

AWS Lake Formationを利用した 企業ネットワーク間のデータインジェスト

2022年3月24日

ヤフー株式会社

石川貴大

自己紹介

石川貴大 (いしかわたかひろ)

- **役職**

- COO事業推進室 事業推進統括部 データディレクター

- **業務**

- コマース全般のマーケティングへのデータ利活用の推進
- PayPayモールやフリマ、ショッピング、トラベルなど

- **その他**

- ソニーコンピュータサイエンス研究所の客員研究員として年金の安定運用に関するAI/MLの応用研究に従事



アジェンダ

1. 背景

- ZHD事業戦略におけるサービス統合

2. 経由流通総額の可視化

- セッションデータ
- 非ログイン/ログインセッション
- アトリビューションモデル

3. 一休様とのデータ連携

- データパイプラインの管理
- AWS Lake Formationを利用したデータ授受
- 従来の連携手法との比較

4. 導入効果と今後の展開

アジェンダ

1. 背景

- ZHD事業戦略におけるサービス統合

2. 経路流通総額の可視化

- セッションデータ
- 非ログイン/ログインセッション
- アトリビューションモデル

3. 一休様とのデータ連携

- データパイプラインの管理
- AWS Lake Formationを利用したデータ授受
- 従来の連携手法との比較

4. 導入効果と今後の展開

ZHDグループについて

ホームページに設定する

Yahoo! BB | きっず版 | アプリ版 | ヘルプ

トラベル ヤフオク! ショッピング LINE Y! PayPay STEP カード メール

ヤフーLINE共同企画
これからは、できること。 検索は、チカラになる。
ヤフーとLINEで「3.11」と検索すると、復興支援のためにそれぞれ10円が寄付されます。 [いますぐ検索 >](#)

ウェブ 画像 動画 知恵袋 地図 リアルタイム 一覧▼

Q 検索

» 寄付はチカラになる 震災復興の支援募金受付中 > 1日1回挑戦 人気賞品100万本がくじで当たる > ロシアのウクライナ侵攻 最新情報

ショッピング PayPayモール ヤフオク! PayPayフリマ ZOZOTOWN LOHACO トラベル 一休.com 一休.comレストラン 出前館 ニュース 天気・災害

ニュース 経済 エンタメ スポーツ 国内 国際 IT・科学 地域

3/11(金) 8:22更新

- 震災11年 予算大幅減で復興新局面 664
- 津波の記憶に苦しむ妹 癒えない傷 **NEW** 79
- EU ウクライナ加盟を見送りへ **NEW** 14
- 政府4回目接種検討へ 今夏開始か
- 正恩氏 ミサイル発射場の拡充指示 29
- 津波で父亡くし 朗希20歳の使命 **NEW** 47
- 元WBC女子王者 男子で復帰準備 439
- 新婚さん卒業に桂文枝 51年は異常 **NEW** 367

もっと見る トピックス一覧

東北グルメ 食べて応援

買うは、チカラになる。 Y! LINE

国内の新型コロナウイルス発生状況 - 3月10日10時時点 > 最新情報を見る

<https://www.yahoo.co.jp/>

ZHDグループ各社（抜粋）

YAHOO!
JAPAN

休.com

ASKUL

GYAO
Entertainment Company

Sportsnavi

LINE

e book
japan

Demaecan

mybest

dely

PayPay

zozo

BuzzFeed
JAPAN

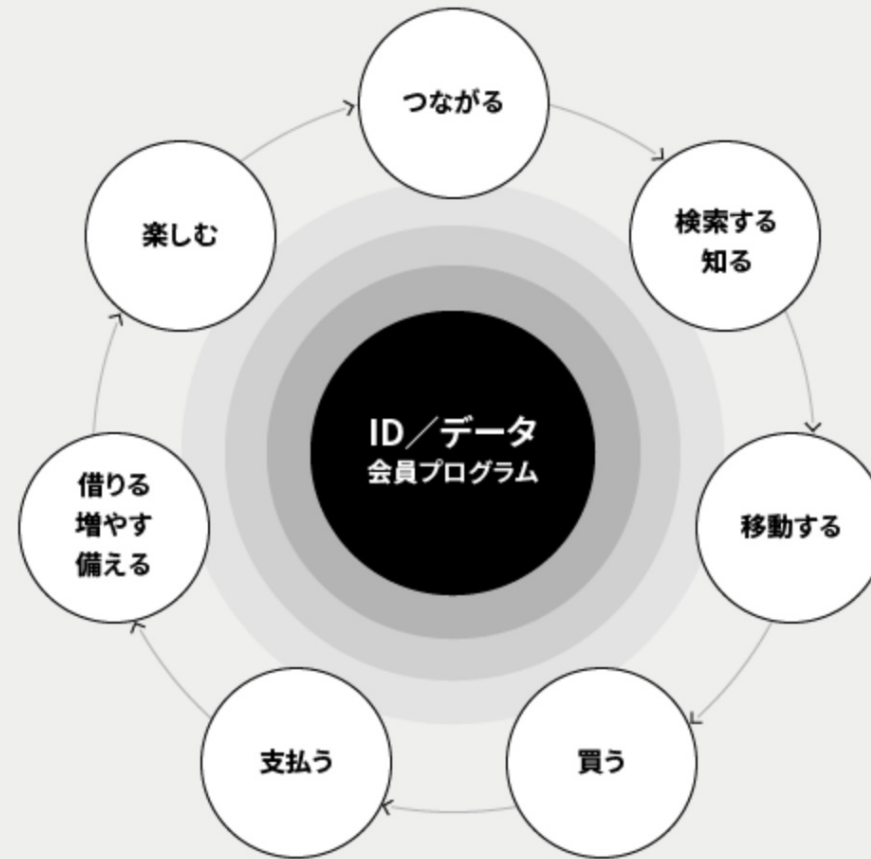
carview™

スタンバク

<https://www.z-holdings.co.jp/company/groupcompanies/>

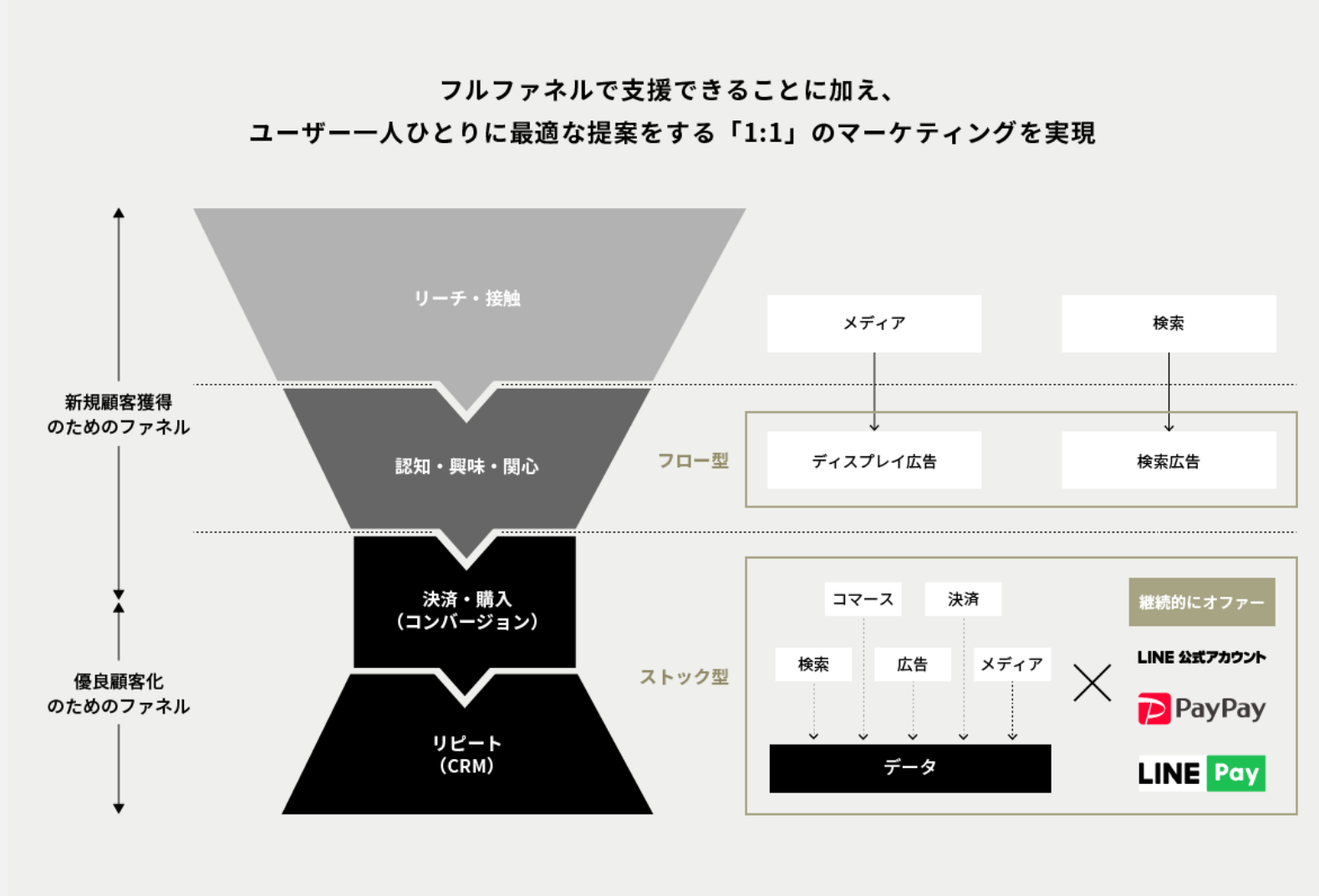
ZHD事業戦略におけるサービスの統合 (1/3)

各サービスの連携を深めて、独自の経済圏を実現



<https://www.z-holdings.co.jp/integrated-report/vision/growth/>

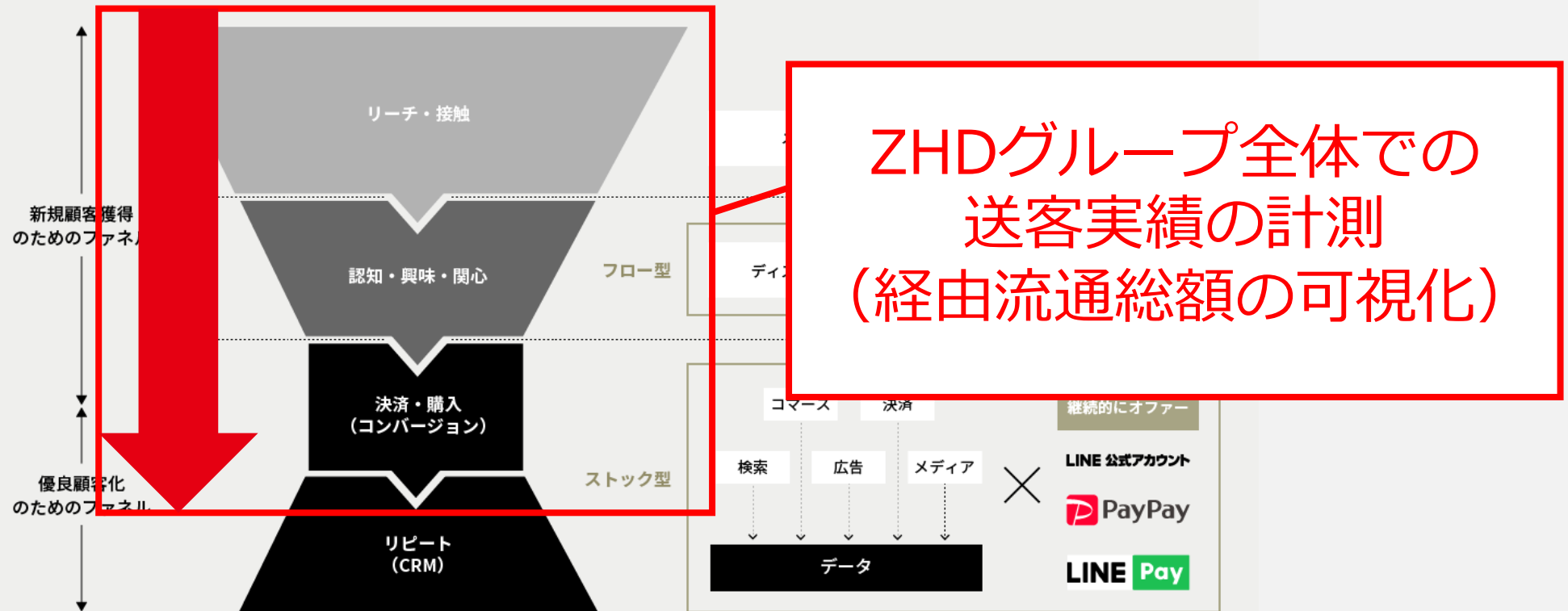
ZHD事業戦略におけるサービスの統合 (2/3)



<https://www.z-holdings.co.jp/integrated-report/vision/growth/>

ZHD事業戦略におけるサービスの統合 (3/3)

フルファンネルで支援できることに加え、
ユーザー一人ひとりに最適な提案をする「1:1」のマーケティングを実現



<https://www.z-holdings.co.jp/integrated-report/vision/growth/>

アジェンダ

1. 背景

- ZHD事業戦略におけるサービス統合

2. 経由流通総額の可視化

- セッションデータ
- 非ログイン/ログインセッション
- アトリビューションモデル

3. 一休様とのデータ連携

- データパイプラインの管理
- AWS Lake Formationを利用したデータ授受
- 従来の連携手法との比較

4. 導入効果と今後の展開

送客実績の可視化の課題

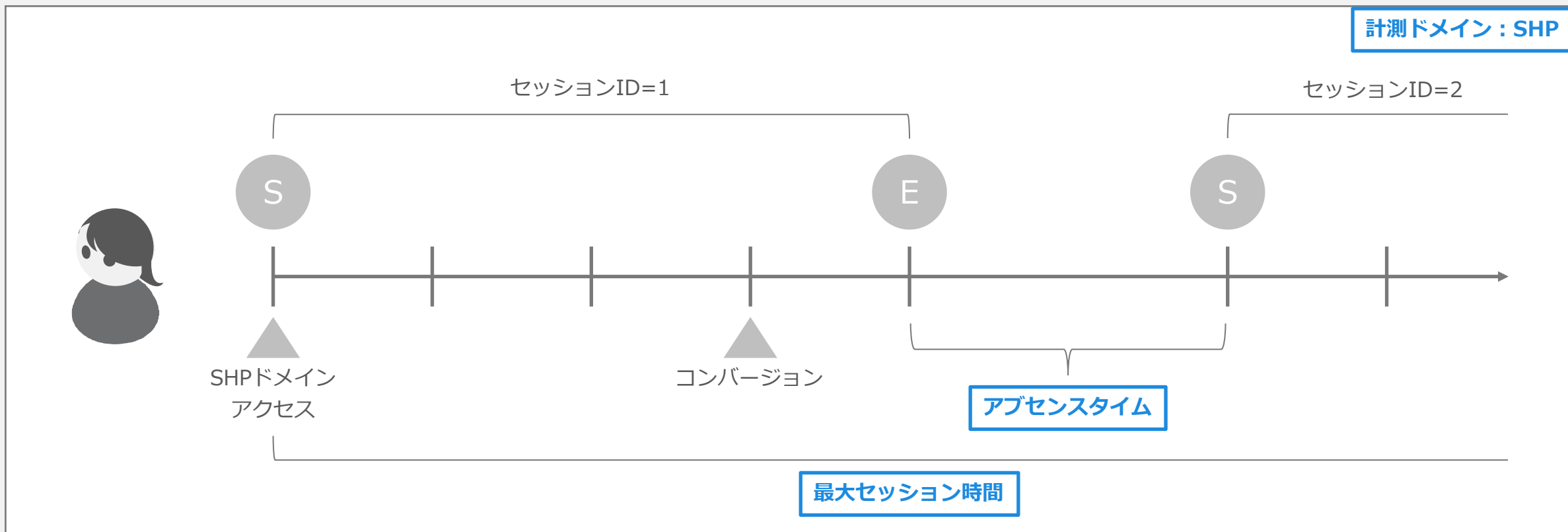
開始時の課題

- 貢献度を計測する定義やツールがサービスで異なる
- サービス横断で貢献度を正しく比較できない

解決するための取り組み

- サービス共通定義のセッションデータを作成する
- 取りうるすべてのチャネルを集計し貢献度を評価する

セッションデータ



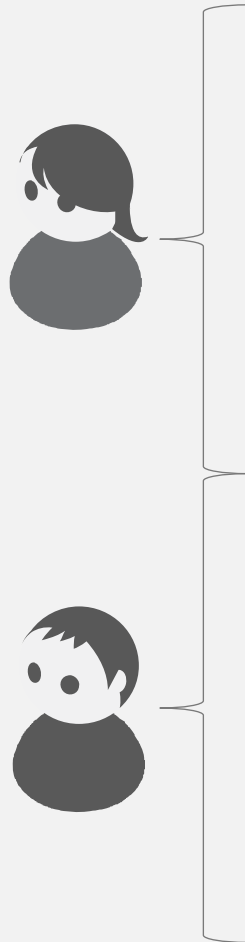
調整可能な条件

- ・計測ドメイン
- ・アブセンスタイム
- ・コンバージョン時間

共通条件

- ・デバイスまたぎあり
- ・非ログイン/ログインまたぎあり
- ・日またぎあり

非ログイン/ログインセッション

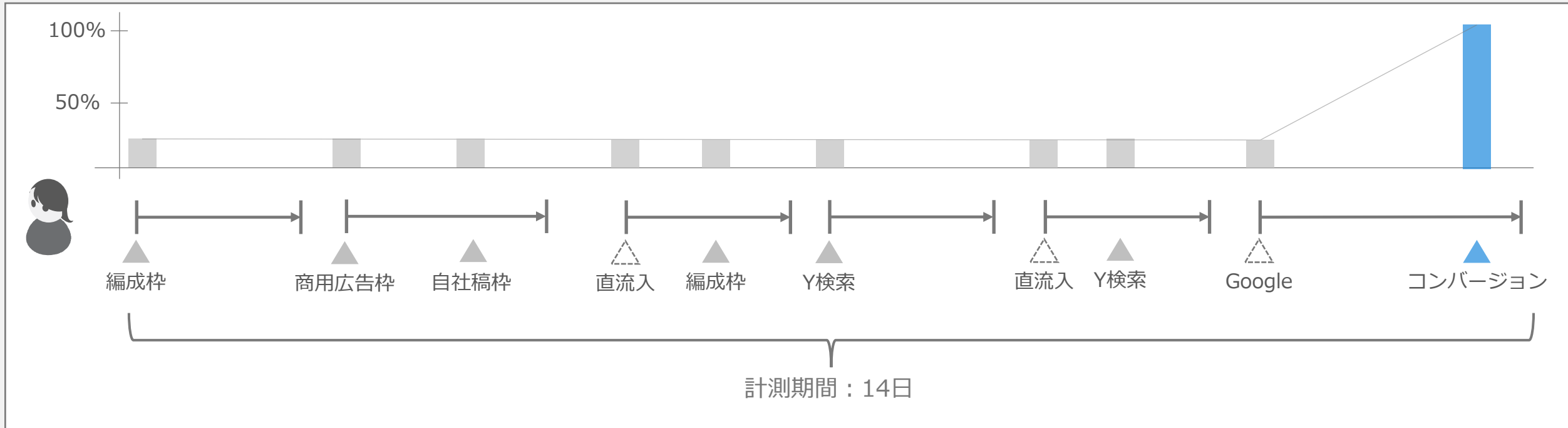


Device	非ログインID	ログインID	TimeStamp	SessionId
PC	aaa		1000	XXX
PC	aaa		1001	XXX
PC	aaa		1002	XXX
APP	bbb		1100	YYY
APP	bbb		1101	YYY
APP	bbb		1102	YYY
PC	ccc	AAA (補完)	2000	ZZZ
PC	ccc	AAA (補完)	2001	ZZZ
PC	ccc	AAA	2002	ZZZ
APP	ddd	AAA	2101	ZZZ
APP	ddd	AAA	2102	ZZZ
APP	ddd	AAA	2103	ZZZ

- 一次的にセッションデータを非ログインIDで作成
- 同一セッションでログインしている場合はログインIDを補完
- 最終的なセッションデータを補完済みログインIDで作成

14日イーブンモデル

一休様への送客実績を同一ロジックで計測のためアクセスログを授受



- Time DecayモデルやLogistic Regressionモデルとの比較検証を実施
- 最終的にコミュニケーション上の目的でシンプルなモデルを採用

アジェンダ

1. 背景

- ZHD事業戦略におけるサービス統合

2. 経由流通総額の可視化

- セッションデータ
- 非ログイン/ログインセッション
- アトリビューションモデル

3. 一休様とのデータ連携

- データパイプラインの管理
- AWS Lake Formationを利用したデータ授受
- 従来の連携手法との比較

4. 導入効果と今後の展開

データパイプラインの管理

Airflowを中心に社内環境とAWSを統合

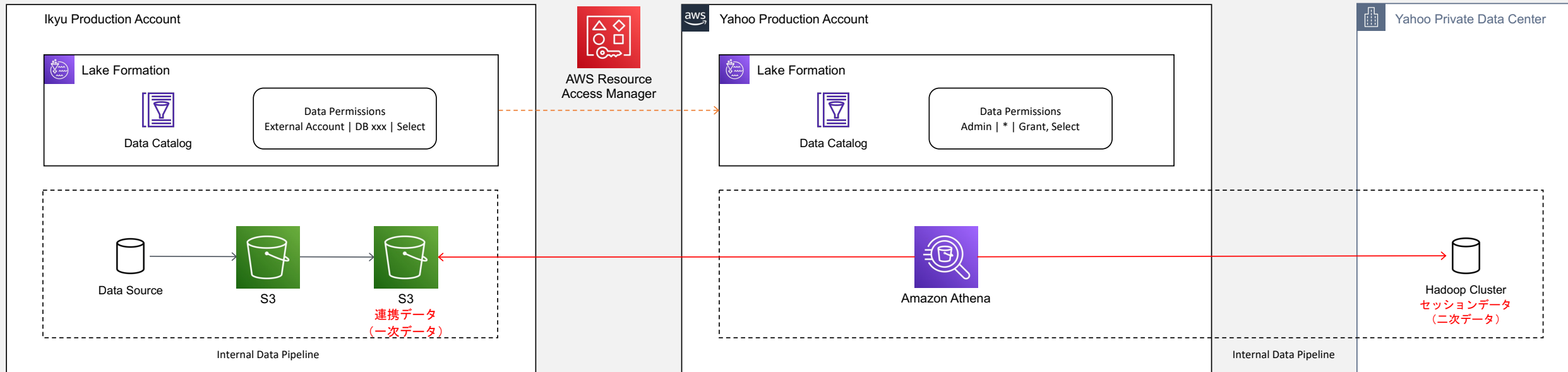
The screenshot shows the Apache Airflow web interface. The top navigation bar includes the Airflow logo, 'DAGs', 'Security', 'Browse', 'Admin', 'Docs', and 'About' menus. The current date and time are 2021-08-19, 17:19:43 JST (+09:00). The main heading is 'DAGs'. Below the heading, there are filters for 'All 35', 'Active 16', and 'Paused 19'. There are also input fields for 'Filter dags', 'Filter tags', and 'Reset', along with a 'Search:' field. The main content is a table of DAGs with the following columns: DAG (with a status toggle), DAG name, Schedule, Owner, Recent Tasks (with a task count and progress indicator), Last Run (with a timestamp and info icon), DAG Runs (with a run count and progress indicator), and Links (with various action icons).

	i	DAG	Schedule	Owner	Recent Tasks i	Last Run i	DAG Runs i	Links
	<input checked="" type="checkbox"/>	01_session_log_clf	03***	psc_insight	73	2021-08-18, 03:00:00 i	522	
	<input type="checkbox"/>	01_session_log_ext	03***	psc_insight	13	2021-08-15, 03:00:00 i	62	
	<input checked="" type="checkbox"/>	01_session_log_ya	04***	psc_insight	21	2021-08-18, 04:00:00 i	523	
	<input checked="" type="checkbox"/>	02_channel_cv_clf	03***	psc_insight	133	2021-08-18, 03:00:00 i	522	
	<input checked="" type="checkbox"/>	03_attribution_model	011***	psc_insight	38	2021-08-18, 11:00:00 i	506	
	<input type="checkbox"/>	04_impression	012***	psc_insight		2020-04-09, 12:00:00 i	9	
	<input checked="" type="checkbox"/>	05_attribution_model_middle_lp	011***	psc_insight	13	2021-08-18, 11:00:00 i	506	
	<input checked="" type="checkbox"/>	06_attribution_model_kpi_manage	011***	psc_insight	10	2021-08-18, 11:00:00 i	39	

- Kubernetes上にコンテナベースのAirflow環境を構築

AWS Lake Formationを利用したデータ授受

連携データをAmazon Athena経由で取得しセッションログを作成



- 社内ストレージに連携データ（一次データ、アクセスログ）は保持しない
- 加工済みデータ（二次データ、セッションログ）のみを保持する

従来の連携手法との比較（1/2）

	従来手法（SFTP）	AWS Lake Formation
メリット	<ul style="list-style-type: none">既存の仕組みであるため学習コストが低い	<ul style="list-style-type: none">連携元のETL開発工数が小さい※1オープンソースのAirflowプラグインが存在する※2連携先のストレージに一次データのコピーが取られない連携データに対して詳細なACL設定ができる
デメリット	<ul style="list-style-type: none">連携元のETL開発工数が大きい連携先のストレージに一次データのコピーが取られる連携データに対して詳細なACL設定ができない	<ul style="list-style-type: none">ACL設定の学習コストが高い

※1 すでに連携元がAWS（S3、Glue、Athenaなど）を導入済みの場合5人日程度

※2 オープンソースの実装を利用できるため効率的にAirflowへの取り込みが可能

従来の連携手法との比較（2/2）

一次データの保持が不要および詳細なACL設定が可能な点は強靱な個人情報保護を実現する上で大きなメリット

The screenshot shows the official website of the Personal Information Protection Commission (PPC). The page title is "令和2年改正個人情報保護法について" (Regarding the 2022 Amendment to the Personal Information Protection Act). The main content area contains the following text:

令和2年3月10日に第201回通常国会に提出されました「個人情報の保護に関する法律等の一部を改正する法律案」は、令和2年6月5日の国会において可決、成立し、令和2年6月12日に「個人情報の保護に関する法律等の一部を改正する法律」が公布されました。

円滑な施行へ向け、引き続き、ガイドライン等の検討を行ってまいります。

- ▶ [「個人情報の保護に関する法律等の一部を改正する法律」の概要等について](#)
- ▶ [「個人情報の保護に関する法律等の一部を改正する法律」の施行日について](#)
- ▶ [改正個人情報保護法の一部施行に伴う法定刑の引上げについて](#)
- ▶ [政令・規則・ガイドライン等の整備に向けた検討の状況について](#)
- ▶ [（参考）個人情報保護法 いわゆる3年ごと見直しについて](#)

<https://www.ppc.go.jp/personalinfo/legal/kaiseihogohou/>

アジェンダ

1. 背景

- ZHD事業戦略におけるサービス統合

2. 経由流通総額の可視化

- セッションデータ
- 非ログイン/ログインセッション
- アトリビューションモデル

3. 一休様とのデータ連携

- データパイプラインの管理
- AWS Lake Formationを利用したデータ授受
- 従来の連携手法との比較

4. 導入効果と今後の展開

導入効果と今後の展開

✓ ビジネスアジリティの観点

- AWS Lake Formationを利用したデータ連携は従来手法と比較して小工数で対応可能※

✓ データガバナンスの観点

- 一次データ不要および詳細なACL設定が可能な点は個人情報保護の観点で大きなメリット

✓ 今後は事業の推進に伴い連携対象を拡大予定

※プラバシーポリシーや契約の調整コストも含めた工数で50%以上削減

End