

職務経歴書（スキルシート）

基本情報

項目	詳細	近影
Name	Akira Iwasaki	
Location	Minato Mirai 21 / Yokohama	
Education	Nagoya University	
Company	株式会社DataMarket	
Qualification	情報セキュリティスペシャリスト 基本情報処理技術者 等	
	Data Engineering Study #4 「データ分析基盤の障害対応事例LT祭り」	
Media	Data Engineering Study #23 「Data orchestration 特集」 等	

Article・SNS

項目	詳細
X(旧twitter)	https://twitter.com/sista05
LinkedIn	https://www.linkedin.com/in/datamarket
Qiita	https://qiita.com/sista05
Zenn	https://zenn.dev/datamarket
Note	https://note.com/datamarket
Medium	https://medium.com/@akisista05
YOUTRUST	https://youtrust.jp/users/datamarket
Facebook	Akira Iwasaki
Wantedly	Akira Iwasaki
Docswell	Akira Iwasaki

Contents

- 自己紹介
- スキル
- 活動期間
- 継続中の開発案件

- 複数プロジェクトのデータ基盤開発・プロジェクト推進 (F社)
- 大手自動車製造会社のナレッジマネジメントシステムのPoc開発 (H社)
- 過去の開発案件
 - 中小企業向けビジネスチャットアプリ (Chatwork株式会社 様)
 - 大手ネット印刷・広告・運送企業 (ラクスル株式会社 様)
 - 大手家電量販店の会員情報移行・DX推進 (U社)
 - ECサイトのCRM環境(Salesforce CDP・MC)構築 (U社)
 - 顧客獲得プロセス可視化サイト構築、広告戦略・分析 (C社)
 - 大規模視聴データの移行支援 (S社)
 - リーガルテック企業の分析基盤構築 (L社)
 - 大手美容サイトデータ基盤構築 (R社)
 - 旅行代理店の新規顧客管理システム (L社)
 - mobility事業 (R社)

自己紹介

データエンジニア・データアナリスト。主にエンタープライズのデータオーケストレーションやデータモデリングなどを中心に仕事をしています。詳細は会社HPや記事などをご覧ください。

スキル一覧

▶ スキル表

Skills • Tools

Programing Language	Python / Go / Ruby / Perl / PHP / TypeScript / C / C++	
Data Enginnering	Workflow Engine & Data Orchestration	Digdag / Treasure Data / AWS Batch / Rundeck / Airflow(Cloud Composer / AWS MWAA) / Dagster
	Data Pipeline & ETL Tools	dbt / Fivetran / Appflow
Marketing Automation		Salesforce CDP・MC / Marketo / Pardot
Data Analytics	BI Tools	Looker / Tableau / GoodData / Data Studio / Metabase / Re:dash
Data Analysis Tools		Pandas / NumPy / Scikit-learn / TensorFlow
Big Data Technologies	Amazon EMR (Hadoop/Spark) / Cloud Dataflow / Treasure Data Snowflake / BigQuery / RedShift	
Data Governance	Cloud Data Catalog / Apache Atlas	
Security & Authentication	Cloud DLP / Cloud Data Catalog / Cloud IAP / AWS Cognito / Auth0 / IAM / KMS / Stackdriver Logging / Stackdriver Monitoring / DataDog	

Skills • Tools

DevOps & DataOps Tools	Git / GitHub / Bitbucket Github Actions / CircleCI Docker / Kubernetes / Terraform / CloudFormation / Ansible
AWS	Cloud Front / API Gateway / Lambda / ALB / ELB / EC2 / ECS / Route53 / Elasticsearch Service / RDS (MySQL / PostgreSQL) / Aurora / DynamoDB / ElastiCache (Redis) / SageMaker / SQS / SNS / SES / Redshift Spectrum / Cloud Watch / VPC Peering / Amazon DMS
Cloud Platforms	GCE / GAE / Cloud SQL / GCS / Datastore / FireStore / AI Platform / Cloud Pub/Sub / Cloud Functions / Cloud CDN
GCP	

活動期間

[!NOTE] 公開許可を得ている企業様は名称を出しています

```
gantt
    title 活動期間
    todayMarker off
    section H社
        RAG :2024-06-01, 60d

    section Chatwork
        Dagster導入 :2023-09-01, 270d

    section raksul
        Dagster導入 : 2022-09-01, 300d
        データ利活用効率化推進 : 2023-07-01, 300d
        データモデリング対応等 : 2023-07-01, 300d
    section U社
        家電量販店 : 2022-10-01, 210d
        ECサイトCRM環境構築 : 2022-03-01, 270d

    section C社
        顧客獲得プロセス可視化サイト構築 :2021-07-01, 400d

    section S社
        データ移行 :2021-07-01, 150d

    section L社
        リーガルテック企業の分析基盤構築 :2020-04-01, 460d

    section R社
        大手美容サイト : 2019-08-01, 240d

    section L社
```

旅行 : 2019-03-01, 120d

section R社

mobility : 2018-12-01, 90d

継続中の開発案件

複数プロジェクトのデータ基盤開発・プロジェクト推進 (F社)

期間: 2024年7月 - 継続中(業務委託)

レコメンドシステムの開発 snowflake、

Data Agentの開発

プロジェクトマネジメン都

期間: 2024年7月 - 継続中(業務委託)

データETL、1st Partyデータの活用、レコメンドエンジンやRAGなどを扱う統合的な複数プロジェクトの大規模データ基盤の、特にDagster部分を中心的に対応している。開発としては、データパイプラインの構築、システム保守、テストなどを担当。また、データチームのサブリーダとして、プロジェクト管理、後進の育成、開発ノウハウの伝授も行なっている。

大手自動車製造会社のナレッジマネジメントシステムのPoc開発 (H社)

期間: 2024年6月 - 2024年8月(業務委託)

RAGを用いた社内情報検索システム構築のPocフェーズの支援でPJ途中より参画。OCRでドキュメントの文字抽出、ドキュメントデータ構造の整備、プロンプトエンジニアリング、などMLエンジニアの支援を担当。

過去の開発案件

中小企業向けビジネスチャットアプリ (Chatwork株式会社 様)

期間: 2023年9月 - 2024年5月(業務委託)

dbt CloudからDagsterへの移行対応

データ分析基盤に複雑性・多重度を追加し、スケールアップを図る目的でDagsterを導入した。[\(関連記事\)](#)

- dbt Cloudの移行先としてDagster環境をAWSにterraformで構築した。その他、CI/CDや処理の自動化、AWS運用コスト削減の仕組みを導入した
- dbtで顧客データのデータモデリングに対応し、同時にsingular/unitテストの対応やdbt elementary導入によるデータ品質改善施策を採用した
- terraformでsnowflakeにユーザデータのデータマスキング処理を施した

大手ネット印刷・広告・運送企業 (ラクスル株式会社 様)

期間: 2022年9月 - 2024年4月(業務委託)

部門横断的なデータETLを実現するパイプラインツール、およびデータ分析利用に関わる各種自動化環境を提供し、全社的なデータオーケストレーションを推進した。

データオーケストレーションツールDagsterの導入

データオーケストレーションツールとしてDagsterを導入し、データアナリストやデータサイエンティストに対して効率的にデータ利活用を進められる基盤を構築した([関連記事](#))

- 従来困難であった本番環境を含めた各種環境における開発環境データを流用したテストを実現し、作業の効率化と品質の向上を果たした
- 柔軟なデータパイプラインの表現により、シリアルな処理をパラレルに実行してボトルネックを減らし、全体的に処理効率を改善した
- dbtとの連携により、処理結果だけでなくデータリネージュやデータアセットなどのデータリソースのUI・UXを改善し、総合的なデータの可視化に大きく貢献した

データ利活用効率化推進

社内データの再整理、最新データ構造を自動でドキュメント化して社内展開する仕組みや、デプロイの自動化といったDataOps活動を推進した

- CI/CDをJenkinsからGithub Actionsに替えることで、コードデプロイをマニュアル操作から自動化し、作業コスト削減に貢献した
- ブランチ命名規則、プルリクエストルール整備、コードのmerge、リソースのデプロイ紐付けなどの開発ルールを整備した
- データテーブル情報のドキュメント化・社内展開、データ分析作業のためのアドホッククエリ整備など、各部署の分析作業に貢献した

データモデリング、データガバナンス対応

ECデータや人事情報などのデータテーブルを整理し、データモデリングとデータ品質向上に貢献した

- dbtのベストプラクティス、dbt evaluatorなど推奨ツールを用いてデータレイク・データウェアハウス・データマート構造を構築しデータテーブルを整理した
- 品質向上施策として、pk/fkテスト、sqlfluffの自動化、elementary、Data Catalogやdbt docsなどへのメタデータ反映など対応した
- データポリシータグを用いたグループ、部署間での列レベルのアクセス制御を適用し、データアクセス管理に貢献

大手家電量販店の会員情報移行・DX推進 (U社)

期間: 2022年10月 - 2023年4月(業務委託)

大手家電量販店のDX化推進として、新旧顧客データのマイグレーション、および統合会員管理システムにおける認証・認可プラットフォームのシステム設計から開発まで担当した。具体的には、新旧顧客データのID統合・マイグレーションにおけるデータ設計/データ基盤のシステム設計開発/ビジネスメタデータのデータカタログ設計/顧客情報の分析環境/データガバナンス環境を構築した。

統合基盤の設計と開発

顧客情報管理システムを全面的に刷新し、新旧環境間におけるマイグレーションと、Salesforceとの有機的な連携のための環境を構築した

- 顧客情報のCRUD操作を行うCMSサーバとAuth0の外部APIを連携させ、Custom DatabaseとAPI Gatewayと連携し認証・認可の仕組みを構築した
- ReactベースのCMSサーバ開発と、API Gateway + Lambda + RDS ProxyによるCRUD構成を構築した
- 旧環境の顧客情報を移行するためにauth0の統一IDを発行し、新旧環境で齟齬なく移行可能な仕組みを構築した
- 顧客情報を蓄積したDBとその情報をシームレスにSalesforce Service Cloudに送信するCDCを構築した

保守・管理

IaCによる環境情報の一元管理と、プロジェクトのルール策定に協力した

- 開発管理としてBacklogの運用方式の提案と、DevOps体制を構築するにあたって開発ルールを整備した
- 複数開発チームにおける総合的なCI/CD環境を構築した
- 統合データ基盤の構成はCloudFormationにより構築し、構成の一元的な管理と運用保守に関しての方針とドキュメントを整備した

データマネジメントとデータガバナンス環境の構築

Salesforceの機能を補うために、リアルタイムなデータ同期するシステムと、ビジネス分析を強化するためにAthenaなどを用いたデータ分析環境と、それらを統合管理するためのメタデータ環境を構築した

- Salesforceだけでは柔軟なクエリや複雑な統計分析が困難であるため、それらを補うためのAthena分析環境を構築した
- 顧客情報のデータ構造の見直しと、データ間の関係性や重要度、セキュリティ情報といったメタデータ設計を行った
- ビジネスサイドのメタデータ管理としてApache Atlas環境を検証。データ分類、データライフサイクル、コンプライアンス用途の設計を担当した

ECサイトのCRM環境(Salesforce CDP・MC)構築 (U社)

期間: 2022年3月 - 2022年9月(業務委託)

大規模ECサイトにおけるSalesforce CDP・MC導入を支援。CDPでは主にデータ取込と変換、システム監視フローの構築を担当。MCではカスタマージャーニーの設定を担当。システム全体の要件定義も一部担当。また、sfdxによるIaC・CI/CDを導入。その他、Tableauによるシステムパフォーマンス・顧客データ可視化、大規模データ・性能速度検証なども担当。

CDP・MC構築

Salesforce CDP・MC環境構築、特にCDPのサービス正常性を維持するための監視システム構築を中心に担当した。

- CDP構築ではデータ取込・変換、データマッピング、データモデリング、セグメント作成・出力までの一連の流れの構築を担当した
- システム状況監視をプロセス自動化フローで実現、MCの検証アクティビティと組み合わせてシステム正常性を監視し、一部自動で修復・通知するシステムを構築した
- MCではシステム監視の他に、LINE通知などデータエクステンションの作成を担当した

要件定義・環境情報のIaC化

Salesforce CDPのシステム運用設計における要件定義を担当。また、システム構成情報をコード化・ドキュメント化しインフラ管理と保守コスト削減に貢献した。

- sfdxによるIaC・CI/CDでの構成情報管理を提言し採用。構成情報の整合性を図ると同時に構成管理コストを大幅に削減した
- 一部最新のシステム構成については、Salesforceの開発チームと並走しながら情報連携し、Salesforce全体の品質向上に貢献した
- 保守作業は作業と並行してドキュメント化するなど、ナレッジの共有を意識しながら作業を進めた

顧客・システム情報・速度性能の可視化

顧客・システム情報可視化のためのBI環境構築を担当した。また、ダッシュボードの構築やデータ突合など整合性テストも実施した。

- エンゲージメント予測、ファネル分析チャート、内部システム監視(システム健常性・性能速度)のダッシュボード構築を担当した
- Tableau OnlineのSalesforce CDPデータや、AWS Athenaと連携したシステム環境構築を担当した
- 一部旧システムからのデータ移行に伴い、抽出データや定義情報のバリデーションチェックやデータインポート後のデータ突合確認などデータ整合性の向上に勤めた

顧客獲得プロセス可視化サイト構築、広告戦略・分析 (C社)

期間: 2021年7月 - 2022年7月(業務委託)

リサーチ・マーケティング専門会社において、顧客獲得プロセス全体の可視化サイトのインフラ構築・CI/CD・ELT・データマート構築の設計・開発を全面的に担当。プロダクトマネージャとして製品知識の提供や他社比較分析、および開発環境全般におけるルール策定やドキュメント整備も担当した。

顧客獲得プロセス可視化ソリューション開発

Salesforce・Marketo・Google Analyticsのデータを統合した独自のインタラクティブなリードファネル・営業リーダーボード・KPIおよび営業活動のダッシュボードサイトを構築した。構築にあたっては、管理、統一された開発手法のなかった会社に一般的なシステム開発手法を導入したいというご希望に応えた。

- システムの技術要件を全面的に検討・決定した。フロントエンドにはReact Chakra UI、バックエンドにはRuby on Rails、その他、CI/CDにGithub Runner、構成管理にTerraformなど一通りの一般的なシステム構築における手法を提言した
- ETL環境として、データ統合にはFivetranを、データ変換にはdbt Cloudを、DB/DWHとしてSnowflakeを用いて従来のデータ構成をモデリングし直した
- 要件定義では、営業用の独自用語を開発の一般的な用語に落とし込み、業務要件をヒアリングして業務フローを作成し、業務における冗長な作業や散逸したリソースを一元管理し、外部仕様として開発の技術選定を行い、必要な技術を決定した

開発環境の整備全般

アジャイル開発を取り入れたプロジェクトやその管理手法、Github導入などDevOps開発手法を取り入れた。

- プロジェクト・ドキュメント管理にJira・Confluenceを採用。slackやGithubなどモダン開発ツールを採用した
- TerraformによるInfrastructure as Codeやgitlab runnerによるCI/CDで構成管理と継続的デリバリーを提供了
- dbt Cloudのドキュメント機能などを自動でクラウドに保存するように、自動でドキュメンテーションを拡充、保守する仕組みを構築した

広告戦略・グロースのためのデータ分析

新製品のBIプロダクトのPdMとして売り上げ拡大に貢献するためのカスタマージャーニーマップ作成や、グロースのための施策としてプロダクトとのフィードバックループを作成した

- ファネルとコホートを強く意識した分析を行い、事業解像度を上げるために注力した
- 分析手法導入にあたってはデータのサイロ化、データガバナンスなどのデータ基盤の根本となるところから整理し、導入における基礎を築くのに貢献した
- 分析手法実現の背景には可視化ソリューションでも利用したSnowflakeやFivetran、dbtなどのモダンなETLツールの導入を役立てることができた

大規模視聴データの移行支援 (S社)

期間: 2021年7月 - 2021年11月(業務委託)

運用費削減のため、Treasure DataからBigQuery/Digdag構成へデータとジョブワークフローをマイグレーションするプロジェクトを支援した。Treasure DataのTD関数とBigQueryの互換性調査や対応、汎用モジュールやワークフローの作成、テスト項目の作成などに従事した。

- Treasure DataのワークフローをDigdagに移植した。移植にあたりBigQueryとHadoopのMapReduce処理との差異などにも留意した
- コスト削減のために最適な処理の置き換えや、Treasure DataとDigdagで共通利用可能な汎用モジュールなども構築した
- テストケース作成や異常時の通知設計、移植後のデータ整合性チェックなどの詳細の確認も行った

リーガルテック企業の分析基盤構築 (L社)

期間: 2020年4月 - 2021年6月(業務委託)

リーガルテック企業におけるデータ連携基盤の構築を担当。基盤設計/開発/スケジュール調整/ドキュメント作成、平時の運用や障害対応まで幅広く対応した。また、パラメータ抽出/顧客データ解析/レポート作成や、機械学習基盤の整備と精度評価指標の提案、等広くデータ分析業務に従事した。

ログ収集基盤設計と構築

ログパラメータが300を超える複雑構成のウェブアプリログを、Fluentdログ収集基盤を大きく刷新することでログ解析可能にした。また、基盤をフルマネージド化することで構成を簡略化、スケーリングを容易にした。これにより従来の収集基盤と比較し年計200万円以上のコスト削減に寄与。併せて分析/運用管理/障害対応コストも削減した。

- firelens、fluentbitを利用したログフォワードの構成の簡略化、モダン化、使用リソースの削減

- ログ格納にCloud Functionsを介することで、複雑なログスキーマ変更にフレキシブルに拡張対応仕組みを作った
- Cloud Pub/Sub、Cloud Dataflowによるフルマネージド構成による運用、保守性の向上、重複/欠損データの削減
- Cloud DLP(Data Loss Prevention)による秘匿情報のマスキング
- Cloud Monitoringおよび拡張機能による基盤環境監視とアラート通知実装

データ処理基盤設計と構築

データのバッチ処理に関わる組織のガバナンス/リソース管理/セキュリティを考慮しつつ、データ処理基盤を再設計、再構築した。これにより従来のバッチサーバ構成よりも著しく耐障害性/スケーラビリティを向上し、またセキュリティに配慮してデータを取り扱う環境を整備した。

- 耐障害性向上のため、Digdag ServerとAgentを分離しAutoScalingで負荷分散、DBにAurora PostgreSQLを使用
- バッチジョブをFargateにすることで並列同時実行を可能とし、処理速度とスケーラビリティを向上
- SSM ParameterStoreをKMSにより管理することで、組織ガバナンスを考慮した構成を採用
- CircleCIによDockerImageによる自動デプロイ、タグ設計によるバージョン管理でデグレ防止を考えた運用
- github actionsを利用したPull Requestの自動化

データ分析とレポーティング

アプリ利用データの解析/アドホック分析/日々のレポーティングや、Salesforce/MarketoなどCRM、MAツールの顧客データ分析を担当した。また、部署間のデータを連携しデータ利用を促進することを進言。DX文化の浸透に貢献した。

- レポーティング、データ分析、ビジュアライゼーションのプロトタイピングとデザインを担当
- 自動化プロセスを効率化（レポート配信、ダッシュボードの作成、データ管理）
- データベーススキーマへのデータマッピング、データクレンジングと処理、データ抽出スクリプトの作成、データ変換のプログラミング、APIを介した複雑なデータ問題の調査
- データソースに構築されたLookerダッシュボードやレポートのパフォーマンスチューニング
- CRM、MAツールなどで顧客セグメントに活用するためのデータ連携の設計、開発

機械学習基盤の精度改善

契約書の自動レビュー(条文検索/比較/マッチング/構文解析/画像解析 等)を実現する機械学習基盤の、特に予測精度改善に貢献。機械学習における継続的デリバリー/自動化のモニタリング/自動化プロセスを向上させた。

- 機械学習の定期的なモデル検証と最適化のワークフローを提供
- モデル精度モニタリングに関する環境整備
- Human In The Loop(HITL:学習に人手を介在させることで継続的に精度向上や維持をするフィードバックループ)の仕組みの導入
- ビジネス要件の理解と分析、予測モデルの構築、実験の設計、仮説の検証、統計結果
- RCT/DIDなどの基礎的な効果検証実験

大手美容サイトデータ基盤構築 (R社)

期間: 2019年8月 - 2020年3月(業務委託)

システムのEC2構築、運用設計、監視、CI/CD部分を担当。バッチサーバ、ログアグリゲータの構築と、それにまつわるシステム全般の設計管理を担当した。

- ・バッチサーバとしてdigdagを用いてECS/EKS上にビッグデータ用の分析環境を構築。また、ログアグリゲータとしてtd-agentを用いた。いずれも負荷に応じてAutoscaleするように構成した。
 - ・システム運用に必要なインフラで使用しているミドルウェアのバージョンアップ方針、データのバックアップ方針、システム監視項目の洗い出し、ログデータの管理方針を設計、検討した。
 - ・CI/CDパイプラインとして、AutoScaleも含めたEC2デプロイ方式を設計、構築を担当。Terraform、EC2 Image Builder、Jenkinsによるシェル実行などでそれぞれシステムのレイヤーに適したデプロイ方式を提供した。
-

旅行代理店の新規顧客管理システム (L社)

期間: 2019年4月 - 2019年7月(業務委託)

"旅行代理店の新規顧客管理システムの立ち上げに従事。システムのマイクロサービス化に貢献。ログ集積基盤の構築と、サービスのコンテナ化を中心的に担当した。

- ・マイクロサービス環境構築 PHP7、Laravel、MySQL、Redisを利用したAPI開発環境をDockerで構築し、APIGateway、Lambda、DynamoDBと連携させたマイクロサービス環境を構築した。
 - ・ローカル開発環境構築 上記のマイクロサービス環境を擬似的に実現するために、API GatewayやLambdaをSAM CLI、DynamoDB Local、Golangでサーバレスアプリケーション化し、開発チームにローカル開発環境として提供した。また、docker for macが持つ潜在的な速度問題にも対応した。
 - ・ログ集積基盤 APIへのアクセスログをkinesisに集積し、s3、Elasticsearch、slackへと送信するログ収集基盤を設計、構築した。集積したログはathena検索に対応し、収集基盤は速度、並列化に対応するためgolangで構築した。また、アラート状況の視覚化や、基盤自体はcloudformationで自動で構築できるように対応した。"
-

mobility事業 (R社)

期間: 2018年12月 - 2019年3月(業務委託)

モビリティサービスプラットフォームの設計、構築を担当。約2万ユーザの緯度・経度をモバイルで管理・把握し、位置情報をpysparkで処理するETL環境を構築した。また、pysparkによる実処理部分も一部担当した。環境構築にあたり自動化の仕組みを整えた。

- ・bastion経由のpublic、private環境の他、認証にCognito/LDAP、バッチ環境としてRundeck/RunCommand、ネットワークはAPI Gateway、NAT Gateway、Route53等を構築した。また、Lifecycle ManagerやSSM Session Managerなど、コスト削減に寄与する新規技術も取り入れた。
- ・ansibleでLDAP設定を一律管理する仕組みを提供。CodeCommit/CodeBuild/CodeDeployを経由するCodePipelineを構築。基本環境、ビッグデータ環境はCloudformationで作成し、再現性、リソース/使用時間管理に役立てた。
- ・ビッグデータ分析基盤AWS IoTの情報をkinesis Data Streamsを経由、またはS3に一時的に格納し、EMR(pyspark)を用いてs3、RedShift、Redshift Spectrumに処理データを格納する基盤を構築した。pysparkの処理にあたって一部実処理部分のコーディングも担当した。