

セミナー情報

SEMINARS

外部主催セミナー

2016年

▶ 「Strata + Hadoop World Singapore 2016」

■開催概要

会期 : 2016年12月6日～8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

【2日目】「IoT and Spark MLlib applications for improving products, services, and manufacturing technologies」

時 間 11:15-11:55

株式会社IHI 鈴木 由宇 氏

NTTデータ OSSプロフェッショナルサービス 土橋 昌

Using ILIPS, a common platform for remote monitoring and maintenance developed in house, IHI has collected data from several kinds of products, including aircraft engines and industrial machineries. To utilize the data for preventive maintenance and operation optimization, IHI started evaluating Spark as a potential scalable and flexible analytics platform in 2014, in collaboration with NTT DATA, an active contributor to Spark. Yoshitaka Suzuki and Masaru Dobashi explain how IHI used PySpark and MLlib to improve its services and share best practices for application development and lessons for operating Spark on YARN.

As a first step, IHI evaluated Spark's characteristics and processing capabilities for time series data through the prediction of a port congestion problem using dummy GIS data. As of this year, IHI has started analyzing real data including sensor data, real GIS data, and system logs. Additionally, IHI has started developing an in-house analytics infrastructure utilizing Spark and leveraging Spark MLlib to get up speed in developing applications for process improvement, product fault diagnosis, and the formalization of highly skilled expert knowledge. Highly skilled experts using MLlib can now analyze the relationships among a large amount of sensor data, enabling IHI to reduce operating costs by automating processes and improve the transmission of knowledge between workers.

【2日目】「Integrated data analytics for consumer electronics using Hadoop and Spark MLlib」

時 間 17:05-17:45

パナソニック株式会社 西川 敬之 氏

NTTデータ OSSプロフェッショナルサービス 山口 永

Panasonic, one of the largest home electronics product manufacturers in the world, has launched a number of IoT home appliances connected to the cloud (air conditioners, rice cookers, microwave ovens, etc.). Along with providing cloud services to these products, Panasonic collects operation and state logs from the connected home appliances.

Takayuki Nishikawa and Ei Yamaguhi offer an overview of Panasonic's recently developed integrated data analytics platform, which uses Hadoop and Hive as core components with NTT DATA (one of the largest system integrators in Japan, which is widely contributing to Hadoop and Spark community with several committers) in order to analyze the ever-increasing log data from these IoT home appliances. The platform makes the complicated data analytics process simple and has achieved scalability for millions of households and yielded an over 10x improvement in processing time. The platform also analyzes and extracts valuable information from multiappliance log data using machine learning with Spark MLlib, enabling Panasonic to get more reliable knowledge about users' lifestyles.

Takayuki and Ei discuss the technical know-how obtained from the use case, designing and building the platform architecture, migrating existing processes, and constructing an analysis model with Spark MLlib.

▶ 「db tech showcase Tokyo 2016」

■開催概要

会期 : 2016年7月14日(木)

会場 : 秋葉原UDX Conference 6F

主催 : 株式会社インサイトテクノロジー

■NTTデータセッション 講演内容

「Apache Spark 2.0がリリースされるって噂だけど何が変わるの？教えて！コミッタさん」

NTTデータ OSSプロフェッショナルサービス 猿田 浩輔

「進化を続けるイマドキのHive ～現場での利用シーンを交えながら～」

NTTデータ OSSプロフェッショナルサービス 吉田 耕陽

▶ 「オープンソースカンファレンス2016 Okinawa」

■開催概要

会期 : 2016年7月2日(土)

会場 : 沖縄コンベンションセンター 会議棟B

主催 : オープンソースカンファレンス実行委員会

■セッション詳細

タイムテーブル

■NTTデータセッション 講演内容

「Hadoop/Sparkの今、そしてこれから」

時 間 13:00 - 13:45

NTTデータ OSSプロフェッショナルサービス 吉田 耕陽

OSSベースの分散処理基盤として、Apache Hadoopは欠かせない技術となりました。

Hadoopと連携する技術も進化を加速しています。特に、Apache Sparkの登場により機械学習や分析などのシーンでもより利用が広がっています。本セッションでは、これまでのHadoop/Sparkの利用を踏まえたうえで、HadoopやSparkが現在どのような開発状況なのか、また今後どのように発展していくかを発表します。さらに、NTTデータのHadoop/Sparkに関する取り組みについてもご紹介します。

▶ 「オープンソースカンファレンス2016 Hokkaido」

■開催概要

会期 : 2016年6月17日(金)～6月18日(土)

会場 : 札幌コンベンションセンター

主催 : オープンソースカンファレンス実行委員会

■セッション詳細

1日目タイムテーブル

2日目タイムテーブル

■NTTデータセッション 講演内容

「Hadoop/Sparkの今、そしてこれから」

時 間 14:00 - 14:45

NTTデータ OSSプロフェッショナルサービス 佐々木 徹

▶ 「OSSユーザーのための勉強会 < OSS X Users Meeting > #14 HadoopとSpark」

■開催概要

会期 : 2016年6月1日(水) 18:30～ (20:15終了予定、開場18時)

会場 : 国立情報学研究所 12階 1208,1210会議室

主催 : SCSK株式会社 R&Dセンター OSS戦略企画室

協賛 : NPO法人トップエスイー教育センター

協力 : 国立情報学研究所 (NII) GRACEセンター、富士通株式会社、株式会社NTTデータ

■NTTデータセッション 講演内容

「Sparkのキホンとはじめ方」

NTTデータ OSSプロフェッショナルサービス 土橋 昌

▶ 「オープンソースカンファレンス2016 Nagoya」

■開催概要

会期 : 2016年5月28日(土)

会場 : 名古屋中小企業振興会館 吹上ホール

主催 : オープンソースカンファレンス実行委員会

■セッション詳細

タイムテーブル

■NTTデータセッション 講演内容

「Hadoop/Sparkの今、そしてこれから」

時 間 13:00 - 13:45

NTTデータ OSSプロフェッショナルサービス 田中 正浩

▶ 「Apache Big Data North America」

■開催概要

会期 : 2016年5月9日(月) - 12日(木)

会場 : Hyatt Regency Vancouver in Vancouver, British Columbia

■NTTデータセッション 講演内容

「Apache Hadoop 3 Current Status」

【1日目】時 間 10:40-11:30

NTTデータ OSSプロフェッショナルサービス 鯨坂 明

Do you want Hadoop 3 release? It is over 4 years since Hadoop 3 and Hadoop 2 were diverged, and there are a lot of great improvements in Hadoop 3, such as Shell Script Rewrite and MapReduce Native Optimization.

Therefore if Hadoop 3 is released, users can enjoy the benefits of the new features. In this session, we will introduce the new features and incompatible changes in Hadoop 3, and how the release is discussed in Apache Hadoop community. In addition, Akira Ajisaka would like to discuss releasing Hadoop 3 with the participants here if possible.

▶ 「Developers Summit 2016」

■開催概要

会期 : 2016年2月18日(木)

会場 : 目黒雅叙園

主催 : 翔泳社

■NTTデータセッション 講演内容

「Apache Sparkに手を出してヤケドしないための基本～「Apache Spark入門より」～」

時 間 13:05-13:50

NTTデータ 基盤システム事業本部 OSSプロフェッショナルサービス 土橋 昌, 吉田 耕陽

▶ 「Hadoop / Spark Conference Japan 2016」

■開催概要

会期 : 2016年2月8日(月)

会場 : 品川区立総合区民会館 きゅりあん

主催 : 日本Hadoopユーザー会

■NTTデータセッション 講演内容

Keynote 10:00-12:00

「ご挨拶、Hadoopを取り巻く環境2016」 濱野 賢一郎

「Hadoopの現在と未来」 鯉坂 明 (Hadoopコミッタ、NTTデータ)、小沢 健史 (Hadoopコミッタ、NTT)

「Spark Conference Japanの開催にあたって」 猿田 浩輔 (Apache Sparkコミッタ)

会場C 12:00-12:40 「Apache Spark 超入門」

猿田 浩輔

会場C 13:45-:14:20 「リクルートライフスタイルの考えるストリームデータの活かし方 ～AWS + Kafka + Spark Streaming～」

車田 篤史 (リクルートライフスタイル) , 堤 崇行 (NTTデータ)

会場B 14:30-15:05 「今あらためて考えるHive ～ユースケースの広がりにより顕在化した課題と対応～」

吉田 耕陽

会場D 15:15-:15:50 「SparkによるGISデータを題材とした時系列データ処理」

鈴木 由宇 (IHI) , 土橋 昌 (NTTデータ)

LT-ランチタイムA 「サポートメンバは見た! Hadoopバグワースト10」

鯨坂 明

LT-ランチタイムA 「本当にあったHadoopの怖い話〜Blockはどこへきえた？」

山下 真一

LT-懇親会 「5分で身に付く! Apache Hadoop開発」

鯨坂 明

▶ 「CROSS 2016」

■開催概要

会期 : 2016年2月5日(金)

会場 : 横浜港大さん橋ホール

主催 : 一般社団法人日本インターネットプロバイダー協会

■NTTデータパネルディスカッション 内容

「コラコラ、Hadoop や Spark よ！君たちそれでよいのかい！？星降る夜空で象に寄り添いたい…そう願うオレ達は何処にたどり着くのか」

時 間 15:50-17:10

NTTデータ 基盤システム事業本部 OSSプロフェッショナルサービス 下垣 徹, 鯨坂 明, 岩崎 正剛, 猿田 浩輔

HadoopやSparkは様々な企業で大規模データ処理に使われていますが、実際に動作しているソースコードの中身を見ると、なぜこんなコードが本番環境で動いているんだろうか？と、思うことがよくあります。

そんな、日頃開発していて「なんじゃこりゃ？」と感じる、ソースコードレベルのお話を持ち出しつつ、いったい Hadoop や Spark はどうなっていくんだろう、なんてことを、Hadoop/Spark のコミッタ陣がただひたすら語ります。