ТЕХНИЧЕСКОЕ ЗАДАНИЕ / ИКТ-РАЗРАБОТКА И ПИЛОТ

Разработка распределённого ядра медиаплатформы следующего поколения для обработки мультимодальных информационных потоков масштаба до 10 млрд публикаций в год

Заказчик	ТОО «IMAS GROUP»
Результат	Развёртываемый сквозной PoC ядра, доказательная нагрузочная модель, два независимых окружения и инженерная документация
Формат	ИКТ-разработка: дизайн двух альтернатив → рабочий PoC → нагрузочные/аварийные испытания → пилот
Предварительный срок	24 недели (предложение к согласованию, отсчёт после договора)
Обоснование	Перспективная инженерная задача предельного масштаба, не описание и не оценка текущей платформы iMAS

Предмет — создание самостоятельного масштабируемого технологического ядра для нового класса геораспределённых медиаплатформ. Задача размещается из-за сочетания экстремальных сценариев роста, мультимодальности и многопользовательской аналитики; не является аудитом или заявкой на замену действующей инфраструктуры iMAS.

1. Вводные разделы

№	Элемент шаблона	Содержание и требования
1.1.1	**Наименование**	Разработка распределённого ядра медиаплатформы следующего поколения для обработки мультимодальных информационных потоков масштаба до 10 млрд публикаций в год (далее — Ядро).
1.1.2	**Краткая характеристика области применения**	Медиамониторинг и аналитика в масштабах многоязычных СМИ, социальных сетей, комментариев, изображений, аудио и видео; многолетние архивы, сложный поиск, независимые клиентские проекты и мультимодальные AI-признаки.
1.2.1	**Документ-основание**	Открытая технологическая задача ТОО «IMAS GROUP» на Astana Hub. Независимая перспективная постановка; параметры и существующая производственная архитектура заказчика не раскрываются.
1.2.2	**Дата утверждения ТЗ**	Дата утверждения назначается после внутреннего согласования заказчиком и выбора исполнителя.
1.2.3	**Орган утверждения**	Уполномоченное лицо ТОО «IMAS GROUP»; решение закрепляется внутренним документом или договором.
1.3.1	**Наименование объекта**	Сквозной программный комплекс: входной событийный поток, управление хранилищами и индексами, жизненный цикл архивов, пользовательские проекты, аналитические представления, AI-расширения, API и средства автономного развёртывания.
1.3.2	**Заказчик / разработчик**	Заказчик — ТОО «IMAS GROUP»; разработчик — отобранный исполнитель. Контакты, реквизиты и распределение прав оформляются в договоре.
1.3.3	**Сроки**	Предварительно 24 недели: формализация 3, исследование архитектур 5, разработка PoC 10, проверки 4, передача/пилот 2. Даты и состав этапов согласуются при контрактовании.

№	Элемент шаблона	Содержание и требования
1.3.4	Финансирование	Коммерческая ИКТ-разработка с этапным финансированием. Бюджет и платежи утверждаются заказчиком отдельно; предварительная стоимость устанавливается в отдельном финансовом согласовании, без публичного обязательства заказчика.
1.4.1	Перечень терминов и сокращений	PoC — проверяемый программный прототип; RPO/RTO — допустимая потеря данных и время восстановления; TCO — полная стоимость владения; multi-tenant — изоляция независимых заказчиков; hot/warm/cold — классы доступа, не обязательно виды физических носителей; P95/P99 — перцентили времени ответа.
1.5.1	Объект	Поток до миллиардов публикаций в год, многолетние тексты и медиавложения, обновляемые AI-обогащения, большое число проектов и сложные пересечения между ними; распределённая инфраструктура с автономными региональными установками.
1.5.2	Цель	Спроектировать и реализовать независимое ядро следующего поколения с доказанной архитектурной масштабируемостью, быстрыми пользовательскими запросами, управляемой стоимостью архивов, восстановлением исторических данных и отказоустойчивостью.
1.5.3	Методы	Сопоставление не менее двух различающихся принципов хранения/индексации, моделирование профиля нагрузки, реализация сквозного PoC, пошаговое увеличение данных, fault injection, измерение P50/P95/P99, TCO и автоматическая установка в двух окружениях.
1.5.4	Новизна	Технологическая новизна проверяется на связке требований: ≥10 млрд материалов/год в перспективном сценарии, разные медиа и AI-данные, непрерывное связывание публикаций с множеством проектов, селективная архивация по фактическому использованию и переносимая автономная установка. Новизна не заявляется за отдельную известную СУБД.
1.5.5	Эффективность	Измеряемые результаты: throughput и lag потока, P95 поиска/фильтрации/аналитики, эффективность хранения и индексов, линейность/нелинейность роста затрат, RPO/RTO на испытательных отказах, время архивирования и восстановления.

2. Назначение, цели, объект

№	Элемент шаблона	Содержание и требования
2.1.1	Назначение системы	Обслуживать полный жизненный цикл материалов и их производных: приём, долговременное хранение, оперативный и архивный поиск, проектные связи, самостоятельную AI-обработку и мультимодальную аналитику с доступом через API.
2.1.2	Показатели достижения цели	Для PoC целевые ориентиры: 100 млн смешанных документов на репрезентативном тестовом стенде; P95 стандартной ленты/поиска по оперативному слою ≤2 с при согласованном профиле запросов; повторное развёртывание ≤1 рабочего дня; тестирование RPO ≤15 мин и RTO ≤60 мин на отказе узла. Пороги согласуются после фиксации конфигурации стенда.
2.2	Назначение разработки	Создать новый переиспользуемый программный фундамент для следующего поколения медиаплатформ, а не проверять квалификацию команды заказчика, не проводить аудит текущего ПО и не обязывать заказчика мигрировать.
2.3	Характеристика предмета	Реалистичные сценарии: тексты на языках Центральной Азии, короткие сообщения и крупные новости, аудио/видео, обновление метрик, частичная AI-разметка, исторические периоды, неоднородные частоты

№	Элемент шаблона	Содержание и требования
		доступа, массовый ретроспективный поиск, множество независимых проектов.
2.4	Существующее состояние	На рынке доступны отдельные классы решений — колоночные, поисковые, документные, объектные и lakehouse-системы. Трудность заключается в сочетании всех нагрузок и ограничений в одной воспроизводимой архитектуре. Текущее устройство iMAS не является предметом конкурсного обследования.
2.5	Научная цель / гипотеза	Техническая гипотеза: событийно-ориентированная гибридная архитектура с независимыми слоями хранения, индексирования, проектных представлений и AI-обогащения, управляемым кешированием и политиками жизненного цикла обеспечит лучшие показатели скорости/TCO при растущем числе проектов, чем универсальный монолитный контур; доказательство — сравнительный PoC.

3. Требования к результату / системе

№	Элемент шаблона	Содержание и требования
3.1.1	Структура и функционирование	Разделённые контуры ingest, хранения исходных объектов, полнотекстового и семантического поиска, project/read-model, мультимодального AI, аналитики, API и observability. Обязательны устойчивые ID, асинхронные обновления и согласованность представлений; выбор СУБД не предписан.
3.1.2	Показатели назначения	Годовые сценарии 1/5/10 млрд публикаций; 10 млрд ≈27,4 млн/сутки в среднем; пиковые окна — до 10× среднего темпа на согласованных выборках. Отдельные оси: 1 тыс./10 тыс./100 тыс. проектов (последнее — перспективная модель), доля перекрытия, срок архива 5/10/20 лет, количество AI-признаков и медиа.
3.1.3	Надёжность	Повторная доставка без неконтролируемых дублей, гарантия сохранности подтверждённо принятого материала, продолжение работы при временной недоступности AI-модуля, восстановление очередей, защита от каскадных отказов и наблюдаемость состояния каждого сервиса.
3.1.4	Безопасность	Проектная и клиентская изоляция, роли, аудит доступа, защищённые API, шифрование каналов, секреты, настраиваемая локализация данных и независимые политики хранения для региональной инсталляции.
3.1.5	Эргономика и техническая эстетика	Полный промышленный UI не требуется; тестовый клиент обязан выполнять ленту, сортировку, пагинацию, множественные фильтры, текстовый и исторический поиск, агрегации и доступ к медиа с замерами P95/P99 при параллельной загрузке.
3.1.6	Эксплуатация и обслуживание	Контейнерная или эквивалентная воспроизводимая установка, конфигурации для облака и on-premise, обновления без потери данных, метрики, логирование, алерты, резервирование, процедуры восстановления и прогнозирование ёмкости.
3.1.7	Защита от несанкционированного доступа	Исключить выдачу объектов другой организации даже при совпадении идентификаторов, поисковых запросов и кешей. Проверить авторизацию по API, сервисным аккаунтам и административным операциям.
3.1.8	Сохранность при авариях	Проверить независимое разрушение узла хранения, поискового узла и обработчика, сбой сети/объектного хранилища, восстановление данных и индексов; результаты должны быть сопоставлены с RPO/RTO и отражать режимы деградации.

№	Элемент шаблона	Содержание и требования
3.1.9	Защита от внешних воздействий	Испытать пики записи, исчерпание диска, неравномерные запросы, недоступность внешней модели, повреждённый ввод, задержанные события и частичную потерю связности между зонами; система должна явно сигнализировать об ограничениях.
3.1.10	Патентная чистота	Сформировать реестр лицензий СУБД, индексов, моделей и библиотек; обозначить ограничения коммерческого использования, поставки третьим лицам, автономного хостинга и трансфера в зарубежные юрисдикции.
3.1.11	Стандартизация / унификация	Открытые документированные API, версионирование схем и признаков AI, форматы экспорта/импорта, инфраструктурные манифесты и переносимость данных; устойчивые ID отделены от физического местонахождения объекта.
3.1.12	Дополнительные требования	Демонстрация клонирования решения в два независимых окружения, эксплуатационная карта инфраструктуры, архитектурные decision records, расчёт TCO и сопровождение согласованной технической передачи.
3.2.1	Функции и критерии отказов	Ввод: тексты/вложения/метаданные, статусы, идемпотентность. Хранение: единая логическая модель с разными физическими слоями. Архив: перенос не только по возрасту, но и по ссылкам из активных проектов и избранного; автоматический rehydrate по ID, дате, запросу, проекту. Проекты: активные/архивные, персональные оценки тональности и избранное, восстановление без потери пользовательских изменений. Поиск: batch retrospective + streaming matching, полнотекстовый, структурный и по желанию семантический. Чтение: быстрая лента, агрегации, фильтры, сортировка при раздельных AI-результатах. AI: OCR/ASR, объекты, тональность по сущностям, векторы, инфоповоды; версии и повторная обработка. Трансфер: независимая установка, export/import, автономная работа. Неприемлемые отказы: потеря подтверждённой записи, межклиентская утечка, невосстанавливаемые связи, отсутствие измерений.
3.3.1	Информационное обеспечение	Синтетический и лицензированный открытый корпус с проверяемыми идентификаторами, разными языками и форматами: минимальный функциональный набор, репрезентативный стенд до 100 млн документов, многоаспектный генератор связей публикация–проект; историческое распределение по 20 годам.
3.3.2	Лингвистическое обеспечение	Поддержка Unicode, кириллицы/латиницы и смешанных сообщений, полнотекстовых словоформ, правил, регулярных выражений по согласованию; языки Центральной Азии без требования повторно обучать существующую классификацию как часть этого PoC.
3.3.3	Программное обеспечение	Выбор технологий свободный: сравнить минимум два существенно разных сочетания SQL/колоночных, документных, поисковых, объектных, lakehouse или иных хранилищ. Оценить отдельные read models и кеши. Не навязывать ни одну конкретную СУБД или брокер.
3.3.4	Техническое обеспечение	Указать и измерить состав CPU/RAM/SSD/объектного хранилища, сеть, резервные копии, отказные домены, стоимость GPU/AI-модулей, стоимость хранения индексов и фактическую производительность на каждом стенде.
3.3.5	Организационное обеспечение	Фиксировать паспорт испытаний, состав выборок, версии и конфигурации, разграничить public/synthetic/обезличенные данные. Не нужен доступ к production или исходной инфраструктуре iMAS.
3.4	Требования к программному изделию	Представить исходный код или развёртываемые компоненты, end-to-end сценарий, раздельное администрирование клиентов, многослойное хранение, сквозные поисковые/аналитические операции, механизм расширения AI, API, средства наблюдаемости и инсталляцию в двух независимых окружениях.

№	Элемент шаблона	Содержание и требования
3.5	Информационная безопасность	Использовать тестовые данные без коммерческой тайны; секреты не в исходниках, минимум привилегий, ролевые права, журналирование, перечень сторонних компонентов, условия безопасной автономной эксплуатации.
3.6	Промышленная безопасность / ESG	Производственная безопасность в узком смысле не относится к предмету; отдельно оценить энергопотребление, избыточность реплик и эффективность CPU/GPU как факторы TCO.

4. Работы, сроки, приёмка

№	Элемент шаблона	Содержание и требования
4.1.1	Стадии, исполнители, документы	Этап 1 (недели 1–3): модель данных и воспроизводимый профиль нагрузки; этап 2 (4–8): два концептуально разных архитектурных варианта и предварительный бенчмарк; этап 3 (9–18): end-to-end PoC; этап 4 (19–22): нагрузки, отказные сценарии и TCO; этап 5 (23–24): пилот в изолированном окружении, передача, документация.
4.2.1	НИР → ОКР → ОПЭ / внедрение	НИР-элемент: инженерное исследование гипотез и альтернатив. ОКР: создание работающего ядра, автотестов и инсталляционных пакетов. Опытная эксплуатация: стенд без доступа к закрытым системам заказчика, воспроизведение KPI и демонстрация трансфера.
4.3.1	Контроль и приёмка	Комиссия заказчика проводит воспроизводимый запуск по утверждённому протоколу, повторяет ключевые поисковые/аналитические запросы, восстановление проектов/избранного, выборочное и пакетное rehydrate, отказ узла и установку во втором окружении. Результат принимается по фактическим метрикам, не презентации.
4.4.1	Технико-экономические показатели	KPI: скорость ingest/indexing, lag доступности, P95/P99 ленты/поиска/аналитики, перекрёстные зависимости от числа проектов, GB/млн документов включая индексы/реплики, стоимость обслуживания 1 млрд материалов/год и 5/10-летнего архива, RPO/RTO, повторная установка. Крупные сценарии 5/10 млрд/год оцениваются с калиброванной и проверяемой моделью масштабирования, а не фиктивным полномасштабным запуском.
4.5.1	Порядок оплаты	Форма вознаграждения — по договорённости: этапный договор на PoC и пилот. Сумма утверждается уполномоченным представителем до заключения договора. Срок 24 недели — проектная оценка.

5. Подготовка объекта и документирование

№	Элемент шаблона	Содержание и требования
5.1	Подготовка к вводу в действие	Подготовить обезличенный стенд, генераторы разнородных публикаций, многоязычные запросы, связи с независимыми клиентами и проектами, архивные политики и профили отказов. Production-данные и схема инфраструктуры iMAS не предоставляются.
5.2.1	Комплект документации	Схемы компонентов и потоков; ADR с анализом двух альтернатив; модель публикаций/проектов/AI-результатов; форматы поиска и API; спецификация hot/warm/cold и реанимации архивов; результаты нагрузок и отказов; воспроизводимый запуск в двух окружениях; TCO, реестр лицензий, руководство эксплуатации и перечень известных ограничений.
5.3.1	Источники разработки	Настоящее ТЗ; публичные спецификации технологий и лицензий; открытые/синтетические тестовые корпуса; предложения разработчика.

№	Элемент шаблона	Содержание и требования
		Использовать внутренние базы и код заказчика на конкурсном этапе не требуется.

6. Квалификационные требования

№	Элемент шаблона	Содержание и требования
6.1.1	Материально-техническая база	Опыт инженерного проектирования распределённых высоконагруженных систем, создание реально работающих прототипов, автоматизированные нагрузки и fault injection, миграция/восстановление данных, воспроизводимая поставка. Допустимы консорциумы исполнителей.
6.2	Разрешительные документы	Специальные разрешения для разработки ПО не предписываются; требуется правомерное использование данных, лицензий и сторонних компонентов, особенно при потенциальном трансграничном трансфере.
6.3	Гарантийные обязательства	Исполнитель устраняет воспроизводимые ошибки PoC, выявленные на согласованных испытаниях, и передаёт инструкции восстановления и контрольный набор тестов. Срок поддержки и гарантий фиксируется договором.

Приложение. Обязательные демонстрационные сценарии

Сценарий	Что предъявить	Критерий проверки
Стресс архива	Разные объёмы, доли текста и медиа, 5/10/20 лет истории	Измерить индекс/архив/резервные копии и обоснованность модели роста
Клиентские проекты	Массовое пересечение одного материала с несколькими независимыми проектами	P95 ленты/фильтров при возрастающем числе проектов
Управляемый архив	Перенос старого материала, который не связан с активным проектом; запрет переноса закреплённого	Целостность связей и воспроизводимое возвратное извлечение
Динамический AI	Добавление новой категории результата и повторная обработка части архива	Доступность в фильтрах, срок обновления read-model, сохранность истории
Аварии	Потеря узла, недоступность AI/поиска, сетевое разделение	Измеренные RPO/RTO, явно документированная деградация
Трансфер	Установка на двух раздельных инфраструктурах с воспроизведением тестов	Независимость от центральных систем и автоматизация развёртывания