セミナー情報

SEMINARS

外部主催セミナー

2017年

▶ 「Strata Data Conference in Singapore」

■開催概要

会期　　：2017年12月5日～7日

会場　　：Suntec Singapore Convention & Exhibition Centre(Singapore)

主催　　：Cloudera, O'REILLY

■セッション詳細

Agenda一覧

■NTTデータセッション 講演内容

【3日目】「Fusing a deep learning platform with a big data platform」

時　間 1:45pm-2:25pm

StarHub YongLiang Xu氏

NTTデータ OSSプロフェッショナルサービス 岩崎 正剛

SmartHub, the analytics division of StarHub, and NTT DATA, a global IT innovator in Japan with committers to Hadoop and Spark, have embarked on a partnership to design next-generation architecture to power the data products that will help generate new insights. YongLiang Xu and Masatake Iwasaki explain how deep learning and other analytics models can coexist on the same platform to address opportunities and challenges in initiatives such as smart cities. Deep learning is the next key-enabler to transform data into actionable analytics products. However, big data platforms using technologies such as Hadoop and Spark remain the backbone for analytic applications. Therefore integrating big data platforms

with deep learning technologies like TensorFlow is crucial to support the development of cutting-edge data analytics products. YongLiang and Masatake present a reference architecture that incorporates distributed deep learning with an existing big data platform through frameworks such as Intel BigDL and TensorFlowOnSpark. This architecture creates an environment in which deep learning workloads can coexist with other existing analytics workloads and continue to leverage the same real-time data pipeline and monitoring frameworks within the platform.

Topics include:
* An overview of the reference architecture that incorporates distributed deep learning with existing big data platforms
* Key considerations on why SmartHub and NTT DATA have chosen the above architecture
* Case studies on the integration and deployment of deep learning models on the integrated architecture
* Future works and plans

▶ 「Global Big Data Conference」

■開催概要
会期　　：2017年8月29日(火) ー 8月31(木)
会場　　：Santa Clara Convention Center, 5001 Great America Parkway, Santa Clara, CA
■セッション詳細
Agenda一覧
■NTTデータセッション 講演内容
「5 Lessons learned from Big Data / IoT platform production case studies」
時　間 13:40-14:20
NTTデータ OSSプロフェッショナルサービス 下垣 徹

NTT DATA provides Big Data / IoT platform related services to various enterprise customers in different fields. Architecture of Big Data / IoT platform tend to be complex and complicated. It is important to come up with simple, efficient and easy enough solutions to manage such complicated platforms. In this session, I would like to present key takeaways based on production case studies.

▶ 「Kafka Summit 2017」

■開催概要
会期　　：2017年8月28日(月)
会場　　：Hilton San Francisco Union Square
■セッション詳細

Agenda一覧

■NTTデータセッション 講演内容

「Worldwide Scalable and Resilient Messaging Services with Kafka and Kafka Streams」

時　間 16:30-17:10

ChatWork株式会社　CTO室 大村 伸吾 氏

NTTデータ OSSプロフェッショナルサービス 土橋 昌

ChatWork is a worldwide communication service, which holds 110k+ of customer organizations. In 2016, we have developed a new scalable infrastructure based on the idea of CQRS and Event Sourcing using Kafka and Kafka Streams combined with Akka and HBase. In this session, we talk about the concept of this architecture and lessons learned in production use cases.

▶ 「第3回Kafka Meetup Japan」

■開催概要

会期　　：2017年7月6日(木) 19:00〜

会場　　：ヤフー紀尾井町オフィス(東京ガーデンテラス紀尾井町 紀尾井タワー)

主催　　：Yahoo! JAPAN

■NTTデータセッション 講演内容

「案件の現場から贈る、Kafka利用の苦労話　〜"こんなとき" に気を付けることは？〜」

NTTデータ OSSプロフェッショナルサービス 佐々木 徹

▶ 「DataWorks Summit 2017」

■開催概要

会期　　：2017年6月13日(火) ー 15日(木)

会場　　：San Jose McEnery Convention Center

■セッション詳細

Agenda

■NTTデータセッション 講演内容

「Worldwide Scalable and Resilient Messaging Services by CQRS and Event Sourcing using Akka, Kafka Streams and HBase」

【3日目】時　間 16:10-17:50

ChatWork株式会社　CTO室 大村 伸吾 氏

NTTデータ OSSプロフェッショナルサービス 土橋 昌

ChatWork is one of major business communication platforms in Japan. We keep growing up for 5+ years since our service inception. Now, we hold 110k+ of customer organizations which includes large organizations like telecom companies and the service is widely used across

200+ countries and regions.
Nowadays we have faced drastic increase of message traffic. But, unfortunately, our conventional backend was based on traditional LAMP architecture. Transforming traditional backend into highly available, scalable and resilient backend was imperative.
To achieve this, we have applied “Command Query Responsibility Segregation (CQRS) and Event Sourcing” as a heart of its architecture. The simple idea of segregation brings us independent command-side and query-side system components and it can subsequently achieve highly available, scalable and resilient systems. It is desirable property for messaging services because, for example, even if command-side was down, user can keep reading messages unless query-side was down. Event Sourcing is another key technique to enable us to build optimized systems to handle heterogeneous write/read load. This means that we can choose optimized storage platform for each side. Moreover, the event data can be the rich source for real-time analysis of user’s communication behavior. We have chosen Kafka as a command-side event storage, HBase as a query-side storage, Kafka Streams as a core library to give eventual consistency between the two sides. In application layer, Akka has been chosen as a core framework. Akka can be a good choice as an abstraction layer to build highly concurrent, distributed, resilient and message-driven application effectively. Backpressure introduced by Akka Stream can be important technology to prevent from overflow of data flows in our backend, which contributes system stability very well.
In this session, we talk about how above architecture works, how we concluded above architectural decisions on many trade-offs, what was achieved by this architecture, what was the pain points (e.g. how to guarantee eventual consistency, how to migrate systems in the real project, etc.) and several TIPS we learned for realizing our highly distributed and resilient messaging systems.
ChatWork is a business communication platform for global teams. Our four main features are enterprise-grade group chat, file sharing, task management and video chat. NTT DATA is one of biggest solution provider in Japan and providing technical support about Open Source Software and distributed computing. The project has been conducted with cooperation of ChatWork and NTT DATA.

▶ 「Apache Big Data North America」

■開催概要
会期　　：2017年5月16日(月) ー 18日(木)
会場　　：InterContinental Miami
■NTTデータセッション 講演内容
「Java9 Support in Apache Hadoop」
【3日目】時　間 9:00-9:50
NTTデータ OSSプロフェッショナルサービス 鯵坂 明

Java 9 is the next major version and will be GA in July 2017, and it's very important for Apache Hadoop to support Java 9 earlier. Hadoop has many downstream projects and it makes the projects to support Java 9 easily. Java 9 has more incompatible changes than any earlier releases. For example, Project Coin (JEP 213) banned '_' as an identifier and Hadoop Web UI is affected. In this session, Akira will introduce what are the incompatible changes and what we need to do to support Java 9 in Hadoop. Classpath isolation is also an important issue for Hadoop. Hadoop has many dependencies, and the developers who write applications running on Hadoop need to be careful not to conflict the classpath. Java 9 Jigsaw feature is expected to solve this 'jar hell' problem but Hadoop does not use the feature for now. Akira will also introduce how Hadoop community solves the problem without Jigsaw.

▶ 「ApacheCon North America 2017」

■開催概要

会期　　：2017年5月16日(月) ー 18日(木)

会場　　：InterContinental Miami

■NTTデータセッション 講演内容

「10 Things to Consider When Using Apache Kafka: Utilization Points of Apache Kafka Obtained From IoT Use Case」

【2日目】時　間 15:30 - 16:20

NTTデータ OSSプロフェッショナルサービス 梅森 直人, 萩原　悠二

We've been working on IoT for around 3 years, and are recently targeting a use case of connected car that the scale is as follows:

* Concurrent connections over a million
* Throughput over 100 Gbps

We thought that Apache Kafka would be effective as a means of data collection function of IoT platform which satisfies these requirements. We built and tested the platform however we faced various issues below:

* Performance saturation of Kafka Producer NOT caused by depletion of computer resources
* Performance deterioration of Kafka Broker due to sudden disk IO
* Crash of Kafka Consumer due to sizing mistake between the Broker and Consumer

To solve the above issues and make full use of capabilities of the Kafka, it is necessary to clarify mechanism of the Kafka and its surroundings. This presentation will introduce pitfalls, solutions and best practices based on our experiences.