



Fundamentals of Creating Analytic Solutions

Analytics 101



Introduction to Analytics 101

Analytics – Are the same problems nagging us all?

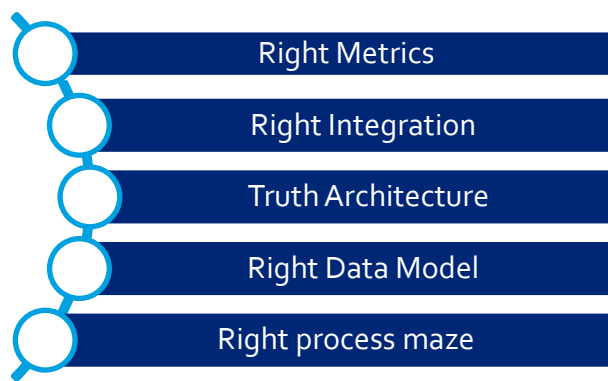
You are a medium sized logistics company and most of your package deliveries are to corporate customers. To increase your profitability you are trying to optimize your business processes as well as increase customer satisfaction. And you are willing to step back and take a holistic approach without leaving any stone unturned. We try and answer some key questions:

- 1) Where should you start?
- 2) What should you do?

And the answer to your surprise is not very different if you are a food product manufacturing company, or medical claims processing company or even a pharmaceutical company.

Decision Support from data can help different organizations in different ways. American Airlines built Electronic Reservations, Otis elevator built predictive maintenance solution, and American Hospital Supply built online ordering systems.

Irrespective of the Industry, from a Decision Support perspective you have three domains at your disposal - Business Processes, Data and Applications. The paper attempts to lay the guide rails for an analytics design procedure that straddles these three domains. Much before an analytics design kicks in, it is the Business Processes and Data which must be obtained right.



Get your process maze right

The first step is to get the lay of the land right. Competing value chains and that means business processes would soon be the only differentiators of your enterprise in the face of increasing commoditization of products and services. A holistic approach is needed, as several of your processes will be interdependent and optimizing one while ignoring the other is like working on one side of the equation while ignoring the other.

For example as a logistics company your business processes might include setting up new customers, get customers package delivery request, drop off and pick up customers packages, sort packages, track packages and invoice customers. A lost package during sorting could result in issues in customer tracking process, package delivery and subsequently deployed payment from the customers. As these processes are *interdependent a holistic view* is critical, to observe the connecting dots.

What do most people get wrong often – it's the Data Model.

The white boarding sessions by the Business Analyst, Data architect, and Integration architect need to be independent so that each serves its unique purpose. The following three artefacts must necessarily have the concurrence of business.

- Process Map
- Conceptual Data Flow
- Conceptual data model (for only the entities in the business process)

A CRUD matrix (Create, Read, Update or Delete) for the data elements should be owned by the data integration architect.

The Conceptual data model is the first step to holistically understand the data that that

your critical business processes are relying on. It is important for the following reasons:

- a) The dimensional model is essentially a subset of this conceptual model. It is based on the data elements documented on in the conceptual model.
- b) The data integration processes use the rules documented via the conceptual model to standardize / cleanse data.

In a nutshell the conceptual data model will help you understand your key data entities and how they are related to each other.

Looking at this data, independent of its physical layouts allows you to identify

and document relationships that are neither enforced at a data or application

level, but are nevertheless critical for the functioning of the business.

Get to the truth architecture, prevent costs from spiraling out

If you are trying to analyse click stream data where you need to slice and dice hits by zip code, product, date, then, you can keep pre-aggregated data on total number of clicks on a product, at a particular date, and from a particular city. But if you wanted to know how long a prospect spent on a particular section of the website you would need data with different attributes and different aggregation levels.

A proper design of the dimensional model is critical to establish the efficacy of your analytic model. In order to get the right performance and reduce infrastructure expenditures the right grain of aggregation must be obtained. **The efficacy of the model is not established by the flexibility - you may have to go to the minutest level, but whether you are architected to have data at the level where action needs to be triggered.**

A dimensional model provides a "star schema" where facts (in the center of the star) contain information that needs to be analyzed or aggregated and dimensions provide the ability to slice, dice and aggregate the information in facts. Dimensional models are basically meant to

arrive at the right granularity or aggregation level where the metrics first need to be actionable (defined to point to a problem). Facts contain metrics or measurements required to create metrics.

Costs of several analytic environments spiral out of control and performance suffers when proper due diligence is not given in understanding the right aggregation needs.

Thus the right architecture of dimensions is actually the Truth Architecture, the architecture that rightly connects the processes at the granularity that brings out the truth.

Leveraging data across organizations (internal as well as external) as well as striking at the right levels of aggregation levels forms the basis of a predictive model. The CRUD matrix and the data model of the applications holds the key to appropriately identifying the Data Sources. Data sources would automatically lead to ERP, CRM data, machine data, trouble tickets, invoices, product demand, weather data, social network data etc.

Getting Data Integration right

Data integration is the process of taking data from various data sources (like databases, files, tables, files, spreadsheets, web server logs etc.), at various levels of granularity, at varied point in times, with varied data quality varied definitions and integrating it to prime it for analysis.

These data integration processes form the backbone of your analytical solution. Typically they will take up 70-80% of your people and infrastructure resources and hence it has the most significant impact on your project costs, complexity and timelines.

Data integration solutions can be real time, batch or real-time / batch combination. Contrary to popular beliefs, simplicity in data integration requires greater effort, more planning and more collaborative efforts. **It requires an approach in design that evaluates various solutions for simplicity and picks the one that is easiest to develop and maintain.** It requires an automated regression testing approach with a constant set of test data ensuring that changes don't break what has been working before.

Did you get your metrics right: Is business taking the right decisions?

A CPG company found most of its Business Intelligence investments meaningless because key metrics like Weeks of Supply were lagging and not predictive. Predictive metrics hold the key to a predictive Decision Support. **If your metrics must help you orchestrate a sense and response system then the indicators need to be predictive.** The truth architecture must necessarily be leveraged to provide the truly predictive metrics such as the futuristic Cost per Case or Weeks of Supply - one must necessarily be able to obtain the metrics at the right aggregation level. When you look at the existing metrics that

are being used, you will often find that the metrics are computed at the wrong aggregation level.

Identifying leading indicators and that too at the right aggregation level is a maturity that requires availability of historical information in order to perform basic linear regression or years of experience to set the empirical rules. Most organizations have reporting and architecture designs that make them look at lagging metrics.

In a nutshell the fundamental process of designing analytic solutions is not very

different across industries. The analytic needs of various industries are different, but the metrics and the data architecture used are very similar. It requires a holistic view, taking into account the core interdependent business processes, having interdisciplinary whiteboard discussions to

map out the processes, analytic needs, data model and data sources. Aided with the appropriate data integration model, this could help you reap benefits of the huge array of data that the information systems within your organization generate.

Need more information?

Email: info@ypointanalytics.com

Phone: **USA:** +1 (703) 880 4128

India: + 91 9892590158

Scan the QR Code to contact an YPoint Professional, or visit www.YPointAnalytics.com



www.YPointAnalytics.com

Y-Point is founded by *IIT-Bombay Alumni*. Together our founders bring 45 years of experience in business consulting. They have successfully delivered strategy and implementation projects related to Analytics, Business Intelligence, Optimization, Data Architecture, Data Integration, Data Migration, Corporate Performance Management, Solution Architecture and Business Strategy. / Planning, Corporate Performance Management. They have been involved in Manufacturing, Government, Telecommunication, Health Insurance, Mortgage, Student Aid / Loan, Logistics, Retail, Media and Banking Industries, CPG, Hi-Tech , Media and Entertainment, CPG, Logistics and Transportation to Electoral Analysis.