



**Проблемы семантической  
безопасности в системах сбора и  
анализа данных АСУТП- BigData,  
МОНИТОРИНГ ИНТЕРНЕТ ВЕЩЕЙ**

ВОЛКОВ ИГОРЬ АЛЕКСЕЕВИЧ

(К.М.Н.,К.Т.Н. ДОЦЕНТ)





Теорема CAP (известная также как теорема Брюера) — эвристическое утверждение о том, что в любой реализации распределённых вычислений возможно обеспечить не более двух из трёх следующих свойств:

согласованность данных (англ. consistency) — во всех вычислительных узлах в один момент времени данные не противоречат друг другу;

доступность (англ. availability) — любой запрос к распределённой системе завершается корректным откликом, однако без гарантии, что ответы всех узлов системы совпадают;

устойчивость к разделению (англ. partition tolerance) — расщепление распределённой системы на несколько изолированных секций не приводит к некорректности отклика от каждой из секций.

Акроним CAP в наименовании теоремы сформирован из первых букв английских наименований этих трёх свойств.

Теорема CAP (известная также как теорема Брюера) — эвристическое утверждение о том, что в любой реализации распределённых вычислений возможно обеспечить не более двух из трёх следующих свойств:

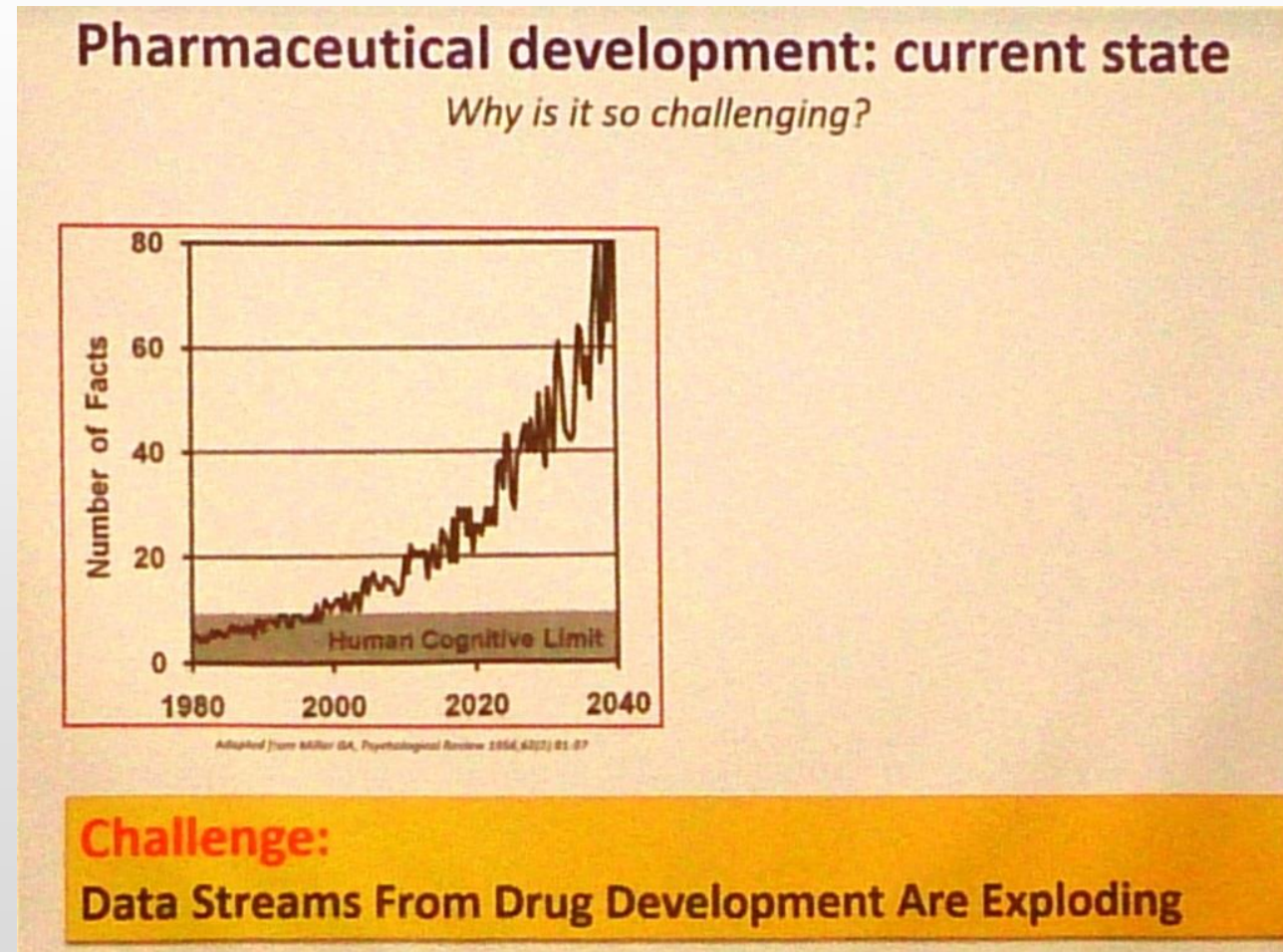
- **СОГЛАСОВАННОСТЬ**
- **ДОСТУПНОСТЬ**
- **УСТОЙЧИВОСТЬ К РАЗДЕЛЕНИЮ**

▪ Eric Brewer



# Информационные революции

- Речь
- Письменность
- Книгопечатание
- Теле- радиовещание
- Персональный компьютер
- Интернет – Информационный поток



# Эпоха больших данных



Информационный поток



Облачные вычисления



Большие данные



# Источники больших данных

- Исследователи: данные исследований и клинических испытаний
- Клиники: данные историй болезней, результатов обследований
- Пациенты: данные геолокации, покупок, персональных устройств
- Учреждения медицинских услуг и розничной/аптечной торговли



# Информационные технологии

- Телемедицина
- Электронное здравоохранение
- Нейронные сети, генетические алгоритмы, искусственный интеллект
- Системы поддержки принятия решений
- Блокчейн (Blockchain)
- Робототехника
- Компьютерное Распознавание образов



# Уровни оценки

- Доказательная медицина (Evidence-based medicine)
- Оценка технологий здравоохранения (ОТЗ, Health Technology Assessment, HTA)
- Доказательная организация здравоохранения (Evidence-based Health Management)



# Оружие против лженауки

- **Наукометрия**
- **Надлежащая медицинская практика  
(Good clinical practice, GCP)**





# ГОСТР 52379-2005

## Надлежащая клиническая практика (GCP)

### 2. Принципы надлежащей клинической практики (GCP)

- Этический принцип
- Оценка рисков
- Права и безопасность пациентов
- Информационный принцип
- Научный принцип
- Ответственность врача
- Добровольное информированное согласие
- Точность, правильность представления, интерпретации и верификации информации
- Конфиденциальность информации

# Осуществление экспертизы и контроля

3. Воздействие на происходящее  
и/или изменение ожиданий



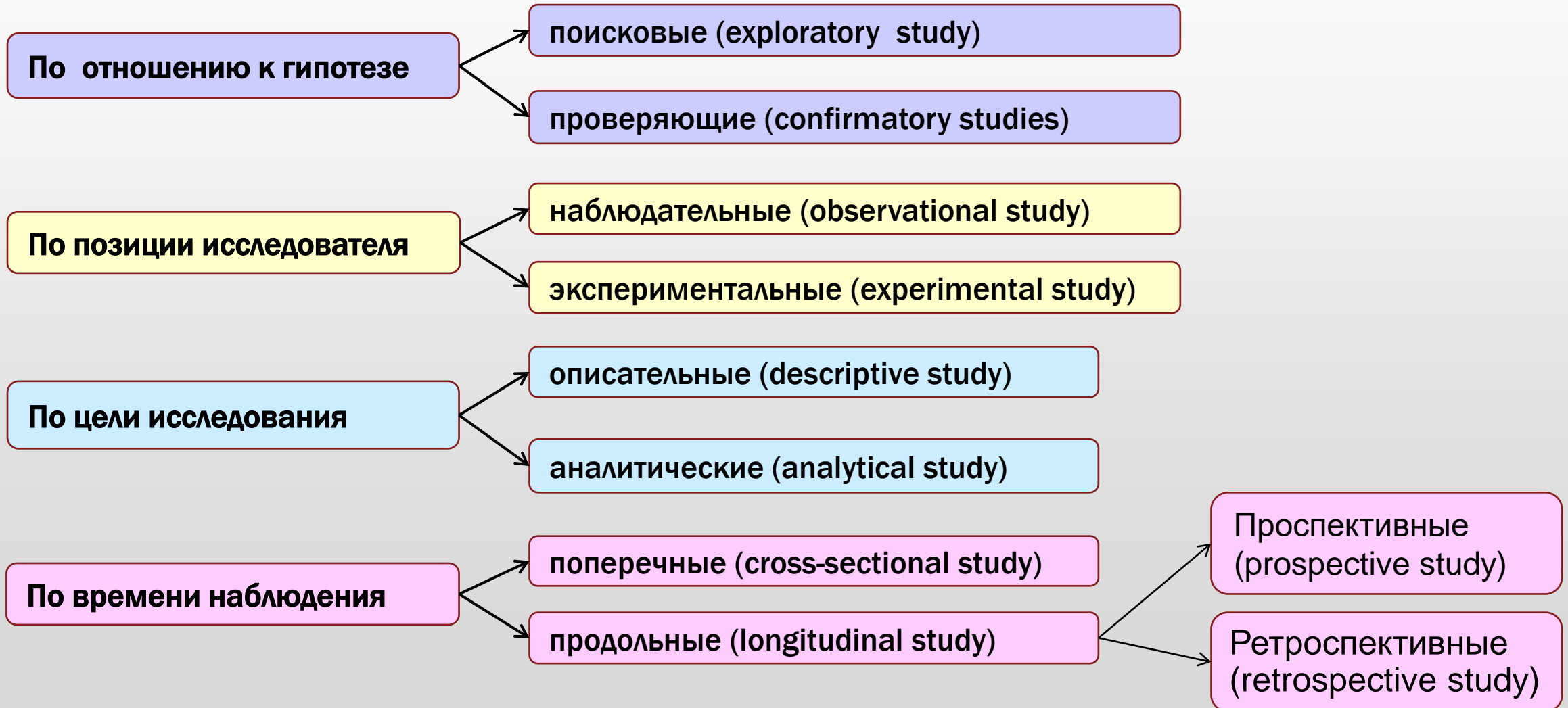
2. Сравнение происходящего с ожиданиями

1. Наблюдение за происходящим

# Виды дизайнов исследований



# Классификация исследований





В настоящее время в ряде случаев более эффективны (как показала эксплуатация – на платформах Unix, Linux, Microsoft, Android, IOS (Apple) ) не реляционные СУБД – NoSQL.

На современных интернет web-сайтах используются традиционные реляционные SQL СУБД ( например, реализующие анализ и управление данными с помощью реляционных SQL запросов в системе 1 C).

Для обеспечения семантической безопасности NoSQL (сохранность/непротиворечивость) данных авторами осуществляется дублирование на разные узлы кластерного сервера

- **1. Используемое оборудование и программное обеспечение**
- Создание программного обеспечения для решения поставленной задачи осуществлялось в среде программирования C++ и NI LabVIEW 8 с модулем NI LabVIEW Real Time. Для реализации системы была использована платформа NI PXI с контроллером реального времени NI PXI-8096 и многофункциональным устройством сбора данных NI PXI-6251 для тестирования (сравнения скорости обработки реляционных SQL запросов и в системе 1 C и не реляционные СУБД – NoSQL Cassandra ).
- **2. Описание решения**
- Сигналы преобразователя и датчика-метки поступают на вход измерительного блока, представляющего собой шасси PXI-1042 со встроенными в него контроллером реального времени PXI-8096 и многофункциональным устройством сбора данных NI PXI-6251.



Результаты исследования показали, что NoSQL позволяют быстро (по сравнению с традиционными реляционными СУБД):

- изменять / вставлять новые данные;
- логировать (замораживать) информацию для анализа данных.

В настоящее время NoSQL используются в системах BigData с большими потоками входных данных:

- рекламные, баннерные и партнёрские сети;
- счётчики интернет статистики;
- телеметрические системы.

- **Реализация NoSQL хранит данные (Базу Данных) в виде колонок (в отличие от строк в реляционных СУБД), которые представляют из себя триаду:**
  - - ключ;
  - - значение;
  - - время сохранения (time stamp).
- **Для обеспечения Семантической безопасности в АСУТП NoSQL (сохранность/непротиворечивость) данных осуществляется дублирование на разные узлы кластера. При настройке Кластера задаётся коэффициент дублирования (уровень репликации). По умолчанию уровень репликации равен 3. Это значит, что одни и те же данные будут храниться на трех разных узлах кластера, и при выходе из строя одного из узлов данные будут взяты с другого. Таким образом обеспечивается Надёжность NoSQL (эта кластеризация технически реализуется с помощью виртуализации и синхронизации NoSQL данных).**



По результатам непрерывной эксплуатации IoT, по мнению авторов, к недостаткам подхода NoSQL следует отнести то NoSQL, что СУБД колоночного типа хорошо (и быстро) работает на Insert/Вставку/Включение колонки (данных), а на выборку/Select - медленнее. в кластерных распределённых системах одновременно невозможно иметь согласованные и не противоречивые данные.

Поэтому при настройке кластера NoSQL выбираются пошагово разные уровни согласованности:

1 шаг - данные посылаются на сервер СУБД, когда они поступают от узлов в соответствии со значением уровня репликации;

2 шаг - аналогичный 1, но на уровне кластера;

3 шаг - данные АСУТП IoT посылаются на сервер СУБД, когда они поступают от всех узлов системы. Таким образом обеспечивается наивысший уровень согласованности/безопасности (запоминаются только самые последние данные по TTL).

- NoSQL хорошо использовать в информационной системе (Web Cloud хранилище) совместно с традиционными реляционными СУБД, в этом случае каждое хранилище будет решать свою ОТДЕЛЬНУЮ задачу (принцип Разделяй и Управляй), эффективно используя преимущества каждого подхода.
- Один из недостатков NoSQL – тройной рост объёма дискового пространства по сравнению с реляционными СУБД. Поэтому в системе используются:
  - - TTL (время жизни данных);
  - - распределение данных по отдельным семействам колонок (наследование);
  - - распределение данных по отдельным кластерам.



NoSQL в кластере обеспечивает CAP - Consistency-Availability-Partition tolerance.

Рассмотрим составляющие CAP:

**Согласованность данных (Consistency)** – во всех вычислительных узлах в один момент времени данные не противоречат друг другу. Иными словами, как только мы успешно записали данные в реализуемое распределенное хранилище, любой клиент при запросе получит эти последние данные.

**Доступность (Availability)** – любой запрос к распределенной системе завершается корректно при наличии операций обновления; на любой запрос к системе мы получаем наши данные или информацию об их отсутствии, если их не сохраняли.

**Устойчивость к разделению (Partition Tolerance)** – расщепление распределенной системы на несколько изолированных секций не приводит к некорректности отклика от каждой из секций. Или, другими словами, если какие-то компоненты выходят из строя, то можно считать, что данные компоненты просто теряют связь со всей остальной системой, а система в целом остается работоспособной.

Под Устойчивостью понимается, что данные внутри кластера не смогут потеряться.

Нарушение принципа Согласованности данных может возникнуть там, где имеется распределенное хранение данных.

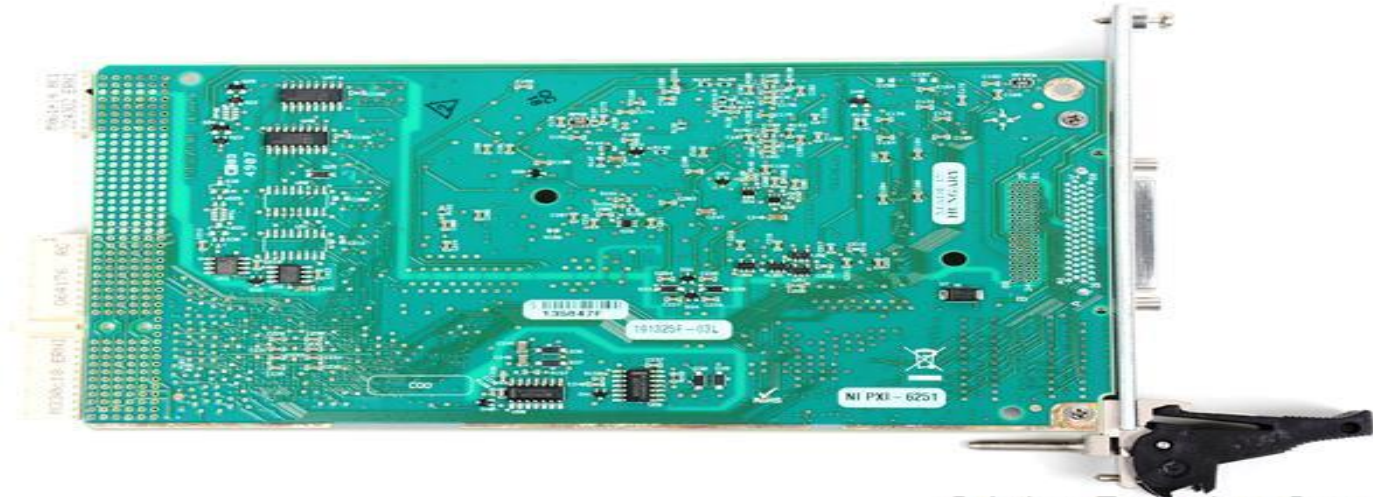
- NoSQL хорошо использовать в информационной системе (Web Cloud хранилище) совместно с традиционными реляционными СУБД, в этом случае каждое хранилище будет решать свою **ОТДЕЛЬНУЮ** задачу (принцип Разделяй и Управляй), эффективно используя преимущества каждого подхода.
- Один из недостатков NoSQL – тройной рост объёма дискового пространства по сравнению с Созданная автоматизированная система АСУТП успешно внедрена в процессе обучения на базе сетевых процессоров Arduino/Raspberry. Это позволило существенно ускорить процесс проведения испытаний и повысить качество результата для проектов Интернет Вещей (IoT).
- Внедрение и его перспективы
- Разработанная система мониторинга успешно внедрена в КИГМ №23 и в Сеченовском университете, на основе применения оборудования «**PXI**» фирмы **National Instruments** для реляционных SQL запросов в системе **1 C** и NoSQL запросов в СУБД **Cassandra**.
- Планируется создание серии аналогичных систем в качестве дипломных работ студентов для внедрения в хозяйственные организации города Москвы на базе IoT и NI Labview.





# NI PXI-8096

# NI PXI-6251 National Instruments



© Artisan Technology Group



© Artisan Technology Group



© Artisan Technology Group

# Cassandra Apache NoSQL Distributed Database

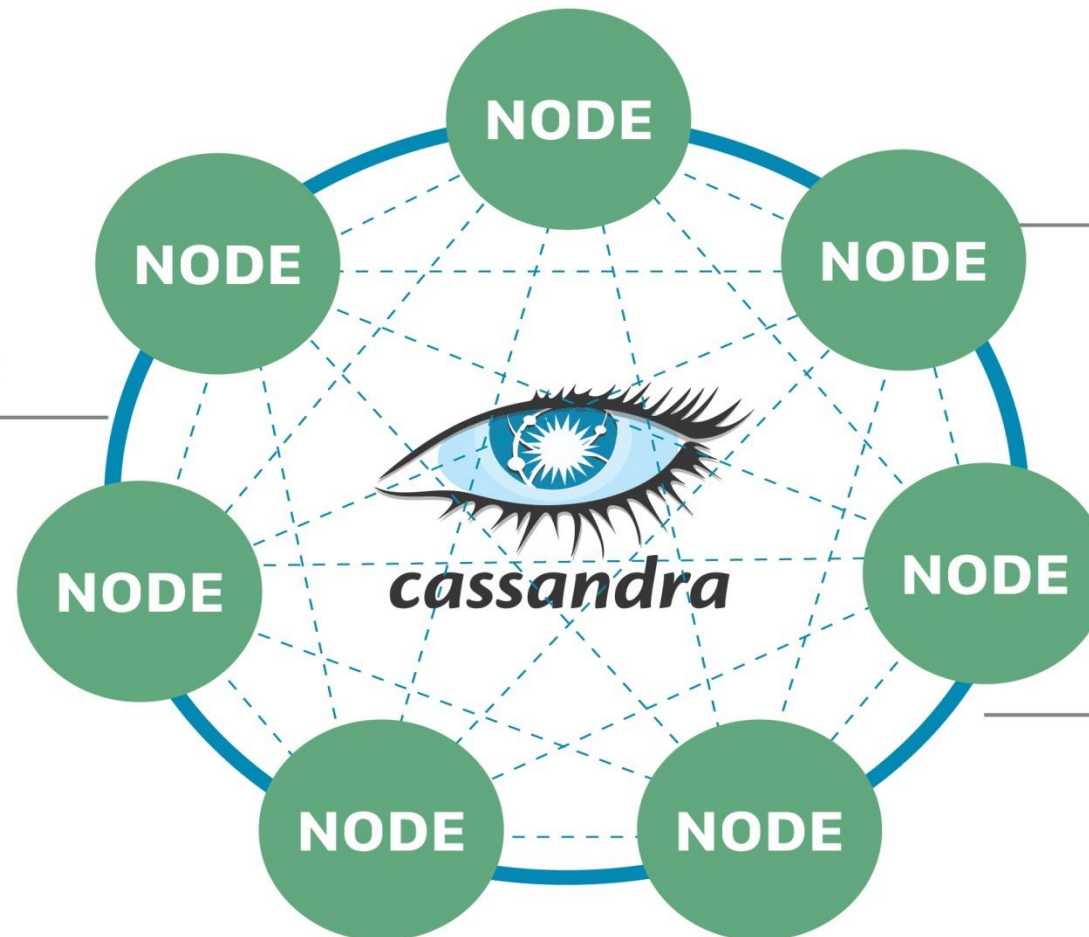
## ApacheCassandra™ = NoSQL Distributed Database

1 Installation = 1 NODE

✓ Capacity + 1TB

✓ Throughput: 3000 Tx/sec/core

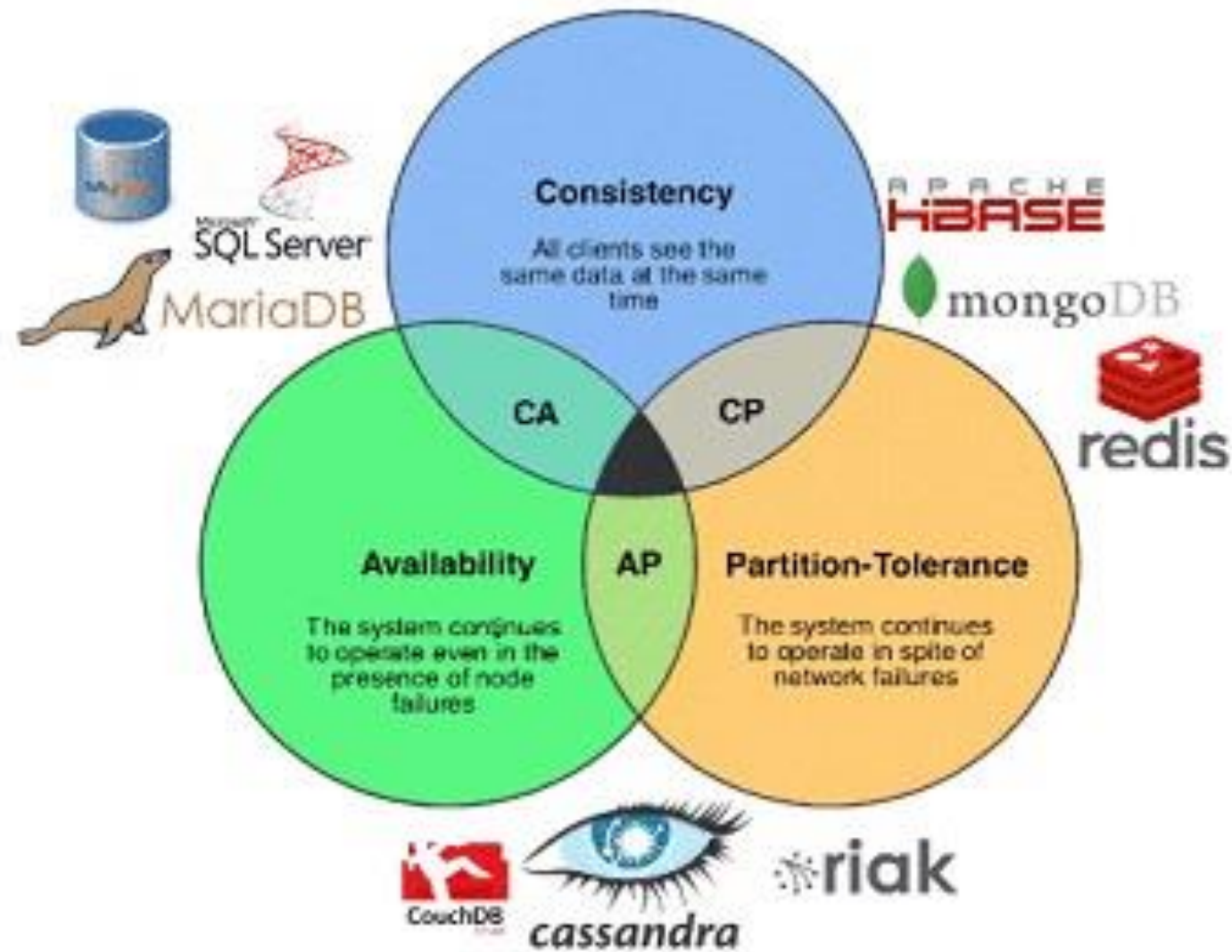
DataCenter | Ring



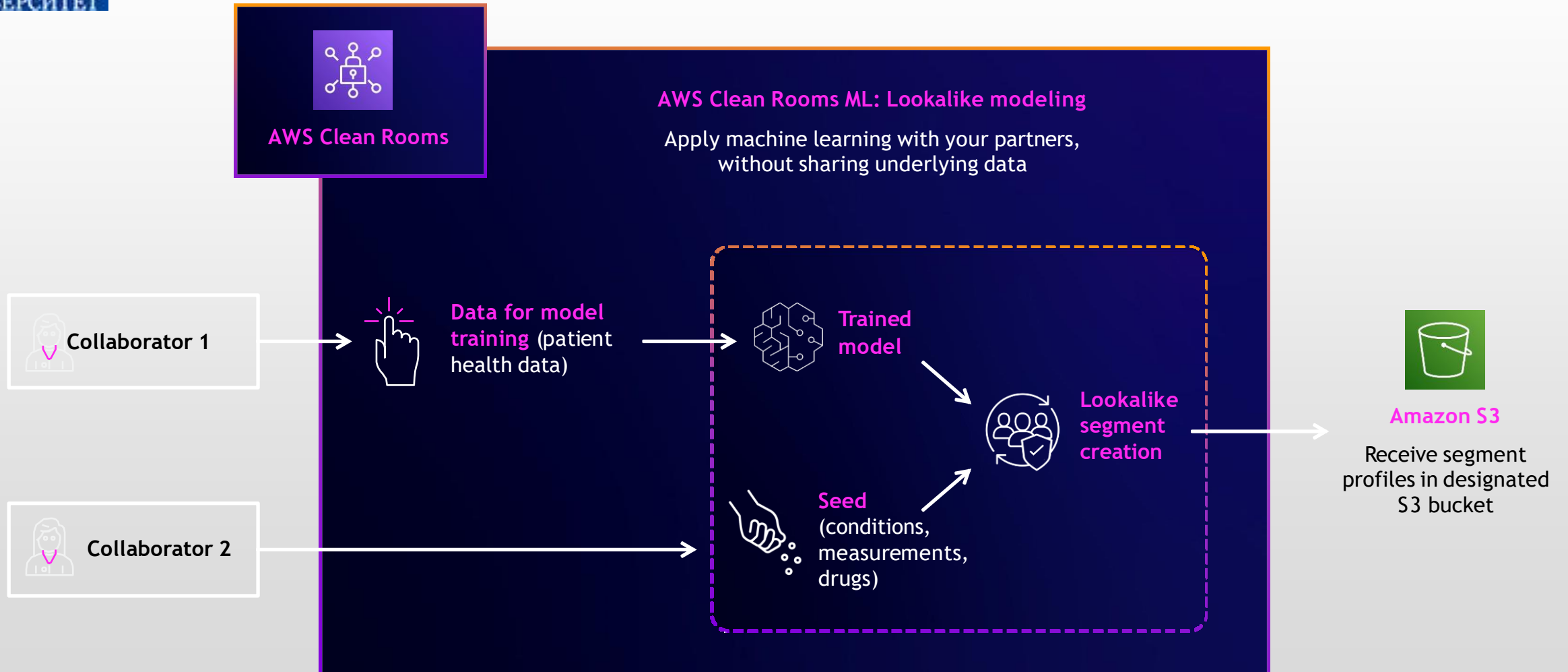
Communication:

✓ Gossiping

# CAP и Менеджмент



# AWS Clean Rooms ML



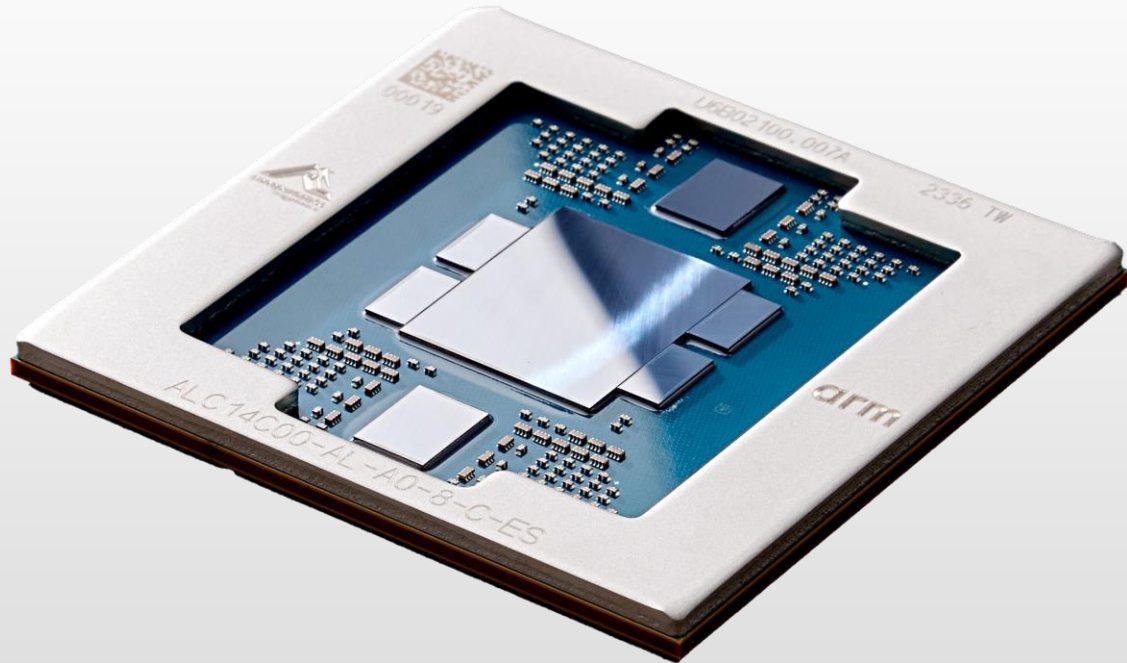
# AWS NVIDIA Partnership



# Amazon EC2 R8g Instances

Core Health IT

MEMORY - OPTIMIZED INSTANCES POWERED BY AWS GRAVITON4 PROCESSORS



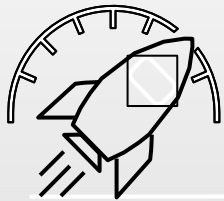
- Up to 30% better compute performance than Graviton3-based R7g instances
- Up to 3x more vCPUs and memory than Graviton3-based R7g instances
- Ideal for memory-intensive workloads such as databases and real time big data analytics

NOW IN PREVIEW

# Amazon FSx for NetApp ONTAP

## Amazon FSx for NetApp ONTAP

### 9X PERFORMANCE



#### INCREASED PERFORMANCE

- Get scale-out levels of performance in just a few clicks
- Up to 9x higher throughput performance
- Up to 7.5x more SSD IOPS for performance-intensive workloads

### INCREASE

#### INCREASED CAPACITY

- Up to 16x more SSD storage capacity to provide fast, scalable storage for a broad spectrum of workflows
- Customers can now store all their data in a single file system—making it even easier to migrate and extend their workloads to AWS



Patent, Member, and  
Clinician Experience

# Guardrails for Amazon Bedrock

Implement application specific safeguards based on your use cases and responsible AI policies

PREVIEW

Apply guardrails to any foundation model (incl. fine-tuned models) and Agents for Amazon Bedrock

Configure harmful content filtering based on your responsible AI policies

Define and disallow denied topics with short natural language descriptions

**COMING SOON**  
Redact sensitive PII information in FM responses





## Model evaluation on Amazon Bedrock

Оценивает модели фундамента, чтобы выбрать лучшую для вашего случая использования.

PREVIEW

Choice of programmatic and human evaluations

Use curated datasets or bring your own

Pre-defined and custom metrics

Leverage your workforce or AWS-managed reviewers