

# Relationship between data categories

# Data categories (Types of data in an organization)

- **Data categories** are groupings of data with common characteristics or features. They are useful for managing the data because certain data may be treated differently based on their classification.

- Reporting data (e.g. aggregated sales data) → tables, graphs
- Transactional data (e.g. order information) →
- Master data (e.g. customer information) →
- Reference data (e.g. order status) →
- Metadata (e.g. created timestamps) →

EMPNO EMP

123	John
345	Ar

EMP-NO	EMP-NAME	LOC
:	:	:

# Data categories

- **Transactional data** relates to the transactions of the organization and includes data that is captured. It is the largest volume of data in the enterprise. The information that documents an exchange, agreement or transfer between organizations or individuals. information directly derived as a result of transactions.

## Example of business events include:

- Purchase by customers,
- A bill for product of services,
- Credits: funds added to an account.
- Interest, payroll, contracts

- Transactional Data is typically handled in operational applications - CRM, ERP, SCM, HR, etc..

customer      supplier      price  
AAA              BBB              1000

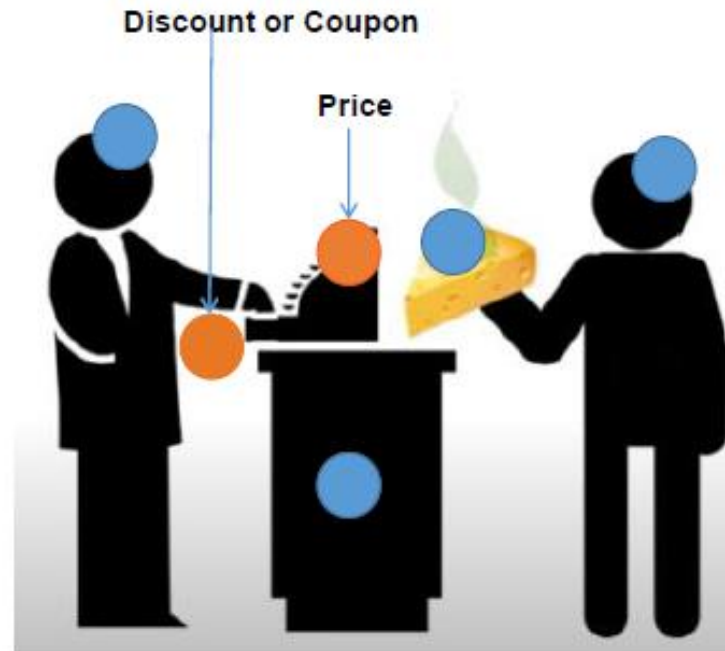
# Data categories

ID	Emp.Name	location
123	John	NAB
...		

- **Master Data:**
- **Master Data is key business information that supports the transactions. information model of business concepts, or entities, and how they relate to each other. identifiers that provides context about business data such as location, customer, product, asset.**
- Master data are key data that are shared by all Service Manager applications, and **are stored as records in support tables**. Master data is the type of data without which any transaction cannot be implemented and therefore it is mandatory for every organization.
- **set of identifiers that provides context about business data.**
- **It is commonly referred to as Places (locations, geography, sites, etc.), Parties (persons, customers, suppliers, employees, etc.) and Things (products, items, material, vehicles, etc.).**
- **Data which are agreed on and shared across the enterprise .**
- Product Data, Employee data, Transactions data, Tickets, Analytical Data.
- all personal attributes can be stored in various SAP standard info types as records with specific validity

# Transactional data vs Master data

## Master Data Versus Transactional Data



### Transaction Data

Represents the event which the master data participates –  
Purchasing of the Cheese

- Price
- Discount or coupon
- Method of payment



Master data to nouns

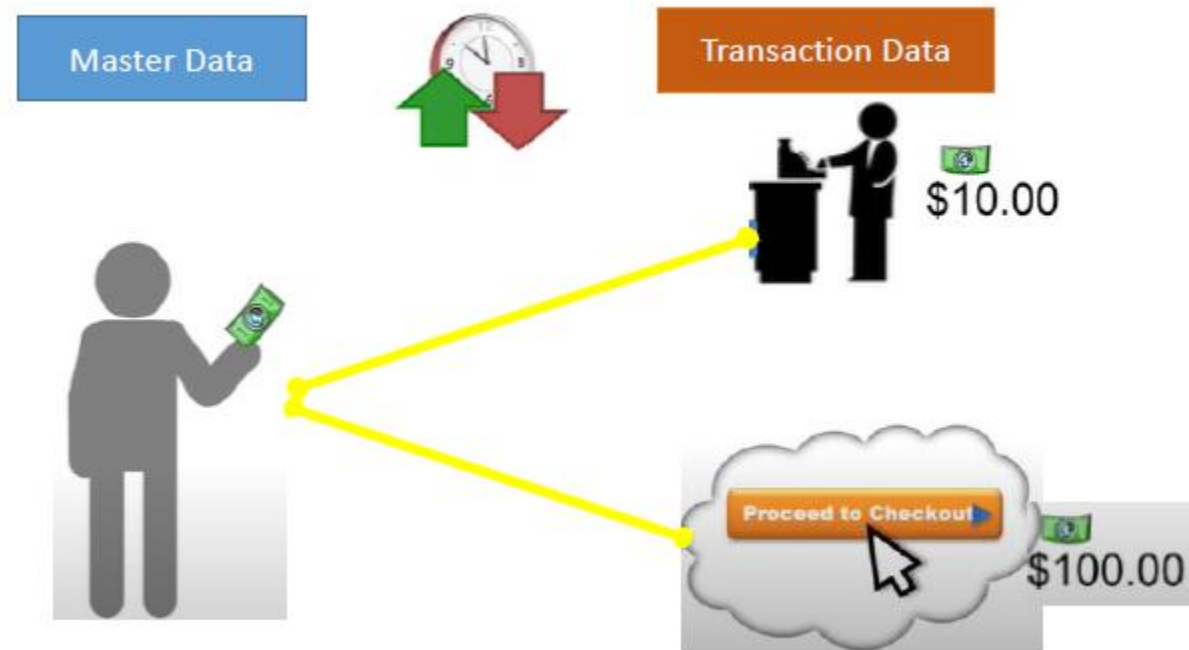
Transaction data  
to verbs

### Transaction Data



# Transactional data vs Master data

## Master Data Versus Transactional Data



# Transactional data vs Master data

Master Data	Transactional Data
<b>Represents the business objects that contain the most valuable, agreed upon information shared across an organization</b>	<b>Data that is periodically updated asynchronously over time as new information becomes available</b>
<b>Represents people, places or things that are related to an organization</b>	<b>Used by master data</b>
<b>Ex. Customer details, product details, employee details</b>	<b>Ex. Price, discount, method of payment</b>
<b>Does not change – not volatile</b>	<b>Change frequently - volatile</b>
<b>Can have issues related to consistency</b>	<b>Can have issues related to logic and quantity</b>



# Reference Data (Master Reference data)

- **Subset of master data that refers to the data that defines the permissible values to be used by other data fields. It is data that is referenced and shared by a number of systems. Reference data** is the information used to define other data.
- Example: country codes, currency, organization unit types.
- Financial product information, Entity information, pricing information.
- Two types: **static data**: won't change throughout the course of the transaction, and includes specifications such as the names.
- **Dynamic data** – information that changes throughout the course of the transaction, such as credit ratings or the pricing of the instrument



# Reference data vs Master data

## Master Data Versus Reference Data

### Structure

Master Data
Many Rows per Table
Many Columns per Table
Few Tables per Database
Changes More Quickly
Highly Shared within Enterprise

Reference Data
Few Rows per Table
Few Columns per Table
Many Tables per Database
Changes More Slowly
Highly Shared within Enterprise and between Enterprises

### Semantics

Master Data
Entities Poorly Defined
No Meaning At Row Level

Reference Data
Entities Well Defined
Individual Rows Have Meaning

### Business Rules

Master Data
Intelligent Keys
Unique Identification

Reference Data
Codes / Acronyms
Use of Standards

### Primary Keys

Master Data
Business Rules Act on Rows

Reference Data
Rows Drive Business Rules

# Reporting data → improve

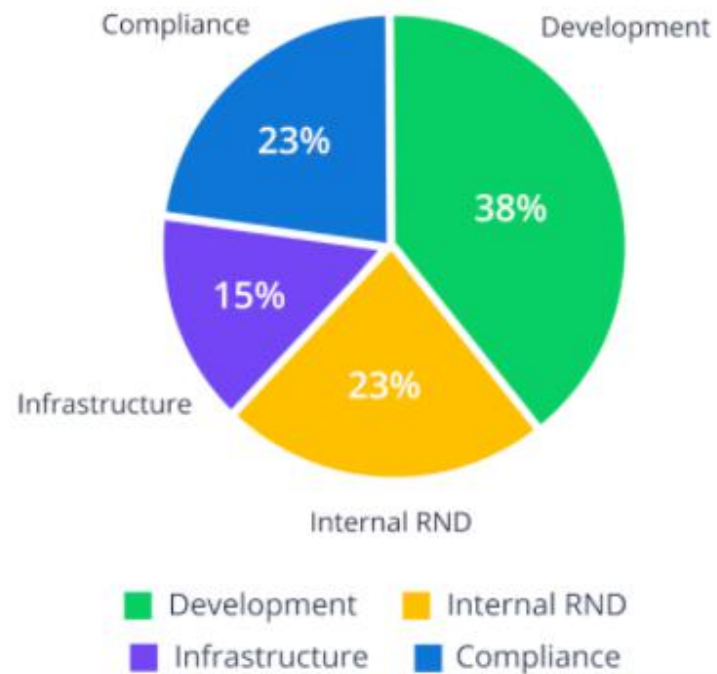
- **Reporting data is data organized for the purpose of reporting and business intelligence.** Data for operational reporting as well as data for enterprise (highly aggregated) reporting belong in this category.
- Reporting data is **created from** the transactional data, master data and master reference data.
- Helps to track the business and evaluate its performance.
- process of **collecting, merging, and visualizing raw data** from all available sources.
- **Form of tables, graphs or charts**

John      pig + v  
      .  
      .  
Arum    :

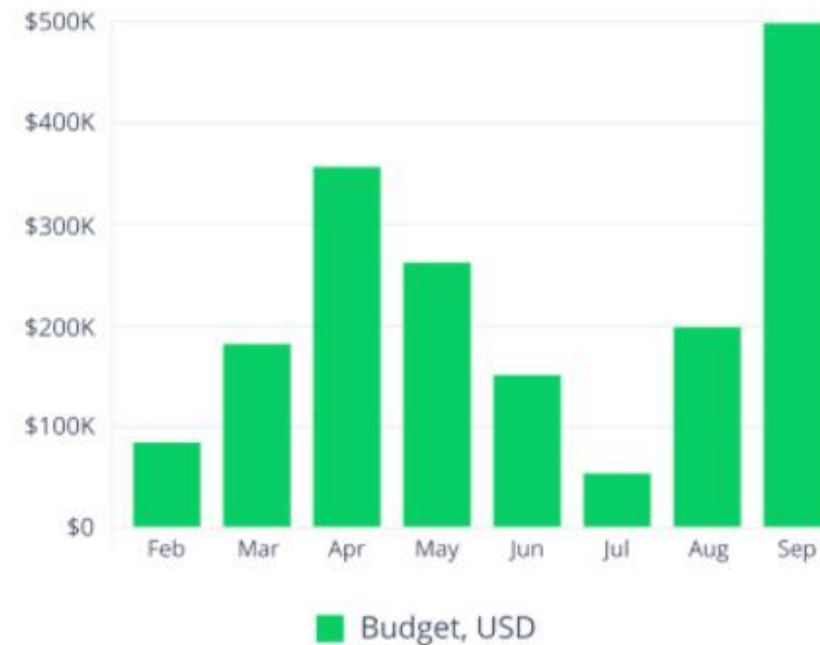
# Reporting data

Active Project Requests	New Project Requests	Total projects
27	13	42

Requests by Project type



Requested Budget by Month



Emp ID Emp Name Title Dept

# Metadata

- Data that describes other data. →
- **Examples: properties file, size, type, resolution, author, and create date. Software applications, documents, spreadsheets, and web pages are all examples that typically have associated metadata. Master data and reference data have related metadata.**

## FOUR MAIN TYPES OF METADATA



# Importance of metadata

- enterprises are increasingly investing in and betting on data to make better decisions, the amount of data we use is only set to increase.
- In order, to increase the shelf life of data, it's important for companies to invest in managing their metadata as well.

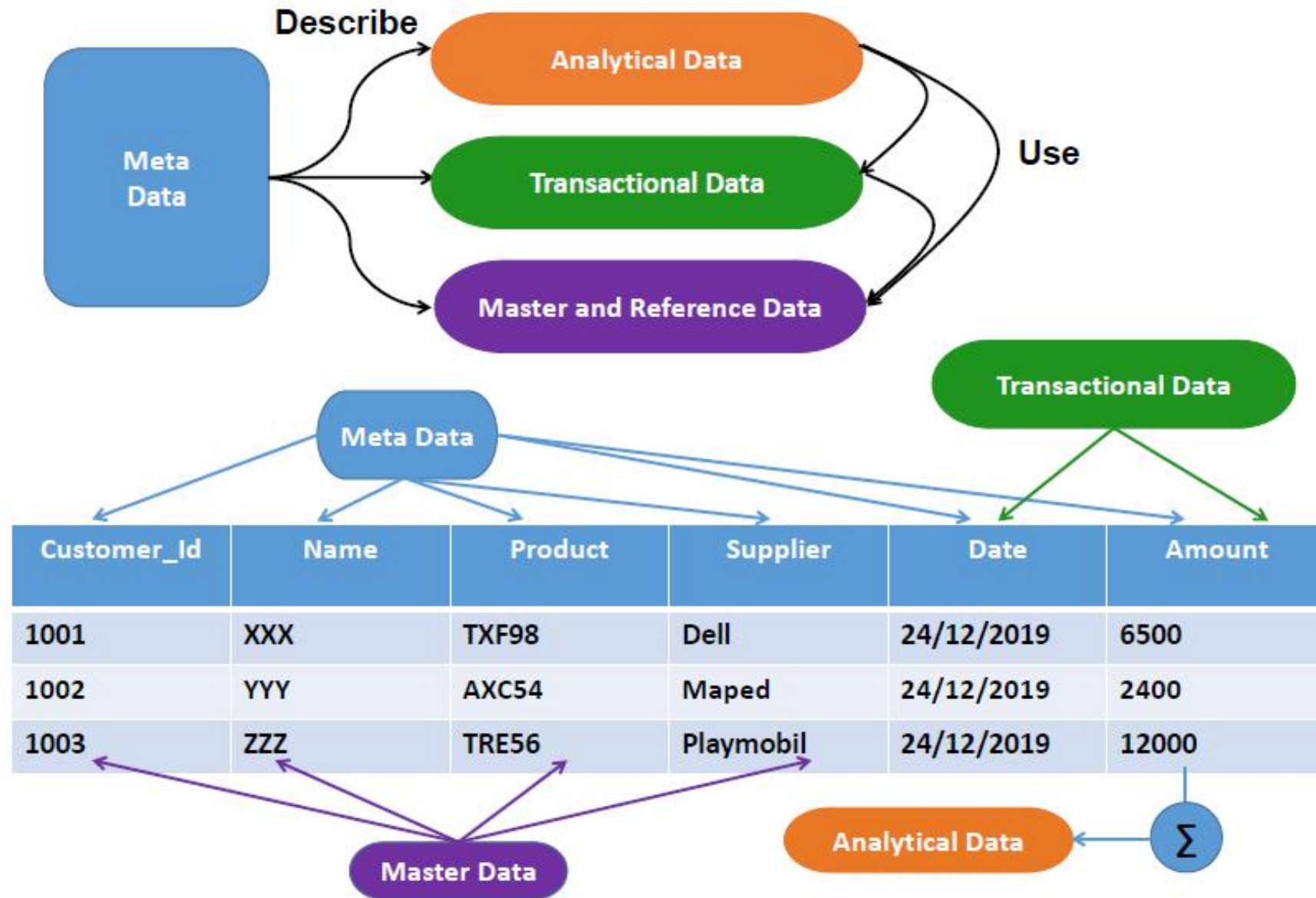
NAME	AGE	GENDER	HEIGHT (CM)
A	20	MALE	172
B	21	MALE	168
C	19	FEMALE	160
D	20	MALE	163

The diagram illustrates the distinction between metadata and data in a table. The first row, containing the column headers (NAME, AGE, GENDER, HEIGHT (CM)), is bracketed on the right and labeled 'METADATA'. The subsequent four rows, containing the data for individuals A, B, C, and D, are bracketed on the right and labeled 'DATA'.

# Additional data categories

- **Historical data:** **significant facts, as of a certain point in time, that should not be altered except to correct an error. They are important to security and compliance.**
- **Example:** point-in-time reports, database snapshots
- **Temporary data:** kept in memory to speed up processing. They are not viewed by humans and are used for technical purposes.
- **Example:** copy of a table that is created during a processing session to speed up lookups.



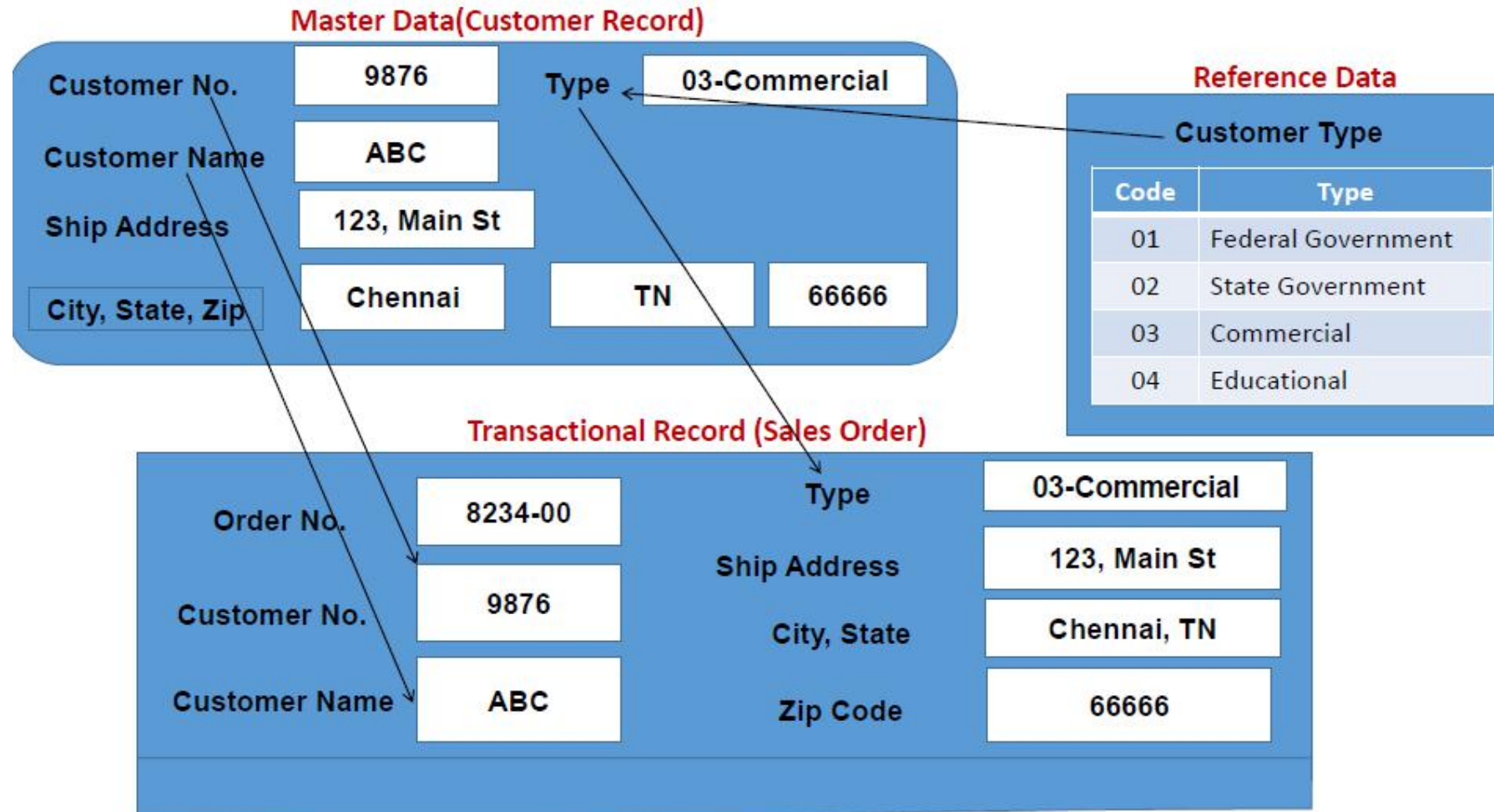


Transactional Data

Order No.	8234-00	Type	03-Commercial		
Customer No.	9876	Ship Address	123, Main St		
		City, State	Chennai, TN		
Customer Name	ABC	Zip Code	66666		
Product #	Description	Qty	Unit Price	Total Price	
90-123	SDV2	4	\$100	\$400	



# Relationship between data



# Relationship between data

## Metadata

Field name	Definition	Field Type	Field Length
Product Number	Unique identifier for a product assigned by the product manager	Numeric	20
Product Name	Short name of a product as it appears in the master list of products	Char	30

## Master Data(Product Record)

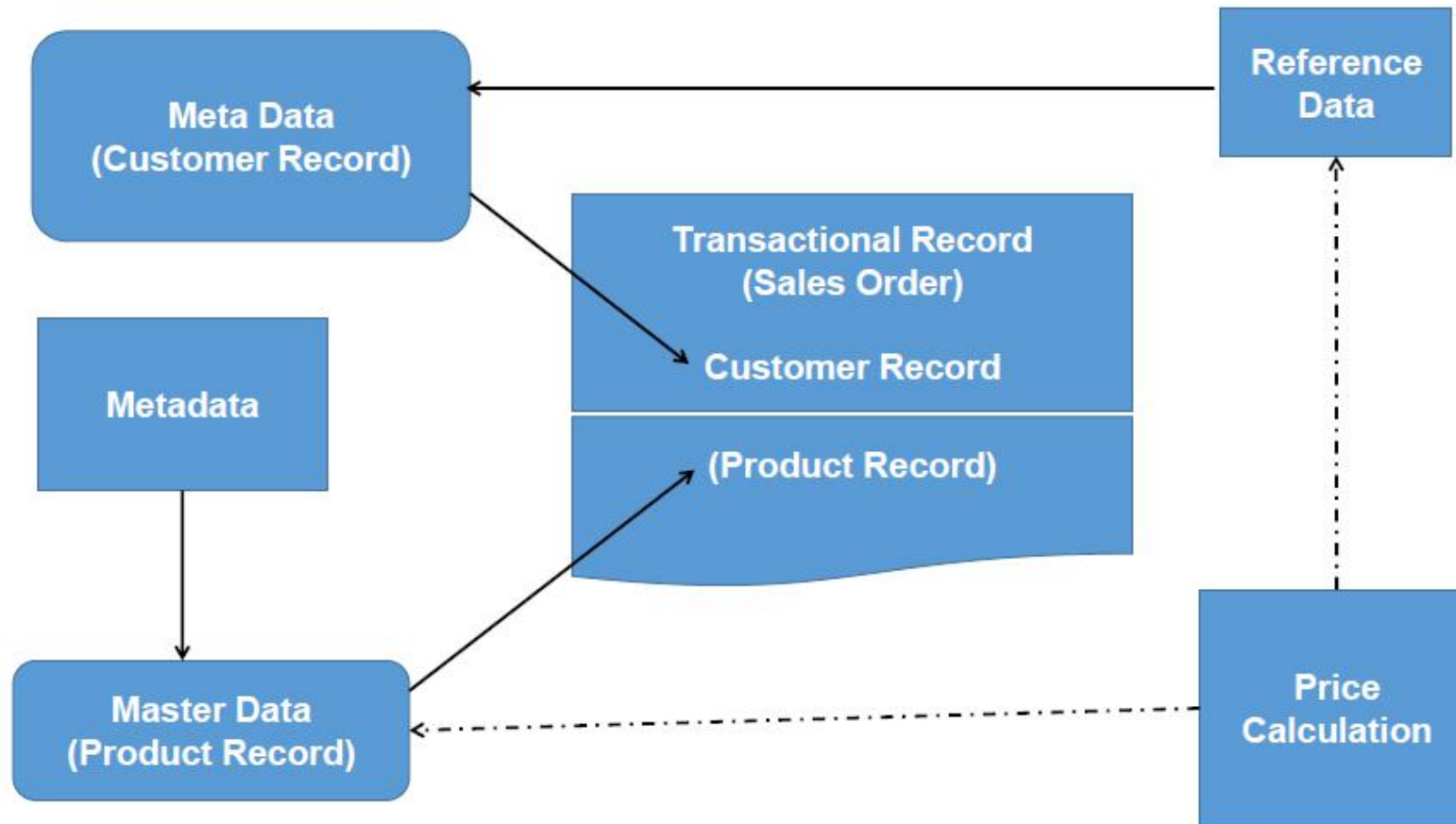
Product #	90-123
Product Name	SDV2

Order No.	8234-00	Type	03-Commercial	
Customer No.	9876	Ship Address	123, Main St	
Customer Name	ABC	City, State	CA	
		Zip Code	66666	
Product #	Description	Qty	Unit Price	Total Price
90-123	SDV2	4	\$100	\$400

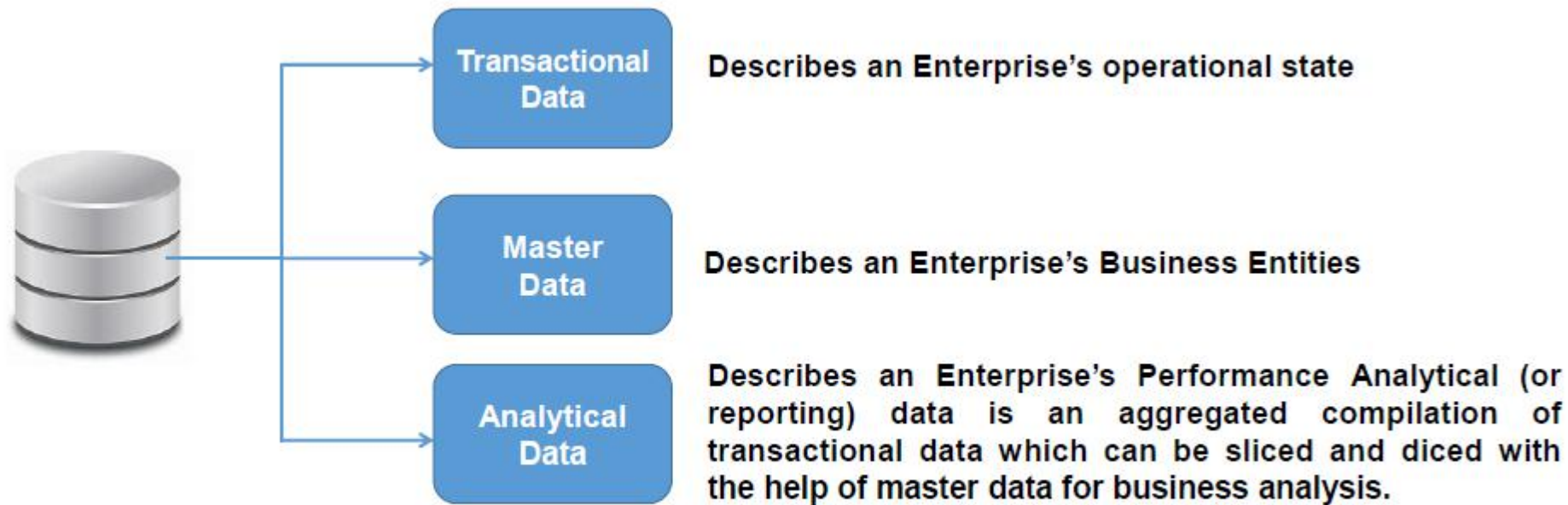
## Price Calculation

Product #	Type	Unit Price
90-123	01	101
90-123	02	101
90-123	03	100
90-123	04	95

# Relationship between data

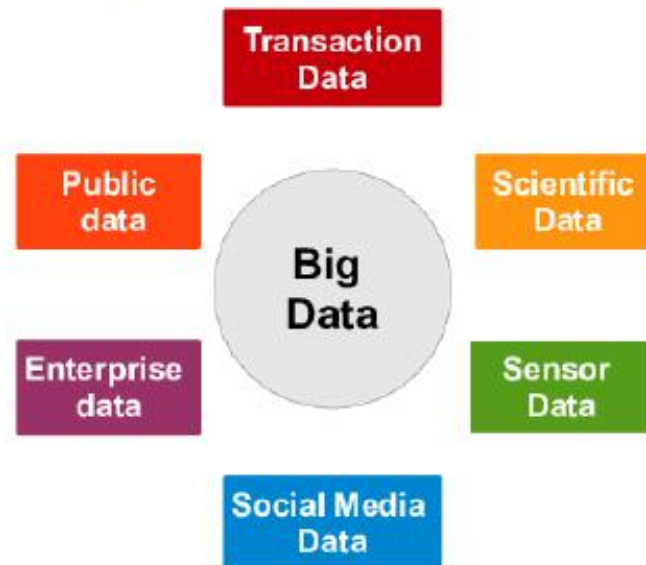


# Enterprise data



## Big Data

- ❑ Big data is the combination of the previous four types of data: transactional data, reference data, and master data.
- ❑ Big Data is high volume, high velocity and high variety information assets that demand cost-effective innovative forms of information processing for enhanced insight and decision making.
- ❑ It cannot be effectively maintained with traditional technology.
- ❑ Big Data is a huge size data and Big Data is a phrase used to define a group of data that is large in size and still increasing exponentially with the period.



Banking. Purchase

.  
.  
.  
.  
.



# Structured Data

Definition

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data.

Example

An Employee Table in a Database

Employee_ID	Employee_Name	Gender	Department	Salary_In_Lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

# Unstructured Data

Definition

Any data with unknown form or the structure is classified as unstructured data.

Example

Google Search Result Page

The screenshot shows a Google search for 'big data'. The search bar at the top contains 'big data' and a magnifying glass icon. Below the search bar, there are tabs for 'All', 'News', 'Images', 'Videos', 'Books', 'Maps', 'Settings', and 'Tools'. The search results are displayed below the tabs, showing a list of links and snippets. The first result is 'Big Data Simplified | Dremio | dremio.com' with a snippet that says 'The open source Data-as-a-Service platform. Open source, free to use. Download for free. Big Data Lakes - Data Lineage - Data Acceleration - Self-Service Data'. The second result is 'IBM® Big Data Analytics | Insights Without Limits | IBM.com' with a snippet that says 'Infrastructure Optimized to Deliver Actionable, Game-Changing Insights. Interactive Data Mining. Improve Efficiency. Accelerate Insights. Types: Servers, Mainframes, Storage Infrastructure. The Future of Analytics. Apply Machine Learning. Extract Hidden Value. Consolidate Storage'. The third result is 'HiveCell: for Edge Computing | Small Servers for Big Data | hivecell.io' with a snippet that says 'HiveCell: The Essential Building Block for Edge Computing. Bring your data & processing to the factory floor, restaurant office, or energy field.' To the right of the search results, there is a large image with the text 'BIG DATA' in the center, surrounded by various icons and graphics. Below the image, the text 'Big data' is written, followed by a definition: 'Big data is a term used to refer to the study and application that are too complex for traditional data-processing applied to adequately deal with. Wikipedia'.

paragraphs  
listings  
descriptions.

# Semi-structured Data

Definition

Any data with unknown form or the structure is classified as unstructured data.

Example

Data Growth over Years

