Crux & Microsoft Azure
Partnership Overview for Microsoft

Background

Crux is a data engineering and operations platform that provides immediate access to thousands of traditional, alternative and public datasets via a cloud-based platform. Crux's data management platform ingests, validates and standardizes datasets from over 90 vendors and counting. Crux delivers the data via the client's preferred delivery mechanism, and in the form they want (e.g. Parquet over API, relational via Snowflake, Excel using ODBC).

Crux's vision is to be the industry's data delivery service; a utility that takes on the onerous but undifferentiated data engineering and operational tasks across consumers and suppliers of data. We achieve economies of scale that firms cannot, by definition, get for themselves, while making the process of acquiring data delightfully easy.

The Crux data management platform is integrated with Microsoft Azure and other cloud platforms for data delivery. Indeed, clients increasingly leverage Crux to accelerate their adoption of the cloud for data management workflows. A Microsoft partnership positions us to jointly approach these opportunities and win these mandates.

Use Case

Every data consumer shares the same fundamental issues with respect to data engineering and management. These are problems that are endemic across data consumers, they are inefficient and costly to address, and firms have been independently and redundantly addressing the same issues over and over:

- Wiring up to every vendor discreetly, and dealing with the multiplicity of ingest technologies, APIs, FTP sites, loader programs, data formats, etc.
- In-house data engineers validating, standardizing and wrangling each dataset so that it's actionable
- A backlog of data to be onboarded because of resource constraints
- Expensive infrastructure to load, store and manage data
- A dedicated data operations team

Crux customers want to fundamentally change the paradigm for ingesting, validating and managing data with:

- A single, cloud-delivered hub for accessing all data sources
- Validated, cleaned and actionable data
- Delivery in the format and method required by the client
- Datasets that can be accessed on demand, as well as flex resources to onboard new data quickly
- The cost efficiencies associated with cloud deployments
- 24x7 data oversight, support, and issue management

What's in it for Crux?

The Crux business model is designed to generate economies for all parties involved as the network of data consumers and suppliers grows. The utilization of data engineering and operations means that data consumers will achieve highly cost effective access to an ever increasing amount of data on Microsoft Azure, leveraging Azure infrastructure and services. Our partnership with Azure enables Crux to:

1. **Increase our sales channel by leveraging Microsoft Azure's sales teams**
2. **Expand our delivery destinations to include Azure accounts**
3. **Leverage world class Microsoft technology to build our utility faster**

What's in it for Microsoft Azure?

There are multiple benefits for Azure in this partnership.

1. **Adoption of Crux accelerates Azure consumption.** Crux's objective is to become the industry utility for data engineering and operations. We will be the central hub through which consumers access, and suppliers deliver, data. More consumers accessing ever increasing amounts of data from more data suppliers, all on the Azure platform, driving revenue.

2. **Drive deeper into the financial services industry.** While Crux is data agnostic, our core client base comprises large asset managers, banks, and other financial institutions. Together, we can drive deeper into this client segment. Moreover, Crux is led by industry veterans and has active relationships with some of the leading buy-side and sell-side firms.

3. **Enable Azure sales teams with a high quality data catalog of thousands of validated datasets.** Data consumers do not want to ingest thousands of datasets over varied delivery mechanisms and in completely different formats. Moreover, Crux validates all ingested data, ensuring delivered data is clean and ready for processing. By working with Crux, Azure sales teams will be able to offer thousands of clean datasets from a single destination, a single delivery mechanism and a consistent format.

4. **Crux takes care of all the work: data validations, delivery, operations, and notifications.** Azure is not expected to do ongoing work or maintenance on Crux services. Crux has an inhouse team dedicated to the execution of our data delivery and operations service.

FAQ

- **1) What vendors/datasets do you have?**
  - Crux has over 700 vendor and 5,000 public datasets from over 85 data suppliers and over 30 public sources. They are actively building hundreds more pipelines with new datasets getting added every day.
- **2) What is the pricing structure?**
  - Crux charges a monthly subscription fee per dataset delivered. The fee covers a company wide subscription (IOW, not a per seat model), and varies in cost depending on the complexity and scale of the datasets getting delivered.
- **3) Who are your competitors?**
  - The Crux solution, central access hub with a validation and operations managed service layer for all datasets, does not have any competitors that cover its full model. There are some outsource ETL businesses and other marketplaces that aggregate various datasets, but Crux does significantly more than ETL processes and it does not sell data, as many marketplaces are usually biased in their selection and curation. By far the most common "competition" we face are internal market data and data engineering teams at customer/prospect firms that have legacy systems and stale processes to getting this work done.
- **4) Do you do transformations (mapping, entity mapping, etc.)**
  - Crux does some standardization, but we do not currently provide custom transformation services for our clients. This is on the roadmap for 2H 2020.
- **5) How do we connect to the platform?**
  - Crux features multiple options for delivering data to best meet your data consumption requirements. Customers can connect via:
    - RESTful API
    - Python SDK
    - ODBC connection
    - SQL queries
    - S/FTP
    - Cloud platforms
      + Custom integration into other data or analytics platforms of your choice
  - The Crux data management platform features a pre-built adapter to Microsoft Azure Storage to automatically and securely push data into a customer's

account. With Crux, users can quickly connect to and leverage Azure technologies to turn large sets of data into intelligent action without the cost and complexity of traditional middleware or custom coding.

- **6) What validations do you perform?**
  - As part of Crux' core data processing service, we perform validations on supplier-provided data to ensure the data received by the supplier is in accordance with their specifications. If any of the below validations fail, Crux will produce a notification and work with the data supplier to resolve.
    - **File:**
      - File Present – File is present within expected time frame
      - File Name – File posted has a name that conforms to the regex pattern established in prior instances
      - File Not Empty – File is not empty (non-zero byte size)
      - Expected Files Present – All expected files are delivered per the supplier's specification
    - **Schema:**
      - Schema Not Changed – Checks data types and order
  - Crux maintains an extensive list of planned validations to implement and will publish updates upon release.

- **7) How do you standardize/normalize?**
  - Our goal at this stage is to become a ubiquitous and neutral utility for the industry: One place to receive all of your data. To accomplish this, we are prioritizing the following:
    1. Increasing our data catalog so that the data you need is covered. Today we have over X number of vendors, large and small, and X number of datasets spanning alternative, traditional, ESG, reference data and more
    2. Ensuring many ways to receive your data: Restful API with python wrappers, cloud store, cloud based data warehouses and traditional methods like SFTP
    3. Offering numerous and flexible file formats
    4. Operating the pipelines at scale: Ensuring maximum platform stability and uptime, detecting and resolving technical file based issues with vendors, on our clients' behalf and more.
  - As we expand, we plan to roll out more validations, standardizations and normalizations and even offer custom business rules that you can create.