

4 сентября

Время в программе указано для часового пояса UTC+3 (Europe/Moscow).

11:00-11:15	Открытие конференции SmartData 2024		
	Михаил Лукин Андрей Дмитриев Судо JUG Ru Group		
11:15-12:00	Использование вероятностных структур данных для оптимизации ETL-процессов	Шардированный не значит распределенный: что важно знать, когда PostgreSQL мало	Как жить с data swamp
	Дмитрий Вертлиб Честный знак	Евгений Иванов Олег Бондарь Яндекс Яндекс	Александр Боргардт bigdataduck
12:00-12:30	Интервью с Дмитрием Устюговым		
	Дмитрий Устюгов Михаил Лукин NTU Singapore Судо		
12:30-13:15	Эффективные сервисы ML Inference нейросетей в Яндекс рекламе	Airbyte. 2 года в продакшене	Assessing Data Pipeline Quality & Sanity with Data Angiograms
	Дмитрий Ульянов Яндекс	Александра Попова СберЗдоровье	Sri Vishnu Chanderraju CueZen
13:15-13:45	Airbyte или NiFi: два подхода, два инструмента, одна цель?		
	Антон Букреев Василий Сафронов Николай Акимов Бронислав Житников Cryptology Независимый эксперт TaskData Positive Technologies		
13:45-14:30	Как построить пайплайн RAG с использованием LlamaIndex	Проблемы обработки Excel-файлов в Apache NiFi и как их решать	От хайпа до продакшена: data mesh на Airflow + dbt
	Алсу Нурутдинова Алина Кочева Positive Technologies Positive Technologies	Игорь Шемаров Axenix	Никита Юрасов Леонид Кожин Toloka Toloka
14:30-15:00	Альтернативные варианты восприятия информации в будущем программирования		
	Павел Филонов Никита Соболев Николай Марков Независимый консультант wemake.services Altener		
15:00-15:20	Оптимизация распределения партиций в последовательности задач распределенной обработки данных	Debezium Engine: практическое руководство по использованию	
	Милена Булкина Т-Банк	Анастасия Сашина Т-Банк	
15:20-15:35	Подведение итогов онлайн-части конференции SmartData 2024		
	Михаил Лукин Андрей Дмитриев Судо JUG Ru Group		

	Зал 1	Зал 2	Зал 3
10:15-10:30	Открытие офлайн-части конференции SmartData 2024 Владимир Озеров CedrusData Олег Кочергин Positive Technologies		
10:45-11:30	Пишем свой cluster manager для Apache Spark Александр Токарев Яндекс	Как мы сократили TTM создания дашбордов Анар Багиров Авито	Как кролик съел зеленую сливу и не умер: сказ о миграции на Iceberg Валентин Пановский BestDoctor
12:00-12:45	Как работает Apache Iceberg на примере Trino Владимир Озеров CedrusData	Как мы Apache Kafka на Redpanda меняли Роман Ананьев Авито	Data Vault 2.0. В каких случаях внедрять, разбор основных проблем применения методологии при построении DWH на Greenplum Денис Лукьянов Ecom.tech
12:45-13:30	Обед		
13:30-14:15	Инструменты Data Quality: как, зачем, почему. Опыт Т-Банка Дмитрий Руднев Т-Банк	dbt в деле: реальные кейсы и лайфхаки Антон Александров Детский мир	Data sketches — быстро, дешево и (почти) точно! Сергей Жемжицкий Arenadata
14:15-15:00	Обед		
15:00-15:45	Data mesh: ожидания vs реальность Петр Гуринов Лемана ПРО (Леруа Мерлен)	CI/CD в большом on-premise Datalake-- проекте Александра Чекмарева (Китченко) Честный знак Никита Благодарный Честный знак	Как мы делаем облачный Greenplum Леонид Борчук Yandex Cloud
16:15-17:00	Каждый байт на вес золота. Опыт построения DMP в рекламе Яндекса Алексей Стыценко Яндекс	Data Lineage: как настроить в зоопарке технологий и зачем это нужно Булат Усманов Кулер (ex-СберМаркет)	Выбрасываем Java икратно ускоряем Spark/Presto... Или пока нет? Павел Солодовников CedrusData
17:30-18:15	Как быстро запустить процесс ведения каталога данных в компании на примере DataHub Наталья Журавлева Ozon	Schema Registry: Ultimate Guide Тимофей Брунько Yandex Cloud	Snowplow: утерянное руководство Олег Кочергин Positive Technologies

18:15-22:00

Нетворкинг

18:30-19:30 Зал 2

[Круглый стол: Data Mesh](#)

18:45-20:45 Зал 1

[Lightning Talks](#)

	Зал 1	Зал 2	Зал 3
11:00-11:45	<p>Apache Arrow: быстрее, ниже, сложнее</p> <p>Евгений Готов Autotech</p>	<p>Как мы тестировали 5 способов загрузки данных в Greenplum и что из этого вышло</p> <p>Татьяна Дидова АЭРО</p>	<p>NiFi. Пишем код для codeless-системы</p> <p>Бронислав Житников Positive Technologies</p>
12:15-13:00	<p>Как мы строим систему распределенного трейсинга, в которой можно терять данные</p> <p>Игорь Балюк Авито</p>	<p>One More Way to Make Backup in Ignite</p> <p>Николай Ижиков Apache Software Foundation</p>	<p>От ручного труда к автоматической генерации проверок качества данных</p> <p>Александр Мадумаров Инновационный центр «Безопасный транспорт» ГКУ ЦОДД</p>
13:00-13:45	Обед		
13:45-14:30	<p>CI/CD для большого хранилища данных</p> <p>Арсений Бирюков Т-Банк</p>	<p>Рецепт платформы потоковой обработки данных на Apache Flink</p> <p>Данил Сабиров Яндекс.Такси Технологии</p>	<p>Обработка событий в Snowplow: от сбора до аналитики</p> <p>Мария Бученкова Детский мир</p> <p>Владимир Седелников Детский мир</p>
14:45-15:15	<p>CDC от источника до хранилища: как в банке Синара построили CDC с применением продуктов Arenadata</p> <p>Иван Клименко Arenadata</p>		
15:15-16:00	Обед		
16:00-16:45	<p>Как навести порядок в двух эксбайтах данных?</p> <p>Максим Гудзикович Яндекс</p> <p>Кирилл Осинцев Яндекс</p>	<p>Автогенерация синтетических данных с использованием алгоритмов ML</p> <p>Данила Самошкин Neoflex</p> <p>Анастасия Коткова Neoflex</p>	<p>Оптимизации сериализатора ВКонтакте</p> <p>Илья Асадуллин VK</p> <p>Илья Кокорин VK</p>
17:15-18:00	<p>Оркестратор пайплайнов для небольшой команды инженеров и аналитиков: как мы используем Dagster</p> <p>Алексей Завальский Bestplace</p> <p>Дмитрий Крылов Bestplace</p>	<p>The State of Data, RU Edition</p> <p>Олег Кочергин Positive Technologies</p>	<p>Storage для lake</p> <p>Александр Казанский Т-Банк</p>
18:30-19:15	<p>Это реальный космос: space-индустрия сегодня</p> <p>Яна Харлан Бюро 1440</p>		
19:25-19:40	<p>Заккрытие конференции SmartData 2024</p> <p>Владимир Озеров CedrusData</p> <p>Олег Кочергин Positive Technologies</p>		