Summary

This document provides an update on the CDO Business Intelligence (BI) team vision, the status of CDO BI metrics, and a summary of key dependencies and challenges faced by the BI team. The document also presents a forward looking H2 roadmap and seeks feedback on roadmap prioritization from the product team.

Team Overview

When CDO was formed in late 2018, the organization did not have a Business Intelligence team. Instead, the organization depended on Smart Home (SH) BI and Alexa BI teams (renamed as AXD BI), as well as tools and metrics built by engineering teams using PMET, Atocha, and ADAC dashboards. In 2020 the CDO organization launched a centralized BI team to obtain greater control over our own BI destiny. Currently we have two BIS plus a summer intern: one L6 Business Intelligence Engineer (BIE), one L5 Data Engineer (DE), and one intern DE.

The establishment of a dedicated BI discipline within CDO follows a pattern throughout the broader Alexa organization. In 2015/2016, when small,. continues to supportalldisciplines H, Each BI team is responsible for establishing a data pipeline and completing the data engineering necessary to enable dashboards/metrics to be quickly generated by its customers.

Customers and Vision

CDO BI customers are CDO PM-T/TPMs/SDMs, Leadership, SDEs & Scientists and other BI teams.

The primary need of the CDO BI customer is to **obtain** **answers to business questions using data** in an efficient, accurate, reliable and self-served manner. Customers are looking to find answers for the following reasons 1) to prioritize the right milestones or deploy engineering resources 2) to know the extent of the defects in the service 3) to measure the success of the project 4) to define goals 5) add data points to justify the funding of new projects.   
  
The customer-facing products that CDO BI is building to support these customer needs are:   
1) **QuickSight Dashboards/Excel decks:** For pre-defined and standard metrics to have access to everyone on their fingertips. It is the standard way to see the output across the Amazon.  
2) **Query Bank for FAQs:** Standard queries often needed to dive deep and get the raw data for frequently asked questions.  
3) **Simplified aggregated tables:** For deep dives and Ad Hoc questions which are not present in dashboards already.

To create these BI products requires the following infrastructure investments: 1) Instrumentation by engineering teams to capture metrics of business interest and store the metrics in a repository such as S3, DynamoDB, or Datamart (Owned by Engineering), 2) A data infrastructure that comprises a) Data pipelining to extract the data from engineering repositories and b) host the data in a cluster such as Redshift with a schema that can be queried using SQL (Owned by Data Engineering), 3) A data front-end that comprises a) Meaningful tables with applied business rules and context to display the data in an insightful way and b) Tools that enable self-service access to the data (BIE & BIA). We elaborate on this data journey below.

Situational Assessment

In 2020, BI team created the part of the foundational BI infrastructure such as (Tool glossaries can be found in appendix 3): AWS account with Redshift, QuickSight and S3 services & internal tools such as Datanet, EDX, Cradle and Hoot to source data from other BI teams to help BI to create the dashboard for available Alexa adoption and engagement metrics for teams like Registry, LAP and Targeting. We have already created 5 QuickSight dashboards. More details can be found in Appendix 1.However, teams like State, Multi agent, Multitasking, Context, DAAPI etc., do not have the data in any BI stores because the data is not yet instrumented by Tech team and publish them to a BI consumable format(usually S3 or EDX). Few teams are in flight to publish data to DataMart and BI team is working on onboarding to SSDG (Self-Service Dataset Generator) platform to source data for these teams directly from DataMart.   
  
Below table provides the BI readiness of CDO product areas. We have identified 24 product areas and out of that 4 teams have a front-end dashboard implemented and 3 teams have WBR excel slides.

Out of 24 products, 14 of them don’t have any intakes, although not all 14 product areas have metric requirements since they are either relatively new e.g. Matter, Alexa Spatial Perception or have OE metrics which are sufficient for these teams e.g. History Store. Also, 5 of them don’t have a PMT owner to request the metrics from BI team. However, in the areas where we have

intakes (8 out of 24), for 3 areas we have an Intake but they don't have any key metrics instrumented by Tech team to S3/EDX , which is the big hurdle for BI team to create dashboards/WBRs for these remaining teams.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Product** | **Front-End Dashboard Implemented?** | **Weekly WBR (Manually or Auto Generated)?** | **Data published to CDO Redshift Cluster?** | **Key Metrics Instrumented by Engineering to S3/ /DataMart?** | **Number of Intakes in BI Queue** | **Metric/Deck Owner** |
| Orchestrator | No | No | No | No | 0 | Shiladitya |
| Cortex | No | No | No | No | 1 | TBH |
| Visual Multi-Tasking | No | No | No | No | 0 | TBH |
| Audio Multi-Tasking | No | No | No | No | 0 | Rohit |
| Alexa Assistants | No | In-Flight | In-Flight | In-Flight | 1 | Yaser |
| LAP | Yes | Yes | Yes | Yes | 1 | Matt |
| Matter | No | No | No | No | 0 | Matt |
| Response Orchestration | No | No | No | No | 0 | Tom / TBH |
| Capabilities Store | No | No | No | No | 0 | Matt |
| State Store | No | No | No | No | 0 | Guy |
| Registry Store | Yes | Yes | Yes | Yes | 0 | Gopi |
| History Store | No | No | No | No | 0 | Sekhar |
| Groups Store | No | No | Yes | Yes | 0 | Gopi |
| Relationships Store | No | No | No | No | 0 | N/A |
| Entity Resolution | No | No | No | No | 0 | None |
| Setup/Discovery | No | No | No | No | 0 | Gopi |
| Echo Spatial Perception | No | Yes | In-Flight | Yes | 4 | Brendon |
| Alexa Spatial Perception | No | No | No | No | 0 | Brendon |
| Targeting | No | Yes | In-Flight | Yes | 2 | Genna |
| I/O Decoupling | Yes | NA | In-Flight | Yes | 1 | Tom / TBH |
| Endpoint Enumeration/Query | No | No | No | No | 0 | Tyler |
| Endpoint Control | No | No | No | No | 0 | Tyler |
| DAAPI Primitives | No | No | No | No | 1 | Tyler |
| TPM Operations(SDE  Intake dashboard) | Yes | No | Yes | Yes | 1 | Ambika |

Challenges and recommendations:

**1) No Intakes from PM-T:** As we saw in the situational assessment, there 14 product areas with no BI intakes and 5 of them don’t have a Product owner. Hence, it is recommended to assign the product owners to each product area and product owners to create the BI intake to the generation of metrics as per the septs in appendix 2.

**2) Dependency on Tech team:** As we saw in the BI journey, if the data is not readily available from other BI sources, BI team is dependent on the Tech team to instrument the data, 16 product areas don’t have any instrumentation yet. BI team and Tech team should have scalable and sustainable solution to publish and consume the data. Currently, we are exploring the SSDG options as recommended by SAIF team who owns the Data Mart. I.e. the tech team will publish the data to DataMart and BI team will use SSDG to extract the data from DataMart to S3 and can have a Redshift spectrum on top of S3 to query the data.

**3) Limited DE bandwidth:** Currently, we have one L5 data engineer in the team, although they are working on building the infrastructure, we as a team still don’t have a solid solution to make the SDE team to publish the data to a BI consumable format with reasonable lead time (current time is more than 6 months). Having a L6 DE can own this part to have scalable and replicable design reduces the dependency and improve the lead time for consuming data from SDE teams.

**4) Privacy regulation:** We are also blocked on accessing Red data with our current infrastructure due to privacy regulations, this is impacting teams like Registry to do the analysis on attributes as the friendly name to the de dup the endpoints. To solve this challenge we are working towards enhancing our knowledge on how to handle Red data and exploring tools like Nexus and Red Cradle.

H1 2021 Highlights:

**MDX Data Engineering:** BI team on-boarded about 6 tables of worldwide data required for automating Targeting and I/O decoupling WBRs. Previously, the WBR was manually generated by PM/SDE and used to take roughly 3 to 4 hours per week. Now, this can be completely automated to save the time and manual efforts.

**Multi device UDR:** Alexa customers are measured on various types of defects i.e. friction, ETER and CPDR. Combining them a Unified Defect Rate (UDR) is created by Gale team. This metric is calculated only for single device customers and there is no tool to calculate the UDR for Multi device customers. To cater for MDX use cases CDO BI team writing a custom SQL query created a UDR metric for Multi device customers.

**GDPR Implementation:** Having the CDO BI Redshift cluster we started onboarding the new data tables to support multiple products like the Registry, ESP, Targeting, I/O decoupling etc. Storing these data we have to comply with the GDPR privacy regulation to delete the customer data if requested to delete by the customer. Hence, CDO BI implemented a Datanet solution to comply with Amazon wide GDPR guidelines.

H1 Lowlights:

**Multi Agent WBR**: In April, BI team was working on Multi Agent WBR since there were no cross-program-level analytics for Agents. BI team had worked with PM to understand the metrics requirement and worked with the Tech team to acquire the required data for BI consumption. The PM of Multi Agent WBR requested to publish the WBR by June 1st. However, while working with Tech team, we realized the data currently instrumented to DataMart by the team but it is not completely accurate, as few domains were not on-boarded to publish the complete data. Hence, the Tech team created a new design to gather accurate data and publish for BI team. Initially, we expected the project could be completed by June but with this change in scope, the timelines got postponed to early August to provide additional time for Tech team instrument the data to include information about all domains.

**CHR Data ownership:** In May, BI team was working with CHR and SH team to onboard the CHR data to CDOBI cluster. Currently, CHR data is consumed by SH team and they want to transfer the ownership for the CDOBI team as CDO is the rightful owner of the data. SH team had put together a new desing doc for CHR tech team to publish the data and CDO BI to consume the data by removing problem areas in current design i.e. like data publishing delay, bottleneck around upstream schema changes etc.However, after the design discussion on how to publish and consume the data, the project is placed on hold as the tech team is working on creating new CHR V2 API which will be replacing V1 API. Developing the new desing on V1 API now would be a throw away work. Hence, the project is on hold till further notice on the CHR V2 API.

Roadmap H2 2021

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Priority No** | **Product area** | **High level Requirements** | **Current Status** | **Work Stream** | **Comments** | **ETA** |
| 1 | Redshift Optimization | Potential Intern project | Yet to Start | DE and BIE | Intern will be joining July 2nd week | End of Q3 |
| 2 | Multi-Agent | The Team wants to create a WBR report to measure the adoption, usage, and impact of Agents 'Powered by Alexa' | Work in Progress | SDE, DE and BIE | We have collected requirements for MA and working with Tech team to publish the data to BI system. | End of Q3 |
| 3 | Wagner/DAAPI | The Team wants the ability to deep dive into the API data and derive insights from joining with other data set to develop the product better | Yet to Start | SDE, DE | Need to explore on how to get DAAPI data in BI consumable format and do the data engineering. | End of Q4 |
| 4 | MDX - Targeting | The Team wants to automate a WBR report to measure the adoption, usage friction metrics for customers impacted with Targeting use cases | Work in Progress | SDE and BIE | Working on automating the WBR for targeting | End of Q3 |
| 5 | MDX - I/O decoupling | The Team wants to automate a WBR report to measure the adoption, usage friction metrics for customers impacted with I/O decoupling use cases | Work in Progress | SDE and BIE | Working on automating the WBR for I/O Decoupling | End of Q4 |
| 6 | LAP | The Team wants to measure the success of the product and derive insights to improve the performance of the product. | Work in Progress | SDE and BIE | Created an initial dashboard working with PM to enhance in coming months. | End of Q4 |
| 7 | MDX - ESP | The Team wants to ensure the ML pipelines are performing well. Also derive insights on the performance of the ML models to improve the efficiency of the models. For matured MDX products, wants to improve the engagement and adoption to improve the DSI | Yet to Start | DE and BIE |  | End of Q4 |
| 8 | Intake Dashboard | The Team wants to enhance the CDO Intake dashboard to reduce the time for TPMs to share the report the data to Leadership | Yet to Start | BIE |  | End of Q4 |

**Appendix:**

Appendix 1: H1 2021 and 2020 look back

Since mid-2020, 1 Sr. BIE CDO BI team started building infrastructure necessary for the team to build dashboards and self-service mechanisms. I.e. we created ACDBI Redshift DW (RA3 4 nodes) and other Data pipelining tools. With the help of our Redshift DW and Data sourced from other teams Redshift DW, we were able to build 5 Quick Sight dashboards, 1 de-normalized table and 35+ AdHoc analysis, on-boarded 10+ tables to CDOBI cluster   
**Key Accomplishments**

* **MDX Data Engineering:** MDX Targeting team reached out to CDBI team to onboard a total of 6 datasets onto our cluster, to support targeting WBR. This would enable us drive more insights by joining targeting tables with already existing data in CDBI cluster.  
  4 of these tables belonged to Alexa and Smarthome teams and were on-boarded through Datanet jobs. For the remaining 2 tables, the data was residing in SSDG S3 and required establishing Redshift Spectrum on our end. Also these data was stored in different regions which was a big blocker, in order to support multi-region data requests, we devised a new solution by setting up 2 new Redshift clusters in EU and FE region and created respective tables in them for targeting and I/O decoupling transactional metrics. Having these data in Redshift will allow BI team to completely automate the WBRs and create dashboards to save manual efforts about 0.5 day per week per WBR.
* **GDPR implementation:** With scaling up the cluster storage and bringing in more tables to support multiple projects, CDOBI implemented a Datanet solution to comply with GDPR guidelines. Till now, we have been doing a full load of tables from the upstream teams, to make sure our cluster was compliant. Having this solution in place, eliminates the dependency on other teams so that we have more complete ownership of cluster compliance.
* **Multi Agent Analytics (MAA):** “Powered by Alexa” is a developer program to let 3P developer build their own voice assistants (agents) using Alexa’s stack (via MAP, the multi-agent processing architecture, that CDO leads), for a fee. Currently, there are no cross-program-level analytics for Agents today. Individual programs have their one-off dashboards (Walken, Alexa-Portal, AVS, and Cortana) and we do not have in-place measures of our customer or business impact with Agents. We want launch measures of customer usage, economic value from agents (e.g. DEV/EVE, compared to control cohorts on same device types), and agent quality (e.g. agent dialogs that end in failed utterances or non-purposeful “dead-ends”). To have these analytical measures we are creating excel based weekly business review report, which will be published on every Monday to the Multi Agent audience over an automated email. More details can be found in Wiki page: <https://wiki.labcollab.net/confluence/display/Doppler/Multi-Agent+Analytics>.**Work in Progress status:** Till date we have aligned on what should be the final metrics and what data should be engineered to support the metrics. Currently working on creating WBR with mock datasets since the data is not yet published by tech team (planned to publish by august) and a beta version of extracting data from DataMart to establish the pipelines. Once we have final dataset we will integrate the WBR with actual data.
* **Data onboarding Proof of Concept (POC) from AWS S3 (Storage service) to Redshift Database:** CDO as an org, we created a Redshift cluster in 2020 to bring and store the datasets from other sources to help feed the WBR/Dashboards. Most of the datasets were present in other BI sources as SmarthomeDB and AlexaDB, helped in the creation of dashboards in 2020 by using Datanet tool. However, for the planned dashboards in 2021, we don’t have data in other BI sources, which leads us to bring data from Non BI sources such as S3, SQS, EDX and DataMart etc. Various teams publish the data using various methods, in majority of use cases it ends up being stored in S3, once the data is present in S3, CDO BI team consume the data from S3 to Redshift for performing analytics. To consume the data from S3 to redshift as well, we have multiple approaches (described in SIM: <https://sim.amazon.com/issues/CDBI-83>). Hence we did proof of concept (PoC) on multiple approaches to know which is the optimal approach in terms of cost, effectiveness, maintenance and also allows to replicate to all future projects of similar nature. We concluded S3 >Cradle > EDX > Datanet > Redshift will the best suited approach for our use case.
* **Multi Device UDR:** Today, for Alexa customers we are measuring various types of defects i.e. friction, ETER and CPDR. Combining them a Unified Defect Rate (UDR) is created by Gale team. This metric is calculated only for single device customers and there is no tool to calculate the UDR for Multi device customers. To cater for multi-device use cases owned by CDO, the CDO BI created a UDR metric for Multi device customers by writing a SQL query for the logic defined with product team. Interestingly, the Multi Device (MD) UDR is less than Single Device (SD) UDR, this is contrary to our expectation of data. I.e. for NA region MD UDR is 16.6 points vs SD UDR is 19.69 points and worldwide MD UDR is 16.85 vs SD UDR is 21.08. Based on this data, Product team came up with following conclusion as MD customers are typically more tenured and more active customers and have an overall lower UDR due to the "training" on how to speak to Alexa. However, no amount of training can avoid failures with basic global requests.
* **Registry Health dashboard:** We created a Registry Health dashboard to provide a comprehensive overview of key Registry related data including overall composition and growth trends. This dashboard was designed to support monthly deep dive meetings and most importantly, to inform and de-risk the Device Deduplication project. The current dashboard offers our PMs, SDMs and BDSA customers a completely self-service dashboard to monitor data, such as unique identifier coverage from our 3P partners and metrics related to a number of existing and merged duplicates. The tool has already helped the project team to proactively identify data irregularities so that they can be solved ahead of the launch.
* **SDE Intake Dashboard:** On a monthly basis CDO TPMs report intake data to answer questions on volume of intakes, accept/reject rates, key customers, tagging and general processing timelines. This was done manually and only once a month with the limited capabilities of excel. We created a CDO Intake dashboard to allow TPMs, SDMs to know if our investments in developer velocity efforts are working and to help them to understand if we need to invest further.
* **LAP Adoption Dashboard:** CD BI created a QuickSight dashboard for the LAP launch to track Local Adoption, specifically the number of customers and devices that will benefit from LAP-enabled local control. The dashboard consists of about 20 metrics, with a plan to expand to 50 metrics. These include # of devices of that are eligible to enable a Local route, # of customers that have disabled cloud control, but local control persists (CX gap) etc. The dashboard is helping the Product and tech teams to monitor how frequently customers experience known CX edge cases so we can take action if this metric moves beyond our estimated thresholds.
* **MDX Adoption and engagement Self-serving query:** MDX team routinely needed to collect data on multi-device spaces to inform prioritization discussions. To assist with this, BI team created a single de normalized table and turned multiple simple queries into one single query. The goal of the project was to create a QuickSight dashboard to refresh on a weekly basis with all the necessary filters to slice and dice the data and have a readily available data for PRFAQs.

Appendix 2: Data life cycle and steps to work with BI team.

A full fledge data project will be in the following stages 1) Customer Data Instrumentation (involves Tech team and PMs) 2) Data engineering (SDEs and DEs) 3) BI engineering (DEs and BIEs) 4) Business analyst/PM (BIEs and PMs) and 5) Data science (DS, BIE, PMs) to help answer business questions effectively. Teams in the CDO org are in different stages of data life cycle, hence scope and steps will vary project to project.

Below are the high level steps:

**Step 1:** Create a SIM using [link](https://wiki.labcollab.net/confluence/display/Doppler/Alexa+Connected+Devices+and+Orchestration+BI#AlexaConnectedDevicesandOrchestrationBI-IntakeRequest:) which includes a requirement document as in samples. [Sample 1](https://quip-amazon.com/9TlWAjlkU4YF/CD-Local-Metrics-Dashboard) & [Sample 2](https://wiki.labcollab.net/confluence/display/Doppler/Multi+Agent+Analytics(MAA)+Requirement+and+Desing+Doc)

**Step 2:** Set up a meeting with the BI team to go over the requirement docs. Based on the requirements, BI team will scope the bandwidth required to work on the request.

**Step 3:** After scoping, if the project requires more than a week to complete, then it needs to be prioritized with @putchas since BI team works on monthly sprint prioritization.

**Step 4:** Facilitate meetings with respective Tech teams if the data is not instrumented yet. Later will BI team will follow the project plan as below.

|  |  |
| --- | --- |
| **Sl. No** | **Milestone Description** |
|  | **Data engineering** |
| 1 | Finalize the Design with SDE team to publish the data |
| 2 | Development of publishing the data to S3 by SDE teams |
| 3 | Proof of concept/Design to publish the data from S3 to Redshift or Redshift Spectrum |
| 4 | Create the pipelines to consume the data from S3 to Redshift or Redshift Spectrum |
| 5 | Test the data is being effectively consuming on agreed cadence |
|  |  |
|  | **BI Engineering** |
| 1 | Understand the metrics definition and review the KPIs received |
| 2 | Create a mock WBR/dashboard for owners review |
| 3 | Review the mock WBR/dashboard with owners and gather feedback |
| 4 | Make the changes and create the final WBR/dashboard with Mock data |
| 5 | Review with owners to get the final go ahead |
| 6 | Integrate with the actual data tables and test the output |
| 7 | Publish the WBR/dashboard by automating the data pipeline and document the details |
| 8 | Test the data is being effectively published |
| 9 | Communicated about the WBR/dashboard |

Appendix 3: Tool glossaries:

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Owner** |
| [DataMart](https://w.amazon.com/bin/view/Datamart/) | DataMart is the secure storage service of the Alexa Machine Learning Platform. DataMart is a custodian of sensitive customer data and upholds the Critical and Restricted Data Protection Model Standard with Strong room/KeyMaster. | SAIF |
| [Cradle(i.e. Dryad)](https://w.amazon.com/bin/view/BDT/Products/Cradle/Overview/) | Cradle provides a low-cost, server-less solution for batch, streaming, and machine learning over all of Amazon's data. Cradle is a fully managed platform with support of using either shared EMR cluster or bring your own EMR cluster. | BDT |
| Redshift | Amazon Redshift is a data warehouse product which forms part of the larger cloud-computing platform Amazon Web Services. | AWS |
| S3 | Amazon S3 or Amazon Simple Storage Service is a service offered by Amazon Web Services that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network. | AWS |
| AlexaDB | Centralized database as primary place to collect source data from various teams - data is usually unaggregated (ex: d\_utterances is at the utterance level, not rolled up to dialogs, customers, etc.) | AXD |
| SSDG | SSDG(Self-Service Dataset Generator) is a platform allowing self-served onboarding of SSDG datasets to filter and transform DataMart streaming events on the fly, to pre-process data in preparation for ML (and BI in the future) related use cases. | SAIF |
| Datanet | [Datanet](https://datanet.amazon.com) is Amazon's internal Big Data Management System and is the gateway to the Data Warehouse. Datanet has a number of web applications that facilitate the process of getting data in and out of the Data Warehouse. | BDT |
| [EDX](https://w.amazon.com/index.php/EDX) | EDX (Elastic Data Exchange) is the centralized and secure metadata store for all Amazon data, established to enable seamless interchange. You can imagine EDX as a metadata layer over S3, and as such its primary use is as a bulk data store S3. The benefits of EDX over S3 include batching, encryption, and compression of your data by default. | BDT |

Appendix 4: FAQs

**Why BI dashboard over PMET/Atocha dashboard?**

PMET is one of the aggregated and reporting back ends for metrics data. It only works with a subset of the data provided by the metrics data model. PMET is not meant for seeing a big picture view of correlated events. In the service world, it is commonly used in monitoring and alarming on increases in specific API latencies or a sudden increase in the number of requests. Where BI dashboard will help in joining data with a verity of other data sets and slice and dice as per the business needs.

**Why a separate CDO Redshift DW, why not use Alexa BI or Others org Redshift DW?**

Historically, Alexa DB was built with the vision of having one DW for entire Alexa when Alexa as a company, size was smaller. Over a period Alexa has grown X times and multiple Orgs have been formed within Alexa. The team built to build Alexa DW is also dismantled and rebranded in Alexa Experience and devices. Hence, they no longer support any new use case which is at individual org level. Similar with SH BI, won’t prioritize the other org needs.

This is a business reason to have a separate CDO Redshift DW. Along with it, technically a CDO Redshift provides flexibility in creating our own tables, connecting to the workbench and Viz tools than using a throttled web based tools to query the data and build dashboards.

**What will be the frequency of data loads?**

Usually depends on the service provider, typical BI team loads data on daily/weekly basis based on the cost involved. However, we strive to be near real time as possible.

**Should there be a centralized BI team in the org or individual BI/DE to a team?**

In a long run it is recommended to have centralized BI team for quicker turnaround of work and support system for running operation seemly in case of leaves and attrition. Centralized team will also help for BI/DE employee’s moral and provides the sense of working with likeminded people.

**Why QuickSight over Tableau?**

In short, QuickSight is an in house tool. For the use cases CDO has 80% of the problems can be solved using QuickSight. Although Tableau is miles ahead in functionality with QuickSight, QuickSight is narrowing the gap for the must have features in a dashboard. In terms of costing QuickSight is free.

**Why S3 for storage?**

As the name says Simple Storage Service, it is Simple. It is widely adopted by all the AWS users. It is low cost and provides easy and multiple ways to access the data stored.

**Who will own the Data dictionary?**

It is expected for the Data producers (SDE Team) to own it, however, it can share with the BI team by working to PMs.

Appendix 5: CDO BI 2021 Goals

1) Create at least 3 CDO owned Data table pipelines in CDO data infrastructure. (Director level goal - Need by December 31, 2021)  
  
**Description**: We will be creating Data pipelines to bring the data from Tech stores (persistence data or Logs) to BI stores (Redshift/S3/Andes) for BI teams to create xBRs and Dashboard.   
Teams need data in BI stores: CHR, ESP (FE), Context, Lifecycle, DAAPI, and LAP & Multi Agent. The goal is dependent on Teach team and Privacy approval hence we will be taking goal for at least 3 teams.  
  
2) Create at least 4 Quick Sight dashboards to have the self-servicing access to data. (Director level goal - Need by December 31, 2021)  
**Description**: Create at least 4 new Quick Sight dashboard/WBR to provide for teams like Multi Agent, ESP, Targeting, LAP, Registry, Multitasking and Context etc.  
  
3) Meet GPDR requirements, Project Spaniel - Privacy compliance, Seller Data Privacy Policy Agreements. (Director level goal - Need by December 31, 2021)  
**Description**: Currently there is no action required, however, once we start plumbing the data in Redshift, we need to ensure the privacy compliance is followed for all the data.  
Wiki: <https://w.amazon.com/bin/view/BDT/Resources/DUAPwiki>, <https://policy.amazon.com/standard/50725>, <https://w.amazon.com/bin/view/Kaizer/Design/SDPPMigration/>

Appendix 6: BI Operational Model

**1. Monthly sprints:** In order to balance requests and support projects across the services, the CDO BI team will have a modified form of Scrum – a framework which combines work in monthly sprint cycles with the visibility of sprint boards. Additionally, BI capacity is calculated monthly to establish the number of requests that can be confidently delivered for a given sprint. BI capacity is the actual working time available to build and deliver BI products and solutions. It excludes business overhead such as team meetings and professional development that reduce available BI time. This model allows the team to (a) prioritize incoming requests based on criticality, urgency, and skill sets, (b) set appropriate expectations for completion while providing regular updates to interested parties, and (c) allocate ample time for CDO BI to focus on creating solutions to identified needs. In advance of each month, CDO BI will prepare a proposed sprint based on requests previously submitted SIM’s, CDO leadership will review and make final decisions regarding which initiatives will be prioritized based on product needs.

**2. T-shirt sizing model: Efforts**

CDO BI will use a T-shirt size framework for estimating effort required to complete a BI request. T-shirt sizes are a general guideline as what to expect in terms of task complexity and effort required. Monthly capacity and sprint objectives are aligned using these T-shirt sizes to ensure the BI can deliver products in a timely fashion. Each request is given 1 (x-small) through 5 (x-large) planning points based on the assigned T-shirt size to simplify capacity and sprint objective alignment. X-Small projects are generally things that can be completed in a matter of hours, whereas X-Large projects may take weeks to months. CD BI collectively estimates T-shirt size using several key questions, prior experience and business insights. Example questions used for T-shirt size estimation as below.

1. Has a similar request been delivered previously?
2. Are all data available in a single Redshift cluster? If not, where is the data source?
3. What is the approximate query complexity?
4. Is the team experienced working with the required technology?
5. Are there dependencies on other teams? If so, what are they?
6. Is the request classified as confidential? (High confidential, Critical, Restricted)

**3. Updated Intake Process and Office Hours:**

CDO BI requests are received via a standardized intake process, where in requestors help define the deliverables/success criteria and potential benefits in order to develop complete request. CDO BI has developed an updated intake process that streamlines intake of requests from stakeholders, and pairs that model with regular office hours designed to guarantee 1:1 attention for our customers while safeguarding CDO BI ad hoc and project capacity.