Проектирование высоконагруженных и аналитических систем

Лекция 17 (33)

Высоконагруженные системы и системы реального времени

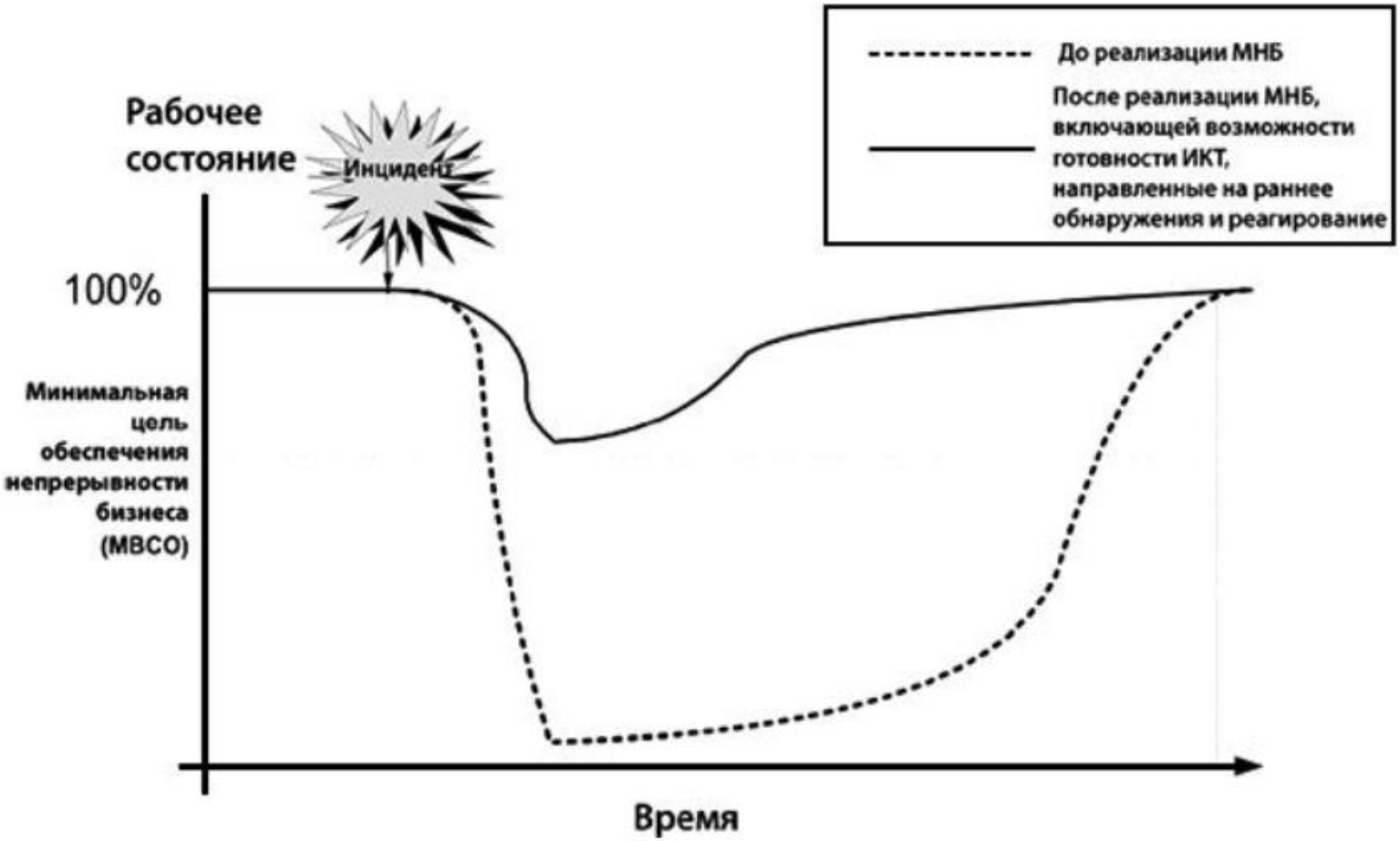
Овчинников П.Е.

МГТУ «СТАНКИН»,

ст.преподаватель кафедры ИС

Терминология: непрерывность бизнеса

ГОСТ Р ИСО/МЭК 27031-2012 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Руководство по готовности информационно-коммуникационных технологий к обеспечению непрерывности бизнеса



ГОСТ Р ИСО/МЭК 27031-2012

Терминология: высокая доступность

ГОСТ Р ИСО/МЭК 27031-2012 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Руководство по готовности информационно-коммуникационных технологий к обеспечению непрерывности бизнеса

В информационно-коммуникационных технологиях "высокая доступность" относится к системам или компонентам, которые непрерывно функционируют в течение желаемого длительного периода времени

Доступность может измеряться по отношению к "100% работоспособности" или "полному отсутствию отказов". Существует широко распространенный, но труднодостижимый эталон доступности систем или продуктов, известный как "пять девяток" (99,999%) доступности

Компьютерная система или сеть состоит из многих компонентов, каждый из которых должен быть в наличии и быть функциональным, чтобы все в целом было работоспособным, и, хотя планирование высокой доступности часто сосредотачивается на резервировании, обработке отказа, хранении данных и доступе к ним, другие компоненты инфраструктуры, такие как энергоснабжение и охлаждение, являются в равной степени важными

ГОСТ Р ИСО/МЭК 27031-2012

Терминология: высоконагруженные системы

Высоконагруженные приложения. Программирование, масштабирование, поддержка

Источники нагрузки:
- пользователи
- роботы
- приложения

Нагрузка:
- трафик
- оперативная память
- процессор
- дисковая память

Негативные последствия:
- замедление работы
- потеря данных
- отказ в обслуживании (DoS)

Высоконагруженные системы (видеокурс)

HighLoad: оптимизация трафика

Подходы к описанию сетевого трафика

"Классические" методы сетевых расчетов и моделирования, основанные на пуассоновских моделях, предполагали, что все поступившие в исследуемую систему вызовы **взаимно независимы** и интервалы времени между приходом двух последующих вызовов распределены согласно экспоненциальному закону.

В то же время **самоподобный трафик** обладает медленно убывающей автокорреляционной функцией, плотность распределения вероятности интервалов между моментами прихода двух последовательных вызовов подчиняется степенному закону

Одно из важных свойств самоподобия трафика - сохранение своей структуры в разные масштабы времени

Из-за таких свойств самоподобного трафика традиционные методы расчета характеристик функционирования сетей дают **слишком оптимистические** результаты и приводят к недооценке реальной нагрузки

HighLoad: кэширование

Кэш или **кеш** (англ. *cache*, от фр. *cacher* — «прятать»; произносится [kæʃ] — «кэш») — промежуточный буфер с **быстрым доступом** к нему, содержащий информацию, которая может быть запрошена с **наибольшей вероятностью**

Доступ к данным в кэше осуществляется быстрее, чем выборка исходных данных из более медленной памяти или удаленного источника, однако её объём существенно ограничен по сравнению с хранилищем исходных данных

Кэширование применяется ЦПУ, жёсткими дисками, браузерами, веб-серверами, службами DNS и WINS

Когда клиент кэша обращается к данным, прежде всего исследуется кэш:

- если в кэше найдена запись с идентификатором, совпадающим с идентификатором затребованного элемента данных, то используются элементы данных в кэше. Такой случай называется ***попаданием кэша***
- если в кэше не найдена запись, содержащая затребованный элемент данных, то он читается из основной памяти в кэш, и становится доступным для последующих обращений. Такой случай называется ***промахом кэша***

Процент обращений к кэшу, когда в нём найден результат, называется *уровнем попаданий*, или ***коэффициентом попаданий*** в кэш.

Кэш (Википедия)

HighLoad: балансировка нагрузки

В терминологии компьютерных сетей **балансировка нагрузки** или **выравнивание нагрузки** (англ. *load balancing*) — метод распределения заданий между несколькими сетевыми устройствами (например, серверами) с целью оптимизации использования ресурсов, сокращения времени обслуживания запросов, горизонтального масштабирования кластера (динамическое добавление/удаление устройств), а также обеспечения отказоустойчивости (резервирования)

В компьютерах балансировка нагрузки распределяет нагрузку между несколькими вычислительными ресурсами, такими как компьютеры, компьютерные кластеры, сети, центральные процессоры или диски

Цель балансировки нагрузки:

- оптимизация использования ресурсов
- максимизация пропускной способности
- уменьшение времени отклика
- предотвращение перегрузки какого-либо одного ресурса

Использование нескольких компонентов балансировки нагрузки вместо одного компонента может повысить надежность и доступность за счет резервирования.

Балансировка нагрузки (Википедия)

HighLoad: управление блокировками

Ведущий узел и блокировки

Зачастую системе необходимо наличие только одного экземпляра чего-либо, например:

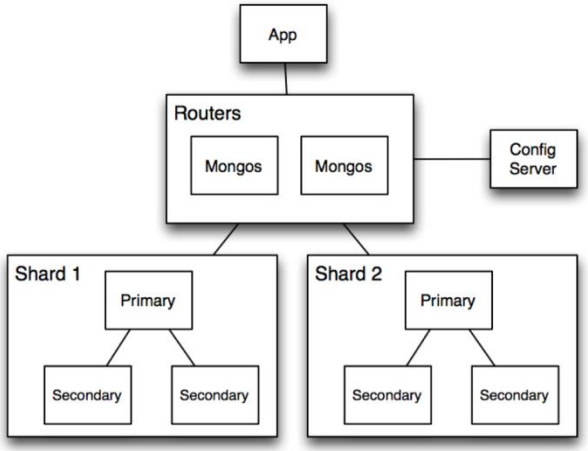
- только один узел может быть ведущим для секции базы данных, чтобы избежать ситуации разделения вычислительных мощностей
- только одна транзакция или один клиент может удерживать блокировку на конкретный ресурс или объект, чтобы предотвратить конкурентную запись в него и его порчу
- только один пользователь может зарегистрировать конкретное имя пользователя, поскольку оно должно идентифицировать пользователя уникальным образом

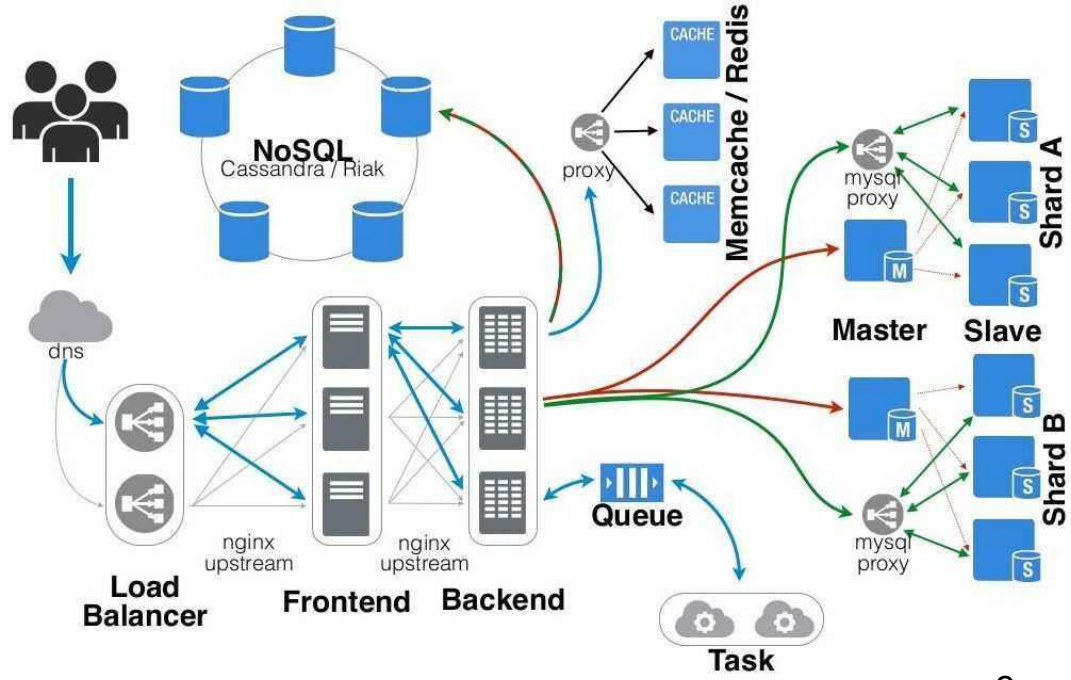
Случай, когда узел продолжает вести себя как «избранный», несмотря на то, что кворум остальных узлов объявил его неработающим, может привести к проблемам в недостаточно тщательно спроектированной системе

Подобный узел способен отправлять сообщения другим узлам в качестве самозваного «избранного», и если другие узлы с этим согласятся, то система в целом может начать работать неправильно.

HighLoad: архитектурные паттерны

- сервисно-ориентированная архитектура
- вертикальное масштабирование
- горизонтальное масштабирование
- отложенные вычисления
- асинхронная обработка
- конвейерная обработка
- использование толстого клиента
- кеширование
- функциональное разделение
- шардинг
- виртуальные шарды
- центральный диспетчер
- репликация
- партиционирование
- кластеризация
- денормализация
- введение избыточности
- нереляционные субд
- толстый клиент
- параллельное выполнение





[HighLoad (хабрахабр)]

Терминология: онлайн и офлайн

Термины **онлайн** (англ. *online*) и **офлайн** (англ. *offline*) имеют значение в отношении к компьютерным технологиям и телекоммуникациям:

- **онлайн** указывает на состояние подключения **,** а
- **офлайн** указывает на отключенное состояние

В современной терминологии это обычно относится к интернет-соединению , но (особенно когда оно выражено «on line») может относиться к любому элементу оборудования или функциональному блоку, который подключен к более крупной системе

Быть онлайн означает, что оборудование или подсистема подключены или что они готовы к использованию

Понятие «онлайн» также описывает действия и данные, доступные в интернете, например: «онлайн-идентификация », «онлайн-игры», «онлайн-шоппинг», «онлайн-банкинг» и «онлайн обучение»

Аналогичное значение также дают префиксы «кибер» и «е», как в словах «киберпространство », «киберпреступность», «электронная почта» и «электронная коммерция»

онлайн и офлайн (Википедия)

Терминология: системы реального времени

Система реального времени (СРВ) — это система, которая должна **реагировать на события** во внешней по отношению к системе среде или воздействовать на среду в **рамках требуемых временных ограничений**

Другими словами, обработка информации системой должна производиться за определённый конечный период времени, чтобы поддерживать **постоянное** и **своевременное** взаимодействие со средой

Естественно, что масштаб времени контролирующей системы и контролируемой ей среды должен совпадать

Под ***реальным временем*** понимается количественная характеристика, которая может быть измерена реальными физическими часами, в отличие от ***логического времени***, определяющего лишь качественную характеристику, выражаемую относительным порядком следования событий

Говорят, что система работает в ***режиме реального времени***, если для описания работы этой системы требуются количественные временны́е характеристики

Системы реального времени (Википедия)

Терминология: системы реального времени

Процессы (задачи) систем реального времени могут иметь следующие характеристики и связанные с ними ограничения:

- дедлайн (англ. *deadline*) — критический **срок обслуживания**, предельный срок завершения какой-либо работы
- латентность (англ. *latency*) — **время отклика** (время задержки) системы на внешние события
- джиттер (англ. *jitter*) — **разброс значений** времени отклика

Можно различить:

- джиттер **запуска** (англ. *release jitter*) — период времени от готовности к исполнению до начала собственно исполнения задачи и
- джиттер **вывода** (англ. *output jitter*) — задержка по окончании выполнения задачи

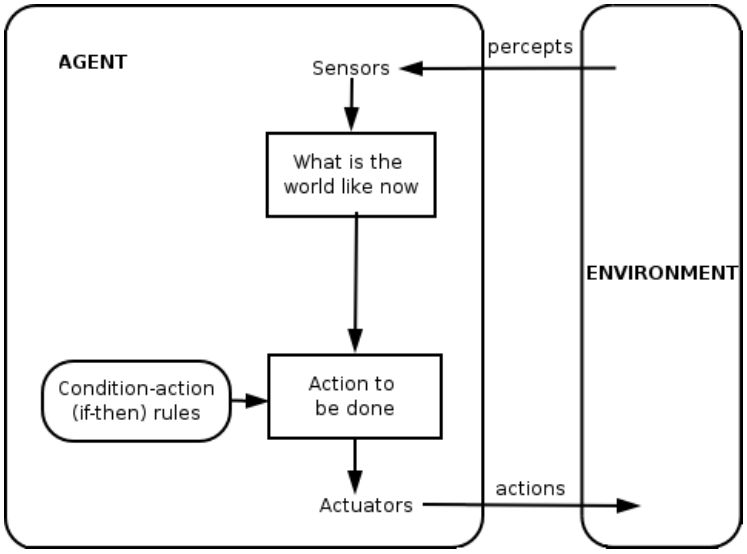
События реального времени могут относиться к одной из трёх категорий:

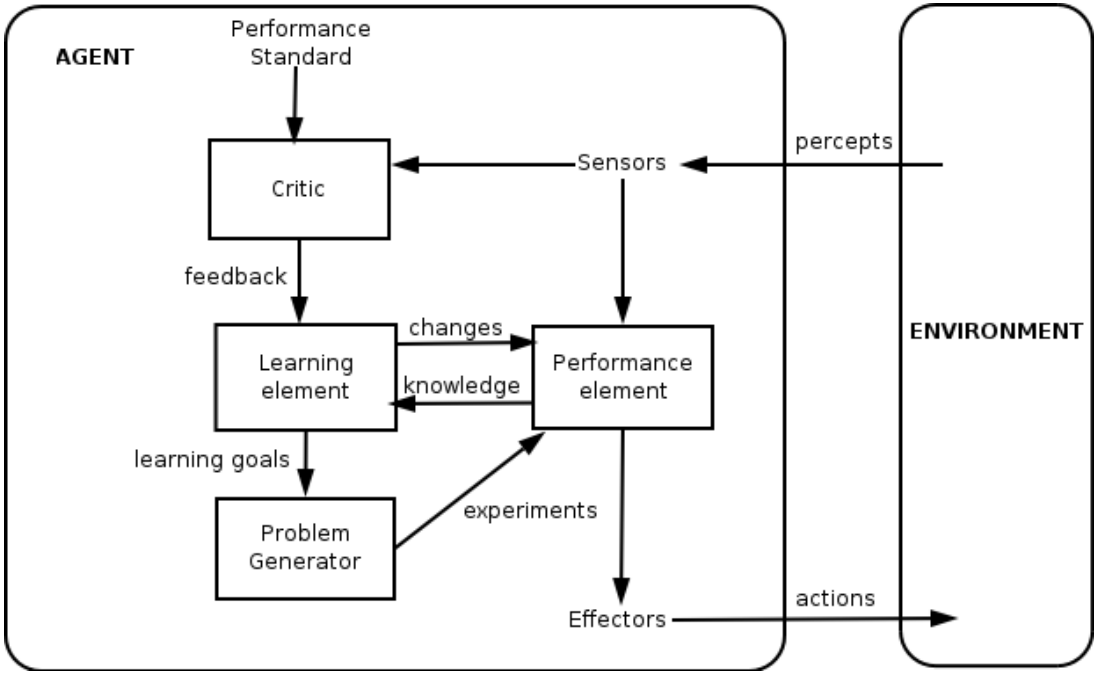
- **асинхронные** *события* — полностью непредсказуемые события
- **синхронные** *события* — предсказуемые события, случающиеся с определённой регулярностью
- **изохронные** *события* — регулярные события (разновидность асинхронных), случающиеся в течение интервала времени

Системы реального времени (Википедия)

Терминология: интеллектуальные агенты

Многоагентная система (МАС, англ. Multi-agent system) — это система, образованная несколькими взаимодействующими интеллектуальными агентами. Многоагентные системы могут быть использованы для решения таких проблем, которые сложно или невозможно решить с помощью одного агента или монолитной системы.

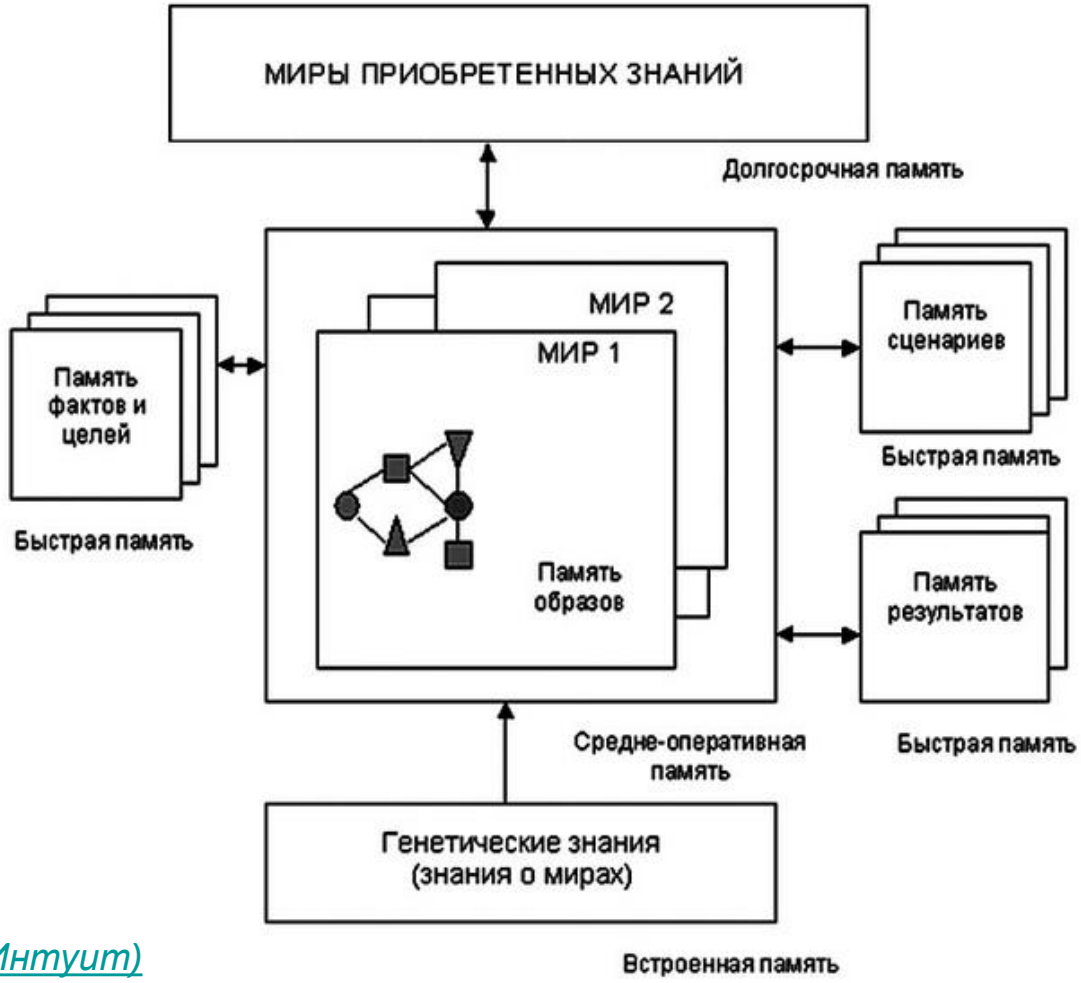




МАС
Мультиагентные системы (Интуит)

Терминология: интеллектуальные агенты

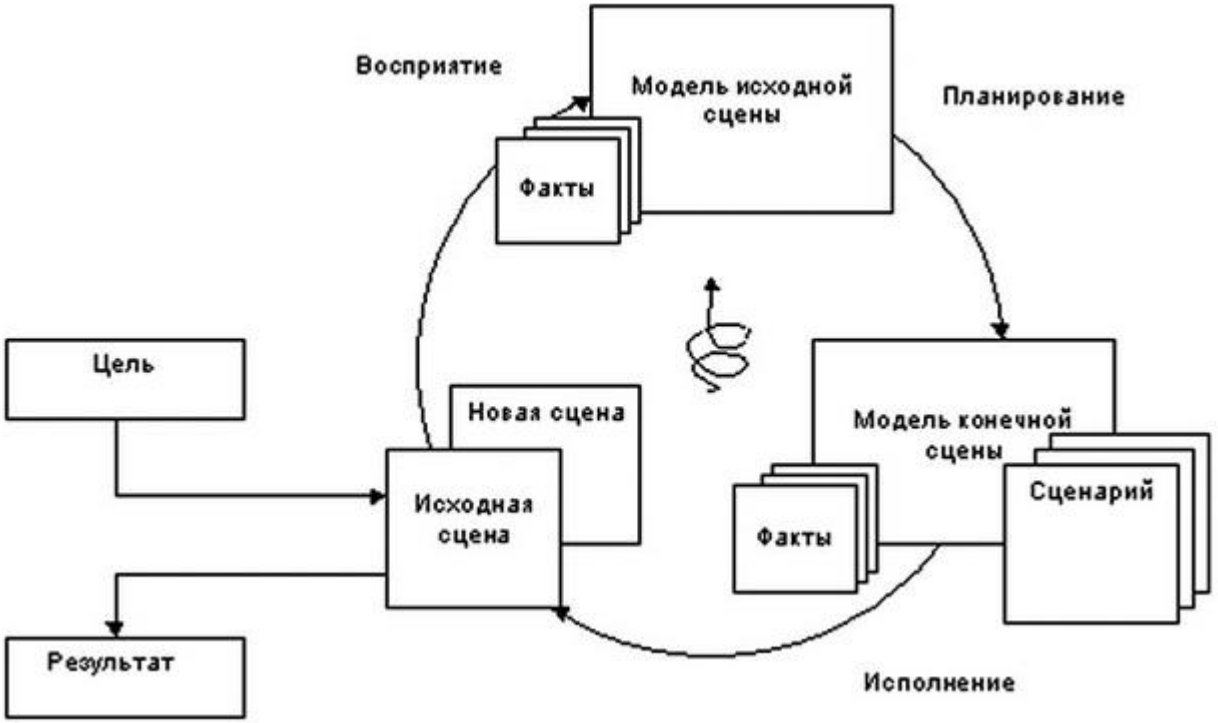
Концепция "Агентов и Миров" реализует формирование общего мира деятельности кооперирующих сторон и миров деятельности каждой из них, путем создания единой комплексной среды. В этом подходе мир действий — это модель среды деятельности, базирующаяся на знаниях



[МАС](#)
[Мультиагентные системы (Интуит)](#)

Терминология: интеллектуальные агенты

Важным элементом при создании мультиагентных систем является язык коммуникации агентов — Agent Communication Language, который определяет типы сообщений, которыми могут обмениваться агенты. В рамках парадигмы коммуникации между агентами, кооперация между ними достигается за счет ACL, языка контента и онтологии, которые определяют набор базовых концепций, используемых в сообщениях кооперации. Онтология здесь выступает синонимом понятия API (Application Programming Interface), т.е. она определяет конкретный интерфейс интеллектуальных агентов.



[МАС](#)
[Мультиагентные системы (Интуит)](#)

Терминология: интернет вещей (IoT)

Интернет вещей (англ. *Internet of Things*, *IoT*) — концепция вычислительной сети физических предметов («вещей»), оснащённых встроенными технологиями для взаимодействия друг с другом или с внешней средой, рассматривающая организацию таких сетей как явление, способное перестроить экономические и общественные процессы, исключающее из части действий и операций необходимость участия человека

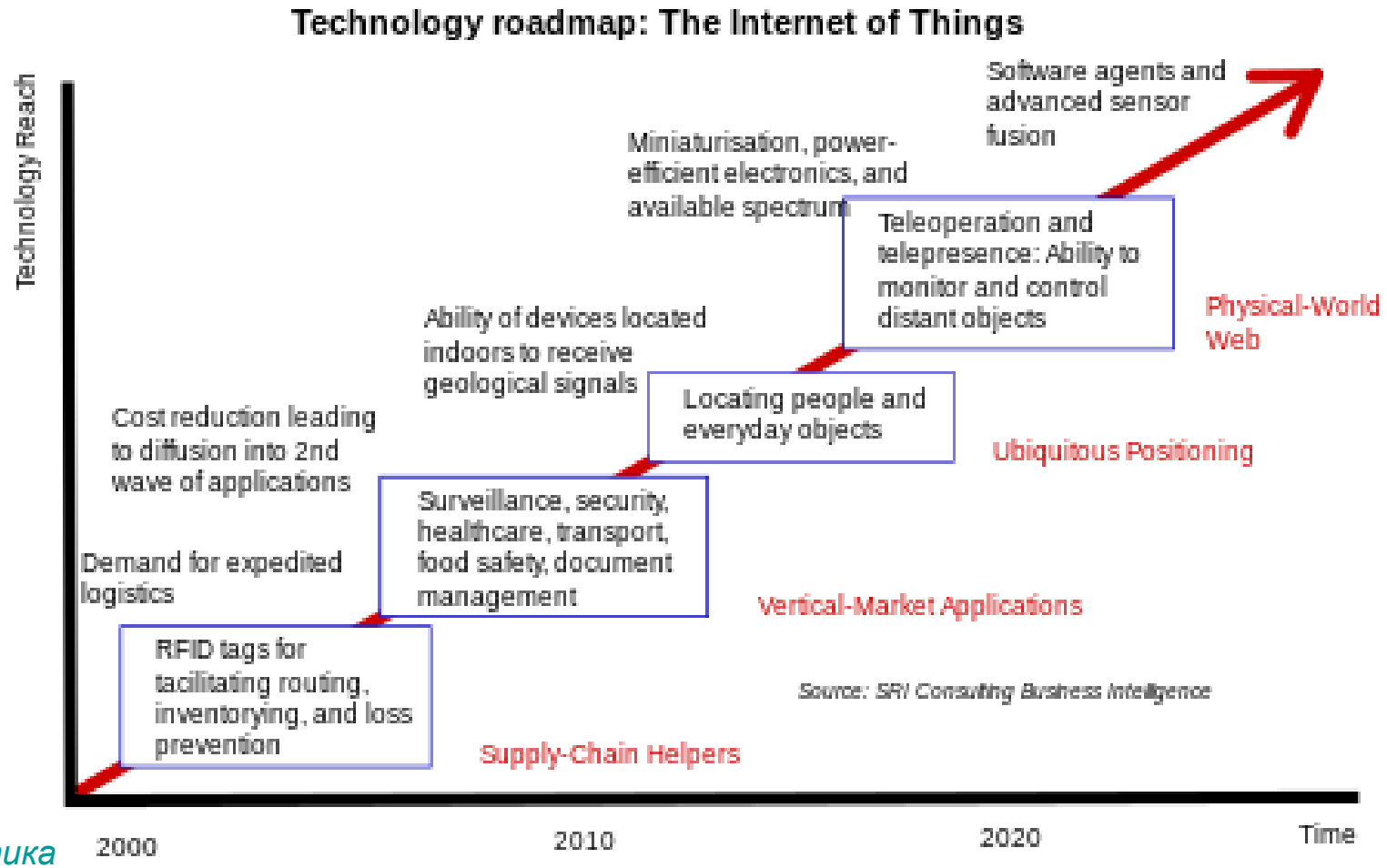
Концепция сформулирована в 1999 году как осмысление перспектив широкого применения средств радиочастотной идентификации для взаимодействия физических предметов между собой и с внешним окружением

Наполнение концепции «интернета вещей» многообразным технологическим содержанием и внедрение практических решений для её реализации начиная с 2010-х годов считается устойчивой тенденцией в информационных технологиях, прежде всего, благодаря повсеместному распространению беспроводных сетей, появлению облачных вычислений, развитию технологий межмашинного взаимодействия, началу активного перехода на IPv6 и освоению программно-конфигурируемых сетей

Интернет вещей (Википедия)

Терминология: интернет вещей (IoT)

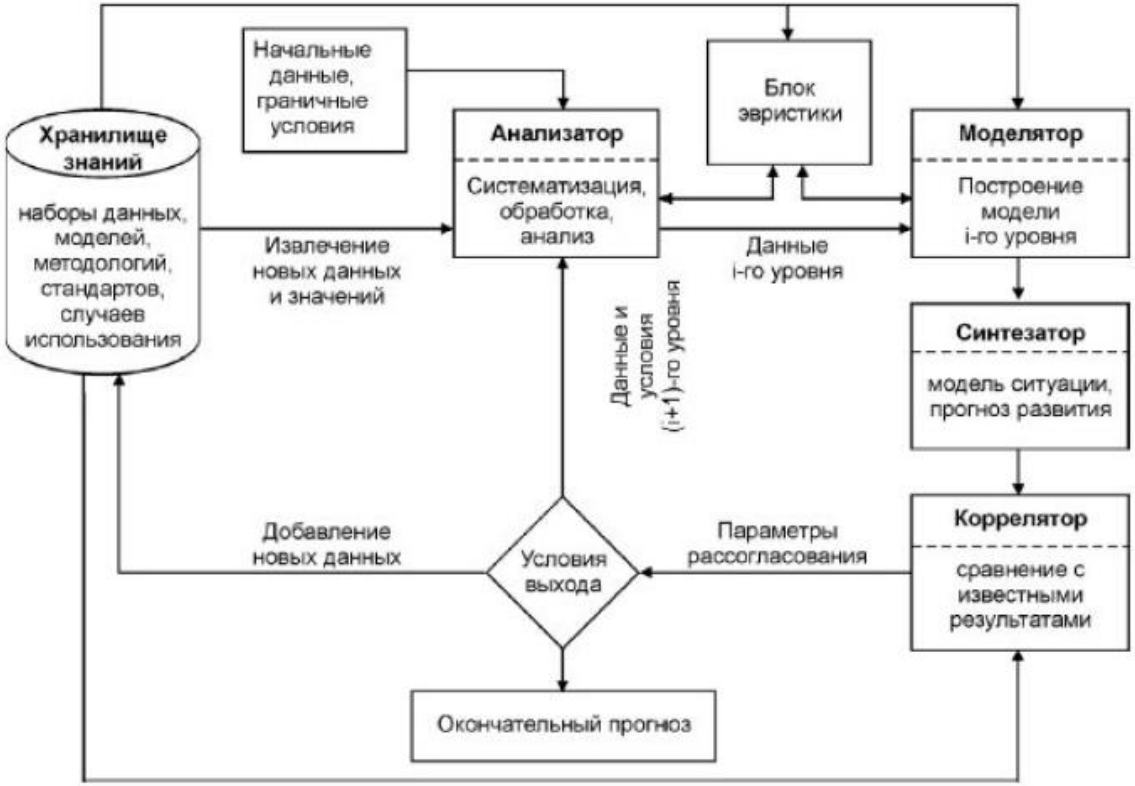
Киберне́тика (от др.-греч. κυβερνητική — «искусство управления») — наука об общих закономерностях получения, хранения, передачи и преобразования информации в сложных управляющих системах, будь то машины, живые организмы или общество

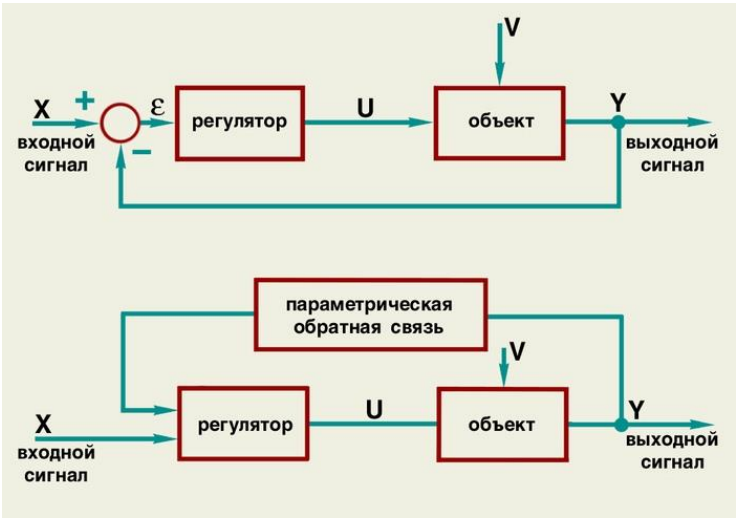


Кибернетика
IoT

Терминология: интернет вещей (IoT)

Кибернетика включает изучение обратной связи, чёрных ящиков и производных концептов, таких как управление и коммуникация в живых организмах, машинах и организациях, включая самоорганизации

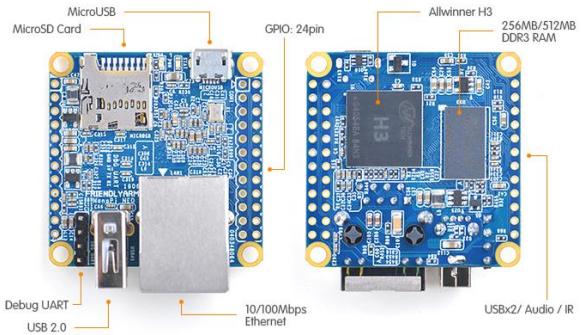




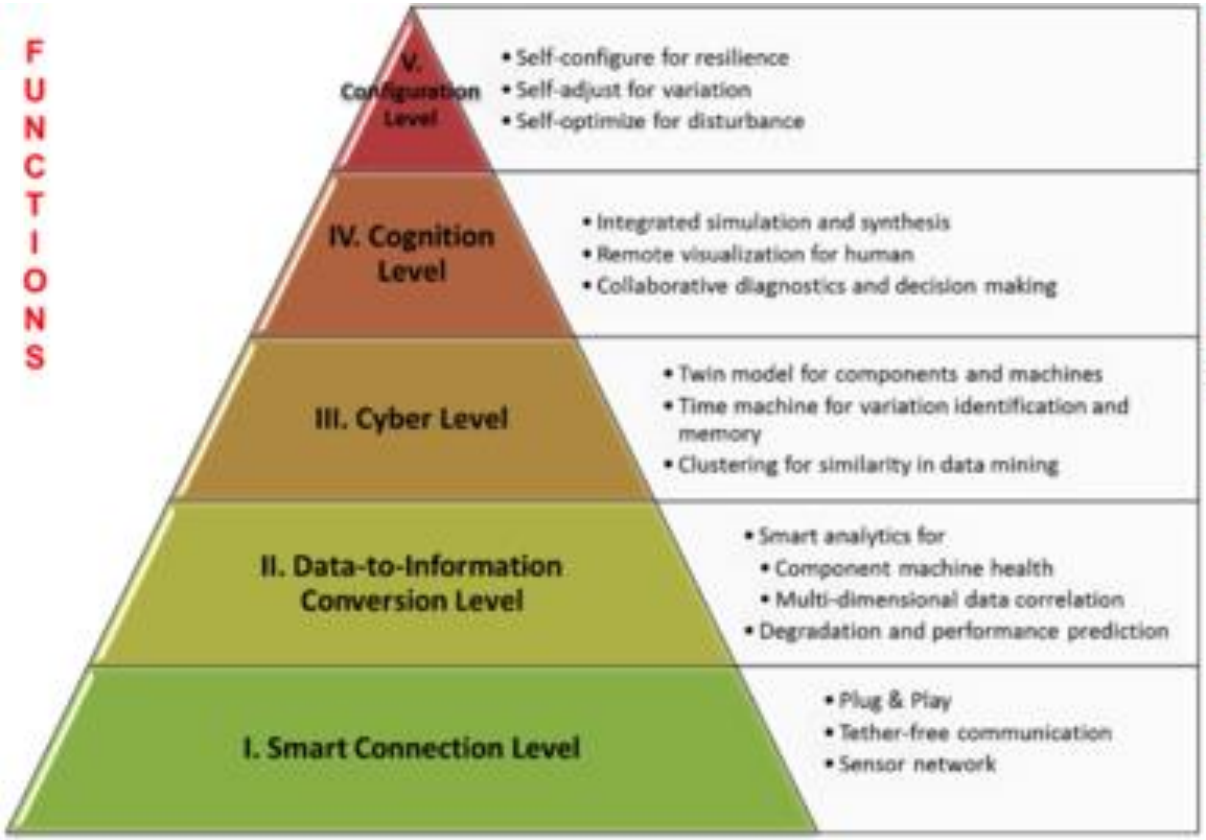
Кибернетика
IoT

Терминология: индустриальный интернет вещей (IIoT)

Промышленный Интернет вещей (англ. *Industrial Internet of Things, IIoT*) - это концепция, при которой различные промышленные устройства, такие как датчики или оборудование, объединены в сеть посредством использования сети Интернет.



FUNCTIONS



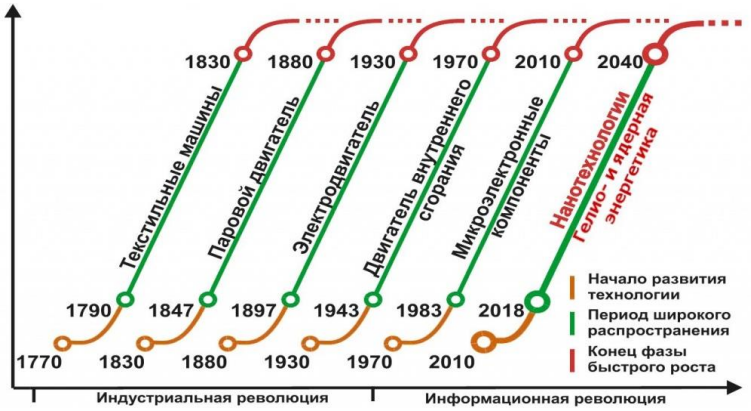
ATTRIBUTES

Индустриальный интернет вещей

Терминология: технологический уклад

Технологический уклад – экономический термин, обозначающий установившийся порядок чего-либо, характеризующийся единым техническим уровнем составляющих его производств, связанных потоками качественно однородных ресурсов, опирающихся на общие ресурсы квалифицированной рабочей силы, общий научно-технический потенциал.

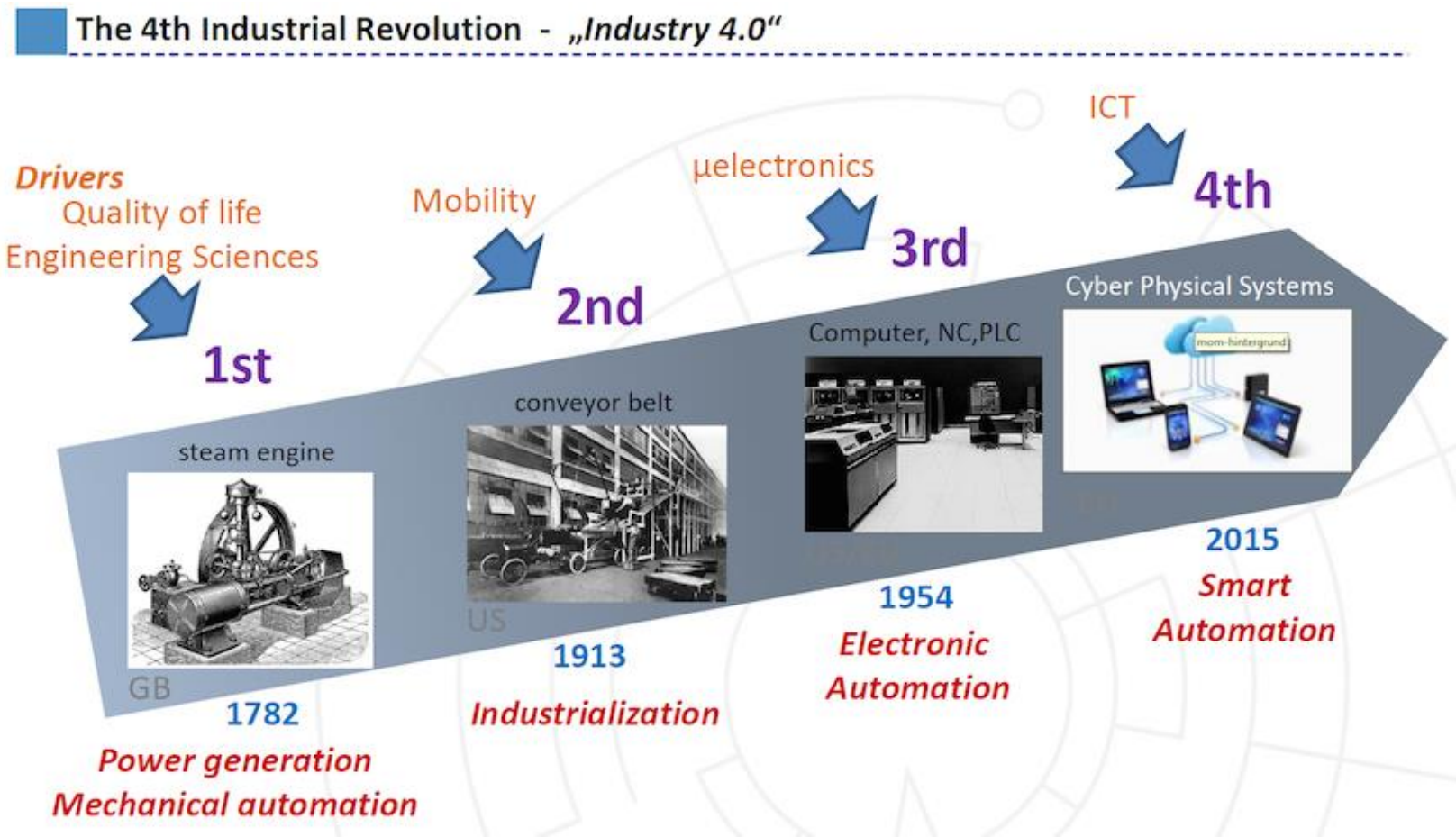
Основной особенностью текущего, **пятого технологического уклада** (1985-2035гг.), является массовый переход от функционирующих разрозненно фирм к появлению единой сети крупных и мелких компаний, соединенных электронными коммуникациями на основе сети Интернет, осуществляющих тесное взаимодействие в области технологий, контроля качества продукции, планирования инноваций и т.д.



Технологический уклад
Импортозамещение

Терминология: Индустрия 4.0

Четвёртая промышленная революция (англ. The Fourth Industrial Revolution) — прогнозируемое событие, массовое внедрение киберфизических систем в производство (индустрия 4.0), обслуживание человеческих потребностей, включая быт, труд и досуг



Четвертая промышленная революция
Индустрия 4.0

Терминология: цифровое производство

Цифровое производство — интегрированная компьютерная система, включающая в себя средства:

- численного моделирования,
- трехмерной (3D) визуализации,
- инженерного анализа и
- совместной работы,

предназначенные для разработки конструкции изделий и технологических процессов их изготовления

Цифровое производство начиналось с таких инициатив, как конструирование с учетом технологичности (DFM), компьютерно-интегрированное производство (CIM), гибкое производство, бережливое производство и других, направленных на расширение совместной работы при конструкторско-технологической подготовке производства изделий

Терминология: робототехника

Робототе́хника (от робот и техника; англ. *robotics* — **роботика**, *роботехника*) — прикладная наука, занимающаяся разработкой автоматизированных технических систем и являющаяся важнейшей технической основой интенсификации производства

Робототехника опирается на такие дисциплины, как электроника, механика, телемеханика, информатика, а также радиотехника и электротехника.

Выделяют строительную, промышленную, бытовую, авиационную и экстремальную (военную, космическую, подводную) робототехнику