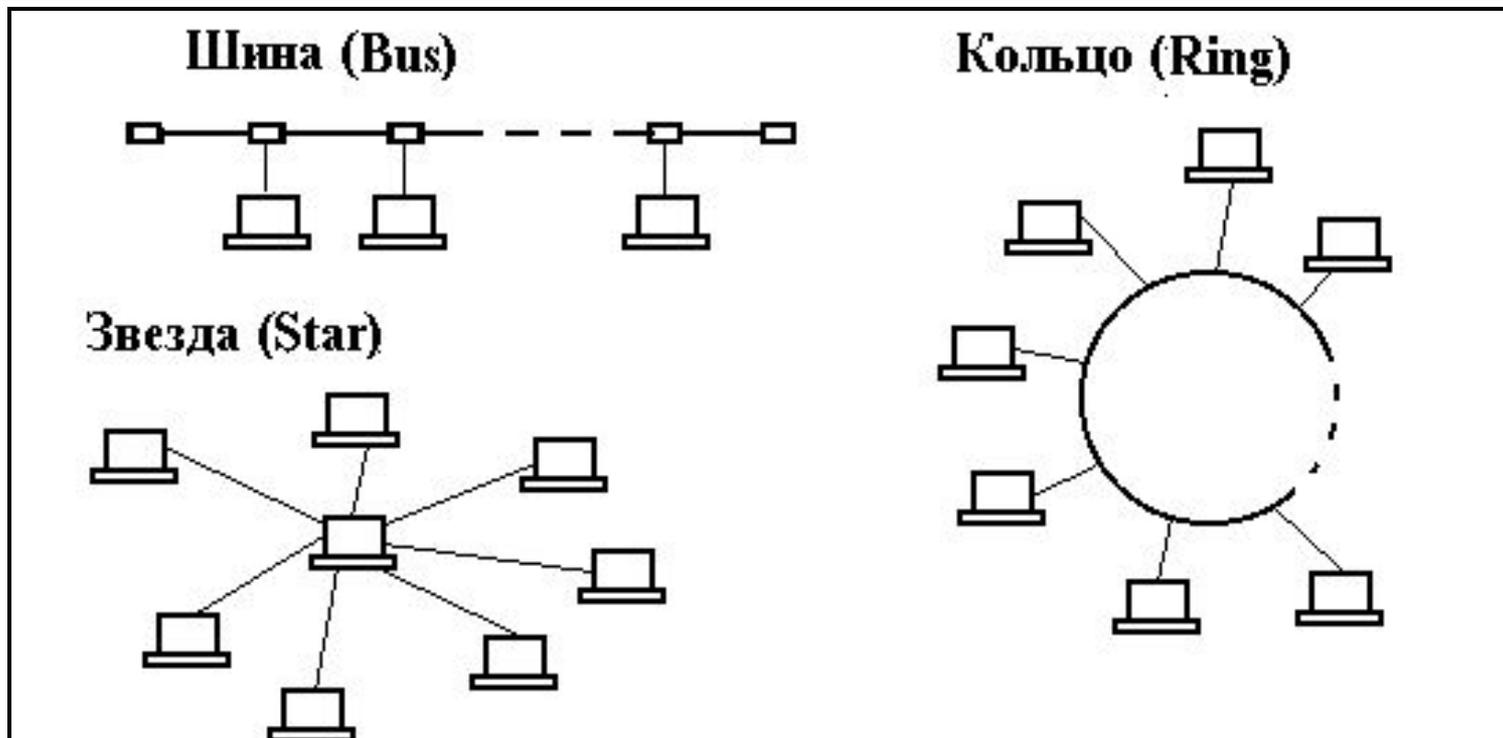
Сетевой адаптер



Топология ЛКС

Существуют три базовые топологии ЛКС:

шина, звезда и кольцо.



Выбор топологии влияет на:

- состав оборудования;

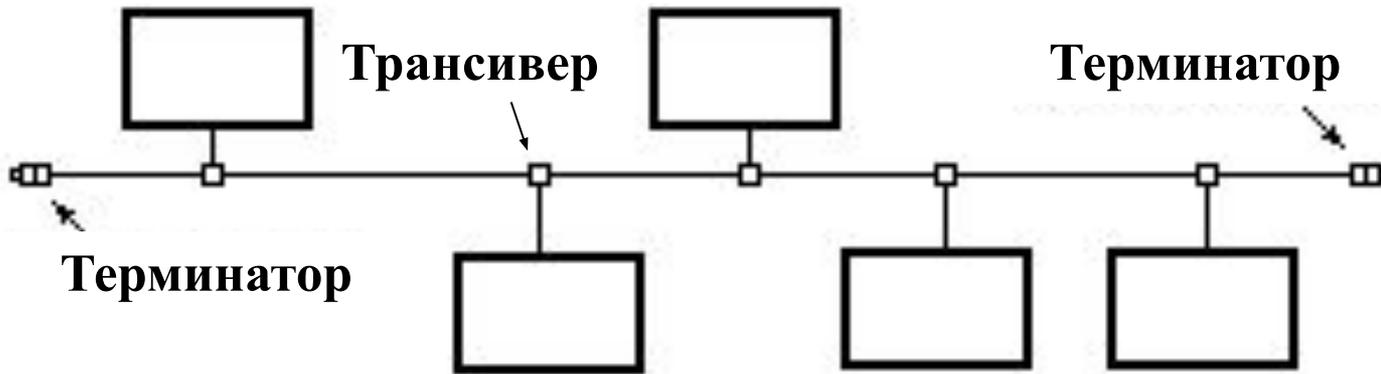
- возможности расширения сети;

- возможности оборудования;

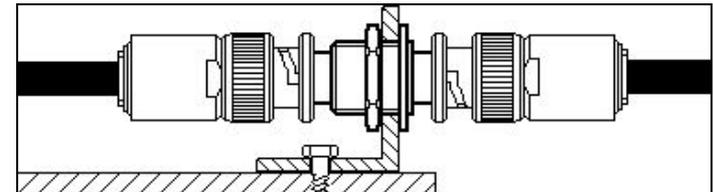
- способ управления сетью.

Топология «Шина»

Один кабель используется всеми рабочими станциями по очереди. Пассивная топология. Сообщение, посылаемое одним компьютером, принимается всеми остальными компьютерами. Компьютеры только «слушают» кабель, но не перемещают сообщения от отправителя к получателю.



Трансивер

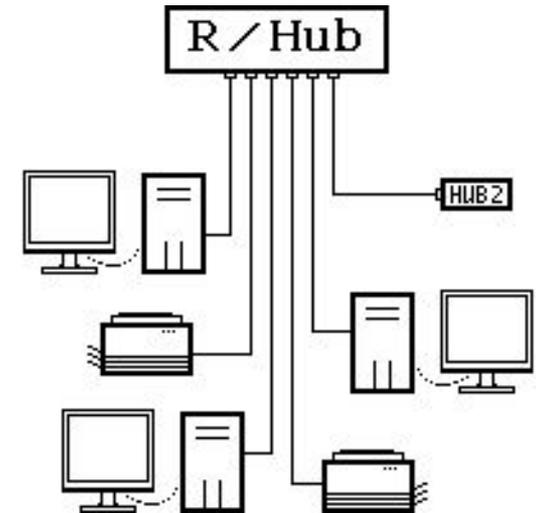
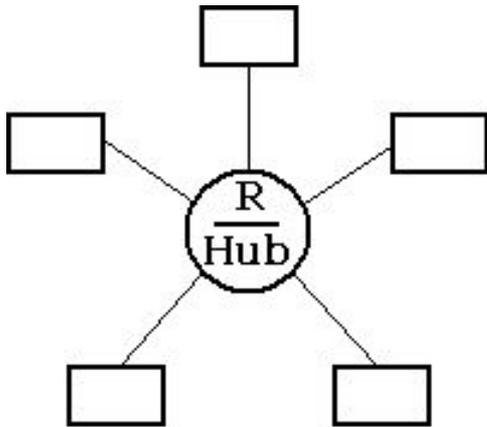


Выход из строя одного компьютера не нарушает работу сети.

Топология «Звезда»

Каждая рабочая станция подключена к объединяющему устройству - **концентратору (*hub*)**. По этой схеме могут быть объединены несколько сетей с образованием разветвленной конфигурации.

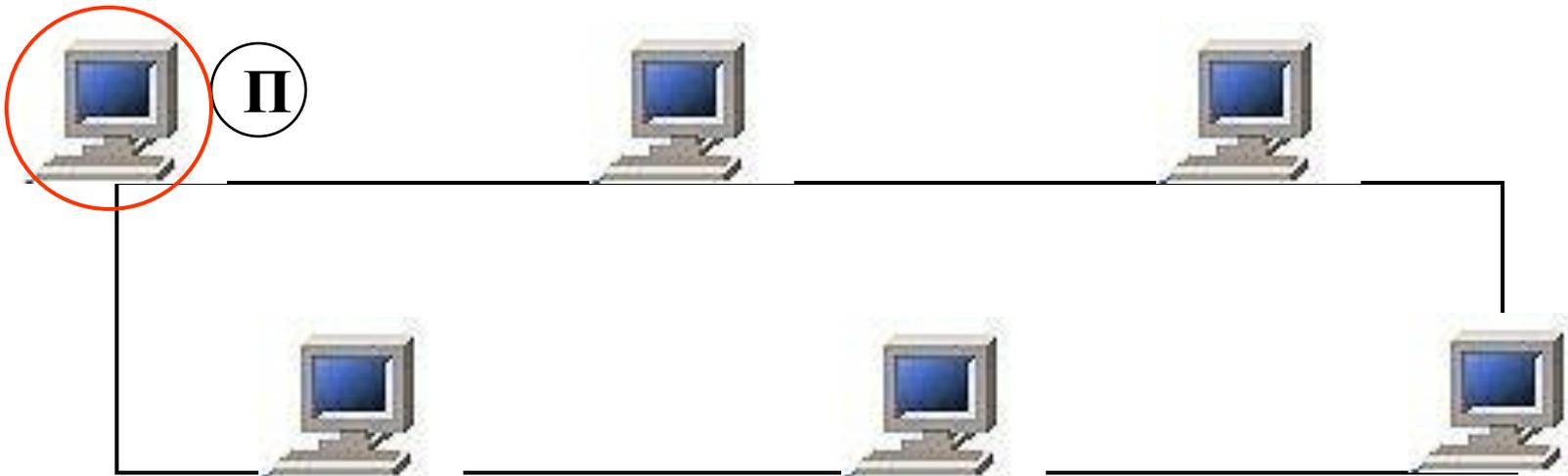
Используются как активные, так и пассивные концентраторы.



Неисправность концентратора выводит из строя всю сеть.

Топология «Кольцо»

Данные, ведомые маркером, передаются последовательно от одной рабочей станции к другой и проходят через каждый компьютер. Пакет, адресованный другой станции, передается дальше до тех пор, пока не достигнет получателя.

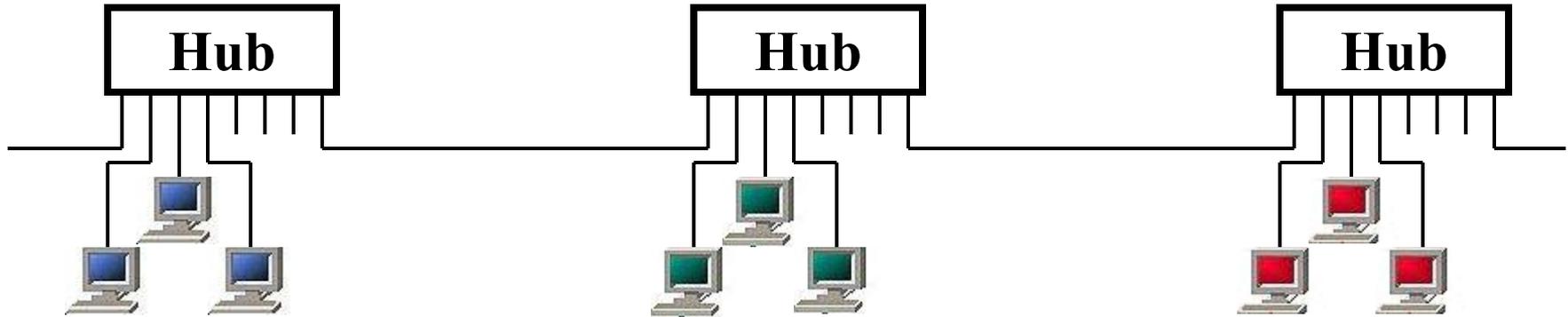


Каждый компьютер выступает в роли повторителя, усиливая сигналы. Неисправность любого компьютера выводит из строя всю сеть.

Комбинированные топологии

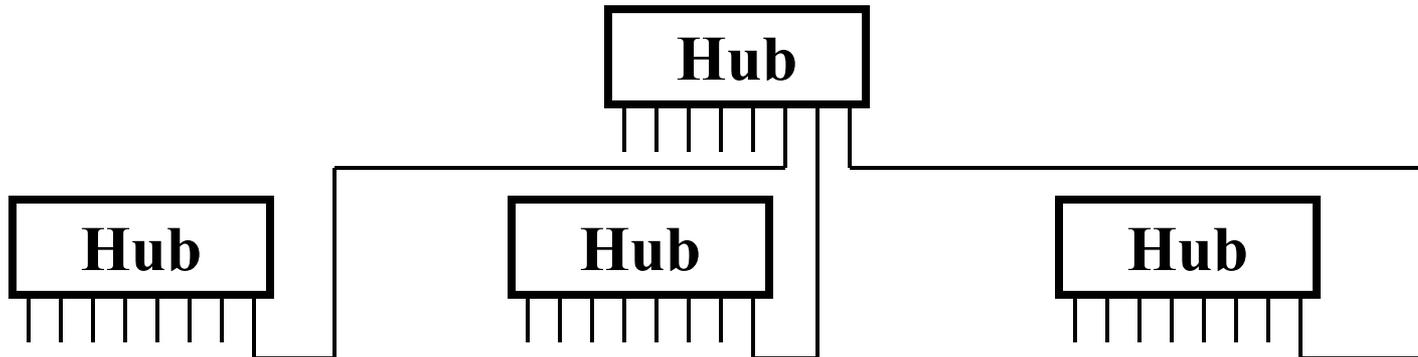
Топология «Звезда-шина»

Сегменты сети с топологией «звезда» объединяются при помощи магистральной линейной шины.



Топология «Звезда-кольцо»

Все концентраторы подключены к главному концентратору, образуя «звезду». Кольцо реализуется внутри главного концентратора.



Типы ЛКС

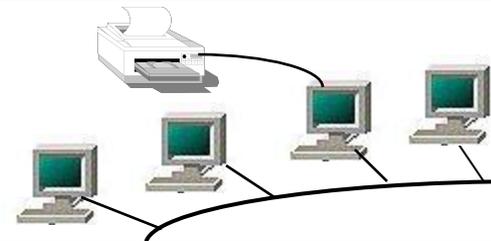
Существует два принципиальных способа организации программного обеспечения ЛКС: **одноранговые** сети и сети с централизованным управлением на основе сервера.

Одноранговые сети

- объединяют обычно не более 10 компьютеров;

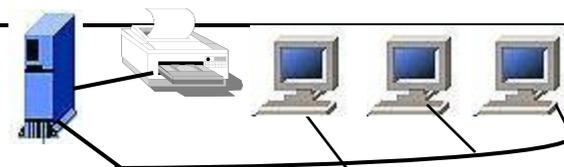
- все компьютеры равноправны, каждый компьютер функционирует как клиент и как сервер, то есть нет иерархии сети и ответственного за всю сеть;

- пользователи сами решают, какие ресурсы на своем компьютере сделать доступными в сети и обеспечивают защиту информации.



Сети на основе сервера

Стали промышленным стандартом. Работают под управлением сетевых ОС. Круг задач, которые выполняет сервер, многообразен и сложен. Чтобы серверы отвечали современным требованиям пользователей, в больших сетях их делают специализированными: серверы файлов и печати; приложений; почтовые; факсов; связи (коммуникационные); служб каталога. Администрирование сети.



Многоуровневый подход. Сетевые протоколы.

Сетевая модель OSI (*open systems interconnection basic reference model*) - базовая эталонная модель взаимодействия открытых систем

В настоящее время основным используемым стеком протоколов является TCP/IP

Семь уровней

- 7. Прикладной** (application). Доступ к сетевым службам
- 6. Уровень представления** (presentation). Представление и шифрование данных
- 5. Сеансовый** (session). Управление сеансом связи
- 4. Транспортный** (transport). Прямая связь между конечными пунктами и надежность
- 3. Сетевой** (network). Определение маршрута и логическая адресация
- 2. Канальный** (data link). Физическая адресация
- 1. Физический** (physical). Работа со средой передачи, сигналами и двоичными данными

Сетевые протоколы. Модель OSI.

Прикладной уровень (уровень приложений; англ. *application layer*) — верхний уровень модели, обеспечивающий **взаимодействие пользовательских приложений с сетью**:

- позволяет приложениям использовать сетевые службы:
 - удалённый доступ к файлам и базам данных,
 - пересылка электронной почты;
- отвечает за передачу служебной информации;
- предоставляет приложениям информацию об ошибках;
- формирует запросы к уровню представления.

Протоколы прикладного уровня: RDP (Remote Desktop Protocol), RTP (Real-time Transport Protocol), **HTTP** (HyperText Transfer Protocol), **SMTP** (Simple Mail Transfer Protocol), SNMP (Simple Network Management Protocol), **POP3** (Post Office Protocol Version 3), **FTP** (File Transfer Protocol), XMPP (Extensible Messaging and Presence Protocol), OSCAR (Open System for CommunicAtion in Realtime), Modbus, SIP (Session Initiation Protocol), **TELNET** и другие.

Сетевые протоколы. Модель OSI.

Уровень представления (*presentation layer*) обеспечивает преобразование протоколов и шифрование/расшифровку данных.

Запросы приложений, полученные с прикладного уровня, на уровне представления преобразуются в формат для передачи по сети, а полученные из сети данные преобразуются в формат приложений. На этом уровне может осуществляться сжатие/распаковка или кодирование/декодирование данных, а также перенаправление запросов другому сетевому ресурсу, если они не могут быть обработаны локально.

Протоколы уровня представления: **AFP** — Apple Filing Protocol, **ICA** — Independent Computing Architecture, **LPP** — Lightweight Presentation Protocol, **NCP** — NetWare Core Protocol, **NDR** — Network Data Representation, **XDR** — eXternal Data Representation, **X.25 PAD** — Packet Assembler/Disassembler Protocol.

Сетевые протоколы. Модель OSI.

Сеансовый уровень (*session layer*) модели обеспечивает **поддержание сеанса связи**, позволяя приложениям взаимодействовать между собой длительное время.

Уровень управляет созданием/завершением сеанса, обменом информацией, синхронизацией задач, определением права на передачу данных и поддержанием сеанса в периоды неактивности приложений.

Протоколы сеансового уровня: ADSP (AppleTalk Data Stream Protocol), ASP (AppleTalk Session Protocol), H.245 (Call Control Protocol for Multimedia Communication), ISO-SP (OSI Session Layer Protocol (X.225, ISO 8327)), iSNS (Internet Storage Name Service), L2F (Layer 2 Forwarding Protocol), L2TP (Layer 2 Tunneling Protocol), **NetBIOS** (Network Basic Input Output System), **PAP** (Password Authentication Protocol), PPTP (Point-to-Point Tunneling Protocol), RPC (Remote Procedure Call Protocol), **RTCP** (Real-time Transport Control Protocol), SMPP (Short Message Peer-to-Peer), SCP (Session Control Protocol), ZIP (Zone Information Protocol), SDP (Sockets Direct Protocol)

Сетевые протоколы. Модель OSI.

Транспортный уровень (*transport layer*) модели предназначен для **обеспечения надёжной передачи** данных от отправителя к получателю.

При этом уровень надёжности может варьироваться в широких пределах. Существует множество классов протоколов транспортного уровня

Например, **TCP** обеспечивает надёжную непрерывную передачу данных, исключая потерю данных или нарушение порядка их поступления или дублирования, может перераспределять данные, разбивая большие порции данных на фрагменты и, наоборот, склеивая фрагменты в один пакет.

Протоколы транспортного уровня: ATP (Apple Talk Transaction Protocol), CUDP (Cyclic UDP), DCCP (Datagram Congestion Control Protocol), FCP (Fiber Channel Protocol), IL (IL Protocol), **NBF** (NetBIOS Frames protocol), NCP (NetWare Core Protocol), SCTP (Stream Control Transmission Protocol), **SPX** (Sequenced Packet Exchange), SST (Structured Stream Transport), TCP (Transmission Control Protocol), UDP (User Datagram Protocol).

Сетевые протоколы. Модель OSI.

Сетевой уровень (*network layer*) модели предназначен для **определения пути передачи данных.**

Отвечает за трансляцию логических адресов и имён в физические, определение кратчайших маршрутов, коммутацию и маршрутизацию, отслеживание неполадок и «заторов» в сети.

Протоколы сетевого уровня маршрутизируют данные от источника к получателю. Работающие на этом уровне устройства (маршрутизаторы) условно называют устройствами третьего уровня (по номеру уровня в модели OSI).

Протоколы сетевого уровня: IP/IPv4/IPv6 (**Internet Protocol**), **IPX** (Internetwork Packet Exchange, протокол межсетевого обмена), X.25 (частично этот протокол реализован на уровне 2), CLNP (сетевой протокол без организации соединений), IPsec (Internet Protocol Security). Протоколы маршрутизации — RIP (Routing Information Protocol), OSPF (Open Shortest Path First).

Сетевые протоколы. Модель OSI.

Канальный уровень (*data link layer*) предназначен для обеспечения взаимодействия сетей по физическому уровню и контролем над ошибками, которые могут возникнуть. Полученные с физического уровня данные, представленные в битах, он упаковывает в кадры, проверяет их на целостность и, если нужно, исправляет ошибки и отправляет на сетевой уровень.

На этом уровне работают коммутаторы, мосты и другие устройства.

Протоколы канального уровня- ARCnet, ATM, Controller Area Network (CAN), Econet, **Ethernet**, Ethernet Automatic Protection Switching (EAPS), Fiber Distributed Data Interface (FDDI), Frame Relay, High-Level Data Link Control (HDLC), IEEE 802.2 (provides LLC functions to IEEE 802 MAC layers), Link Access Procedures, D channel (LAPD), IEEE 802.11 wireless LAN, **Local Talk**, Multiprotocol Label Switching (MPLS), Point-to-Point Protocol (PPP), Point-to-Point Protocol over Ethernet(PPPoE), StarLan, **Token ring**, Unidirectional Link Detection (UDLD), x.25.

Сетевые протоколы. Модель OSI.

Физический уровень (*physical layer*) — нижний уровень модели, который **определяет метод передачи данных**, представленных в двоичном виде, от одного устройства (компьютера) к другому.

Осуществляют передачу электрических или оптических сигналов в кабель или в радиозэфир и, соответственно, их приём и преобразование в биты данных в соответствии с методами кодирования цифровых сигналов.

На этом уровне также работают концентраторы, повторители сигнала и медиа конвертеры.

Протоколы физического уровня: IEEE 802.15

(**Bluetooth**), IRDA, EIA RS-232, EIA-422, EIA-423, RS-449, RS-485, DSL, ISDN, SONET/SDH, **802.11 Wi-Fi**, **Etherloop**, GSM Um radio

interface, ITU, ITU-T, TransferJet, ARINC 818 и G.hn/G.9960.

Сетевые протоколы. Модель OSI.

Семейство TCP/IP. Имеет три транспортных протокола: TCP, полностью соответствующий OSI, обеспечивающий проверку получения данных; UDP, отвечающий транспортному уровню только наличием порта, обеспечивающий обмен датаграммами между приложениями, не гарантирующий получения данных; и SCTP, разработанный для устранения некоторых недостатков TCP, в который добавлены некоторые новшества.

Семейство IPX/SPX. В семействе IPX/SPX порты (называемые сокетами или гнёздами) появляются в протоколе сетевого уровня IPX, обеспечивая обмен датаграммами между приложениями (операционная система резервирует часть сокетов для себя). Протокол SPX, в свою очередь, дополняет IPX всеми остальными возможностями транспортного уровня в полном соответствии с OSI.

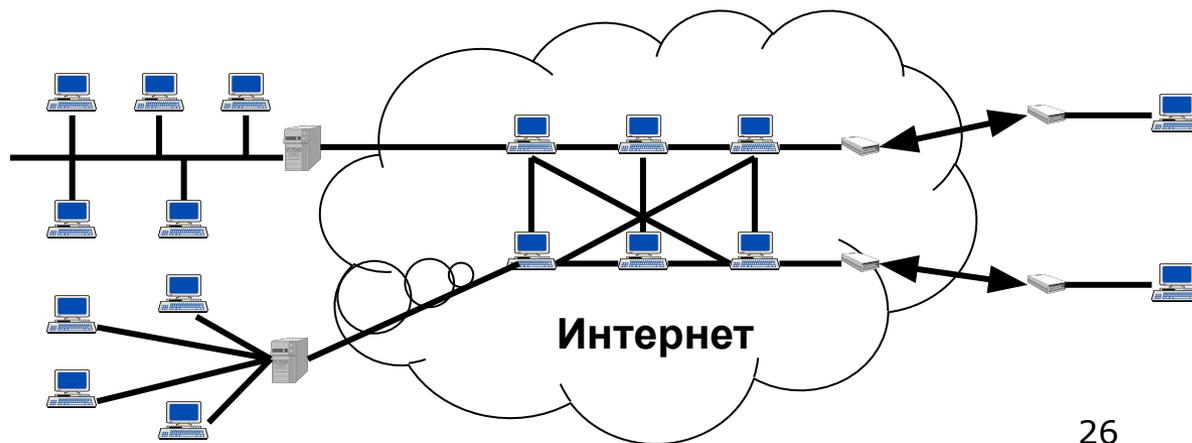
В качестве адреса хоста IPX использует идентификатор, образованный из четырёхбайтного номера сети (назначаемого маршрутизаторами) и MAC-адреса сетевого адаптера.

1.3. Глобальная компьютерная сеть Интернет

В **1969** году специалистами из Пентагона была создана крупная децентрализованная компьютерная сеть *Advanced Research Project Agency Network (Arpanet)*. Спустя некоторое время доступ к Arpanet получили и гражданские учреждения США. В **1977** году началось слияние Arpanet с другими сетями как внутри США, так и в других развитых странах. В результате такого слияния и родилась глобальная компьютерная сеть *Internet*.

Интернет - это глобальная компьютерная сеть, объединяющая многие локальные, региональные и корпоративные сети и включающая десятки миллионов компьютеров.

Интернет фактически является **сетевой базой данных**. Гиперссылки связывают между собой сотни миллионов документов в единую сетевую базу данных.



Адресация в Интернете

Каждый компьютер, подключённый к Интернету, имеет свой уникальный 32-битовый **IP-адрес** (Internet Protocol). Возможно $2^{32} = 4\,294\,967\,296$ IP-адресов, записываемых в виде четырех десятичных чисел от 0 до 255, разделенных точкой: **123.45.67.89**.

IP-адрес состоит из двух частей, одна из которых является **адресом сети**, а другая **адресом компьютера** в данной сети.

Все адреса подразделяются на три класса: **A, B, C**. Первые биты адреса отводятся для идентификации класса, а остальные разделяются на адрес сети и адрес компьютера в этой сети.

Класс А	0	Адрес сети (7 бит)		Адрес компьютера (24 бит)	
Класс В	1	0	Адрес сети (14 бит)		Адрес компьютера (16 бит)
Класс С	1	1	0	Адрес сети (21 бит)	
				Адрес компьютера (8 бит)	

Класс сети	Класс А	Класс В	Класс С
Адрес	число от 1 до 126	число от 128 до 191	число от 192 до 223
Кол. сетей	126	16 384	2 097 152
Комп. в кажд. сети	16 777 214	65 534	254 ²⁷

Доменная система имён

Доменная система имён (**DNS - Domain Name System**) ставит в соответствие числовому IP-адресу компьютера уникальное доменное имя.

Иерархическая структура доменной системы имён:

№	Домен верхнего уровня	Комментарий
1	com	коммерческие организации
2	ru	Россия
...		
264	zr	Заир

DNS-сервер первого уровня

№	Домен второго уровня	Домен первого уровня
1	keldysh	ru
2	mipkro	ru
...		
11147		ru

№	Домен второго уровня	Домен первого уровня
1	microsoft	com
2	intel	com
...		
1884466		com

DNS-сервер второго уровня

№	Доменное имя компьютера	IP-адрес компьютера	Комментарий
1	www.mipkro.ru	213.171.37.202	WWW-сервер
2	ftp.mipkro.ru	213.171.37.203	FTP-сервер

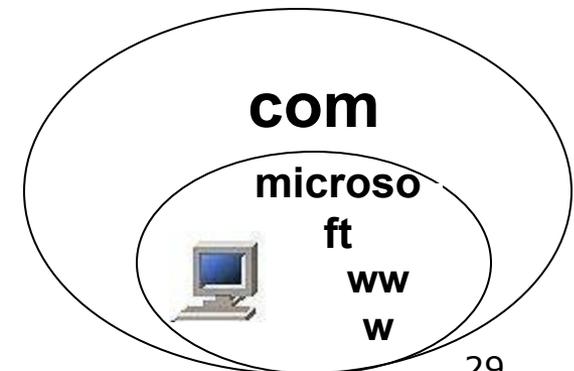
DNS-сервер третьего уровня

Доменная система имён

Домены верхнего уровня бывают двух типов: **географические** (двухбуквенные - каждой стране соответствует двухбуквенный код) и **административные** (трёхбуквенные).

Административные	Тип организации	Географическая	Страна
com	Коммерческая	ca	Канада
edu	Общеобразовательная	de	Германия
gov	Правительственная США	jp	Япония
int	Международная	ru	Россия
mil	Военная США	su	бывший СССР
net	Компьютерная сеть	uk	Англия / Ирландия
org	Некоммерческая	us	США

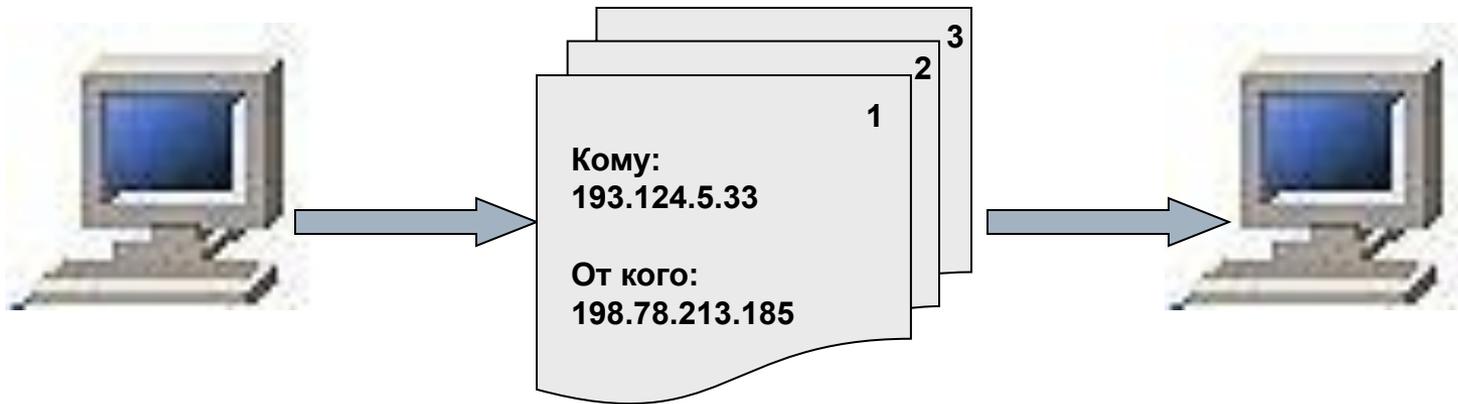
Основной сервер
компании Microsoft
www.microsoft.com



Протокол передачи данных TCP/IP

Сеть Интернет функционирует и развивается благодаря использованию единого протокола передачи данных *TCP/IP*.

Internet Protokol (IP) - протокол маршрутизации - обеспечивает маршрутизацию IP-пакетов, т.е. доставку информации от компьютера-отправителя к компьютеру-получателю.



Transmission Control Protocol (TCP) - транспортный протокол – обеспечивает разбиение передаваемых файлов на IP-пакеты в процессе передачи и сборку файлов в процессе получения.

Информационные системы Интернет

World Wide Web (WWW) =
спецификация URL (*Uniform Resource Locator, Унифицированный указатель ресурса*) + спецификация HTTP
(*HyperText Transfer Protocol, Протокол передачи гипертекста*) +
спецификация HTML (*HyperText Mark-Up Language, Язык разметки гипертекста*)

Термины WWW

WWW – это **единое информационное пространство**, состоящее из множества **взаимосвязанных электронных документов**, хранящихся на **Web-серверах**.

Все серверы используют специальный язык *HTML*. *HTML-документы* представляют собой текстовые файлы, в которые встроены **специальные команды**, называемые **тегами**.

Отдельные документы, составляющие *пространство Web*, называют **Web-страницами**.

Группы, тематически объединенных Web-страниц, называют **Web-узлами** (*Web-сайт* или просто *сайт*).

Программы для просмотра Web-страниц называют **Web-браузерами**.

Унифицированный указатель ресурса

Адрес любого файла в WWW определяется **унифицированным указателем ресурса – URL**. Этот адрес состоит из **трех частей**:

- 1. Указание службы**, которая осуществляет доступ к данному ресурсу. Например, для службы WWW прикладным является **протокол HTTP**. После имени протокола ставиться двоеточие и два знака косой черты, что выглядит так: ***http://...;***
- 2. Указание доменного имени хост-сервера**, на котором хранится данный ресурс:
http://armath.spbu.ru...;
- 3. Указания полного пути** доступа к файлу данного хост-ПК, например, так:

Понятие URL: (унифицированный указатель ресурса)

http://ru.wikipedia.org/wiki.ru/lab1.htm#_ftn1

1

2

3

- 1. условное наименование протокола взаимодействия (http, ftp...);**
- 2. полное доменное имя компьютера;**
- 3. полный путь доступа к ресурсу на компьютере.**

Вид сообщений в протоколе TCP/IP

Уровень приложений
(FTP, HTTP)

Блок данных

Уровень
TCP

Заголовок
TCP

Блок данных TCP,
включающий сообщение

Уровень
IP

Заголовок
IP

Заголовок
TCP

Блок данных TCP,
включающий сообщение

Службы и сервисы ГВС

- **сервис DNS**, или система доменных имен, обеспечивающий возможность использования для адресации узлов сети мнемонических имен вместо числовых адресов;
- **электронная почта (E-mail)**, обеспечивающая возможность обмена сообщениями одного человека с одним или несколькими абонентами;
- **сервис IRC**, предназначенный для поддержки текстового общения в реальном времени (chat);
- **телеконференции**, или группы новостей (**UseNet**), обеспечивающие возможность коллективного обмена сообщениями;

Службы и сервисы ГВС

- **сервис FTP** — система файловых архивов, обеспечивающая хранение и пересылку файлов различных типов;
- **сервис TelNet**, предназначенный для управления удаленными компьютерами в терминальном режиме;
- служебная **программа Ping** – позволяет проверять качество связи с сервером;
- **программы Whois** и **Finger** – используются для нахождения координат пользователей сети или определения пользователей, работающих в настоящий момент на конкретном хосте.

Услуги Интернет

- Всемирная паутина

- Веб-форумы
- Блоги
- Вики-проекты (и, в частности, Википедия)
- Интернет-магазины
- Интернет-аукционы
- Социальные сети

- Электронная почта и списки рассылки

- Группы новостей (в основном, Usenet)

- Файлообменные сети

- Электронные платёжные системы

- Интернет-радио

- Интернет-

- телевидение IPTV

- IP-телефония

- FTP-серверы

- IRC (реализовано также как веб-чаты)

- Поисковые системы

- Интернет-реклама

- Интернет-коммерция

- Удалённые

- терминалы

- Удалённое

- управление

- Многопользовательские игры

2. Облачные вычисления

Облако (*cloud*) - **модель организации ИТ-инфраструктуры**, состоящая из распределенных и разделяемых конфигурируемых аппаратных и сетевых ресурсов, а также ПО, развернутых на ЦОД поставщиков.

Облачные вычисления (*cloud computing*) - **модель предоставления** возможности повсеместного и удобного сетевого **доступа** по требованию к общему пулу конфигурируемых вычислительных ресурсов, которые могут быть оперативно предоставлены и освобождены с минимальными затратами или обращениями к провайдеру

Облачные технологии – **модель предоставления** потребителю **ИТ как сервиса** через Интернет

Облачные вычисления

1960 г. Дж. ван Неуман предложил архитектуру компьютера, когда-нибудь кто-нибудь производит

1999 г. По подписка в

2002 г. На вычислительном магазине

2006 г. [Amazon](#) эластичное Computing

Отсылка к «облаку» использовалась как метафора, основанная на изображении Интернета на диаграмме компьютерной сети



Облачные вычисления

Ключевые составляющие «облака»

Характеристики

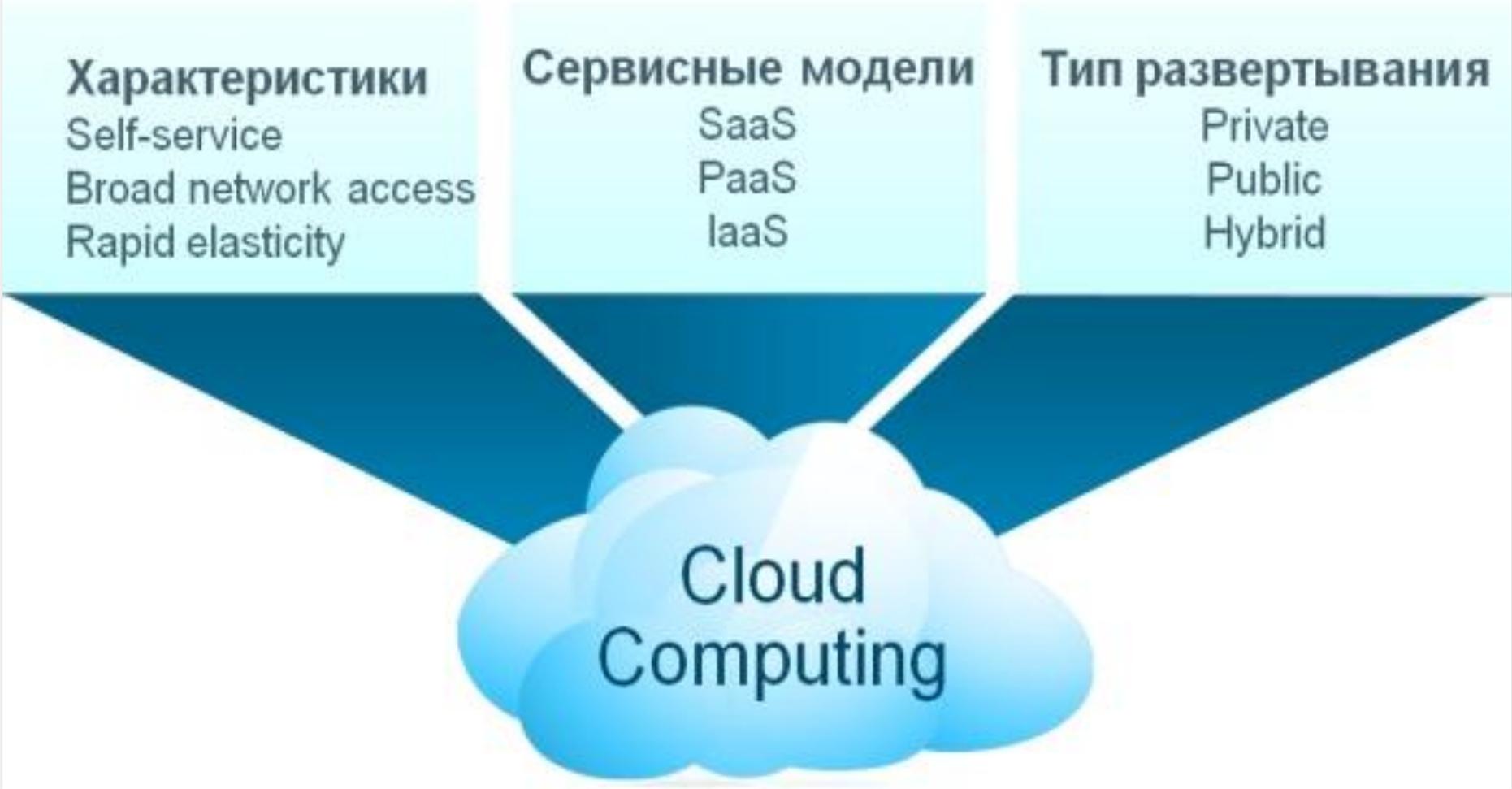
Self-service
Broad network access
Rapid elasticity

Сервисные модели

SaaS
PaaS
IaaS

Тип развертывания

Private
Public
Hybrid



Cloud
Computing

Облачные вычисления



Модели

Гибридное облако (Hybrid cloud)

Композиция из двух или более различных инфраструктур облаков, связанных между собой стандартизированными или собственными технологиями, которые позволяют переносить данные или приложения между компонентами

Предприятие

Традиционная
ИТ служба

Private cloud

Hybrid cloud

Public cloud

Модели работы с «облаком» для разных групп пользователей

Модели обл

Программ



SaaS



SaaS (облачное приложение) Возможность предоставления потребителю в **использование приложений провайдера**, работающих в облачной инфраструктуре. Приложения доступны из различных клиентских устройств или через интерфейсы тонких клиентов, такие как веб-браузер (например, веб-почта) или интерфейсы программ. Потребитель при этом не управляет базовой инфраструктурой облака, в том числе сетями, серверами, операционными системами, системами хранения и даже индивидуальными настройками приложений за исключением некоторых пользовательских настроек конфигурации приложения

Облачные вычисления

Достоинства и недостатки



Облачные вычисления

Мифы и заблуждения

- **«Облако» основано только на программном обеспечении**

Объединяем несколько виртуальных устройств и получаем «облако».

- **«Облако» и эластичные ресурсы**

Понятие «эластичных ресурсов» не должно ограничиваться только серверами и системами хранения данных, оно должно распространяться на все «облако», в том числе его сетевые элементы.

Облачные вычисления

Мифы и заблуждения

□ «Облако» и объединенные ресурсы

Это еще одна область, которую многие связывают только с серверами и системами хранения данных. В течение многих лет, до появления «облака», сеть представляла собой комплексный ресурс, который предоставлял услуги (полоса пропускания, безопасность, сегментация или изоляция, функция QoS, и т. д.) отдельным группам систем.

Облачные вычисления

Мифы и заблуждения

□ **Виртуализация делает «облако» более гибким**

Абсолютная правда, в отношении динамического распределения ресурсов и оперативного включения новых виртуальных машин с помощью шаблонов и клонов.

Однако, точно также как в свое время сеть стала поддерживать передачу различных мультимедийных данных в масштабе реального времени, теперь она должна приспособливаться к изменениям, связанным с виртуализацией и динамичными компьютерными средами.

Облачные вычисления

Мифы и заблуждения

□ Виртуальная машина – это тот же сервер, только программный

Это не совсем так.

Действительно, приложение и гостевая операционная система, наверное, не видят никакой разницы, но сетевым администраторам и службе безопасности это совсем не безразлично.

Виртуальная машина больше не начинается и не заканчивается на одном конце кабеля Ethernet.

Облачные вычисления

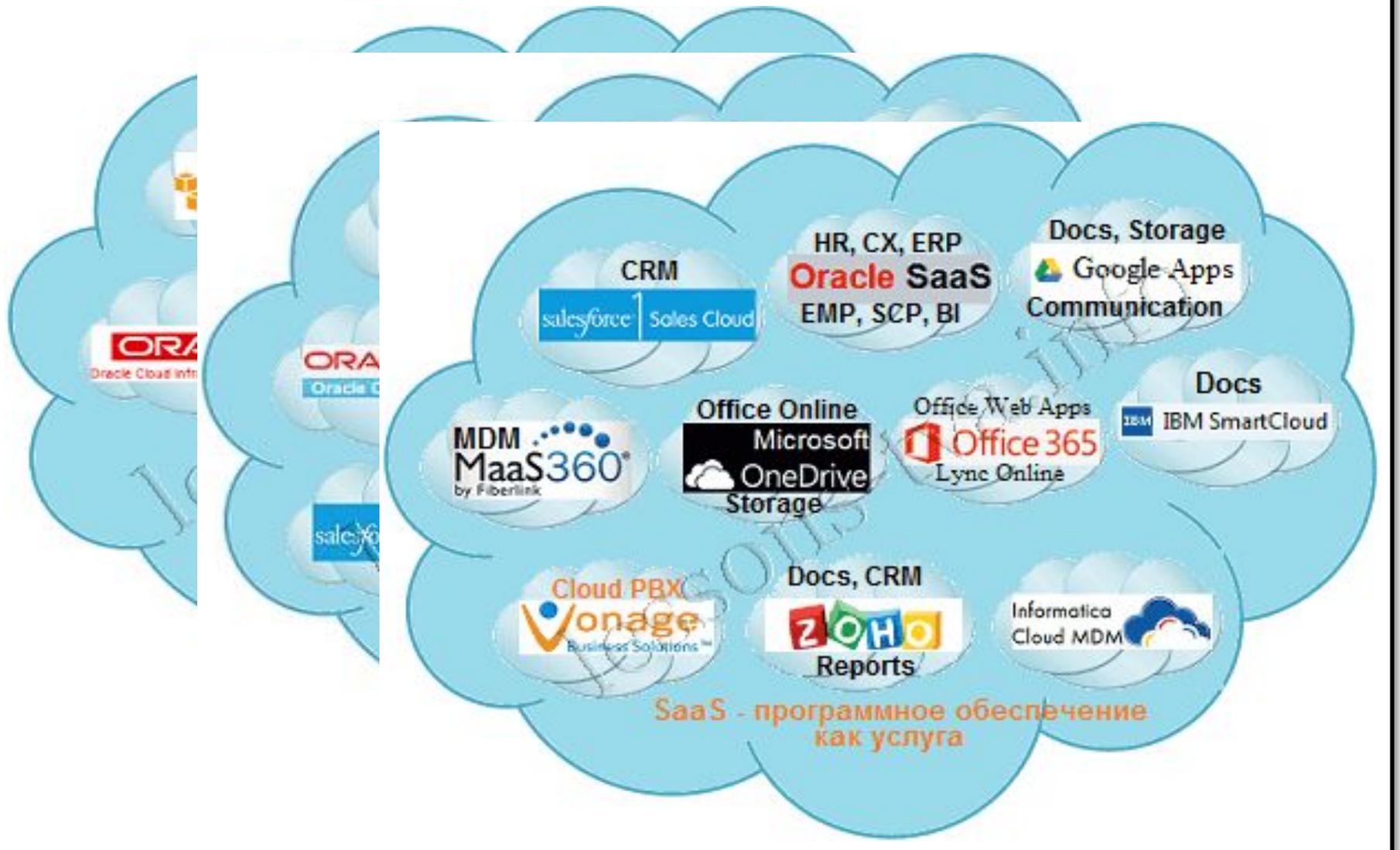
Мифы и заблуждения

□ «Облака» планируют, а сети администрируют

Это подразумевает, что ваша компьютерная среда по-прежнему изолирована от вашей сети. Такие технологии, как виртуализация требуют не только более тесной интеграции вычислений и сетевых операций, но и более высокого уровня автоматизации и контроля, такого уровня автоматизации, который обеспечивал бы открытый доступ ко всем автономным системам и функциональным модулям с заданными политиками.

Облачные вычисления

Основные разработчики

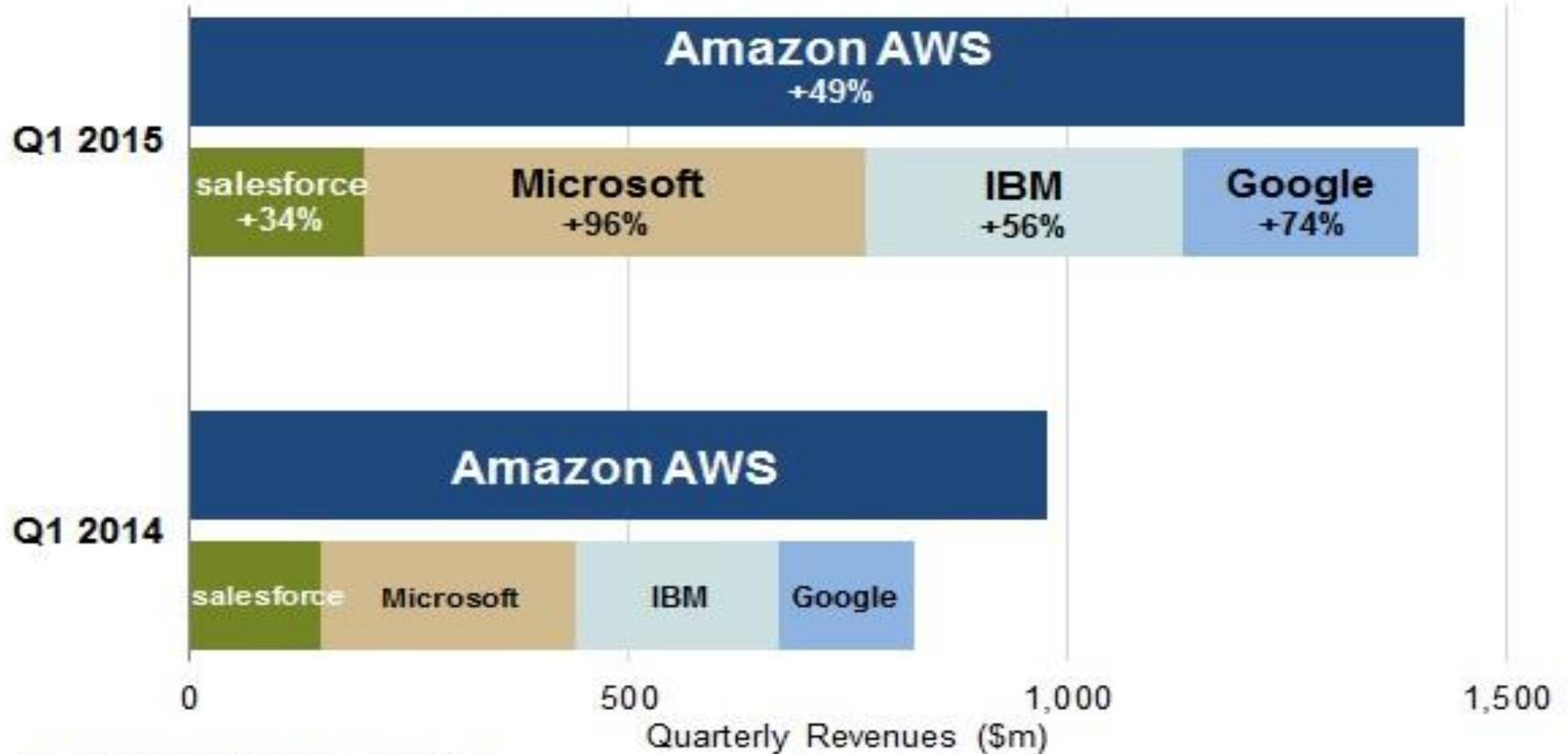


Облачные вычисления

Основные игроки рынка

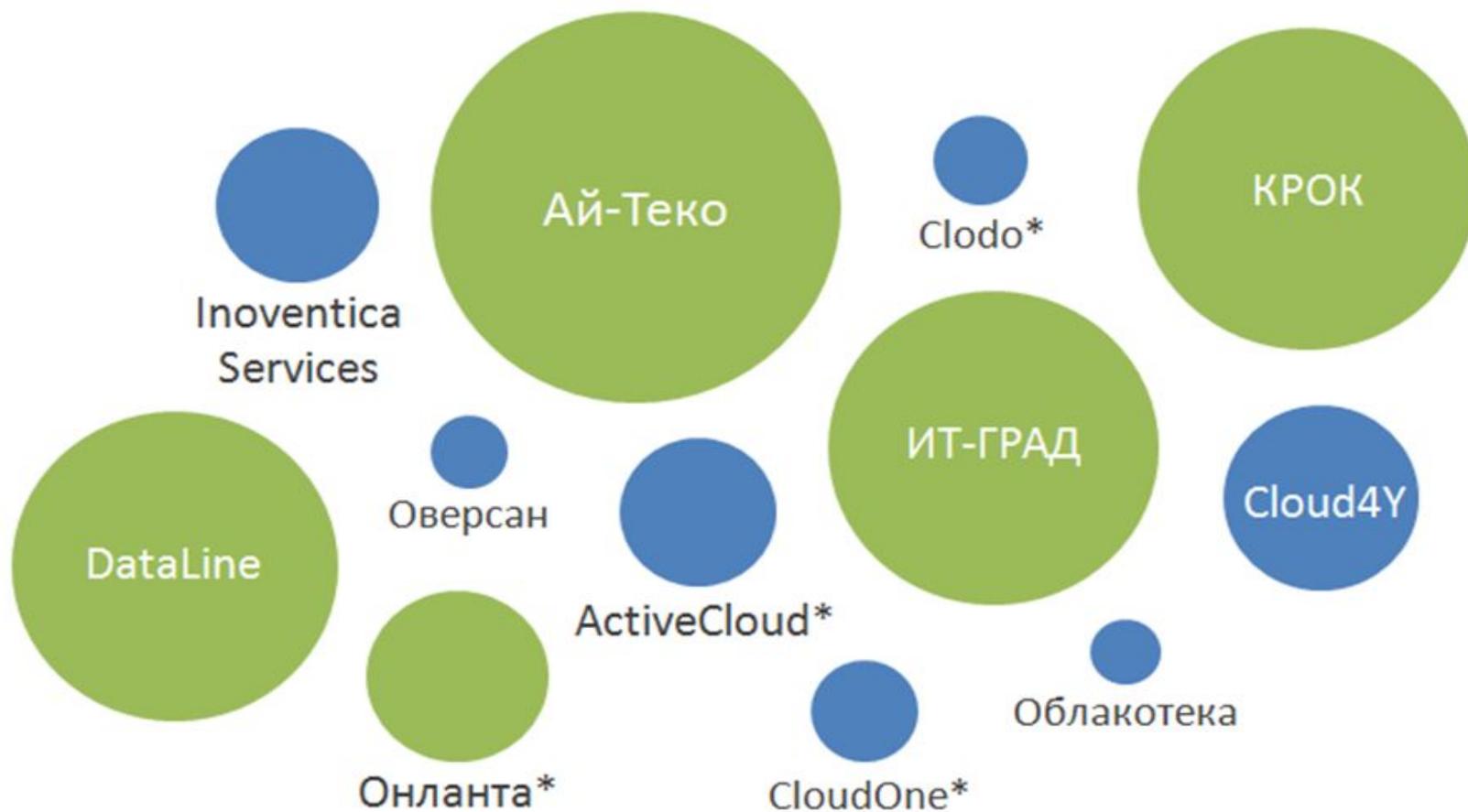
Cloud Infrastructure Services Revenue Growth

(IaaS, PaaS, Private & Hybrid services)



Source: Synergy Research Group

Рис. 2. Крупнейшие игроки российского рынка по выручке, 2014 г.



- годовой ARPU по услугам IaaS более 500 тыс. руб.
- годовой ARPU по услугам IaaS менее 500 тыс. руб.

* По оценкам J'son & Partners Consulting

Источник: J'son & Partners Consulting

4. Перспективные информационные технологии, понятия и проблемы

Data Mining – интеллектуальный анализ данных

Data Mining переводится как "добыча" или "раскопка данных". Нередко рядом с Data Mining встречаются слова "**обнаружение знаний в базах данных**" (knowledge discovery in databases) и "**интеллектуальный анализ данных**". Их можно считать синонимами Data Mining. Возникновение всех указанных терминов связано с новым витком в развитии средств и методов обработки данных.

Перспективные информационные технологии, понятия и проблемы

Колоссальные потоки информационной «руды» в самых различных областях. Что делать с этой информацией? Без продуктивной переработки потоки сырых данных образуют никому не нужную свалку.

Специфика современных требований к такой переработке следующие:

- Данные имеют **неограниченный объем**
- Данные являются **разнородными** (количественными, качественными, текстовыми)
- Результаты должны быть **конкретны и понятны**
- Инструменты для обработки сырых данных должны быть **просты** в использовании

Перспективные информационные технологии, понятия и проблемы

Методы математической статистики— *концепция усреднения по выборке*

Методы математической статистики оказались полезными главным образом для проверки **заранее сформулированных гипотез** (verification-driven data mining) и для “грубого” разведочного анализа, составляющего основу **оперативной аналитической обработки данных** (online analytical processing, **OLAP**).

Перспективные информационные технологии, понятия и проблемы

В основу современной технологии **Data Mining** (discovery-driven data mining) положена **концепция шаблонов** (паттернов), отражающих *фрагменты* многоаспектных взаимоотношений в данных.

Эти шаблоны представляют собой **закономерности**, свойственные *подвыборкам данных*, которые могут быть компактно выражены в понятной человеку форме.

Поиск шаблонов производится **методами**, не ограниченными рамками априорных предположений о структуре выборке и виде распределений значений анализируемых показателей.

Перспективные информационные технологии, понятия и проблемы

Примеры формулировок задач для OLAP и Data Mining

OLAP	Data Mining
Каковы средние показатели травматизма для курящих и некурящих?	Встречаются ли точные шаблоны в описаниях людей, подверженных повышенному травматизму?
Каковы средние размеры телефонных счетов существующих клиентов в сравнении со счетами бывших клиентов отказавшихся от услуг телефонной компании)?	Имеются ли характерные портреты клиентов, которые, по всей вероятности, собираются отказаться от услуг телефонной компании?
Какова средняя величина ежедневных покупок по украденной и не украденной кредитной карточке?	Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?

Перспективные информационные технологии, понятия и проблемы

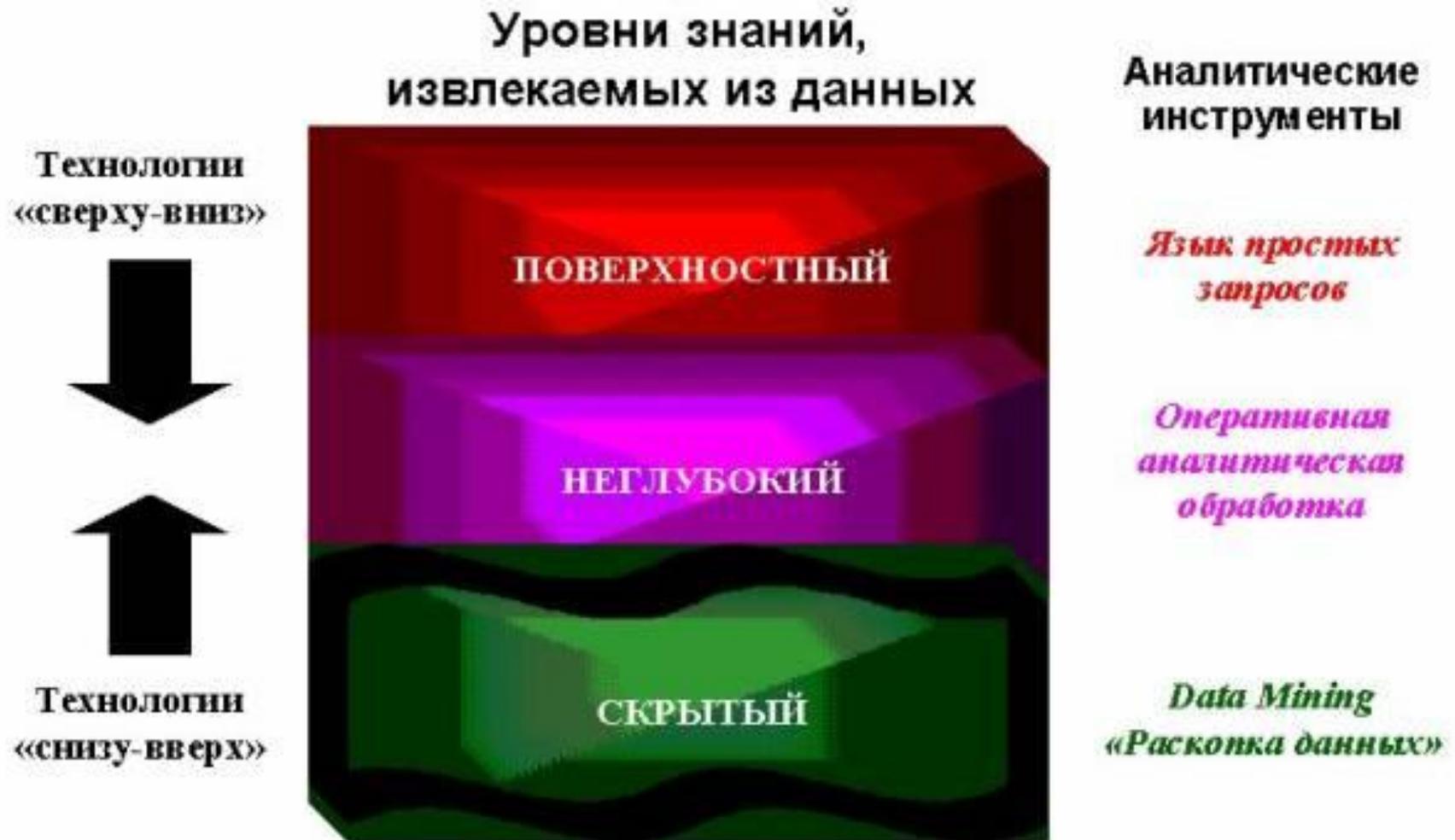
Важное положение **Data Mining** — **нетривиальность разыскиваемых шаблонов.**

Это означает, что найденные шаблоны должны отражать неочевидные, неожиданные (unexpected) регулярности в данных, составляющие так называемые **скрытые знания** (hidden knowledge).

К обществу пришло понимание, что *сырые данные* (raw data) содержат глубинный пласт знаний, при грамотной раскопке которого могут быть обнаружены настоящие самородки

Data Mining — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

Перспективные информационные технологии, понятия и проблемы



Перспективные информационные технологии, понятия и проблемы

Классы систем **Data Mining**

Data Mining является **мульти-дисциплинарной областью**, возникшей и развивающейся на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз **данных** и др.

Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data Mining. Многие из таких систем интегрируют в себе сразу несколько подходов.

Тем не менее, как правило, в каждой системе имеется какая-то ключевая компонента, на которую делается главная ставка.

Перспективные информационные технологии, понятия и проблемы классификация ключевых компонент



Перспективные информационные технологии, понятия и проблемы

1. Предметно-ориентированные аналитические системы

Предметно-ориентированные аналитические системы очень разнообразны. Наиболее широкий подкласс таких систем, получивший распространение в области исследования финансовых рынков, носит название "**технический анализ**". Он представляет собой совокупность нескольких десятков методов прогноза динамики цен и выбора оптимальной структуры инвестиционного портфеля, основанных на различных эмпирических моделях динамики рынка. Эти **методы часто используют несложный статистический аппарат, но максимально учитывают сложившуюся своей области специфику** (профессиональный язык, системы различных индексов и пр.) На рынке имеется множество программ этого класса. Как правило, они довольно дешевы (обычно \$300–1000).

Перспективные информационные технологии, понятия и проблемы

2. Статистические пакеты

Последние версии почти всех известных статистических пакетов включают наряду с традиционными статистическими методами также элементы Data Mining. Но **основное внимание в них уделяется все же классическим методикам — корреляционному, регрессионному, факторному анализу** и другим.

Недостатком систем этого класса считают требование к специальной подготовке пользователя. Также отмечают, что мощные современные статистические пакеты являются слишком "тяжеловесными" для массового применения в финансах и бизнесе. К тому же часто эти системы весьма дороги — от \$1000 до \$15000.

Большинство методов, входящих в состав пакетов опираются на статистическую парадигму, в которой главными фигурантами служат усредненные характеристики выборки. А эти характеристики при исследовании реальных сложных жизненных феноменов часто являются фиктивными величинами.

В качестве примеров наиболее мощных и распространенных статистических пакетов можно назвать SAS (компания SAS Institute), SPSS (SPSS), STATGRAPICS (Manugistics), STATISTICA, STADIA и другие.

Перспективные информационные технологии, понятия и проблемы

3. Нейронные сети

Это большой класс систем, архитектура которых имеет аналогию (как теперь известно, довольно слабую) с построением нервной ткани из нейронов (**М:М**)

Основным недостатком нейросетевой парадигмы является необходимость иметь очень большой объем обучающей выборки. Другой существенный недостаток заключается в том, что даже натренированная нейронная сеть представляет собой черный ящик. Знания, зафиксированные как веса нескольких сотен межнейронных связей, совершенно не поддаются анализу и интерпретации человеком.

Перспективные информационные технологии, понятия и проблемы

4. Системы рассуждений на основе аналогичных случаев

Идея **систем case based reasoning — CBR** — на первый взгляд крайне проста. Для того чтобы сделать прогноз на будущее или выбрать правильное решение, эти системы находят в прошлом близкие аналоги наличной ситуации и выбирают тот же ответ, который был для них правильным. Поэтому этот метод еще называют **методом "ближайшего соседа"** (nearest neighbour).

Главным их минусом считают то, что они вообще не создают каких-либо моделей или правил, обобщающих предыдущий опыт, — в выборе решения они основываются на всем массиве доступных данных, поэтому невозможно сказать, на основе каких конкретно факторов CBR системы строят свои ответы.

Другой минус заключается в произволе, который допускают системы CBR при выборе меры "близости".

Перспективные информационные технологии, понятия и проблемы

5. Деревья решений (decision trees)

Деревья решения являются одним из наиболее популярных подходов к решению задач Data Mining. Они создают иерархическую структуру классифицирующих **правил типа "ЕСЛИ... ТО..." (if-then)**, имеющую вид дерева.

Для принятия решения, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Вопросы имеют вид "значение параметра А больше х?". Если ответ положительный, осуществляется переход к правому узлу следующего уровня, если отрицательный — то к левому узлу; затем снова следует вопрос, связанный с соответствующим узлом

(Такие ИС реализуют наивный принцип последовательного просмотра признаков и "цепляют" фактически осколки настоящих закономерностей, создавая лишь иллюзию логического вывода).

Перспективные информационные технологии, понятия и проблемы

6. Эволюционное программирование

Гипотезы о виде зависимости целевой переменной от других переменных **формулируются в виде программ** на некотором внутреннем языке программирования. Процесс построения программ строится как эволюция в мире программ. Когда система находит программу, более или менее **удовлетворительно выражающую искомую зависимость**, она начинает вносить в нее небольшие модификации и отбирает среди построенных дочерних программ те, которые повышают точность. Таким образом **система "выращивает" несколько генетических линий программ**, которые конкурируют между собой в точности **выражения искомой зависимости**. Специальный модуль переводит найденные зависимости с внутреннего языка системы на понятный пользователю язык (математические формулы, таблицы и пр.).

Другое направление эволюционного программирования связано с **поиском зависимости целевых переменных от остальных в форме функций какого-то определенного вида.**

Перспективные информационные технологии, понятия и проблемы

7. Генетические алгоритмы

Мощное средство решения разнообразных комбинаторных задач и задач оптимизации. Первый шаг при построении генетических алгоритмов — это **кодировка исходных логических закономерностей в базе данных**, которые именуют хромосомами, а весь набор таких закономерностей называют популяцией хромосом. Далее для реализации концепции отбора вводится **способ сопоставления** различных хромосом. Популяция обрабатывается с помощью **процедур репродукции**, изменчивости (мутаций), генетической композиции. Эти процедуры **имитируют биологические процессы**. Наиболее важные среди них: случайные мутации данных в индивидуальных хромосомах, переходы (кроссинговер) и рекомбинация генетического материала, содержащегося в индивидуальных родительских хромосомах (аналогично гетеросексуальной репродукции), и миграции генов. В ходе работы процедур на каждой стадии эволюции получают популяции со все более совершенными индивидуумами.

Перспективные информационные технологии, понятия и проблемы

Генетические алгоритмы удобны тем, что их **легко распараллеливать**. Например, можно разбить поколение на несколько групп и работать с каждой из них независимо, обмениваясь время от времени несколькими хромосомами. Существуют также и другие методы распараллеливания генетических алгоритмов.

Генетические алгоритмы имеют **ряд недостатков**. **Критерий отбора** хромосом и используемые процедуры **являются эвристическими и далеко не гарантируют нахождения "лучшего" решения**. Как и в реальной жизни, эволюцию может "заклинить" на какой-либо непродуктивной ветви. И, наоборот, можно привести примеры, как два неперспективных родителя, которые будут исключены из эволюции генетическим алгоритмом, оказываются способными произвести высокоэффективного потомка. Это особенно становится заметно при решении высокоразмерных задач со сложными внутренними связями. Примером может служить система GeneHunter фирмы Ward Systems Group. Его стоимость — около \$1000.

Перспективные информационные технологии, понятия и проблемы

8. Алгоритмы ограниченного перебора

Эти алгоритмы **вычисляют частоты комбинаций простых логических событий** в подгруппах данных.

Примеры простых логических событий:

$X = a$; $X < a$; $X > a$; $a < X < b$ и др.,

где X — какой либо параметр, "a" и "b" — константы.

Ограничением служит длина комбинации простых логических событий.

На основании **анализа вычисленных частот** делается заключение о полезности той или иной комбинации для установления ассоциации в данных, для классификации, прогнозирования и пр.

Перспективные информационные технологии, понятия и проблемы

9. Системы для визуализации многомерных данных

Средства для графического отображения данных

В подобных системах основное внимание сконцентрировано на дружелюбности пользовательского интерфейса, позволяющего ассоциировать с анализируемыми показателями **различные параметры диаграммы рассеивания объектов (записей)** базы данных.

К таким параметрам относятся цвет, форма, ориентация относительно собственной оси, размеры и другие свойства графических элементов изображения. Кроме того, системы визуализации данных снабжены удобными средствами для масштабирования и вращения изображений.

4. Перспективные информационные технологии.

Резюме

1. Рынок систем **Data Mining** развивается экспоненциально. В этом развитии принимают участие практически все крупнейшие корпорации.
2. Системы **Data Mining** применяются по **двум основным направлениям**: **1)** как массовый продукт для бизнес-приложений; **2)** как инструменты для проведения уникальных исследований (генетика, химия, медицина и пр.).

В настоящее время стоимость массового продукта от \$1000 до \$10000. Количество инсталляций массовых продуктов, судя по имеющимся сведениям, сегодня достигает десятков тысяч. Лидеры Data Mining связывают будущее этих систем с использованием их в качестве интеллектуальных приложений, встроенных в корпоративные хранилища данных.

3. Несмотря на обилие методов **Data Mining**, на сегодня **приоритет смещен в сторону логических алгоритмов поиска в данных if-then правил**. С их помощью решаются задачи прогнозирования, классификации, распознавания образов, сегментации БД, извлечения из данных “скрытых” знаний, интерпретации данных, установления ассоциаций в БД и др. Результаты таких алгоритмов эффективны и легко интерпретируются.

4. Главной проблемой логических методов обнаружения закономерностей является **проблема перебора вариантов за приемлемое время**. Другие проблемы связаны с тем, что известные методы поиска логических правил **не поддерживают функцию обобщения найденных правил и функцию поиска оптимальной композиции** таких правил.

Удачное решение указанных проблем может составить предмет новых конкурентоспособных разработок.

БЛАГОДАРЮ ЗА ВНИМАНИЕ !

ВОПРОСЫ ?

