Final Report

CSDA 1030 - Class Project

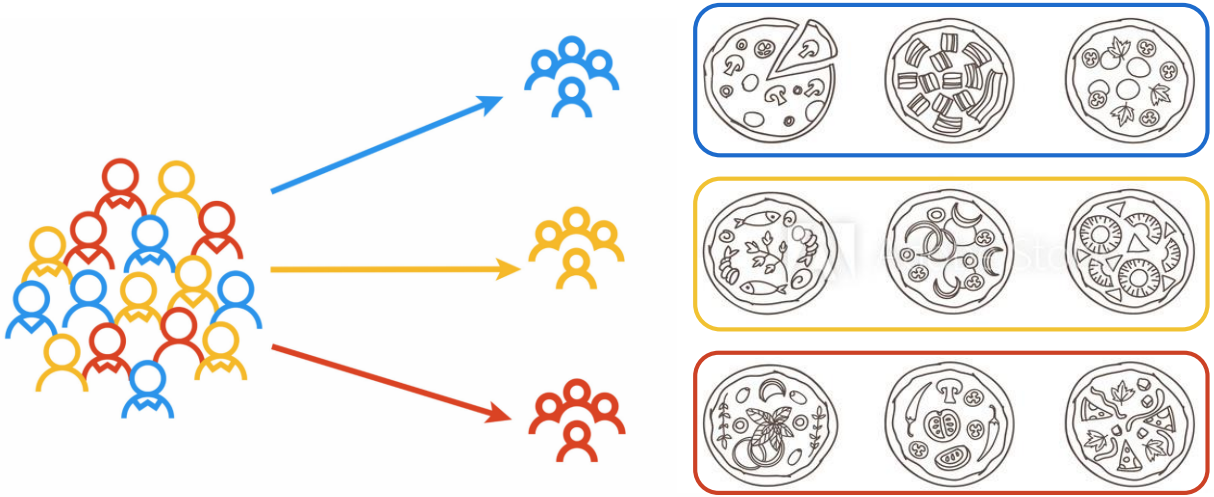
Report Outline

Introduction	• Client profile • Problem statement • Approach and Success criteria • Assumptions
Strategic Decisions	• Analytics workloads and Value Chain Analysis • Data requirements, sources, and assessment • Recommended provisioning solution architecture by component
Tactical Decisions	• Analytics workloads and Value Chain Analysis • Data requirements, sources, and assessment • Recommended provisioning solution architecture by component
Operational Decisions	• Analytics workloads and Value Chain Analysis • Data requirements, sources, and assessment • Recommended provisioning solution architecture by component
Conclusion	• Overall provisioning solution • Recommended next steps
Appendices	• Documenting process • Resources

Introduction

- Problem statement
- Client profile and objectives
- Approach and Success criteria
- Assumptions

Problem Statement



Our client's business ***strategy*** is to identify **high value customers**, position their brand with **high mark-up product offerings** and target suitable **locations for expansion** that will provide access to their desired customers in markets that are not already saturated.

In this report, we will recommend data provisioning solutions to support specific strategic, tactical and operational decisions, and an overall architecture that will bring together their existing data with complimentary additional data sources to enable a data-driven expansion plan.

Client Profile

- Industry: Food Services
- Sector: Pizza Restaurant
- Business Model: Full Service (comprising customer take-out, Home Delivery & Dine-in)
- Geography: North America
- Size: Currently (Assumption: 1 outlet)
- Menu: Pizza & Salads, alcoholic and non-alcoholic beverages
- Business Strategy: Utilize Data-Driven Analysis to inform the Company’s growth objective

Client Objectives

- Expansion
  - Acquire a lease on a new location within the year
  - The restaurant will acquire 3 brick & mortar locations within 2 years and will distribute variable costs across its locations to benefit from economies of scale
  - Additional items to be added to the menu

- New Markets
  - Will target upscale market with gourmet pizzas (regular and healthy options)
  - Begin online order processing

Client Objectives

- Finances
  - The restaurant will begin to charge premium prices
  - Utilizing inventory to its maximum potential, generating maximum revenue
- Relevance
  - Grow online presence via social media (Twitter, Facebook, Instagram, etc)
  - Through its online presence, always be “in the know”, regularly adding novel types of ingredients
  - Brand products as healthy and nutritious
- Health Conscience
  - The restaurant will place a higher importance on health and nutrition
  - Change image by offering healthier options to its patrons

Project Approach

1. Conduct Value Chain Analysis to identify which strategic, tactical and operational decisions are required and the analytic workloads to support these decisions.
2. Gather sample data sets and recommend complementary data sources to ensure our recommended solution enables these decisions and workloads.
3. Consider potential metrics to recommend for these decisions.
4. Refine and consolidate value chain analysis, recommended data sets and metrics to ensure alignment with requisite decisions.
5. Design draft solutions for each component.
6. Design an overall solution, iterate, and refine components to ensure continuous alignment with client’s business strategy.
7. Summarize recommendations in a presentation and report.

Project Success Criteria for Recommended Architecture

- ✓ Comprehensive and integrated
  - Provides overall design advice on how the system can work as a whole.
  - Provides component-level design advice on how to optimize, align and extract value from data.
- ✓ Relevant, feasible and actionable
  - Aligned to client business objectives and context (e.g. available data, business processes).
  - Offers clear and practical next steps.
- ✓ Adds value, drives continual improvement and innovation
  - Identifies new ways for client to achieve business value through expansion, increased profit margins, greater system alignment and operational efficiencies.
  - Creates an analytics-driven system that enables the client to continue to generate insights into customers, product offerings, locations, operations, profit margins to align with current and future strategic direction.

Assumptions

- We are able to speak for the client in terms of business objectives, metrics, success criteria, normally this would require significant iteration.
- Data sets listed in the assignment as ‘assumed’ are available, of sufficient quality and currently in excel spreadsheets due to small scale of the business,
- Client has relevant data on customer preferences and satisfaction.
- Client has engaged user experience experts to provide experience maps for customer personas.
- Client has safety cameras on location and uses data-enabled systems (e.g. fire alarm, security alarm)
- Client has a relatively high level of analytics capability, or is able to engage contractors with this skill set on an ongoing basis as required (noting that some reports will require a simple user interface for non-expert users).

Strategic Decisions

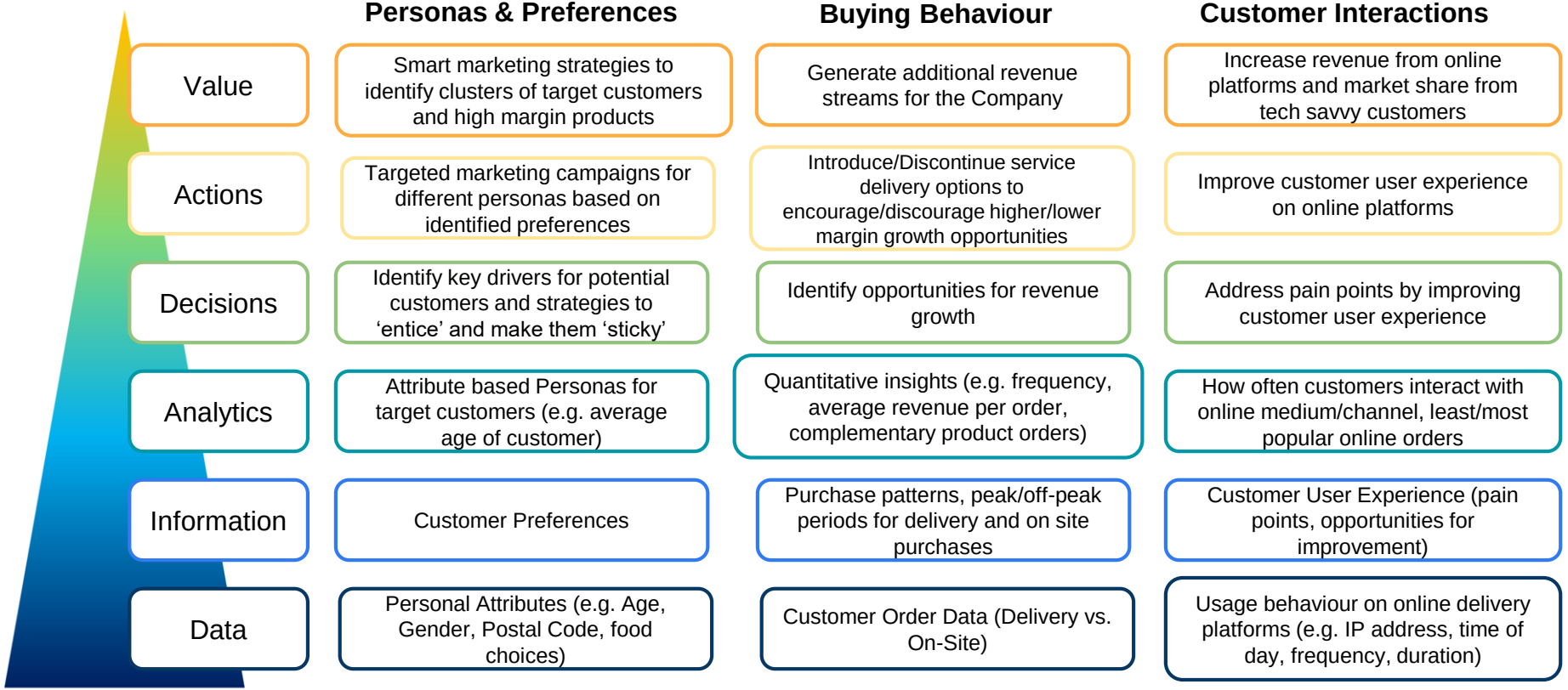
- Analytics workloads and Value Chain Analysis
- Data requirements, sources, and assessment
- Recommended provisioning solution architecture by component

Value chain analysis - Strategic decisions

Tactical Decisions

- Analytics workloads and Value Chain Analysis
- Data requirements, sources, and assessment
- Recommended provisioning solution architecture by component

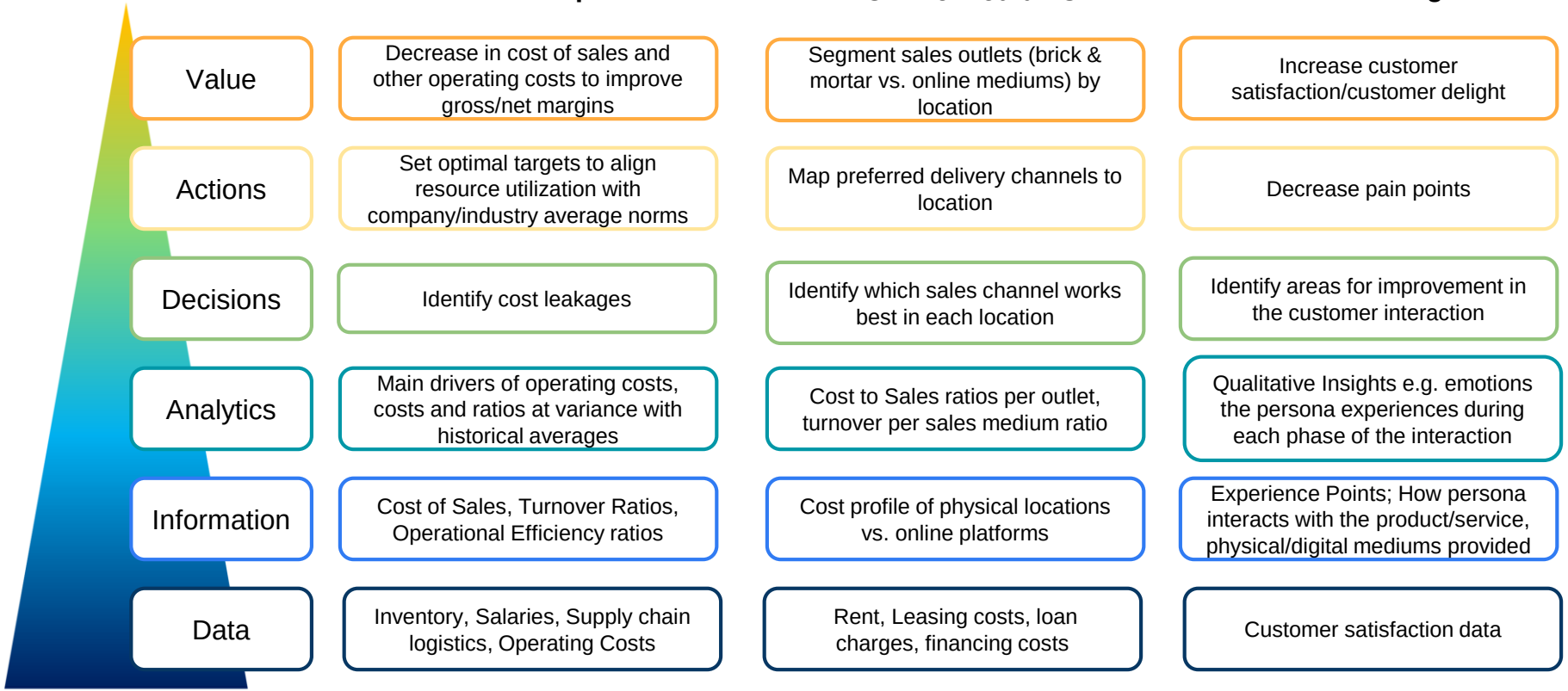
Value chain analysis - Tactical decisions



Operational Decisions

- Analytics workloads and Value Chain Analysis
- Data requirements, sources, and assessment
- Recommended provisioning solution architecture by component

Value chain analysis - Operational decisions



Matrix of Analytic Categories by Function

Functional Levels	Analytics Categories				
	Descriptive	**Diagnostic**	**Discovery**	**Predictive**	**Prescriptive**
Strategic			What products to promote? What Deals to offer?	How much revenue can we generate from organic traffic?	
Tactical	How well are we utilizing the supply inventory	What is the hourly capacity usage of our ovens		How many orders we can expect every hour?	
Operational		What percentage of orders are delivered within 40 minutes of the order?		How is our customer satisfaction rating going to change if we share tracking updates?	Which off-shelf items go out of stock the most?

Strategic Decisions

- Past Sales data
- Web scraping of competitors web data
- Search keywords and online search ranking data

Tactical Decision

- Machine Utilization and maintenance data
- Calls and order volume data
- Inventory and Finance Data (BS)

Operational Decision

- Delivery data
- Customer inquiry data

Decision Support

- Recommended Metrics
- Approach

Recommended Metrics

- Revenue and profits
  - Measured by increase or decrease in orders
  - Increased profit margin by client segment, product offering, client location
  - Increased average order values and repeat order percentage
- Expansion into new markets (products and locations)
  - Measured by number of orders or new products and additional menu items
  - Measured by number of online pizza orders
  - Measured by increase in orders in areas targeted for expansion
- Online presence via mediums such as Instagram, Facebook, Twitter, etc
  - Measured by number of online orders
  - Measured by number of followers with increased engagement metrics (likes,shares,comments)
  - Higher ranking on organic search results

Recommended Metrics (cont’d)

- Maximize inventory potential
  - Measured by decrease in costs
  - Measured by decrease of inventory lifecycle
  - Measured by high inventory turnover
- Client satisfaction
  - Measured by increase or decrease of orders
  - Measured by customer feedback and surveys
- Loyalty Program/Repeat Customers
  - Measured by revenue generated from the Loyalty Program
  - Measured by number of active program members
  - Measured by repeat order percentage from customers

Recommended Approach

- We are considering a relational approach for our data modeling and to build the schema for the following reasons:
  - ⇨ Schemas and queries are easier to work with for the end user
  - ⇨ Structural independence and improved performance
  - ⇨ Scalable and easier to use high level query tools (e.g. MySQL)
  - ⇨ Each table will have a primary key and a timestamp to store data efficiently
  - ⇨ Better suited for dynamic and time-variant data
  - ⇨ Better security, easier access, operations are based on set theory making storage efficient

- The analytic workloads we have identified will be split along the following:
  - ⇨ Strategic decision making using predictive analytics. This would include predictions for expected/average growth rate in revenue, profits, customers and orders.
  - ⇨ Tactical decision making using prescriptive or diagnostic analytics for marketing. For example which product or customer segments to target and how much discounting to apply for products, orders and bundled purchases.
  - ⇨ Operational decision making using descriptive and discovery analytics. Examples include average inventory utilization, minimum/maximum duration of delivery, frequency of commingled purchases/orders.

Recommended Approach (cont'd)

- We plan to follow Inmon's Top Down approach for the following reasons:
  - ⇨ Our data is subject oriented and has dependencies which are best addressed using a top down approach.
  - ⇨ Our data is non-volatile. For example once an order has been placed we cannot go back and change the information even in future.
  - ⇨ Changes in data need can be closely tracked. For example status of orders, reconciliations of supplies to inventory, website and online interactions.
  - ⇨ Integrated view at an enterprise level for multiple stores.
  - ⇨ Data is closely linked and residing in multiple domains - finance, marketing, operations, sales, human resources among others making this approach more suitable.

- Certain disadvantages of a top down approach include
  - ⇨ At inception, the cost is higher and complex to design. To mitigate this, the costs can be amortized amongst the locations after the initial setup so maintenance & operations costs are spread evenly.
  - ⇨ Schemas are complex to build but they ensure data integrity and errors such as duplicate records and null values.

Data Requirements

- Internal Data Sets
- Complementary open data sets
- Structure, format, size, quality, etc.

Internal Data Sets

Data Set	Assumed Contents	Volume
Financial Information Data	General Ledger	8 Years
Staff hiring and performance data	Staff names, years of service, salary, performance	8 Years
Supplier performance data	Fulfilled orders, completeness of orders, delivery times	8 Years
Energy use data	Energy usage	8 Years
Customer satisfaction data	Feedback from customers surveys	8 Years
Safety data	History of injuries	8 Years
Marketing and Promotions data	Flyers, advertising	8 Years
Health inspection data	History of inspections	8 Years
Inventory management data	Products, volumes	8 Years
Menu items data	History of menu items	8 Years
Customer order data	History of which menu items were ordered by customers	8 Years
Customer delivery data	History of which menu items were ordered for delivery and addresses delivered to	8 Years

External (Open) Data Sets

Data Set	URL	Description	Format	Volume
Yelp Reviews	www.yelp.com/dataset/challenge	businesses, reviews, and user data	JSON	6,685,900 reviews 192,609 businesses 1,223,094 tips by 1,637,138 users
Pizza restaurants and Pizzas on their Menus	www.kaggle.com/datafiniti/pizza-restaurants-and-the-pizza-they-sell	A list of pizza restaurants, 3,500 pizzas, and their menu prices. The dataset includes the category, name, address, city, state, menu information, price range, and more for each pizza restaurant.	CSV	5.51 MB
Restaurant & consumer data	archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data	Effects of relevant contextual features in the performance of a restaurant recommender system.	CSV	5.51 MB
Real Estate	www.kaggle.com/marianeth/real-estate-across-the-united-states-inventory/version/1	Track and manage the government's real property assets. To store inventory data, building data, customer data, and lease information. Consists of both owned and leased buildings.	CSV	1.26 MB
Income by Location	www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations	Records on US Household Income Stats & Geo Locations. Statistical information - such as Mean, Median, Standard Deviation. Geographic location.	CSV	4.79 MB
Location Crime data	www.kaggle.com/knitemblazor/crime-data-of-us-from-2010	Crime information by location. Description and severity with date and time	CSV	422 MB

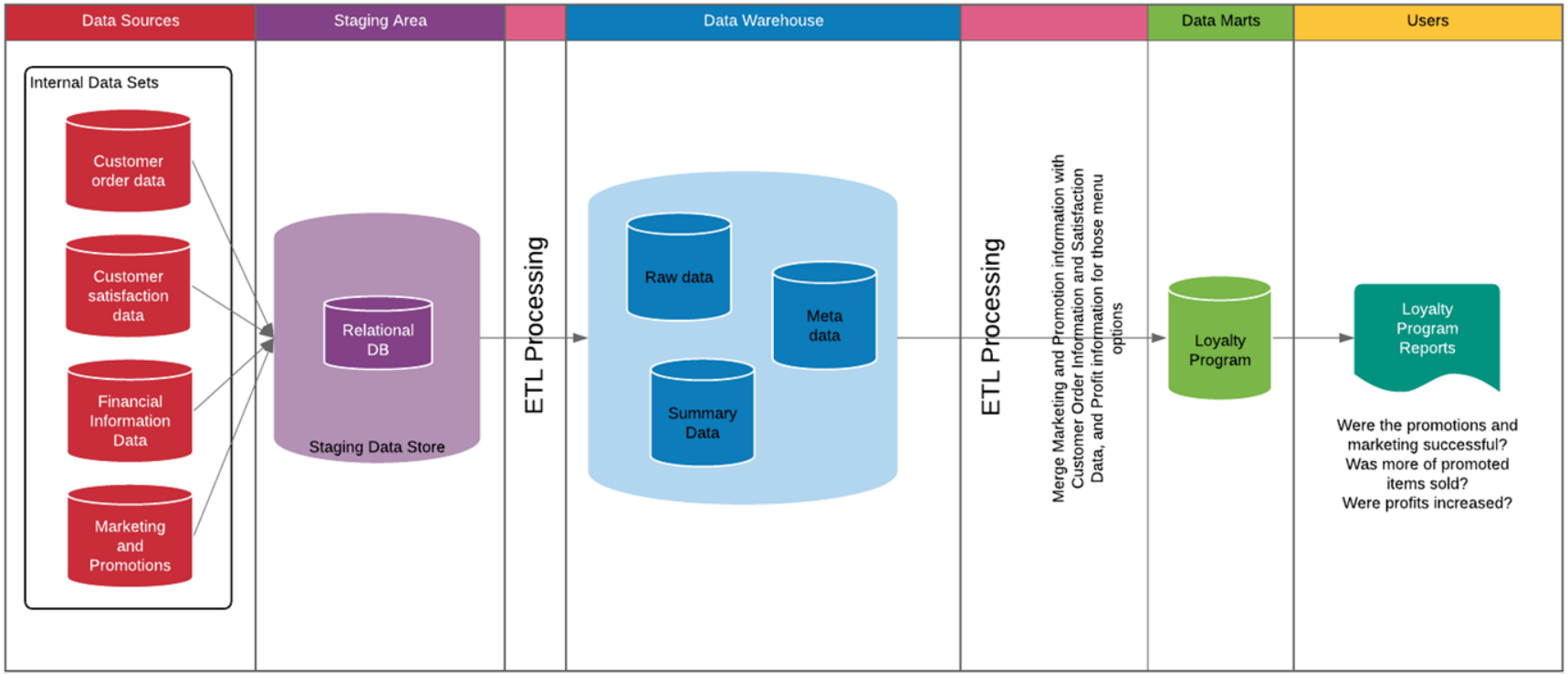
Data Provisioning Architecture

- For each of the Nine Workloads
- Overall Recommended Data Provisioning Architecture

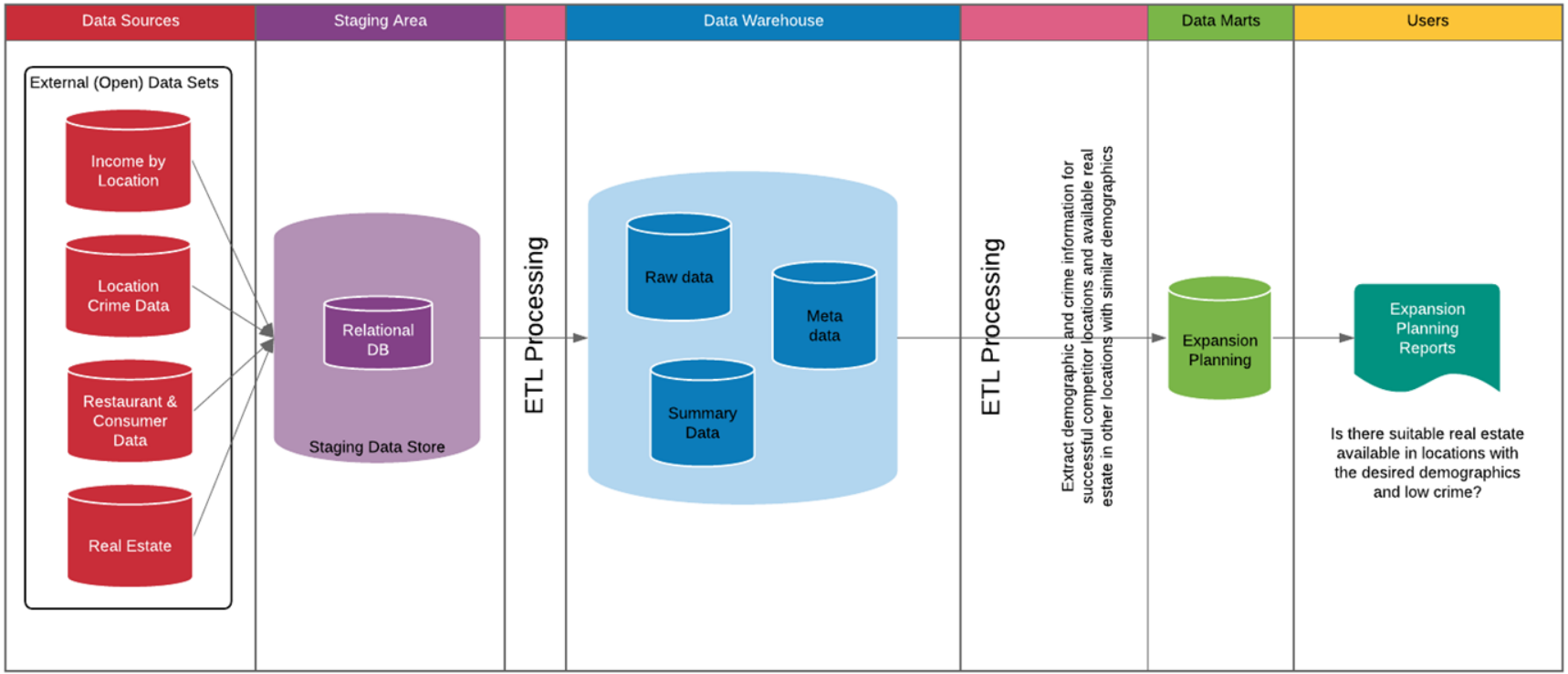
Workload Provisioning Architecture

- Data will be brought into the Staging Area as is
  - CSVs, spreadsheets, and relational DB tables will be stored in a relational DB in the staging area
  - JSON documents will be stored in a Document DB
- Data relevant to each of the workloads will be copied to the Data Warehouse thru an ETL process, the data will also be
  - Cleaned
  - Summarized as required
  - Linked to other data for that workload
- The data will then be copied (through an ETL process) to a Data Mart specifically for that workload where it can be analyzed

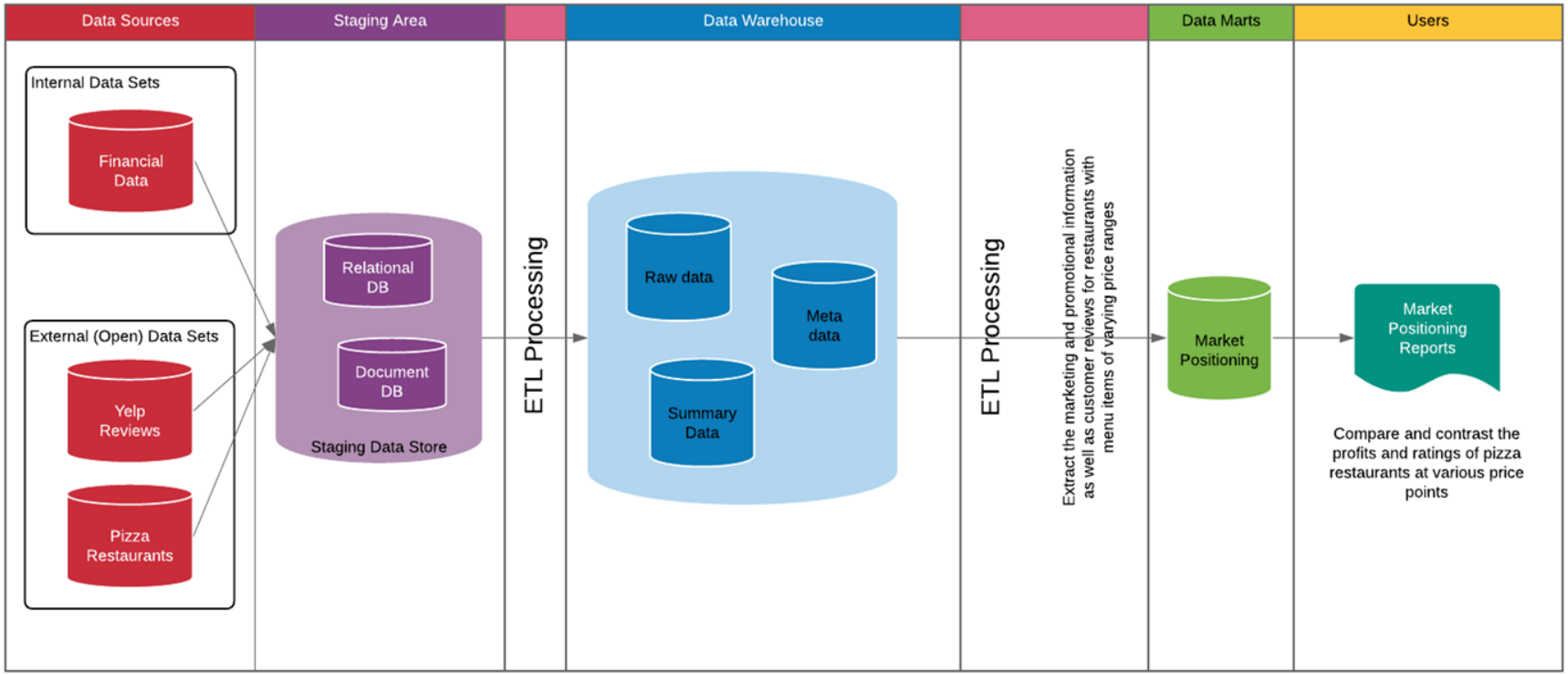
Loyalty Program



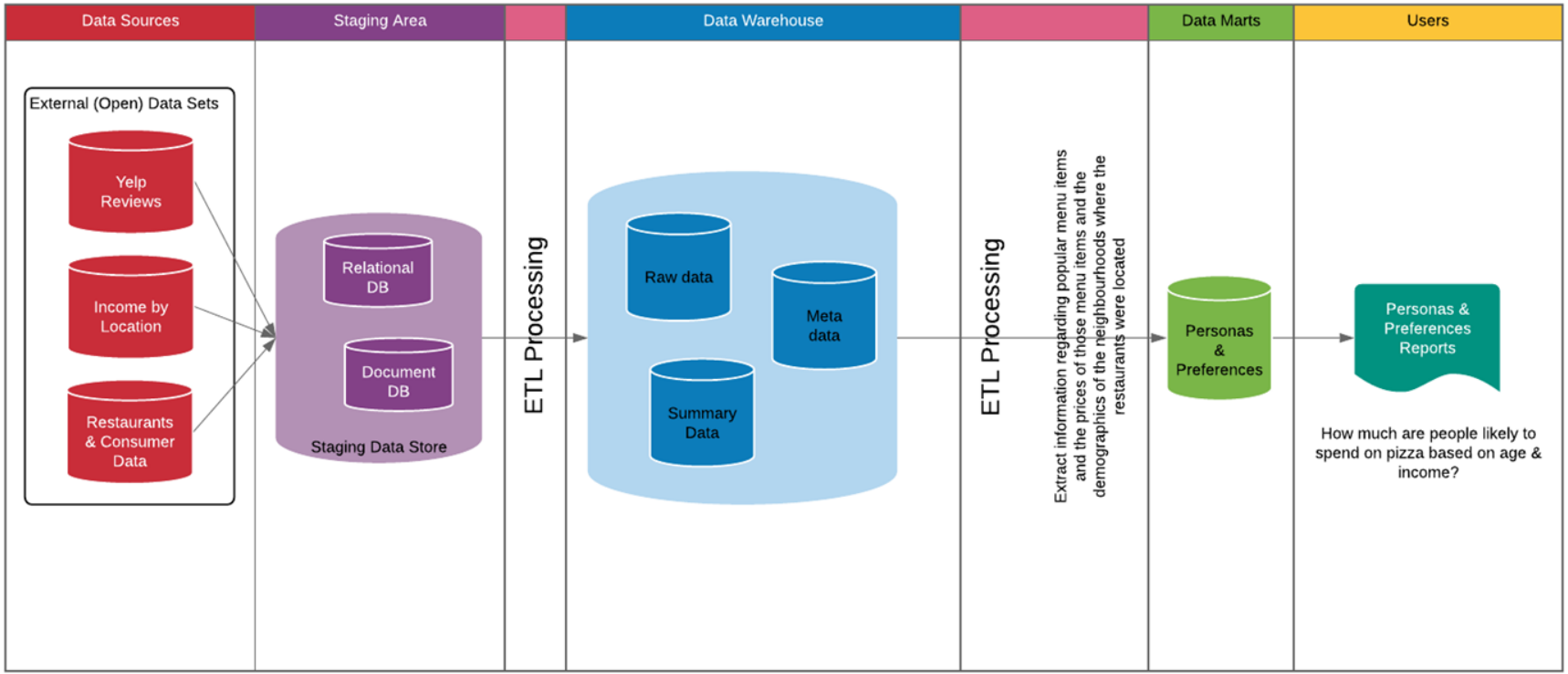
Expansion Planning



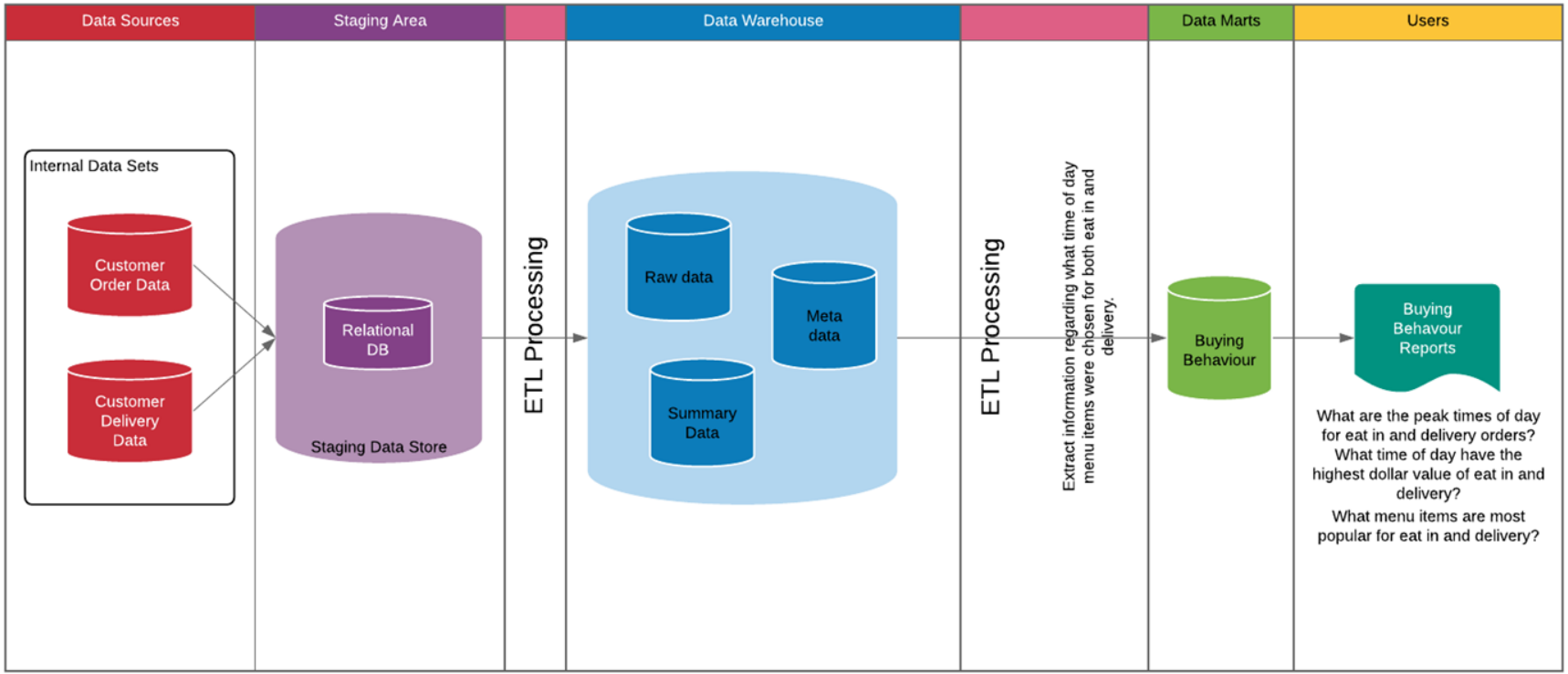
Market Positioning



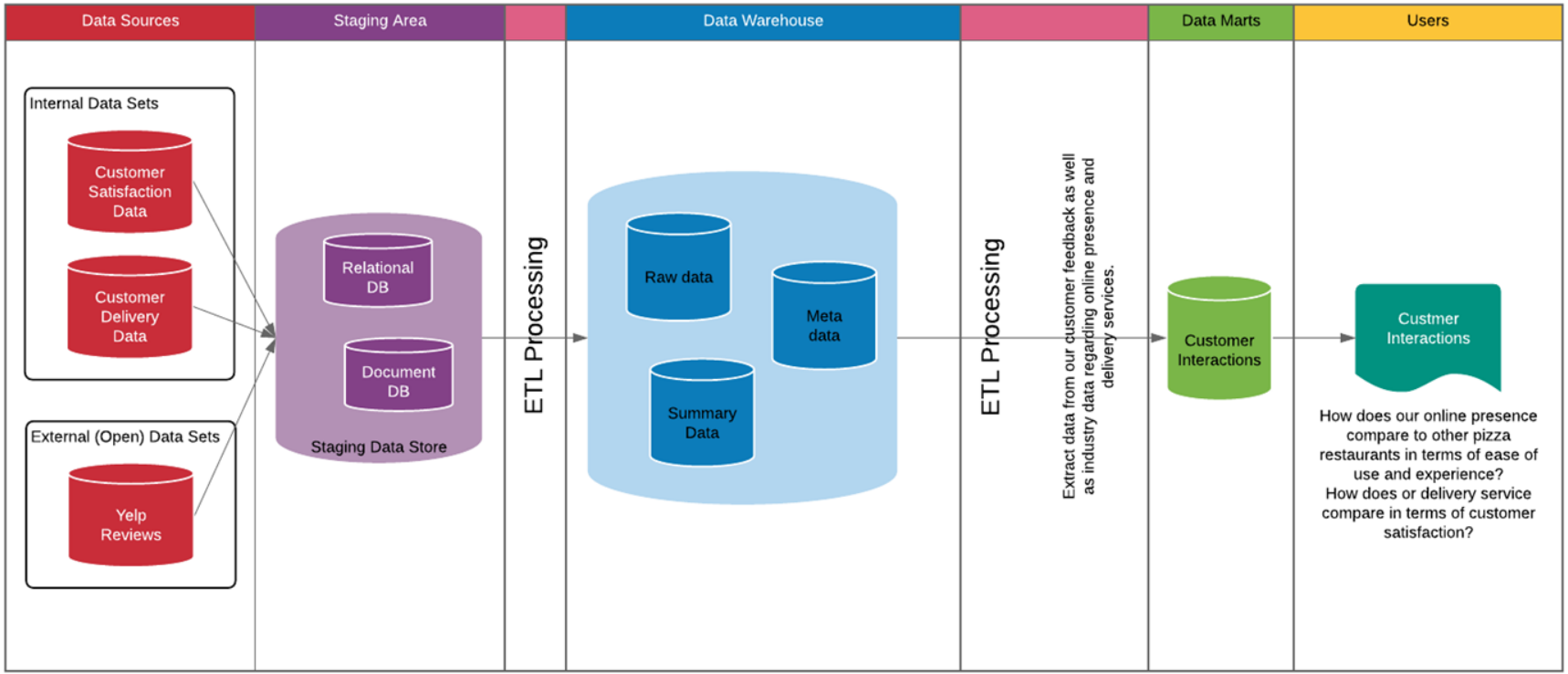
Personas and Preferences



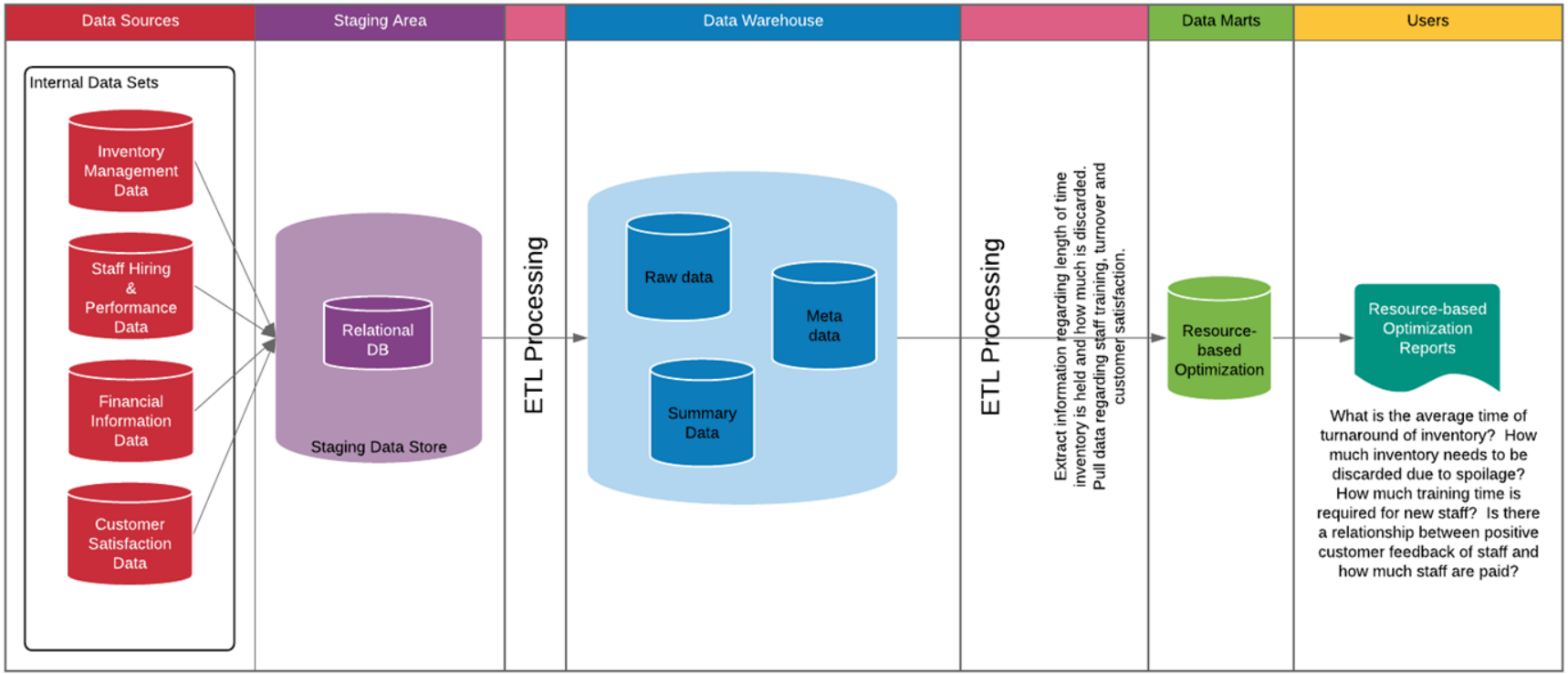
Buying Behaviour



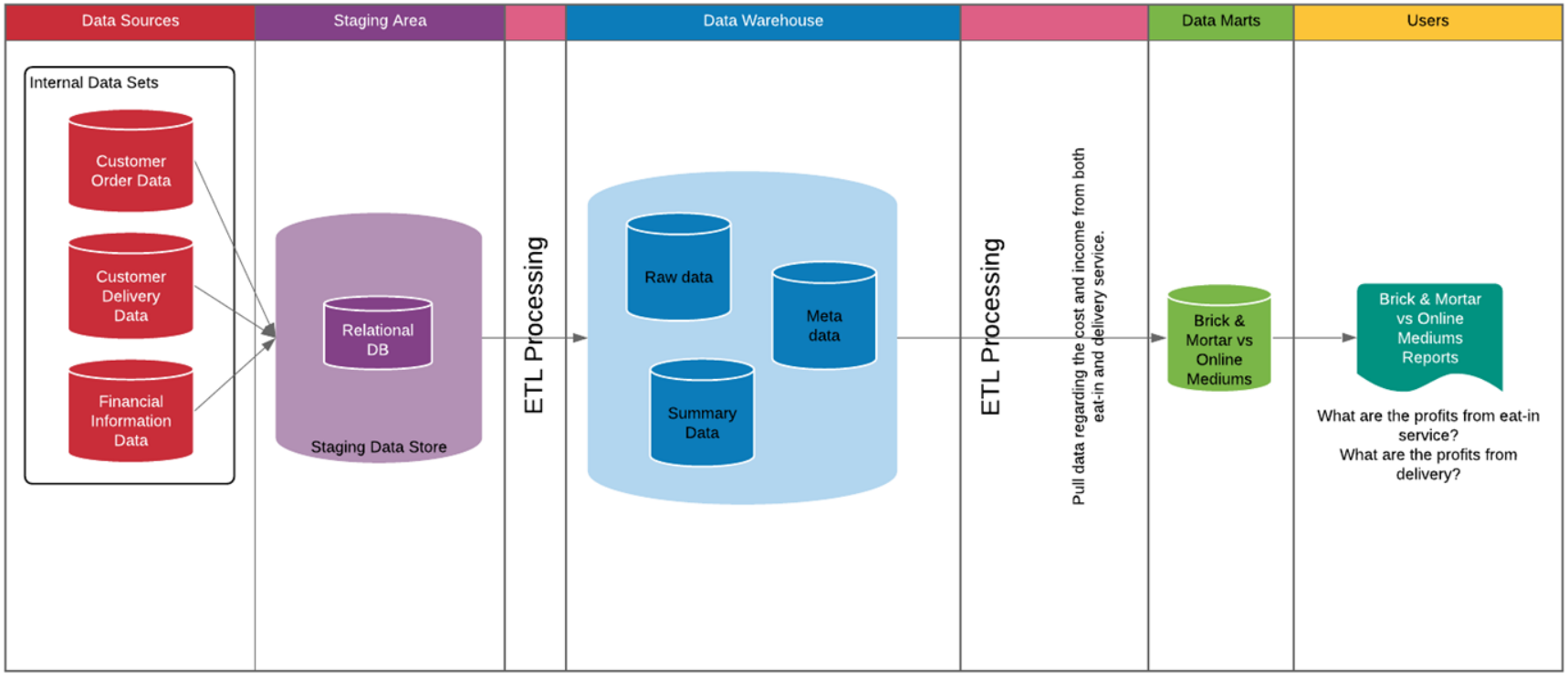
Customer Interactions



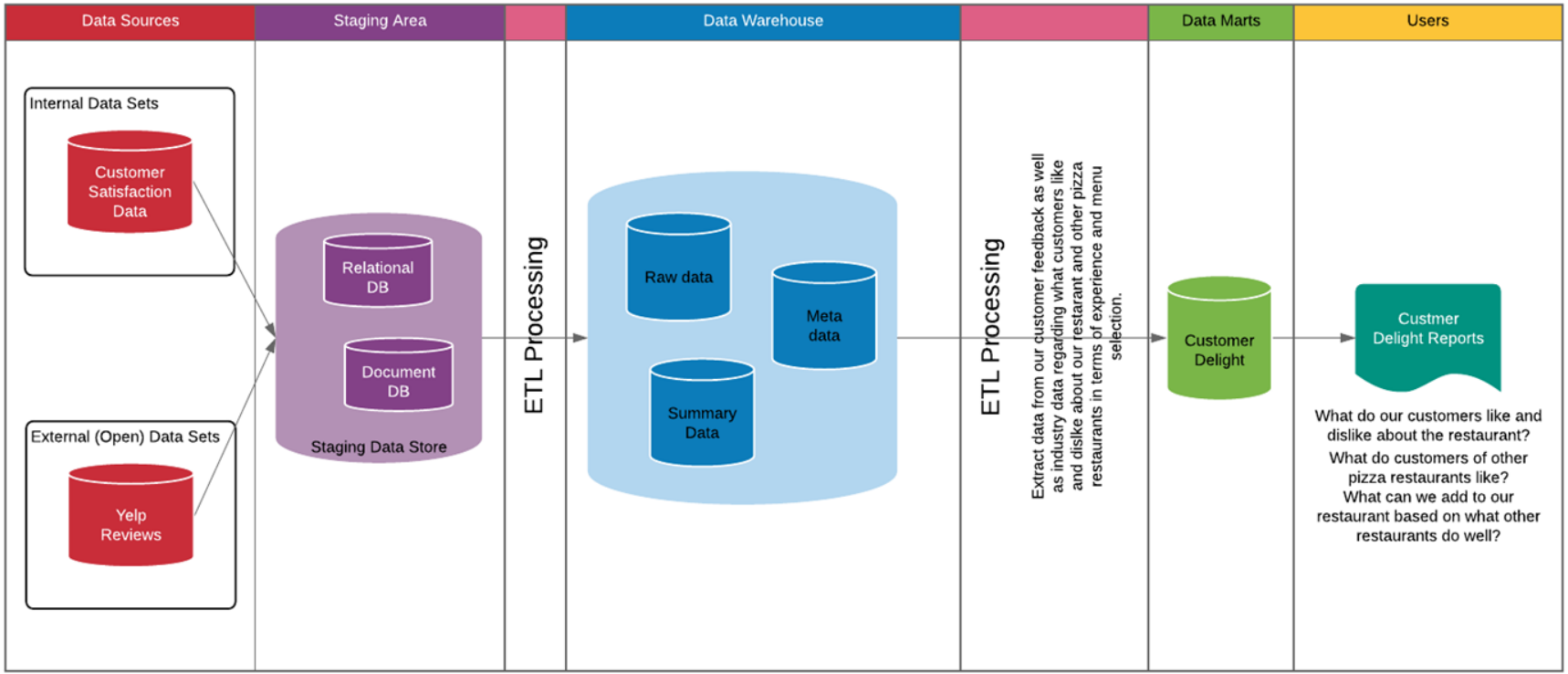
Resource-based Optimization



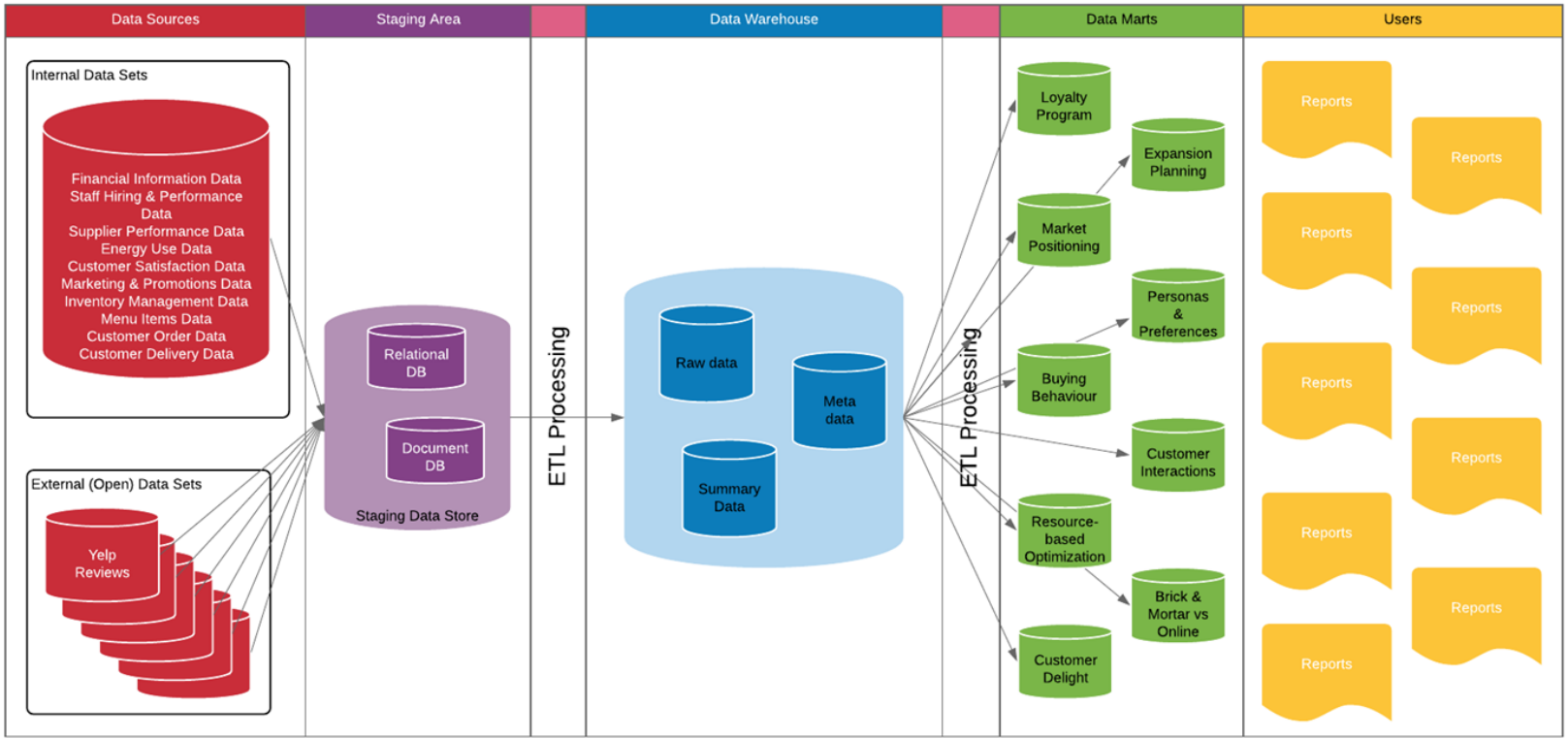
Brick & Mortar vs. Online Mediums



Customer Delight

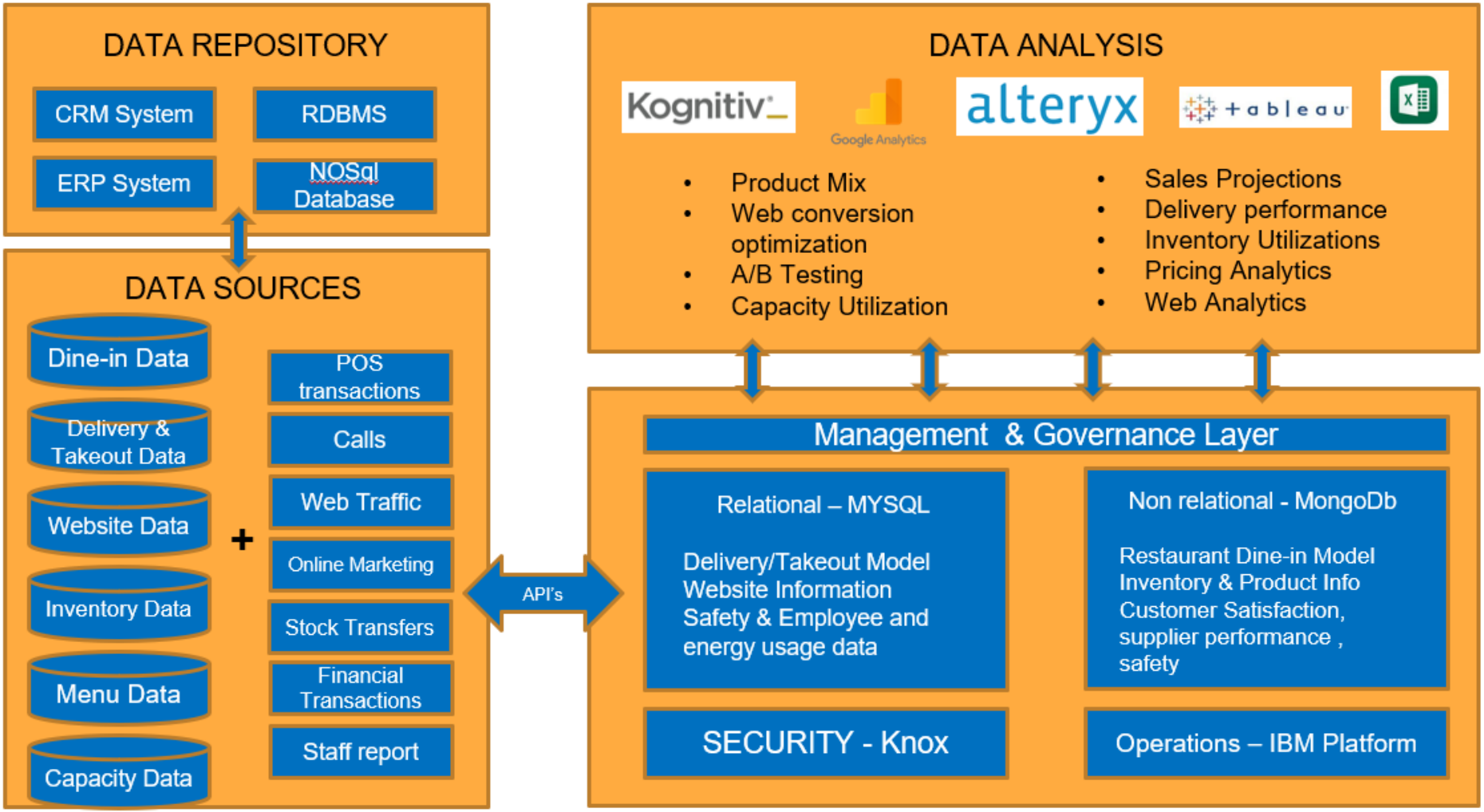


Overall Data Provisioning Architecture



Data Architecture

- Overall provisioning solution


DATA REPOSITORY
CRM System
RDBMS
ERP System
NOSql Database
DATA SOURCES
Dine-in Data
Delivery & Takeout Data
Website Data
Inventory Data
Menu Data
Capacity Data
+
POS transactions
Calls
Web Traffic
Online Marketing
Stock Transfers
Financial Transactions
Staff report
API's
DATA ANALYSIS
Kognitiv
Google Analytics
alteryx
tableau
Product Mix
Web conversion optimization
A/B Testing
Capacity Utilization
Sales Projections
Delivery performance
Inventory Utilizations
Pricing Analytics
Web Analytics
Management & Governance Layer
Relational – MYSQL
Delivery/Takeout Model
Website Information
Safety & Employee and energy usage data
Non relational - MongoDb
Restaurant Dine-in Model
Inventory & Product Info
Customer Satisfaction, supplier performance , safety
SECURITY - Knox
Operations – IBM Platform

Relational vs Dimensional

We chose a relational approach as compared to dimensional because

- It provides structural independence and improved performance
- Easily accessible and most commonly available in the market
- Better suited for dynamic and time variant data
- Allows querying through high level languages (mysql)

Relational vs Non relational

We chose a combination of relational and non relational approach because

- Relational is scalable vertically, Non-relational is scalable horizontally
- Different business models have different data requirements
- Speed + Efficiency (Mongo db) vs Size and non redundancy (Mysql)

Data Management

- Data Governance
- Meta Data
- Security

Data Governance

- Employ a Data Governance Model that conforms to Industry Best Practices e.g. a Virtual Data Governance Team.

- Forge a strong commitment and partnership between Business and IT Departments.

- Periodic (e.g. annual assessments) of Data Governance Models to address identified gaps or areas for improvement e.g. data quality, data ownership/custodianship.

Metadata

- Employ a robust Metadata Management Tool get a holistic view of the data flow and how the data evolves.

- The output of the Metadata Management Tool should be a data tracking map detailing the evolution of data through from the different source to targets systems.

- Metadata Management must be sacrosanct and incorporated in the Data Governance Model.

- The Categories of Metadata (Business, Process, Technical & Application) should be included in the scope of Metadata Management.

Security

- An important step in the Data Asset Lifecycle to protect the integrity of data as a Corporate Asset. Knox was selected as our Security Tool.
- SDP - Sensitive Data Protection tool for Enterprises
- EXT4 encryption FBE mechanism
- FIPS compliant Kernel crypto module

Business Intelligence

- **Alteryx -** An R-based tool used for predictive modeling, classification modeling, AB tests and Clustering
- **MS - Excel -** Standard data preparation and simple report making
- **Tableau -** Visualizations and Dashboards
- **Google Analytics -** To analyze website and ecommerce data
- **Kognitiv -** Marketing Intelligence

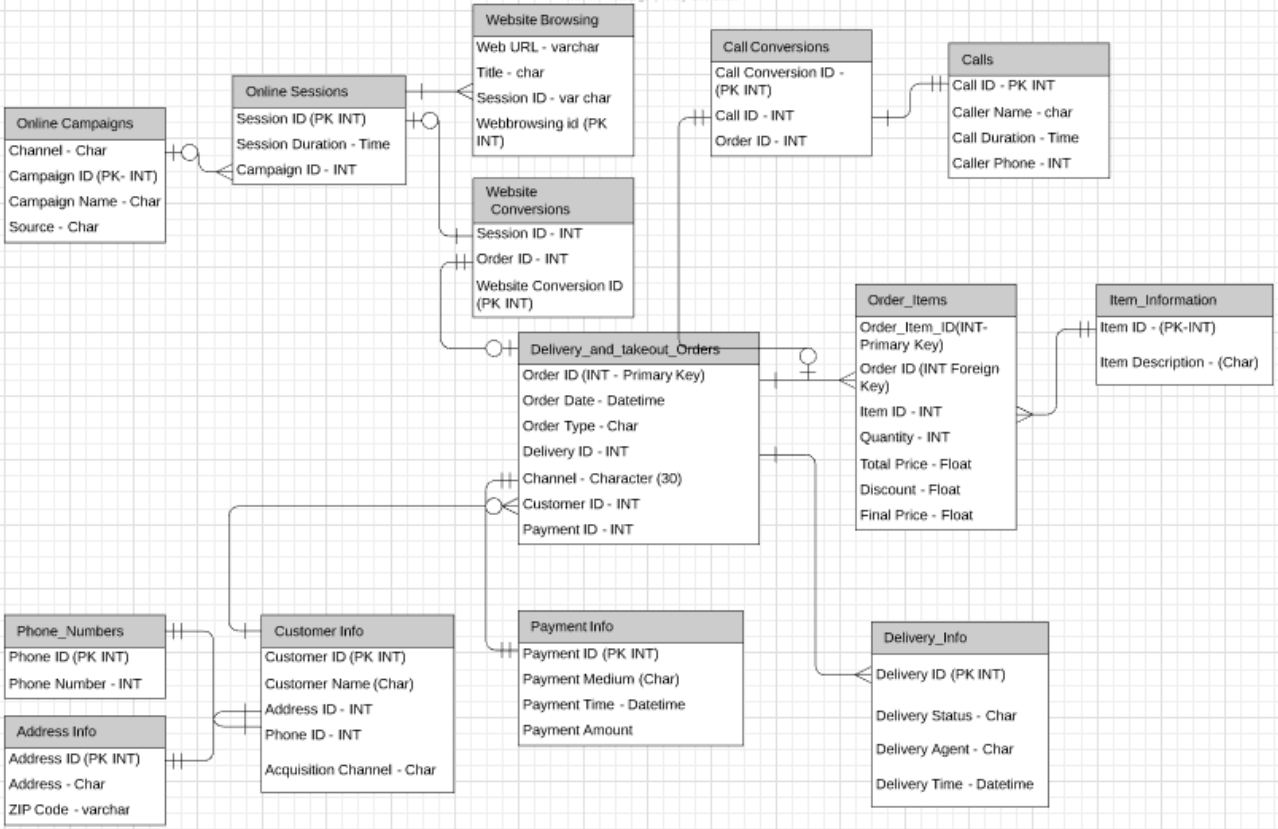
Appendices

- Documenting process
- Resources
- ERD Diagram

ERD Diagram and Code

Entity Relationship Diagram - Schema

Vaibhav Rastogi | July 26, 2019



```
CREATE TABLE `Item_Information` (
  `Item ID - (PK-INT)` <type>,
  `Item Description - (Char)` <type>);

CREATE TABLE `Call Conversions` (
  `Call Conversion ID - (PK INT)` <type>,
  `Call ID - INT` <type>,
  `Order ID - INT` <type>);

CREATE TABLE `Delivery_Info` (
  `Delivery ID (PK INT)` <type>,
  `Delivery Status - Char` <type>,
  `Delivery Agent - Char` <type>,
  `Delivery Time - Datetime` <type>);

CREATE TABLE `Online Campaigns` (
  `Channel - Char` <type>,
  `Campaign ID (PK- INT)` <type>,
  `Campaign Name - Char` <type>,
  `Source - Char` <type>);

CREATE TABLE `Website Conversions` (
  `Session ID - INT` <type>,
  `Order ID - INT` <type>,
  `Website Conversion ID (PK INT)` <type>);
```

```
CREATE TABLE `Address Info` (
  `Address ID (PK INT)` <type>,
  `Address - Char` <type>,
  `ZIP Code - varchar` <type>);

CREATE TABLE
`Delivery_and_takeout_Orders` (
  `Order ID (INT - Primary Key)` <type>,
  `Order Date - Datetime` <type>,
  `Order Type - Char` <type>,
  `Delivery ID - INT` <type>,
  `Channel - Character (30)` <type>,
  `Customer ID - INT` <type>,
  `Payment ID - INT` <type>);

CREATE TABLE `Calls` (
  `Call ID - PK INT` <type>,
  `Caller Name - char` <type>,
  `Call Duration - Time` <type>,
  `Caller Phone - INT` <type>);

CREATE TABLE `Phone_Numbers` (
  `Phone ID (PK INT)` <type>,
  `Phone Number - INT` <type>);
```

```
CREATE TABLE `Customer Info` (
  `Customer ID (PK INT)` <type>,
  `Customer Name (Char)` <type>,
  `Address ID - INT` <type>,
  `Phone ID - INT` <type>,
  `Acquisition Channel - Char` <type>);

CREATE TABLE `Online Sessions` (
  `Session ID (PK INT)` <type>,
  `Session Duration - Time` <type>,
  `Campaign ID - INT` <type>);

CREATE TABLE `Payment Info` (
  `Payment ID (PK INT)` <type>,
  `Payment Medium (Char)` <type>,
  `Payment Time - Datetime` <type>,
  `Payment Amount` <type>);

CREATE TABLE `Website Browsing` (
  `Web URL - varchar` <type>,
  `Title - char` <type>,
  `Session ID - var char` <type>,
  `Webbrowsing id (PK INT)` <type>);

CREATE TABLE `Order_Items` (
  `Order_Item_ID(INT- Primary Key)` <type>,
  `Order ID (INT Foreign Key)` <type>,
  `Item ID - INT` <type>,
  `Quantity - INT` <type>,
  `Total Price - Float` <type>,
  `Discount - Float` <type>,
  `Final Price - Float` <type>);
```

Resources

- https://www.propellercrm.com/blog/ideal-customer-profile-icp
- https://rapidboostmarketing.com/restaurant-marketing-defining-customer-profiles-for-your-restaurant/
- https://joinposter.com/en/post/how-to-write-a-target-customer-profile-for-a-restaurant
- https://www.franchisedirect.com/foodfranchises/pizzafranchises/pizzaindustryreportbusinessmodelsandmarketleaders2/80/276
- Sample company used as proxy for project scenario – Good Fellas Wood Oven Pizza: https://www.goodfellaspizza.ca//