



УНИВЕРСИАДА ПО ИННОВАТИКЕ.
Отборочный этап (заочный)

КЕЙС

«ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ В ОБЛАСТИ ХРАНЕНИЯ ДАННЫХ»

Стремительное развитие информационных технологий приводит к возрастанию значения проблемы хранения информации. Объём используемой информации каждый год превосходит прогнозы предыдущих лет и темпы роста продолжают увеличиваться. Таким образом, необходимо не только совершенствовать имеющиеся технологии, но и разрабатывать качественно новые.

Сбор данных и его обработка началась еще 8 тыс. лет с зарождением культуры и письменности. Со временем расширялись задачи сбора информации и области применения, это перестало быть прерогативой отдельных структур. Но информации становилось слишком много как для хранения, так и для обработки. На обработку переписи населения США в 1880 году ушло 8 лет, что сподвигло Германа Холлерита создать первый статистический табулятор в 1888 году.

С тех пор скорость обработки данных постоянно увеличивалась, новые изобретения многократно превосходили по показателям предыдущие.

На сегодняшний день можно сказать, что мы находимся в информационной эпохе, в которой информационные коммуникативные технологии, базирующиеся на электронике и генной инженерии, являются главным инструментом достижения цели. В области обработки данных человечество добилось внушительных результатов – суперкомпьютеры, решающие сложнейшие задачи и даже квантовые вычисления уже не просто мечта футуристов, а современная действительность. К примеру, в конце 2021 года ученые из Российского квантового центра запатентовали новую архитектуру квантового процессора на основе кудитов, аналоги которого есть только у трех стран, и на его основе к концу 2024 г. В России собираются построить универсальный квантовый компьютер с облачным доступом. До 2024 г. на его создание «Росатом» планирует выделить более 23 млрд руб. [1]. Исследования в области искусственного интеллекта также продолжаются и требуют все больших скоростей обработки информации. Каждую секунду в Google выполняется около 99 000 поисковых запросов [2], а по скромным подсчетам, 148 ГБ — это объем данных, которые один интернет-пользователь в среднем генерирует за день [3]. По некоторым оценкам, 90% всех данных, накопленных в мире, были созданы за два года, и, согласно прогнозам, объем данных будет ежегодно увеличиваться примерно на 40%. В 2020 году было создано порядка 64,2 зеттабайта данных, что на 314 % больше, чем в 2015 году. Существенная доля этих данных — это пассивно собранные данные, полученные в результате повседневных цифровых операций [4]. К 2025 по разным оценкам в мире будет до 200 зеттабайт данных [5]. Следовательно, потребность в хранении и обработке информации будет ещё больше возрастать.

Закономерный вопрос, который возникает – где и как хранить такие объемы информации. Технологии центров обработки данных постоянно совершенствуются. Увеличивается емкость HDD накопителей, SSD накопители на протоколах NVMe становятся все быстрее. Большая часть усилий тратится на совершенствование твердотельных накопителей, однако и жесткие диски имеют потенциал для совершенствования.

В сфере современных решений систем хранения данных в 2019 году впервые в истории лидером по объему поставленной емкости стала российская компания – Yadro [6]. Компания занимается разработкой и производством систем обработки и хранения данных — высокопроизводительных серверов Vesnin и семейства систем хранения данных Tatlin. Таким образом, отечественные производители имеют достойное решение в области современных систем хранения данных, но, учитывая темпы роста объёма данных в мире, становится очевидно, что существующие технологии не смогут в долгосрочной перспективе в полной мере решить проблемы хранения и обработки информации. Нужен новый технологический прорыв.

Одной из наиболее невероятных технологий является хранение данных в ДНК. В отличие от используемых сегодня оптических и магнитных форм технологий хранения данных, в ДНК данные не представляются в виде бинарных единиц (1 и 0). Спираль ДНК состоит из множества чередующихся пар четырех уникальных оснований - аденин (А), цитозин (Ц), гуанин (Г) и тимин (Т). На основе этих оснований и кодируется и хранится информация. Эти молекулярные цепочки, будут

затем при необходимости преобразовываться в двоичные цифры. Эксперименты по практическому хранению данных ДНК проводятся уже несколько лет с предложениями по методам, охватывающим как живые организмы, такие как бактерии, так и использование синтезированной ДНК. В ноябре 2021 года группа исследователей добились первого синтеза произвольных последовательностей ДНК с наномасштабными размерами элементов, которые могут обеспечить достаточно высокую плотность, чтобы их можно было использовать как ДНК-хранилища [7].

Твердотельная перспективная технология в этой области - пятимерное (5D) оптическое хранилище, команда Университета Саутгемптона занимается ей уже более десяти лет. Данные записываются с помощью лазера, который излучает невероятно короткие, но мощные световые импульсы, создавая крошечные структуры в стекле, размеры которых измеряются в наномасштабе. Эти структуры содержат информацию об интенсивности и поляризации лазерного луча, в дополнение к своим трем пространственным измерениям, поэтому ученые называют их хранилищем данных 5D. Несмотря на отсутствие компактности, которую теоретически можно добиться с записью на ДНК, пятимерные оптические хранилища выдерживают до 1000 градусов по Цельсию и способны храниться 13,8 миллиардов лет при комнатной температуре без ухудшения качества [8].

Еще одна инновационная технология – метод хранения данных с использованием флуоресцентных красителей, который может быть намного более долговечным и экономичным, чем традиционные методы хранения. Химики из Гарварда показали, что цифровые данные могут храниться в смесях молекул флуоресцентных красителей, нанесенных струйным принтером. Распечатанные данные можно прочитать с помощью микроскопа, который определяет длину волны света, излучаемого каждым красителем, и позволяет различать красители в смеси. Наличие или отсутствие красителя может представлять значение в двоичном формате — основу современных вычислений. Исследователи сравнивают эту технологию концепцией ДНК-хранилищ и утверждают, что у нее есть преимущество в том, что молекулы просто смешиваются, а не синтезируются сложные структуры, поэтому скорость значительно выше [9].

Тем не менее, все эти технологии пока находятся в стадии разработки и еще длительное время не смогут использоваться повсеместно. Более того, еще неизвестно какая из этих технологий в итоге может стать решением для постоянно возрастающих потребностей в обработке и хранения информации.

Несмотря на привлекательность новых технологий необходимо помнить, что к новым технологиям хранения много требований в условиях информационной эпохи. Помимо возможности хранения огромного количества данных в единице объема, должна сохраняться возможность эти данные быстро считывать и обрабатывать, без потерь. Сохранность данных в различных условиях хранения и транспортировки также является важным условием выбора той или иной технологии. Система хранения должна быть отказоустойчивой, а сама технология должна быть доступной, потому что переход к новым технологиям случится только тогда, когда это будет экономически выгодно.

Источники:

1. Интернет-издание в сфере высоких технологий в России и странах СНГ «CNews», «Ученые из России запатентовали новую архитектуру квантового процессора», URL: https://www.cnews.ru/news/top/2022-05-26_uchenye_iz_rossii_zapatentovali
2. Открытый счетчик активности в Интернете «Internet Live Stats», Url: <https://www.internetlivestats.com/>
3. Интернет-платформа «FinancesOnline», «53 Important Statistics About How Much Data Is Created Every Day» Url: <https://financesonline.com/how-much-data-is-created-every-day/>
4. Организация объединенных наций, «Инновации в сфере данных для целей развития» Url: <https://www.un.org/ru/global-issues/big-data-for-sustainable-development>
5. Ведущий мировой исследователь мировой киберэкономики «Cybersecurity Ventures», «The World Will Store 200 Zettabytes Of Data By 2025», 08.06.2020 Url: <https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/>
6. Интернет-издание «ComNews», «Рынок СХД на подъеме», 09.01.20, URL: <https://www.comnews.ru/content/203813/2020-01-09/2020-w02/rynok-skhd-podeme>
7. Bichlien H. Nguyen et all. «Scaling DNA data storage with nanoscale electrode wells» SCIENCE ADVANCES, 24.11.21 Url: <https://www.science.org/doi/10.1126/sciadv.abi6714#tab-contributors>
8. Интернет-издание «Engadget», «'5D' storage could fit 500TB on a CD-sized glass disc», 01.11.2021, Url: <https://www.engadget.com/5-d-storage-could-fit-500-tb-on-cd-sized-disc-095039455.html>
9. Интернет издание «The Harvard Crimson», «Harvard Chemists Use Dyes as a New Data Storage Method», Url: <https://www.thecrimson.com/article/2021/10/21/dye-storage-method-explored/>

Задание:

1. Проведите сравнительный анализ представленных в задании технологий хранения данных. Какая из них на Ваш взгляд является наиболее перспективной? Обоснуйте ответ.
2. Разберите одну из трех предложенных технологий более подробно с точки зрения перспектив использования и имеющихся проблем на текущий момент.

Требования к решению кейса

Решение кейса должно быть представлено в виде двух файлов:

- 1) Презентация (формат .pdf или .pptx) с основными положениями решения и выводами (не более 15 слайдов);
- 2) Текстовый файл (формат .pdf или .docx) с дополнительной информацией (не более 1 страницы формата А4 12 шрифтом): расчеты, аналитические данные, ссылки на источники информации.

В презентации и текстовом файле должны содержаться разные материалы.

Файлы с решением кейса должны быть отправлены в срок до **20 марта 2022 года 23:59 мск.** по следующему электронному адресу: **olympicmsu@mail.ru**. Позже решения кейса приниматься не будут.

Основные критерии оценки

При выставлении оценок за решение кейса будут использоваться следующие критерии:

- Качество проведенного анализа и аргументированность сделанных выводов.
- Логика и структура изложения.
- Качество оформления презентации.
- Нестандартность мышления при выработке решения.
- Учет современных экономических особенностей и условий.