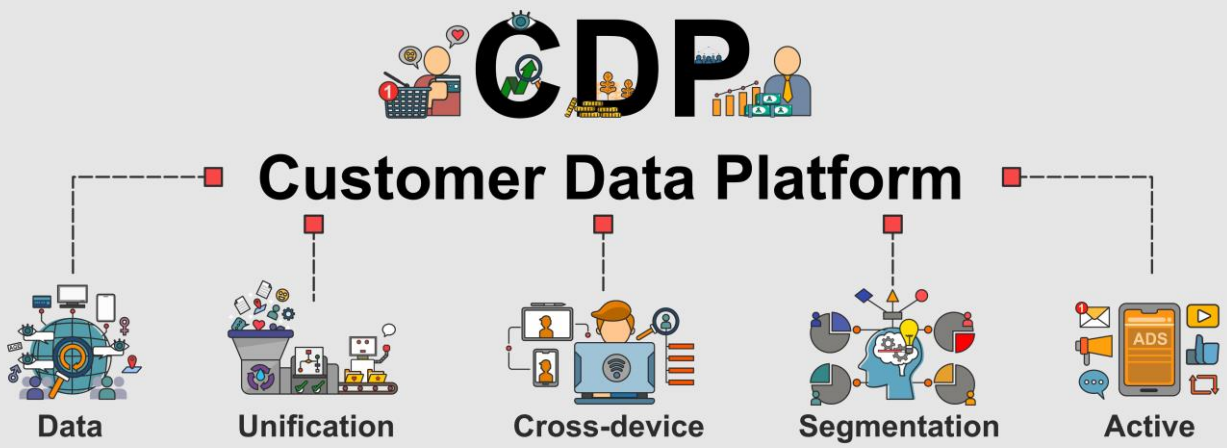
Financial Data Management – A Detailed Step-by-Step Approach



Date: January 2021

INVESTMENT RESEARCH & ANALYTICS

Executive Summary

Digital transformation is constantly changing the way we treat data, which has become a precious commodity. Yet, few experts have the requisite knowledge about the correct way to store and process data.

In the past few years, innovations have transformed the way data is treated. A key turning point for the entire data processing industry was the advent of cloud computing. The days of legacy databases are past; cloud databases have taken over. While cloud computing and cloud databases have seen acceptance over these past few years, the cost involved and a lack of relevant skills have delayed its widespread adoption.

Other technologies, such as artificial intelligence (AI) and digitalization, have made faster inroads into the data processing space. While they ease data analysis – helping managers make informed decisions – these technologies still require manual data amalgamation. The issues of data accuracy and cost also remain.

Challenges appear right from the inception of data sourcing and need to be managed until final evaluation.

However, manual data processing remains a critical step in overall **data research and management processes**. With the advent of new technologies, the dependence on manual intervention may lessen, but would not disappear entirely; manual intervention would still be required to facilitate AI and machine learning (ML) platforms.

What are the optimum ways to collect, store, and convert data into actionable insights? In this whitepaper, we attempt to answer this question and outline the steps of the complex data research and management process.

Data sourcing: Source and authenticity matter

Data sourcing refers to the identification and tapping of the primary location where data originates. Data can be present as hard copy, or physical information, or as soft copy. The source could be a flat file, web data, or the even the cloud.

Massive amounts of data have been generated with the widespread penetration of the Internet coupled with online businesses and the emergence of digital natives. This has helped companies derive important information and take strategic decisions.

Informed decision-making necessitates a tight data-building process. This process defines how data is to be collected, stored, transformed, circulated, and used. It would also outline solutions for the proper storage of data and file systems and the procedure to be followed to create structured plans. It is critical to plan these processes by ensuring the correlation between data and business strategy is leveraged fully. This allows for the conversion of data into facts that the business can apply while making decisions (strategic long-term or short-term plans).

Though companies have attempted to ensure a consolidated, controlled approach for data and information building, challenges related to reliable sourcing, accuracy, and usefulness persist. Furthermore, data sourcing is frequently ignored and thought of as a tedious, time-consuming process, especially by companies lacking a data warehousing background. However, the validity and quality of data is crucial for any data warehousing project; thus, cleansing and verification of data cannot be disregarded. Yet, some companies skip this step and concentrate on data entry applications, often skipping the pre-process stage.

The following example emphasizes the importance of data source verification.

Case study

A private investment firm intends to partner with management teams to build leading companies. The firm seeks companies with potential to grow either organically or via acquisitions. It targets firms with focused management and strong offerings as well as cash-generation prospects.

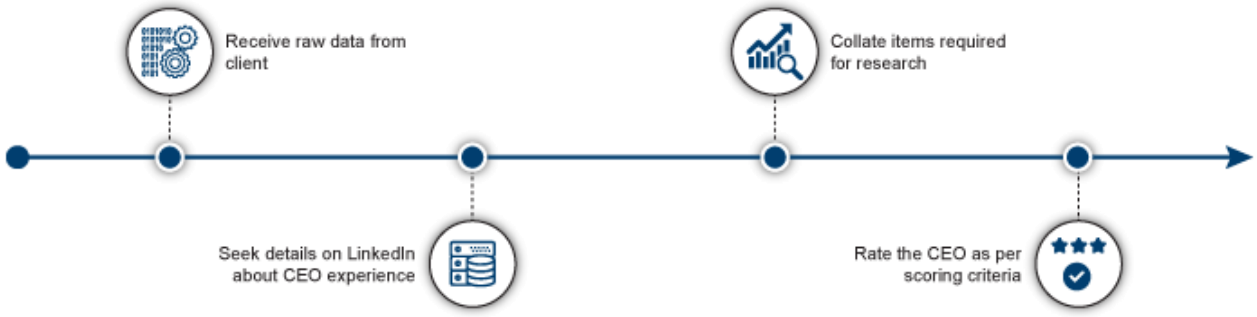
It is imperative that the investment firm carefully evaluates curated information about the performance of the CEOs of the target companies and their background, so it can take a strategic decision. As such information is generally present on multiple channels, its authenticity should be verified. For professional data, the recommended websites include the following:

- LinkedIn
- Bloomberg
- Capital IQ
- Reuters
- Regulatory websites
- Company websites
- News wires

Additional data can be searched for and added while ensuring the reliability of the source.

The process flow detailed subsequently shows the steps involved in gathering data. The analyst receives a set of C-Level executive names from the client, generally a spreadsheet, or via a cloud-based platform. Specific information on each executive is looked up, e.g., the CEO's experience, tenure in a company as a CEO, location, number of private-equity exits that this CEO was involved with, and ROI during their tenure as a CEO. All this information is collated from multiple sources to help the analyst arrive at a rating for the CEO, based on a pre-defined scoring matrix.

Process Flow



Data amalgamation: Collection of information

The next step of data processing is data collection or amalgamation. The process refers to gathering or grouping of scattered, unstructured data from multiple sources. This data then goes through cleansing, deduplication, validation, and segmentation processes.

The application of technologies in data collection makes the process effective and efficient. Modern data management uses ML coupled with human expertise and agile methodologies to collate data. The need for manual data processing thus remains relevant, to an extent. AI is also progressing rapidly in this space, but the dependence on humans remains, especially to train AI to gather data.

Relevance of Data Amalgamation

Consistent Improvement	Helps streamline company-wide data processes; provides accurate data
Visibility of Data	Provides in-depth insights to enhance core decision-making
Improved Performance	Offers quicker data analysis, simplified data structures
Easy data access for robust data management	Reduces data complexities, thus enhancing easy, secure access to data
Organized data security	Offers transparency: GDPR in EU requires companies to evaluate data with greater transparency
Streamlined processes and workflow management	Offer benefits of data unification, data management: Helps companies streamline large datasets and processes, thus enabling smooth workflows

Data cleansing: An essential aspect to maintain data integrity

Data cleansing is the next step in the process of data management. Data is important across industries and sectors for business strategy and planning. It is imperative to have access to error-free, complete data to make informed decisions. Data cleansing aims to improve the quality of data by cross-checking and removing inaccurate, incomplete data and correcting errors and omissions. Formatted, checked, and cleansed data enables optimum utilization of the available information.

Data cleansing is a complex, time-intensive process. The process entails painstaking checks of compiled data, one spreadsheet at a time, to eliminate duplication or blank spaces, update details, complete missing points, and format the entire set into a single file. Even a minor error in the compilation can cascade into a large-scale mistake and result in a loss for the company.

The importance of data cleansing cannot be over-stated. Data cleansing can be outsourced to a specialist agency as this would save the company countless hours in trying to fix data. The company will have access to well-organized, updated, correct data, without putting in massive man hours.

The steps involved in data cleansing are as follows:

1) Identifying correct data set after scrutiny – focused approach to critical data is essential to avoid errors

2) Identifying problems present in critical data – considering the most systematic approach to cleansing

3) Choosing either manual workflow scripts and/or systematic automation for data cleansing

    - Manual cleansing requires a setting of resources and exceptions to rules, after which cleansing can start. Manual intervention has limitations and is directly related to acceptable levels of data quality.

    - In real-time automation, workflow scripts run in batches (daily, weekly, monthly), depending on the quantity and quality of data.

4) Reviewing working process periodically, irrespective of the workflow chosen – issues are detected and resolved before they escalate; bounce rates can also be tracked

Some standard concepts in data cleansing are:

- Data deduplication – a process that eliminates multiple copies of data and significantly decreases storage capacity requirements
- Data analysis – a process for cleaning, transforming and segmenting data such that a story can be derived out of it, which would help in decision making
- Data standardization – a process that involves making data internally consistent, i.e., each data type follows a standard format and content style
- Data normalization – a process where data within a database is reorganized such that users can use the same database for other queries and analysis
- Quality check – a key step in data cleansing; the process involves checking of data by another pair of eyes to avoid errors

The main objective of the data cleansing process is to bring uniformity to data sets merged from multiple sources and make them error-free. Once the final data is ready, it is easier to apply algorithms and mathematical models to it to extract the relevant information.

Data segmentation: Understanding your data well

A critical aspect of data processing is correct segmentation, which means classifying the available data by parameters. This procedure allows a systematic arrangement of unstructured data to aid in correct analysis and estimation of scenarios. Accurate data segmentation enables the company to identify high-opportunity groups within an existing customer base. Customer groups are classified by purchasing patterns, demographics, and preferences, and categorized into segments. This allows the company to gain insights from the vast amounts of data that businesses collect and use these to plan strategy.

The Harvard Business Review published a case study on Hill-Rom. The study showcases how accurate data segmentation helped Hill-Rom cut costs, ramp up its sales, and consequently boost its revenue. With an objective to improve its overall economics, the company focused on creating an appropriate customer segmentation model. This helped it determine the requirements of each segment. Aided by this information, Hill-Rom tailored sales techniques that served different customer needs. With this customized offering, Hill-Rom was able to extend its value proposition to multiple client sets, enhance customer satisfaction, increase its revenue outlook, and in turn generate profits.

Data Reporting or Publishing: Making your data work for you

Data reporting is a key aspect in data research and management. After all processes are complete, data reporting or publishing makes the final step. The data could be published over client's online platforms or could be shared in the form of MIS reports. The data collected and analyzed is structured in a format that makes it easier for the reader to understand the analysis. Behind every data analysis lies a story about a certain product or process in the company. This helps key stakeholders make critical decisions about the business. Interpreting data correctly is crucial at all times. Customized reports are prepared to meet client requirements in multiple formats, such as slide presentations and documents. Research output is generally published in these formats, either on the client's internal network or in an external environment.

Emerging trends in data management

- Data protection with AI
  - Businesses are using AI to study customer behavior as well as perform other analyses. While it is considered a complex technology, AI-based tools, if implemented correctly, help not only collect but also protect data.
  - AI can enhance data security. There are touchpoints from which businesses collect data, such as websites, devices powered by the Internet of things (IoT), mobile devices, scanned images, and other documents. With AI, users can ensure that data security remains uncompromised.
- Augmented data management
  - The main objective of most companies is to easily collect and verify data. Augmented data management helps users with tasks such as schema recognition, regulatory compliance, and easy utilization. Many business owners try to achieve this manually, but end up with costly errors. Augmented data tools address precisely this problem.
  - Various studies indicate that dependence on manual data collection and validation will reduce by nearly 50% in the next couple of years. It would therefore be beneficial for companies to adapt new technologies.
- Residual element of manual data management
  - Even though companies embrace new technologies like AI and ML, some element of manual intervention remains. Humans intervention is needed to create and train these programs to process data accurately. Over time, these programs would then process data seamlessly, without human intervention.
- Reliable analysis and intelligence
  - Data amalgamation facilitates systematically organized data that is accurate and allows problem solving and real-time decision making. It also helps with quicker decisions, resolving issues faster, thus leading to accelerated growth that falls in line with business demands.

How can we help at ARANCA?

Data Sourcing

Our data research team follows a systematic approach while sourcing quantitative or qualitative data, as per clients' needs. We are proficient in sourcing data sets from company websites, filings, reports and presentations from industry associations, news feeds, articles, and through keyword searches.

Data Research and Analysis

Our data research analysts research, collect, and manage data to build a robust data warehouse to support powerful front-end analysis. Our analysts use LEAN Six Sigma methodologies to bring in efficiencies and deliver high accuracy. We offer further value additions in the form of automated solutions.

Data Segmentation

Our data research analysts segment large data sets into sub-sets on a single platform by integrating, filtering, organizing, and summarizing the points collected from fragmented, diverse sources. Data can be further segmented by region, product, segment, task, etc., using structured frameworks as per clients' requirements.

Data Validation

Our data research team cleanses and normalizes large data sets collected as per clients' requirements. For this, we deploy a robust maker-checker process to ensure that the data finally published on clients' platforms is of the highest quality. Our overall review mechanism involves checklist-based reviews, peer reviews, high-level audits by senior staff, and automation-based audits, wherever possible.

Data Publishing

Our data research team publishes validated data by pushing it through the client's platform, both online and offline, whereby this data goes live for customers to view directly. This also involves formulating and presenting data in Excel spreadsheets, Word documents, PDF files, and slides, as per clients' requirements.

About Aranca

Founded in 2003, Aranca is a global research & advisory services firm working with clients worldwide across financial markets, industry sectors and technology domains. Aranca brings to play the strong combination of best data and best talent to empower decision makers with intelligence and insights, enabling them to reach better business decisions. Our multi-disciplinary expertise is designed to cater to clients of all sizes across a wide spectrum, from Fortune 500 companies and financial institutions to private equity and high potential start-ups. In the MENA region, Aranca works with some of the top family groups, private equity and investment management firms with strong focus on strategic corporate and financial advisory services.

Disclaimer

The information contained herein is believed to be obtained from the reliable sources and Aranca disclaims all warranties as to the accuracy, adequacy and completeness of such information.

Since Aranca is not a law firm, it (including its directors, employees and representatives) does not and cannot render legal services/advice to any individual or entity in any part of the world.

Author

Ryan Fernandes

Manager

Financial Data Management, IR division

Aranca

Unit 201 & 301, Floor 2 & 3, B Wing, Supreme Business Park, Hiranandani Gardens, Powai, Mumbai – 400 076

+91 22 3937 9999

www.aranca.com