

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 Дмитрий Вертлиб Честный знак | Шардированный не значит распределенный: что важно знать, когда PostgreSQL мало Евгений Иванов Олег Бондарь Яндекс Яндекс | Как жить с data swamp 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 Дмитрий Ульянов Яндекс | Airbyte. 2 года в продакшене Александра Попова СберЗдоровье | Assessing Data Pipeline Quality & Sanity with Data Angiograms 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 Алсу Нурутдинова Алина Кочева Positive Technologies Positive Technologies | Проблемы обработки Excel-файлов в Apache NiFi и как их решать Игорь Шемаров Axenix | От хайпа до продакшена: data mesh на Airflow + dbt 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 Милена Булкина Т-Банк | Debezium Engine: практическое руководство по использованию 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 | Как мы сократили TTM создания дашбордов | Как кролик съел зеленую сливу и не умер: сказ о миграции на Iceberg |
| | Александр Токарев Яндекс | Анар Багиров Авито | Валентин Пановский BestDoctor |
| 12:00-12:45 | Как работает Apache Iceberg на примере Trino | Как мы Apache Kafka на Redpanda меняли | Data Vault 2.0. В каких случаях внедрять, разбор основных проблем применения методологии при построении DWH на Greenplum |
| | Владимир Озеров CedrusData | Роман Ананьев Авито | Денис Лукьянов Ecom.tech |
| 12:45-13:30 | Обед | | |
| 13:30-14:15 | Инструменты Data Quality: как, зачем, почему. Опыт Т-Банка | dbt в деле: реальные кейсы и лайфхаки | Data sketches — быстро, дешево и (почти) точно! |
| | Дмитрий Руднев Т-Банк | Антон Александров Детский мир | Сергей Жемжицкий Arenadata |
| 14:15-15:00 | Обед | | |
| 15:00-15:45 | Data mesh: ожидания vs реальность | CI/CD в большом on-premise Datalake-- проекте | Как мы делаем облачный Greenplum |
| | Петр Гуринов Лемана ПРО (Леруа Мерлен) | Александра Чекмарева (Китченко) Честный знак Никита Благодарный Честный знак | Леонид Борчук Yandex Cloud |
| 16:15-17:00 | Каждый байт на вес золота. Опыт построения DMP в рекламе Яндекса | Data Lineage: как настроить в зоопарке технологий и зачем это нужно | Выбрасываем Java икратно ускоряем Spark/Presto... Или пока нет? |
| | Алексей Стыценко Яндекс | Булат Усманов Кулер (ex-СберМаркет) | Павел Солодовников CedrusData |
| 17:30-18:15 | Как быстро запустить процесс ведения каталога данных в компании на примере DataHub | Schema Registry: Ultimate Guide | Snowplow: утерянное руководство |
| | Наталья Журавлева Ozon | Тимофей Брунько Yandex Cloud | 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 | Apache Arrow: быстрее, ниже, сложнее Евгений Готов Autotech | Как мы тестировали 5 способов загрузки данных в Greenplum и что из этого вышло Татьяна Дидова АЭРО | NiFi. Пишем код для codeless-системы Бронислав Житников Positive Technologies |
| 12:15-13:00 | Как мы строим систему распределенного трейсинга, в которой можно терять данные Игорь Балюк Авито | One More Way to Make Backup in Ignite Николай Ижиков Apache Software Foundation | От ручного труда к автоматической генерации проверок качества данных Александр Мадумаров Инновационный центр «Безопасный транспорт» ГКУ ЦОДД |
| 13:00-13:45 | Обед | | |
| 13:45-14:30 | CI/CD для большого хранилища данных Арсений Бирюков Т-Банк | Рецепт платформы потоковой обработки данных на Apache Flink Данил Сабиров Яндекс.Такси Технологии | Обработка событий в Snowplow: от сбора до аналитики Мария Бученкова Детский мир Владимир Седелников Детский мир |
| 14:45-15:15 | CDC от источника до хранилища: как в банке Синара построили CDC с применением продуктов Arenadata Иван Клименко Arenadata | | |
| 15:15-16:00 | Обед | | |
| 16:00-16:45 | Как навести порядок в двух эксбайтах данных? Максим Гудзикович Яндекс Кирилл Осинцев Яндекс | Автогенерация синтетических данных с использованием алгоритмов ML Данила Самошкин Neoflex Анастасия Коткова Neoflex | Оптимизации сериализатора ВКонтакте Илья Асадуллин VK Илья Кокорин VK |
| 17:15-18:00 | Оркестратор пайплайнов для небольшой команды инженеров и аналитиков: как мы используем Dagster Алексей Завальский Bestplace Дмитрий Крылов Bestplace | The State of Data, RU Edition Олег Кочергин Positive Technologies | Storage для lake Александр Казанский Т-Банк |
| 18:30-19:15 | Это реальный космос: space-индустрия сегодня Яна Харлан Бюро 1440 | | |
| 19:25-19:40 | Заккрытие конференции SmartData 2024 Владимир Озеров CedrusData Олег Кочергин Positive Technologies | | |