

ОПИСАНИЕ И УСЛОВИЯ ПРЕДОСТАВЛЕНИЯ УСЛУГИ «ВИРТУАЛЬНЫЙ ЦОД С GPU»

1. ОБЩАЯ ИНФОРМАЦИЯ И ОПИСАНИЕ УСЛУГИ

- 1.1. Настоящий документ содержит описание состава Услуги, ее базовой функциональности, возможных сопутствующих и дополнительных услуг, общего порядка подключения, изменения и отключения Услуги, условий предоставления и ограничений.
- 1.2. Виртуальный ЦОД с GPU является услугой по предоставлению базовых информационно-технологических ресурсов на основе совокупности функционирующего под управлением Исполнителя серверного и сетевого оборудования, систем хранения данных и специализированного программного обеспечения. Услуга построена на основе модели обслуживания IaaS и реализована средствами КУ ENT. Услуга представляет собой совокупность ресурсов внутри тенанта, привязанных к типу процессора и предоставленных в пользование Заказчику для создания и бесперебойного функционирования одного или группы виртуальных серверов. В рамках Услуги Исполнитель предоставляет Заказчику Виртуальный ЦОД, имеющий в распоряжении согласованный между Исполнителем и Заказчиком набор виртуализированных вычислительных мощностей процессора (vCPU), виртуальной памяти (vRAM) и дискового пространства (vHDD), а также аппаратные мощности GPU и средства управления Виртуальным ЦОД, достаточные для создания и управления виртуальными серверами в требуемой Заказчику конфигурации в пределах выделенных виртуализированных мощностей.
- 1.3. Управление Виртуальным ЦОД осуществляется Заказчиком при помощи КУ ENT.
- 1.4. Руководство Пользователя по Услуге доступно по электронному адресу: <https://docs.sbercloud.ru/gpu/ug/index.html>.
- 1.5. **Состав Услуги и основные компоненты:**

Табл.1. Состав и основные компоненты

Ресурсы	
Наименование группы	Содержание
Вычислительные ресурсы	<ul style="list-style-type: none"> – виртуальные процессорные ядра (vCPU); – виртуальная оперативная память (vRAM); – графические ускорители (GPU); – виртуальное дисковое пространство (vHDD); – Edge Gateway.
Сетевые сервисы и компоненты	<ul style="list-style-type: none"> – подключение к сети интернет (в общем канале); – один публичный IP-адрес.
Основные компоненты ENT ¹	<ul style="list-style-type: none"> – платформа виртуализации; – КУ ENT; – платформа виртуализации сети (SDN).

- 1.6. В целях обеспечения защиты Инфраструктуры ENT реализовываются следующие меры и механизмы защиты:

Табл.2. Обеспечение защиты Инфраструктуры ENT

Уровни защиты	Мероприятия
Защита Облака Cloud и средств его управления	
Физический	<p>Обеспечивается:</p> <ul style="list-style-type: none"> – размещение всего оборудования инфраструктуры в ЦОД, соответствующих требованиям надежности по категории Tier 3; – контроль и управление доступом к оборудованию; – наличие системы видеонаблюдения на объектах информатизации ЦОД.
Сетевой	<p>обеспечивается защита периметров ЦОД и их сегментирование с использованием межсетевых экранов нового поколения (NGFW), осуществляющих в том числе выявление и предотвращение компьютерных атак.</p>

¹ Описание составных компонентов ENT представлено в Приложении № 1.ENT.1.1. к Договору

Инфраструктурный	Обеспечивается: – антивирусная защита инфраструктуры с использованием антивирусных средств для облачных сред; – управление доступом к инфраструктуре с использованием средств двухфакторной аутентификации подключающихся к ней администраторов; – контроль действий привилегированных пользователей с использованием специализированных средств; – регулярный контроль и анализ защищенности инфраструктуры с использованием специализированных средств по выявлению уязвимостей в используемом ПО и его некорректной конфигурации, влияющей на уровень защищенности ПО, с устранением выявленных уязвимостей и/или недостатков; – сбор и анализ событий информационной безопасности.
Дополнительный	Осуществляются периодические тестирования на проникновение и аудит информационной безопасности Инфраструктуры ENT с привлечением сторонних организаций. Выявленные в ходе соответствующего тестирования и/или аудита недостатки устраняются по факту их выявления.
Защита КУ ENT	
Приложения	Защита с использованием специализированного межсетевого экрана уровня приложений (Web Application Firewall)
Дополнительный	Осуществляются регулярные сканирования консоли на наличие актуальных уязвимостей и его периодические тестирования на проникновение с привлечением сторонних организаций. Выявленные уязвимости и/или недостатки устраняются по факту их выявления.
Изоляция «Организаций» Заказчика	
ENT	Осуществляется встроенными средствами КУ ENT и на сетевом уровне средствами SDN.
Дополнительный	В рамках периодических тестирований на проникновение всей Инфраструктуры проводятся тестирования на возможность проникновения потенциального нарушителя из одной «Организации» в другую с преодолением используемых механизмов защиты.

1.7. Распределение ролей, обязанностей и ответственности Исполнителя и Заказчика в области ИБ в отношении Услуги приведено в Таблице 3 Приложения № 1.ENT.1.1. к Договору.

1.8. **Типы ресурсов, рекомендации и ограничения:**

Табл.3. Типы ресурсов, рекомендации и ограничения

Тип ресурса: виртуальные процессорные ядра (vCPU)	
Требования	Рекомендации и ограничения
Ядра обслуживаются физическими процессорами Intel	В рамках Услуги Заказчик может использовать только vCPU с одинаковой частотой (обслуживаемые процессорами одного типа). Ограничения на количество vCPU указаны в разделе 3, таблице «Параметры предоставляемых услуг».
Тип ресурса: виртуальная оперативная память (vRAM)	
Требования	Рекомендации и ограничения
При формировании Заказа Заказчик указывает требуемый объем vRAM. При заказе Услуги требуемый объем vRAM должен быть дополнительно учтен в рамках заказываемого объема Виртуального дискового пространства (vHDD) выбранного профиля для размещения swar-файлов виртуальных серверов	Ограничения на количество vRAM указаны в разделе 3, таблице «Параметры предоставляемых услуг».
Тип ресурса: графические ускорители (GPU)	
Требования	Рекомендации и ограничения
Используются NVIDIA Tesla V100 32ГБ, NVIDIA Tesla A100 40 Гб, NVIDIA Tesla A100 80 Гб, NVIDIA Tesla A40 48 Гб и NVIDIA Tesla T4 16Гб	-
Тип ресурса: виртуальное дисковое пространство (vHDD)	
Требования	Рекомендации и ограничения

<p>В рамках услуги предоставляется два дисковых профиля, отличающихся по скорости обмена данными (количеству операций ввода-вывода (IOPS)) и времени отклика: SATA/NLSAS и SSD. Каждый дисковый профиль соответствует своему типу дисков на системе хранения данных.</p> <p>В рамках одного Виртуального ЦОД можно использовать дисковые профили различного типа.</p> <p>При заказе Виртуального дискового пространства отдельно должен быть учтен объем vRAM для размещения swap-файлов виртуальных серверов. Выделенная емкость дискового пространства занимает все виртуальными машинами Заказчика (включенными и выключенными), созданными Заказчиком Snapshot, а также swap-файлами vRAM всех виртуальных машин Заказчика.</p>	<p>Минимальное значение для экземпляра Виртуального ЦОД – 100 Гб. Суммарный объем виртуального дискового пространства каждого дискового профиля должен быть кратен 100 Гб.</p>
---	--

1.9. Для подключения к Услуге Заказчик может выбрать один или несколько типов подключения:

Табл.4. Типы подключения к сети и сетевые сервисы

Тип подключения	Описание
Подключение через выделенный ¹ канал Интернет	Заказчику предоставляется отдельная полоса для доступа к Услуге, которая не разделяется с другими заказчиками) или подключение через общий канал Интернет (shared), который предполагает логическое подключение к общему для всех заказчиков Услуги каналу передачи данных (скорость сетевого соединения для каждого заказчика не является гарантированной и зависит от загрузки общего канала передачи данных).
Подключение через выделенный канал связи	Позволяет обеспечить взаимодействие сетей Заказчика с сетью в облаке с помощью выделенных каналов связи стороннего провайдера. Для данного подключения могут быть задействованы выделенные каналы Заказчика, организованные с использованием «темной оптики», либо каналы, предоставляемые Исполнителем.

Подробная информация по доступным подключениям приведена в Приложении № 1.ENT.6. к Договору.

1.10. Сетевые сервисы:

1.10.1. Edge Gateway предоставляется по умолчанию в конфигурации Compact, достаточной для создания небольших вычислительных сред с малым количеством сервисов, и может быть изменен в случае необходимости обработки большого количества сетевых функций. Ознакомиться с возможными конфигурациями и предельными значениями по эксплуатации виртуального шлюза Edge Gateway можно в документации по ссылке <https://docs.sbercloud.ru/vdc/ug/topics/networks.html>.

1.10.2. Edge Gateway предоставляет следующие сетевые сервисы:

- межсетевое экранирование (Firewall);
- маршрутизация (Routing);
- преобразование адреса (NAT);
- виртуальные частные сети (VPN) IPSec;
- динамическое распределение адресов DHCP.

Развертывание виртуального шлюза возможно в отказоустойчивой конфигурации (HA), при этом создается дополнительный шлюз в режиме ожидания, который перенимает на себя нагрузку в случае выхода из строя основного шлюза Edge Gateway.

С целью организации сложных сетевых взаимодействий, например, прямого сетевого взаимодействия между Edge Gateway, не строя IPSec туннели между ними необходимо использовать Виртуальный Edge Gateway T0. При необходимости увеличения пропускной способности доступа в интернет и из интернета предоставляется Выделенный сервер Edge Gateway T0.

1.11. Описание Шаблонов VM и образов ОС представлено в пункте 1.10. Приложения № 1.ENT.1.1. к Договору.

1.12. Описание Программной платформы представлено в пункте 1.11. Приложения № 1.ENT.1.1. к Договору.

1.13. Аппаратная платформа:

Табл. 5. Компоненты и характеристики аппаратной платформы

Компоненты	Характеристики
Серверная платформа	В качестве вычислительной платформы используются серверные решения корпоративного уровня, базирующиеся на процессорах архитектуры x86.
Графический ускоритель	В качестве GPU предоставляются графические ускорители от производителя NVIDIA.
СХД	Для организации сервиса предоставления виртуальных дисков применяются системы хранения данных уровня middle-range с резервированием основных компонент, таких как блоки питания, контроллерные модули.

¹ Заданная скорость гарантируется внутри сети Исполнителя начиная от порта пограничного маршрутизатора узла связи Исполнителя.

Сеть	<p>Базируется на оборудовании ведущих мировых производителей, которое обеспечивает:</p> <ul style="list-style-type: none"> - высокий уровень контроля и безопасности благодаря потоковой телеметрии и упреждающему анализу на линейной скорости передачи; - высокую производительность приложений благодаря интеллектуальным буферам и отсутствию потери пакетов; - высокую производительность и масштабируемость благодаря мультискоростным портам 1/10/25/50/100G. <p>Сетевая подсистема реализована с применением топологии Leaf - Spine, которая обеспечивает следующие преимущества:</p> <ul style="list-style-type: none"> - предсказуемость задержек; - высокий уровень масштабируемости без прерывания работы сети; - защиту от появления петель; - высокий уровень автоматизации управления и поддержки.
------	--

- 1.14. Предоставление доступа к программному обеспечению осуществляется в соответствии с условиями и порядком, предусмотренными в пункте 1.13. Приложения № 1.ENT.1.1. к Договору.

2. БАЗОВАЯ ФУНКЦИОНАЛЬНОСТЬ И МЕТРИКИ УСЛУГИ

- 2.1. Услуга Виртуальный ЦОД с GPU доступна только в конфигурациях, описанных в Таблице 6.
- 2.2. Вычислительные ресурсы vCPU, предоставляемые в рамках Услуги, резервируются на 100% и не разделяются с другими заказчиками.

Табл.6. Параметры предоставляемых Услуг

Сервис	Тарифицируемые единицы	Характеристики и метрики	Допустимые значения
Вычисления	Набор вычислительных ресурсов (шт.)	Базовая частота процессора vCPU	Не ниже 2,4 ГГц
		Host CPU Ready time	Менее 5%
		Кол-во vCPU	8/12/16/24/32 шт.
		Объем vRAM	96/192/256/384/512 Гб
		RAM Swapped	0%
		Графический ускоритель (GPU) Nvidia Tesla V100 32 Гб	1/2 шт.
	Набор вычислительных ресурсов (шт.)	Базовая частота процессора vCPU	Не ниже 2,4 ГГц
		Host CPU Ready time	Менее 5%
		Кол-во vCPU	4/8/16/32 шт.
		Объем vRAM	16/32/64/128 Гб
		RAM Swapped	0%
		Графический ускоритель (GPU) Nvidia Tesla T4 16 Гб	1/2/4 шт.
	Набор вычислительных ресурсов (шт.)	Базовая частота процессора vCPU	Не ниже 2,4 ГГц
		Host CPU Ready time	Менее 5%
		Кол-во vCPU	8/16/24/32 шт.
		Объем vRAM	96/192/256/384/512 Гб
		RAM Swapped	0%
		Графический ускоритель (GPU) Nvidia Tesla A100 40 Гб	1/2 шт.
	Набор вычислительных ресурсов (шт.)	Базовая частота процессора vCPU	Не ниже 2,4 ГГц
		Host CPU Ready time	Менее 5%
		Кол-во vCPU	16/24/32 шт.
		Объем vRAM	192/256/384/512 Гб
		RAM Swapped	0%
		Графический ускоритель (GPU) Nvidia Tesla A100 80 Гб	1/2/4 шт.
Хранилище данных	Виртуальный жесткий диск SSD (Гб)	HDD IOPS. Эталонные значения	2000 IOPS/1 000GB
		Среднее время доступа к SSD Storage на виртуальной машине	0 мс - 5 мс
		Допустимый объем одного виртуального жесткого диска SSD на виртуальный сервер	1 – 4096 Гб
		Шаг увеличения размера виртуального диска в допустимом диапазоне	1 Гб
	Виртуальный жесткий диск SATA (Гб)	HDD IOPS. Эталонные значения	100 IOPS/1 000GB
		Среднее время доступа к SATA Storage на виртуальной машине	0 мс - 30 мс
		Допустимый объем одного виртуального жесткого диска SATA на виртуальный сервер	1 – 4096 Гб
		Шаг увеличения размера виртуального диска в допустимом диапазоне	1 Гб
Сетевые сервисы	Доступ в Интернет в общем канале	Полоса пропускания	Не тарифицируется: не более 100 Мб/с на Виртуальный ЦОД

	Пропускная способность на виртуальный сервер	Средняя сетевая задержка в пределах сети передачи данных Исполнителя	0 мс - 5 мс
		Процент потерянных пакетов в пределах сети передачи данных Исполнителя	0% - 0,2 %
	Виртуальный шлюз (шт.)	Средняя сетевая задержка в пределах сети передачи данных Исполнителя	0 мс - 5 мс
		Пропускная способность	Не более 10 Гб/с
Гостевая ОС	Доступ к шаблону Серверная операционная система: <ul style="list-style-type: none"> • VM размером 4 и менее vCPU: VM (шт.)/ч.; VM (шт.)/мес.² • VM размером более 4 vCPU: vCPU (шт.)/ч.; vCPU (шт.)/мес.² 	Шаблоны Серверной операционной системы	Серверная операционная система 2016 Серверная операционная система 2019 Серверная операционная система 2022

² Минимальный период тарификации – календарный месяц. Начало использования, начиная с первой минуты, или продолжение использования Услуги в Отчетном периоде предполагает списание стоимости за полный календарный месяц. Неполный календарный месяц использования Услуги, начиная с первой минуты, округляется до полного календарного месяца пользования Услугой.

3. ТАРИФИКАЦИЯ УСЛУГИ

- 3.1. Тарификация Услуги статическая (Allocation).
- 3.2. Величина ежемесячного платежа за пользование услугой определяется в соответствии с заказанным объемом перечисленных ниже ресурсов и опций:
- Набор вычислительных ресурсов;
 - Виртуальная память³;
 - Виртуальный жесткий диск SATA;
 - Виртуальный жесткий диск SSD;
 - Предоставление публичного IP адреса;
 - Виртуальный EDGE Gateway T0;
 - Выделенный сервер EDGE Gateway T0;
 - Доступ к шаблону Серверная операционная система.
- 3.3. Методика расчетов потребляемых ресурсов и опций предполагает тарификацию суммы значений выбранных ресурсов и опций за Отчетный период в соответствии с тарифами. На основе суммы значений выставляется счёт. Методика расчёта потребляемого дискового пространства предполагает оплату за выбранный Заказчиком объём ресурсов дискового пространства.

4. ИНЫЕ УСЛОВИЯ, ПРИМЕНИМЫЕ К УСЛУГЕ

- 4.1. **Возможные виды подключения / изменения / отключения Услуг:**
- 4.1.1. Посредством подписания Заказа (с учётом п. 4.5. Приложения).
- 4.2. **Возможный порядок расчётов по Услуге:**
- 4.2.1. Постоплата.
- 4.3. **Возможные способы оплаты / порядок пополнения Баланса:**
- 4.3.1. Оплата в безналичном порядке на основании выставленного Исполнителем счёта.
- 4.4. Заказчик самостоятельно несет ответственность за работоспособность программного обеспечения, устанавливаемого на ВМ.
- 4.5. Стороны установили следующий порядок заказа Услуги по настоящему Приложению:
- 4.5.1. Заказ на подключение Услуги по настоящему Приложению должен быть направлен Исполнителю не позднее, чем за 6 (шесть) рабочих дней до даты начала оказания Услуги;
- 4.5.2. В течение 3 (трех) рабочих дней Исполнитель или его уполномоченный представитель обязуется рассмотреть Заказ на Услугу и направить лицу, направившему Заказ, ответ (подписанный со своей стороны Заказ или отказ в предоставлении Услуги с обоснованием причины);
- 4.5.3. В случае согласования Сторонами Заказа Услуга по такому Заказу предоставляется в дату начала оказания Услуги, зафиксированную в Заказе, с 10:00 по московскому времени.
- 4.6. Заказчик самостоятельно несет ответственность за сохранность данных и принимает самостоятельно меры по их сохранению при отказе от Услуги. При отказе от Услуги Исполнитель вправе удалить данные Заказчика по истечении 5 (пяти) рабочих дней после отказа от Услуги.

³ Объем Виртуальной памяти (vRAM) должен быть дополнительно учтен при заказе Виртуального дискового пространства для хранения swap-файлов виртуальных серверов в соответствии с пунктом 2.4.2 настоящего Приложения.