

6 сентября ONLINE

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

16:00-16:15 [Открытие конференции SmartData 2023](#)

Максим Стаценко
Яндекс

Андрей Кузнецов
Одноклассники

16:15-17:00 [Как перевести отчетность в чат-бота Microsoft Teams, используя Apache Airflow](#)

Станислав Горчаков
Независимый эксперт

[DataFrame — настоящее и будущее](#)

Александр Боргардт
duckstax.com

Евгений Селиверстов
Agro Software

Николай Марков
Rayo Data

Павел Филонов
Независимый консультант

[От потоков данных до высококачественных ML-моделей. Секрет конструирования и тестирования признаков](#)

Юрий Гусев
Независимый консультант

16:55-17:35 [Что не так с NoSQL?](#)

Александр Волочнев
Datastax

Алексей Стукалов
HAULMONT

Максим Стаценко
Яндекс

17:30-18:15 [Эволюция схемы данных. Носим данные из реляционной СУБД в Hadoop](#)

Михаил Благов
билайн

[Как мы переезжали с PostgreSQL на Data Lake в AWS](#)

Павел Сивохин
Whoosh

Никита Зеленский
Whoosh

Константин Малыхин
Whoosh

[Телескопы XXI века: от железа и софта к данным и результатам](#)

Александр Плавин
АКЦ ФИАН

18:10-18:50 [Data Maturity: 18+](#)

Максим Стаценко
Яндекс

Андрей Кузнецов
Одноклассники

Бронислав Житников
Тинькофф

Андрей Филипьев
Dodo Brands

18:45-19:30 [Apache Flink под капотом: distributed, stateful, realtime](#)

Валентина Предтеченская
Авито

[Как навести порядок в логировании продуктовых событий](#)

Алексей Балехов
Окко

[Планирование миллиардов задач каждый день](#)

Игнат Колесниченко
YTsaurus

19:25-20:05 [Ну как там с рынком труда? Взгляд изнутри](#)

Семён Осипов
Gett

Екатерина Фирсова
Altenar

Сергей Михалев
VK

Максим Стаценко
Яндекс

20:00-20:45 [ACID-транзакции в Apache Cassandra 5.0](#)

Александр Волочнев
Datastax

[CDC. От баззворда к реализации в Data Transfer](#)

Тимофей Брунько
Yandex Cloud

[Примеры реальных аналитических решений и дата-команд в западных компаниях](#)

Дмитрий Аношин
Surfalytics

10:35-10:45 [Открытие второго дня SmartData 2023](#)

Максим Стаценко
Яндекс

Михаил Лукин
Судо

10:45-11:30 [Платформа как продукт: разработать и внедрить сложное технологическое решение внутри компании](#)

Наджим Мохаммад
MTC Digital

Дмитрий Бодин
MTC Digital

Максим Бартенев
MTC Digital

[Глубокое погружение в производительность запросов](#)

Петр Зайцев
Percona

[dbt — ядро современной платформы данных](#)

Евгений Ермаков
Toloka Ai Inc

11:25-12:05 [Тренды Data Engineering or DEorDIE](#)

Максим Стаценко
Яндекс

Михаил Лукин
Судо

Ксения Томак

Никита Бочкарев
Toloka.ai

Дарья Буланова
HelloFresh

12:00-12:45 [Обеспечение качественных ETL на Vertica](#)

Александр Крашенинников
Тинькофф

[ML System Design Interview](#)

Павел Филонов
Независимый консультант

Аркадий Василенко
Одноклассники

12:40-13:20 [Своя пирамида](#)

Михаил Благоев
билайн

Александр Шорин
Сбер

Максим Бартенев
MTC Digital

Максим Стаценко
Яндекс

Михаил Лукин
Судо

13:15-14:00 [От сырого кликстрима к чистым датасетам, или История развития Feature Storage в Lamoda](#)

Дана Злочевская
Lamoda Tech

Михаил Нестеров
Lamoda Tech

[Что делать, если DWH растет слишком быстро](#)

Александр Филатов
Авито

14:00-14:30 [Викторина и подведение итогов online-части конференции SmartData 2023](#)

Максим Стаценко
Яндекс

Михаил Лукин
Судо

Мария Лаврова
JUG Ru Group

14:30-15:15 [Моделирование потоков событий в эволюционирующем окружении](#)

Николай Голов
ManyChat

	Зал 1	Зал 2	Зал 3
11:00-11:15	Открытие офлайн-части конференции SmartData 2023 Михаил Марюфич Алексей Федоров Одноклассники JUG Ru Group		
11:30-12:15	Я изменю ваш взгляд на хранилище данных за 30 минут Максим Стаценко Яндекс	Возможна ли жизнь с open source BI в большой компании Илья Аникин Авито	Common Data Index. Как построить поисковую систему по открытым данным такую же, как Google Dataset Search, но проще и быстрее Иван Бегтин API Crafter
12:45-13:30	Как мы адаптировали динамические таблицы YTsaurus для хранения блобов Максим Бабенко Яндекс	Методы обезличивания данных Алексей Даньшин Neoflex	Как сделать так, чтобы вашему Apache NiFi было плохо Бронислав Житников Тинькофф
13:45-14:15		Построение катастрофоустойчивых хранилищ данных Александр Тарасов Arenadata	
13:45-14:45	Graph challenge от «Поиска и рекламных технологий» Яндекса		
14:15-15:00	Обед		
15:00-15:45	Hadoop в Облаке — это ОК Михаил Марюфич Одноклассники	Архитектура распределенного SQL--движка для аналитических запросов Алексей Озерицкий Яндекс	Пара слов о том, как мы Data Observability готовим Олег Кочергин
15:45-16:30	Обед		
16:30-17:15	Разгоним запросы: как быстро готовить ClickHouse Кузьма Лешаков Yandex Cloud	Apache Flink на примере задачи дедупликации Александр Бобряков MTC Digital	Mage: волшебный инструмент оркестрации Валентин Пановский more.tv
17:45-18:30	Внеатмосферная астрономия и новый космический телескоп «Джеймс Уэбб» Владимир Сурдин МГУ		

18:30-22:00
Нетворкинг

18:30-18:45
[Подведение итогов Graph challenge от «Поиска и рекламных технологий» Яндекса](#)
Илья Шишов
Яндекс

19:30-21:00 **Зал 2**
[Квиз](#)
Программный комитет Smart-Data

19:30-20:15 **Зал 3**
[VOF-сессия: Наука и программисты в космической сфере](#)
Владимир Сурдин
МГУ
Сергей Космос SR Data Михаил Лукин Судо

	Зал 1	Зал 2	Зал 3
11:00-11:45	<p>Быстрая обработка данных в Data Lake с помощью Trino</p> <p>Владимир Озеров Querify Labs</p>	<p>Как обрабатывать данные с помощью Spark в облаке</p> <p>Максим Зиналь Дмитрий Рыбалко Yandex Cloud Yandex Cloud</p>	<p>Streaming Data Integration – ETL-инструмент для создания near real-time-процессов</p> <p>Василий Мельник GlowByte</p>
12:15-13:00	<p>Слабоумие и отвага, или Как мы за два месяца объединили данные Delivery Club и Яндекс Еды</p> <p>Ольга Титова Яндекс Еда</p>	<p>Движение к универсальности: гибридная OLTP-база с поддержкой OLAP-запросов</p> <p>Алексей Дмитриев Яндекс</p>	<p>Путь Model Serving: от Flask к своей платформе</p> <p>Алина Кочева Positive Technologies</p>
13:30-14:15	<p>Сжатие, шифрование и не только: меняем поведение и гарантии распределенной базы данных</p> <p>Антон Виноградов Apache Software Foundation</p>	<p>Развитие инструмента BI-аналитики, DataOps.BI, на основе open source-решения Apache Superset</p> <p>Павел Шестаков MTC Digital</p>	<p>Kafka Connect: что за зверь этот ваш Single Message Transform?</p> <p>Даниэл Рачич</p>
14:15-15:00	Обед		
15:00-15:45	<p>Визуализация для ELT-процессов в DWH</p> <p>Виталий Бодренков СберМаркет</p>	<p>Платформа управления данными вокруг YTsaurus</p> <p>Владимир Верстов Яндекс Go</p>	<p>Чего стоит достижение линеаризуемости в распределенной системе</p> <p>Сергей Петренко Tarantool</p>
15:45-16:30	Обед		
16:30-17:15	<p>Spark Streaming: брать или не брать?</p> <p>Евгений Ненахов MTC Digital</p>	<p>Предиктивный анализ паразитной нагрузки на кластерах GreenPlum</p> <p>Павел Тернюк Марк Лебедев Data Sapience GlowByte Consulting</p>	<p>Применение TLA+ для эффективного тестирования распределенных систем</p> <p>Никита Синяченко ВКонтакте</p> <p>Евгений Чернацкий ВКонтакте</p>
17:45-18:30	<p>Создание группы сервисов по анализу космических снимков с помощью машинного обучения</p> <p>Сергей Космос SR Data</p>		
18:40-18:55	<p>Заккрытие конференции SmartData 2023</p> <p>Максим Стаценко Михаил Марюфич Яндекс Одноклассники</p>		