



Технологические основы Интернета Вещей

Лекция 4- Облачные платформы Интернета Вещей

Жматов Дмитрий Владимирович
кандидат технических наук, доцент
доцент кафедры Математического обеспечения и
стандартизации информационных технологий

Определение облачной IoT платформы

IoT-платформа – это программное обеспечение системы интернета вещей для подключения конечных устройств (датчиков, сенсоров, контроллеров и т.д.) к облаку и удаленного доступа к ним.

Целью IoT-платформы является обеспечение бесшовной интеграции различных аппаратных средств с помощью специальных интерфейсов, протоколов связи, сетевых топологий, а также средств хранения, обработки и интеллектуального анализа данных.

Основные компоненты IoT платформы

1. Связь и нормализация (Connectivity & normalization) – сведение различных протоколов и форматов данных в один программный интерфейс, гарантирующий точную передачу информации и взаимодействие со всеми конечными устройствами.

2. Управление устройствами (Device management) – обеспечение корректной и бесперебойной работы конечных устройств, их конфигурирование и обновление программных приложений на них и пограничных шлюзах.

3. **База данных (Database)** – обеспечение масштабируемого и надежного хранения информации.

4. Обработка и управление действиями (Processing & action management) – мониторинг текущего и прогнозирование будущего состояния технологического оборудования на основании данных с конечных устройств, которые на нем установлены, а также выработка команд для изменения состояний оборудования и передача этих сигналов на исполнительное устройство.

5. **Аналитика(Analytics)** – интеграция и кластеризация данных, а также прогнозирование значений искомых параметров, в т.ч. с использованием методов машинного обучения (Machine Learning);

6. визуализация(Visualization) – наглядное представление собранной, обработанной и проанализированной информации в виде графиков, диаграмм, таблиц и других понятных представлений;

7. **дополнительные инструменты** (Additional tools) – средства, которые позволяют разработчикам ПО и DevOps-инженерам расширить функциональные возможности платформы Internet of Things с помощью графического интерфейса и программирования;
8. **внешние интерфейсы** (External interfaces) – API, SDK и шлюзы для интеграции с другими сервисами, системами и платформами.

Основные компоненты IoT платформы

База данных Репозиторий, в котором хранятся важные наборы данных	Внешние интерфейсы API, SDK и шлюзы, которые служат интерфейсами для сторонних систем (ERP, CRM)	
	Аналитика Алгоритмы для сложных вычислений и машинного обучения	Вспомогательные инструменты Прототипирование приложений, управление доступом, отчёты
	Визуализация данных Графическое отображение показаний датчиков (возможно, в реальном времени)	
	Обработка данных и управляющие воздействия Механизм правил, который позволяет в реальном времени совершать действия, основываясь на показаниях сенсоров и состояниях устройств	
	Менеджер устройств Инструмент для управления статусами устройств, удаленной развертки ПО и обновлений	
	Связность и нормализация данных Агенты и библиотеки, которые позволяют поддерживать постоянную целостность и связность данных	

Требования к IoT платформе:

Интеллектуальный анализ информации в реальном времени с использованием следующих инструментов Big Data:

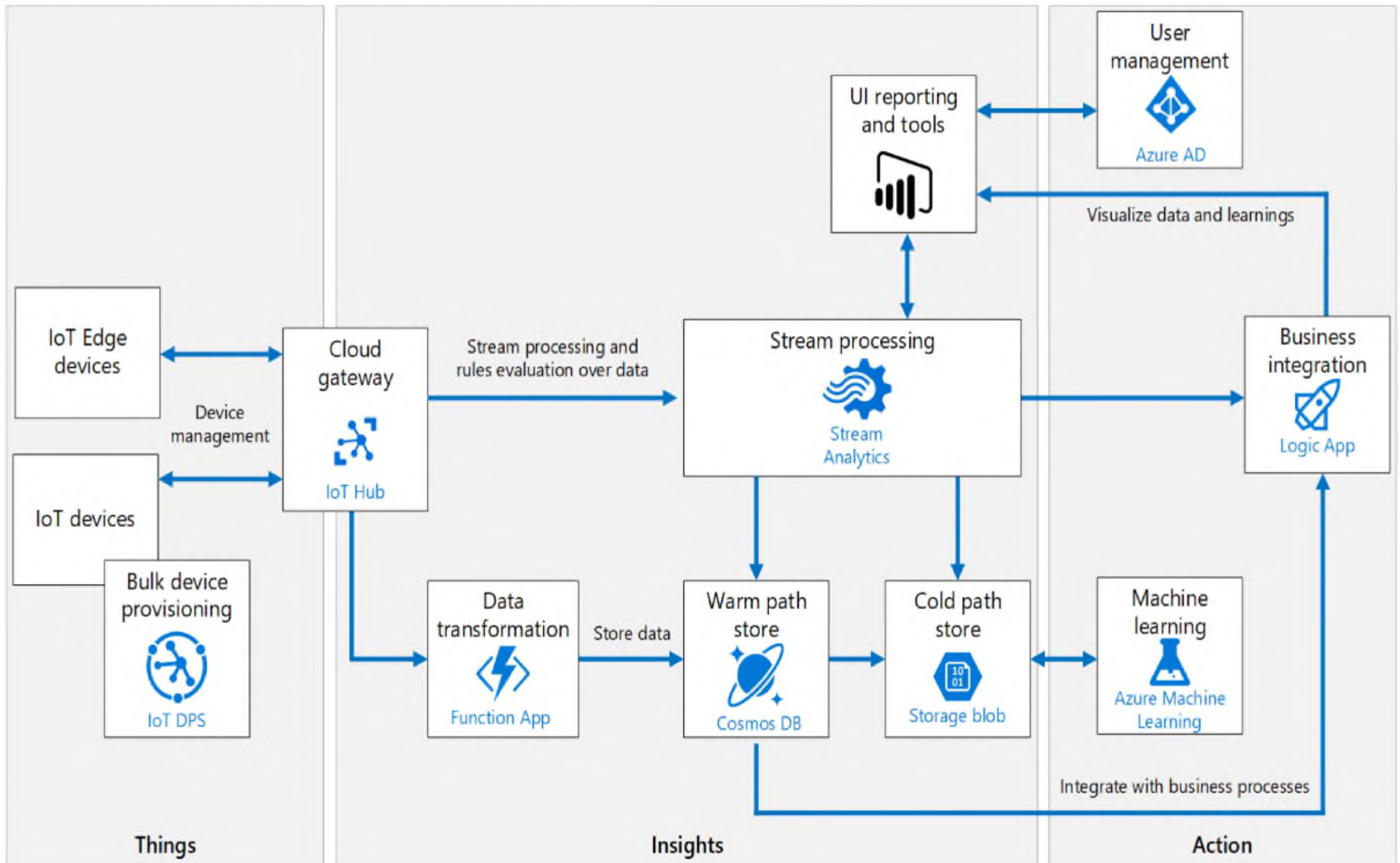
- агрегирование и фильтрация потоков данных (Storm, Samza);
- поддержка пакетных операций с накопленным набором Big Data (средствами Hadoop, Spark);
- предиктивная аналитика с использованием методов Machine Learning потоковых и пакетных данных (Spark, MLLIB).

Требования к IoT платформе:

- прикладные протоколы семейства TCP/IP — CoAP, HTTP/HTTPS;
- протоколы обмена сообщениями в концепции «издатель-подписчик» (MQTT, AMQP, XMPP, DDS), реализованные в программных брокерах RabbitMQ, Apache Qpid, Apache ActiveMQ, а также Apache Kafka, который считается наиболее масштабируемым инструментом управления очередью;
- средства быстрой загрузки потоковых данных со шлюза и конечных устройств (Apache NiFi, Apache MiNiFi, Apache Flume).

Архитектура IoT платформы:

*на примере MS Azure



Требования к IoT платформе:

На высшем уровне существует два способа обработки данных телеметрии: "горячий" и "холодный" путь. Разница связана с требованиями к задержке и доступу к данным.

"Горячий" путь позволяет анализировать данные в режиме реального времени по мере их поступления. При применении этого способа данные телеметрии должны обрабатываться с очень низкой задержкой.

Требования к IoT платформе:

На высшем уровне существует два способа обработки данных телеметрии: "горячий" и "холодный" путь. Разница связана с требованиями к задержке и доступу к данным.

"Холодный" путь позволяет выполнить пакетную обработку в течение более длительных интервалов (ежечасно или ежедневно). "Холодный" путь обычно работает с большими объемами данных, однако результаты не требуются также своевременно, как в "горячем" пути.

Термин «Анализ данных», или «Интеллектуальный анализ данных» - перевод английского термина «Data Mining», т.е. буквально «добыча данных» или даже «раскапывание данных».



Сырые данные: данные о валютных операциях за определенный промежуток времени, хранящиеся в файле (текстовом, XLS, XML...) или в базе данных – сами по себе они бесполезны.

Информация: результат первичной обработки и визуализации данных – например, построенный по сырым данным график, отражающий динамику изменений котировок акций.

Знание: Выявленная (с помощью математических методов и специального программного обеспечения) тенденция динамики котировок акций, позволяющая сделать вывод о том, когда нам нужно продать акции.

«Data Mining — это процесс обнаружения знаний (в сырых данных), которые являлись бы:

- неизвестными ранее,
- нетривиальными,
- доступными для интерпретации;
- практически полезными;
- необходимыми для **принятия решений** в различных сферах человеческой деятельности».

Прогнозирование (Forecasting)

Нахождение будущих состояний объекта на основании предыдущих состояний (исторических данных).

Примеры:

- прогнозирование ситуаций на валютных рынках,
- прогнозирование цен на рынке недвижимости,
- прогнозирование демографических процессов,
- прогнозирование климатических процессов....

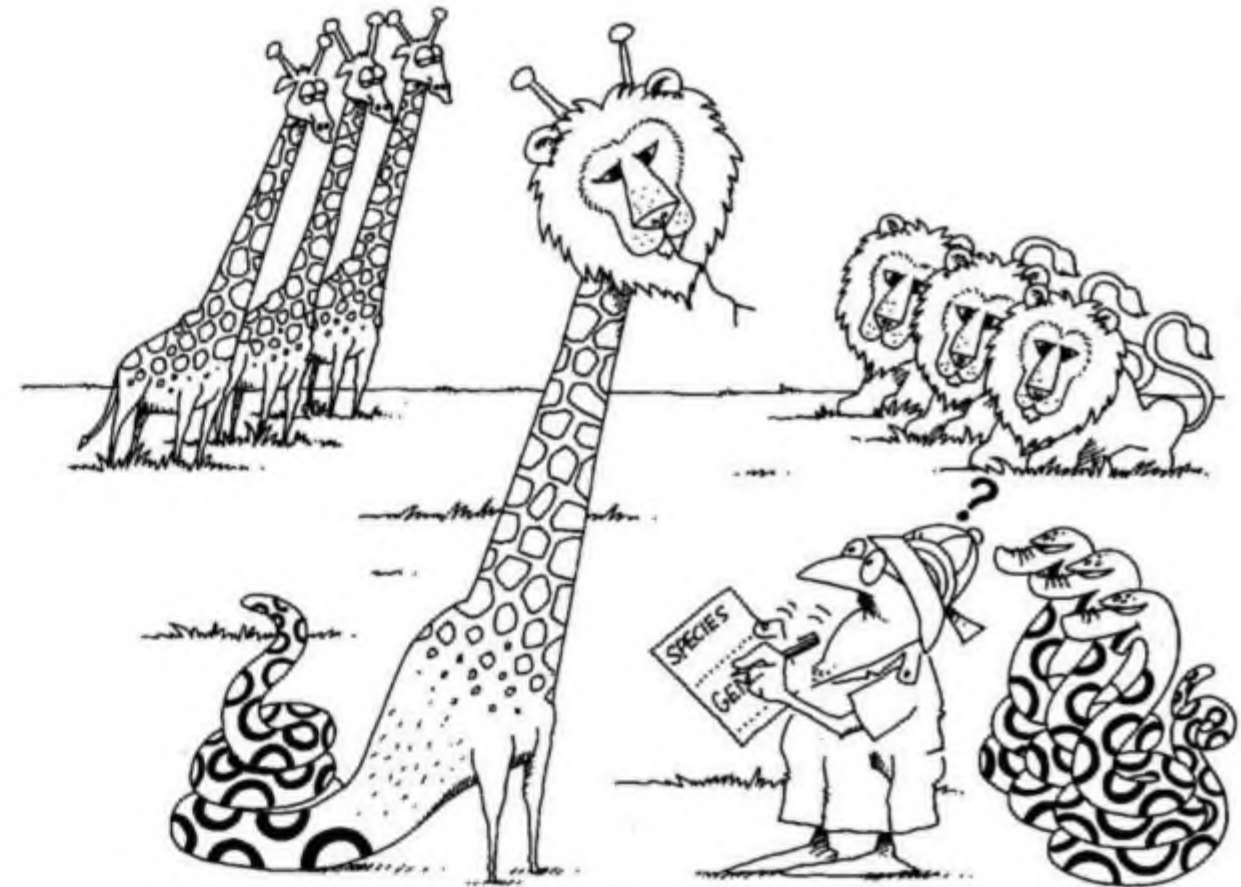
Классификация (Classification)

Нахождение правила, позволяющее отнести объект к тому или иному классу (выбрать класс из числа известных заранее классов) на основе информации о том, к какому классу относятся другие объекты.

Классификация (Classification)

Примеры:

- Задачи распознавание образов (распознавание рукописного текста, фотографии (например, номера автомобиля по фото), идентификация личности по фото, голосу, видео...);



Классификация (Classification)

- Задачи атрибуции (определение авторства / периода создания / страны происхождения ... произведений искусства, археологических находок);
- Задачи диагностики (в медицине и технике).

Кластеризация (Clusterization)

Нахождение правила для автоматического разделения имеющихся объектов на классы на основании сходства тех или иных характеристик (факторов) этих объектов. При этом ни сами классы, ни их количество заранее неизвестны.

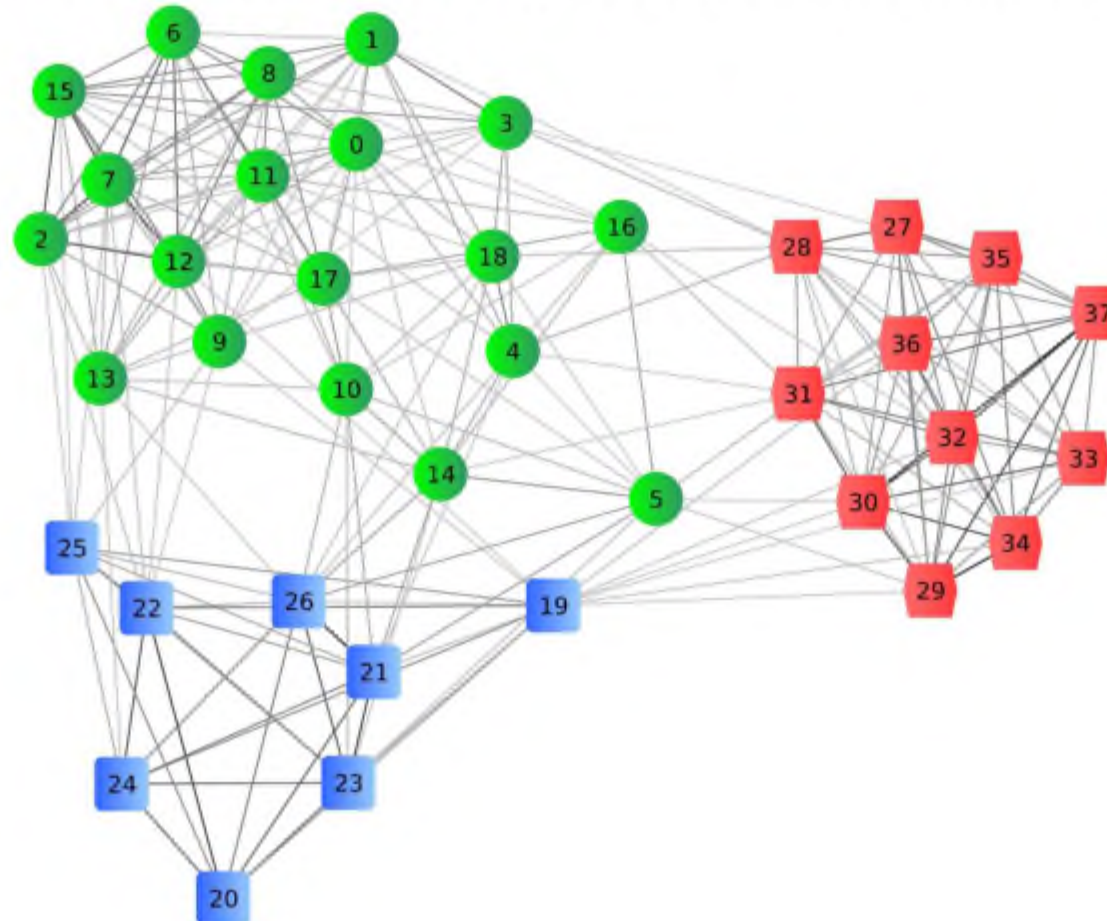
Примеры:

- Сегментация рынка (разделение всех потенциальных потребителей на кластеры для последующего целевого воздействия, например, создания целевой рекламы);
- Задачи разбиения множества индивидов на группы кластеры (в социологии, психологии, биологии)

Кластеризация (Clusterization)

Примеры:

- Сегментация рынка
- Задачи разбиения множества индивидов на группы кластеры (в социологии, психологии, биологии)



Ассоциация (Associations)

Поиск устойчивых закономерностей между случайными событиями, наступающими одновременно.

Пример:

- Анализ покупательской корзины – поиск «устойчивых связей в корзине покупателя» (осуществляется с целью учёта их при планировании расположения отделов в супермаркете).

Последовательность ассоциация, Нахождение шаблонов) (Sequence, Sequential association, Sequential pattern) (Последовательная последовательных

Поиск устойчивых закономерностей между случайными событиями, связанными во времени, т.е. правил вида: после события X через время t происходит событие Y .

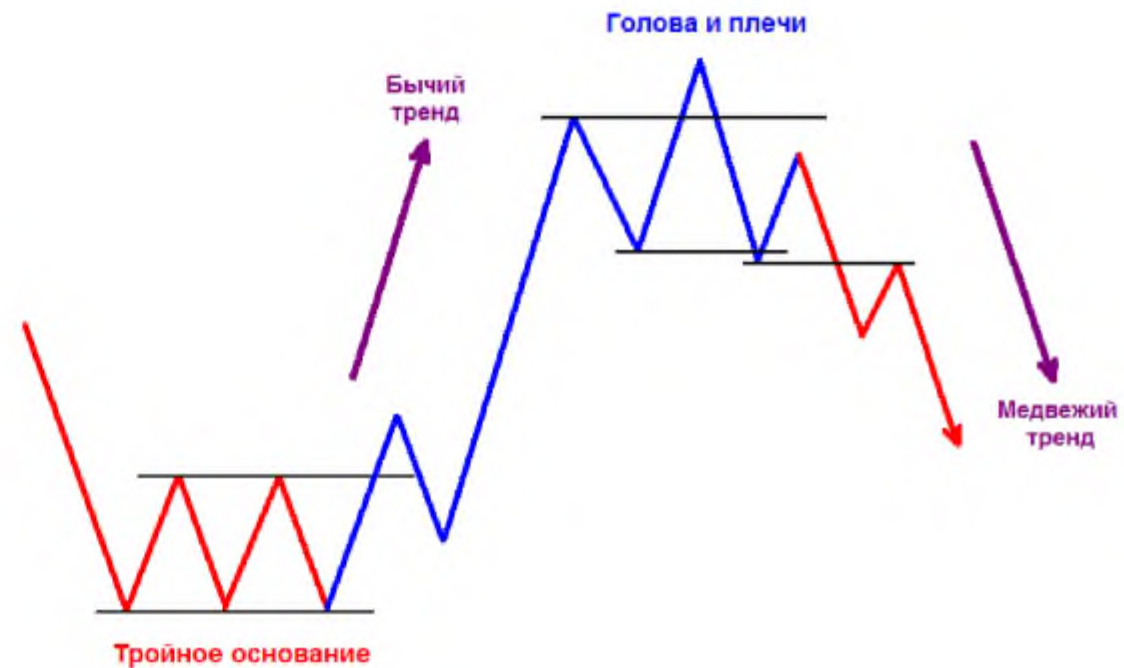
Последовательность (Последовательная ассоциация, Нахождение последовательных шаблонов) (Sequence, Sequential association, Sequential pattern)

Пример:

- После покупки квартиры жильцы в 60% случаев в течение двух недель приобретают холодильник, а в течение двух месяцев в 50% случаев приобретается телевизор.

Визуализация данных (Data Visualization)

Графическое изображение данных (2D и 3D диаграммы, гистограммы, графики, облака точек...)



Анализ отклонений (Deviation Detection)

Обнаружение и анализ данных, наиболее отличающихся от общего множества данных.

Примеры:

- выявление нетипичной сетевой активности позволяет обнаружить вредоносные программы;
- выявление мошенничества с кредитными карточками.

Большие данные – was ist das?

Big data – термин возник в 2011 году

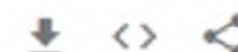
По всему миру ▼

2004 – настоящее время ▼

Все категории ▼

Веб-поиск ▼

Динамика популярности (?)



Примечание

Большие данные – was ist das?

Big data – термин возник в 2011 году

Популярность по регионам ?

Регион ▾ ⬇ ⏪ ⏩ 🔗



1	о-в Св. Елены	100	<div><div></div></div>
2	Сингапур	69	<div><div></div></div>
3	Индия	68	<div><div></div></div>
4	Гонконг	53	<div><div></div></div>
5	Республика Корея	39	<div><div></div></div>

Большие данные – was ist das?

- Big Data – это когда данных больше, чем 100Гб (500Гб, 1ТБ, кому что нравится)
- · Big Data – это такие данные, которые невозможно обрабатывать в Excel
- · Big Data – это такие данные, которые невозможно обработать на одном компьютере
- Big Data – это вообще любые данные.
- Big Data не существует, ее придумали маркетологи.

Большие данные – was ist das?

Большие данные (англ. big data) — серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence.

Горизонтальная масштабируемость.

Поскольку данных может быть сколь угодно много – любая система, которая подразумевает обработку больших данных, должна быть расширяемой. В 2 раза вырос объём данных – в 2 раза увеличили количество железа в кластере и всё продолжило работать.

Отказоустойчивость.

Принцип горизонтальной масштабируемости подразумевает, что машин в кластере может быть много.

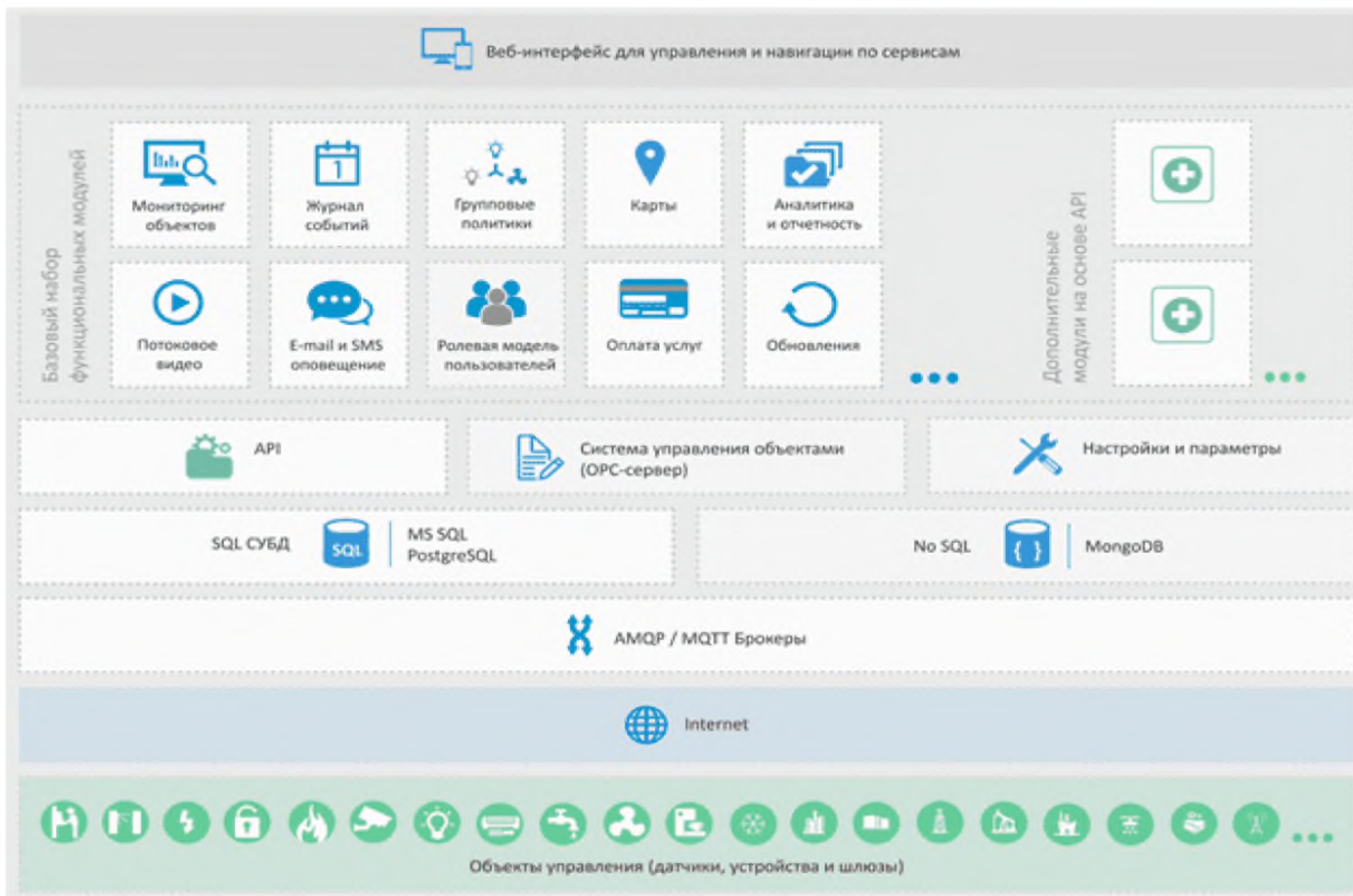
Например, Hadoop-кластер Yahoo имеет более 42000 машин. Часть этих машин будет гарантированно выходить из строя. Методы работы с большими данными должны учитывать возможность таких сбоев и переживать их без каких-либо значимых последствий.

Локальность данных.

В больших распределённых системах данные распределены по большому количеству машин. Если данные физически находятся на одном сервере, а обрабатываются на другом – расходы на передачу данных могут превысить расходы на саму обработку.

Принцип локальности данных – по возможности обрабатываем данные на той же машине, на которой их храним.

И еще раз архитектура





Технологические основы Интернета Вещей

Лекция 4- Облачные платформы Интернета Вещей

Жматов Дмитрий Владимирович
кандидат технических наук, доцент
доцент кафедры Математического обеспечения и
стандартизации информационных технологий