



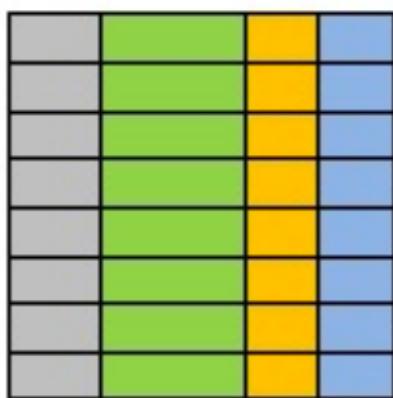
PABLO FIGUEROA

MINERÍA DE DATOS

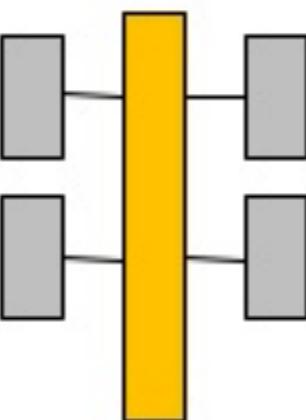
TIPOS DE BASES DE DATOS

Six Types of Databases

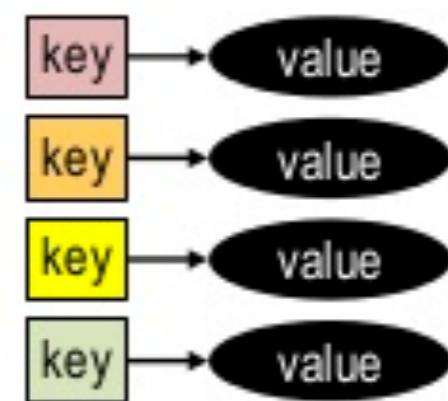
Relational



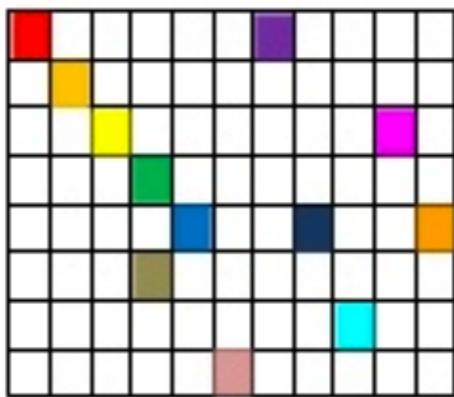
Analytical (OLAP)



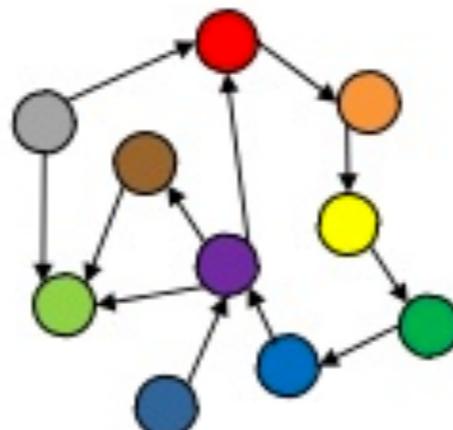
Key-Value



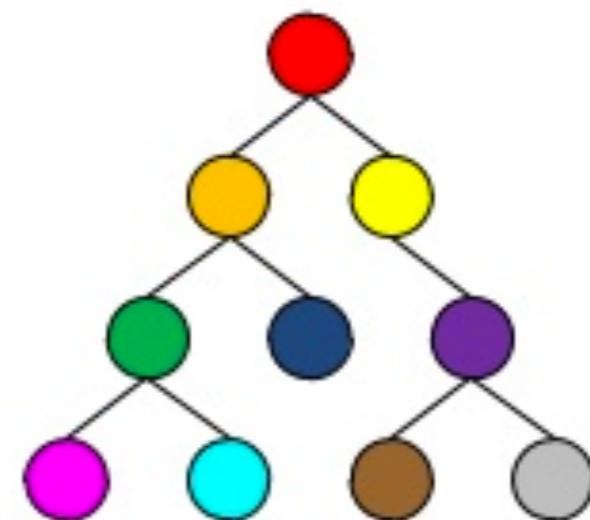
Column-Family



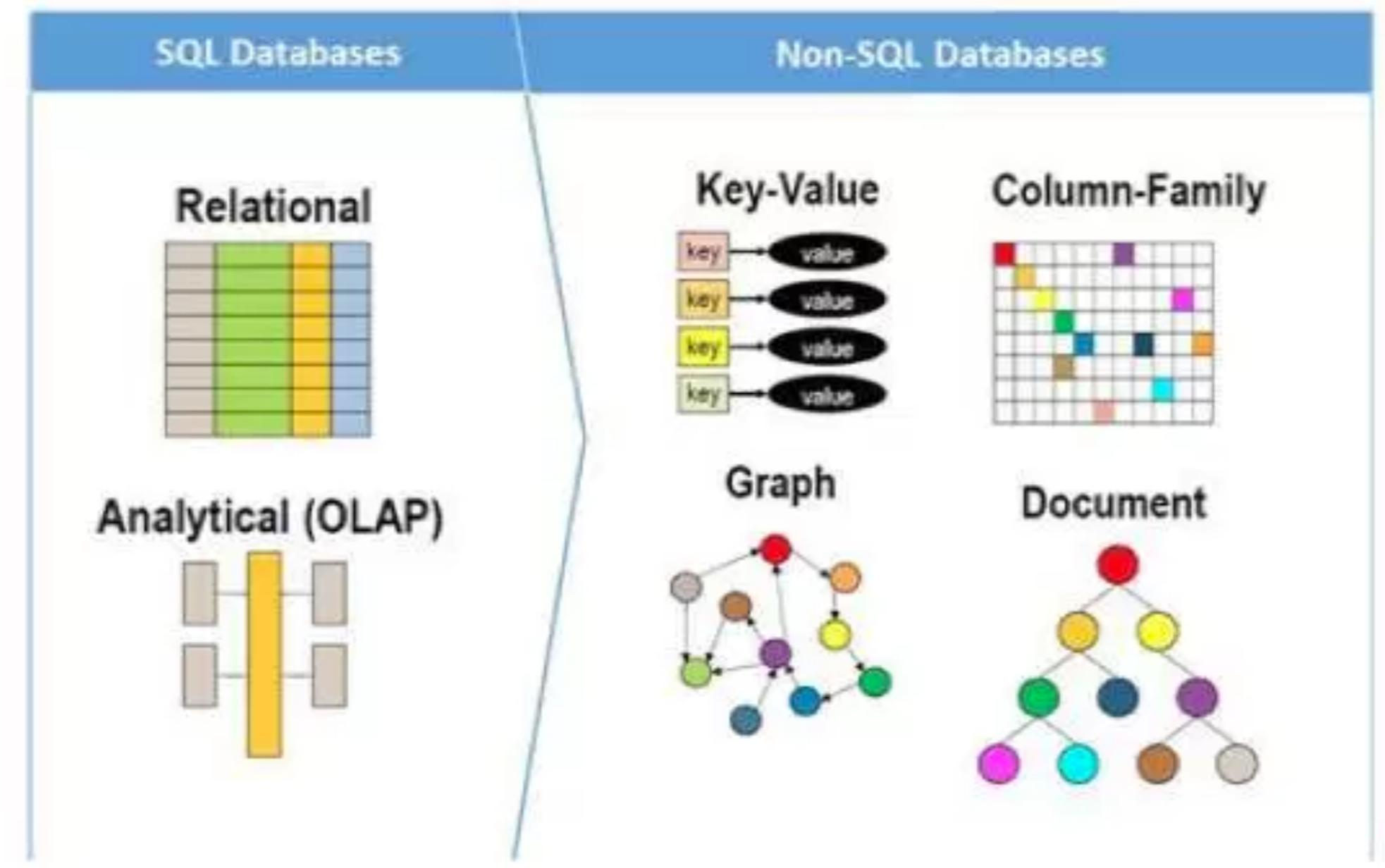
Graph



Document



noSQL: “Not Only SQL”



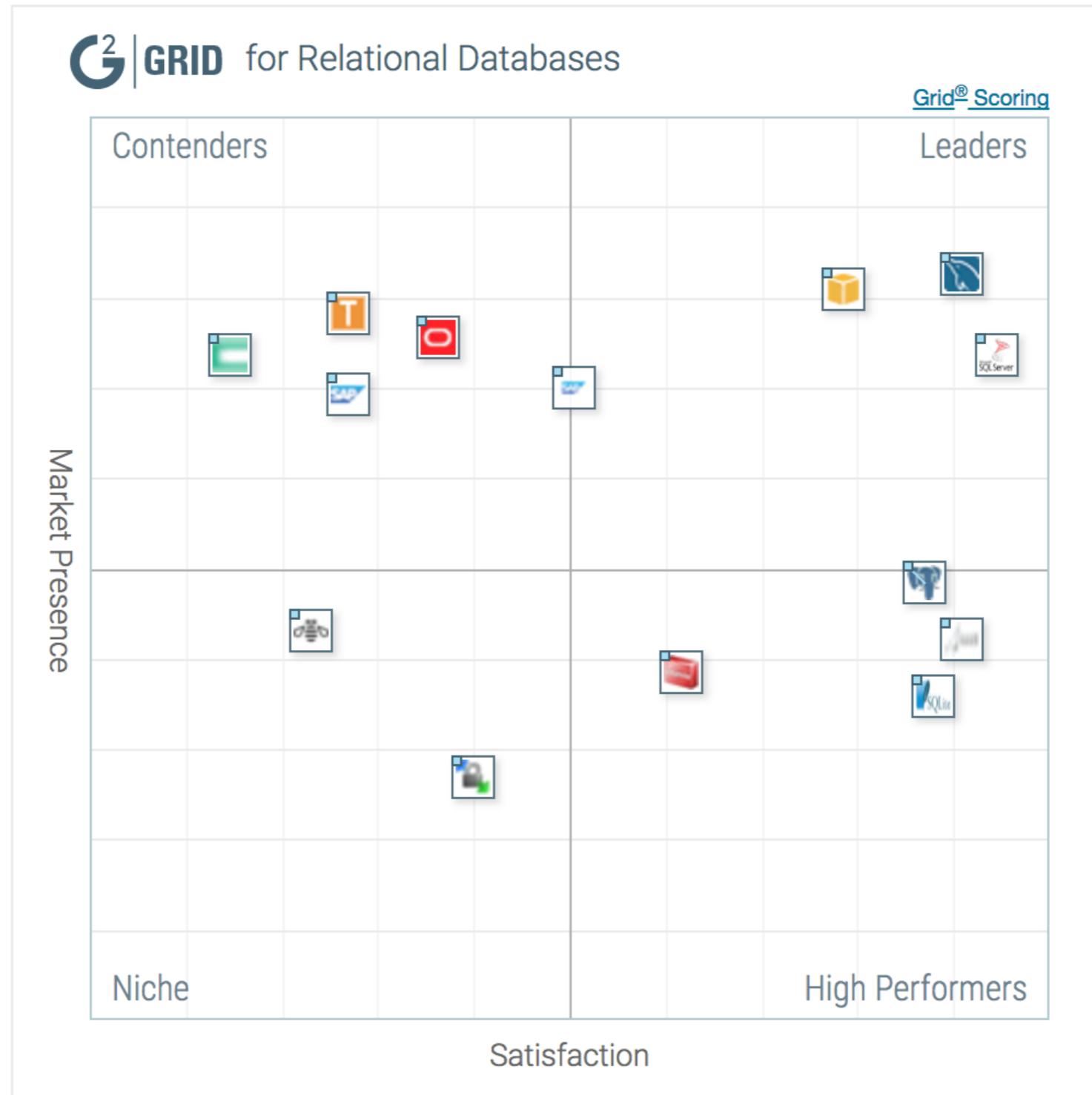
RELACIONAL

- ▶ **Oracle**, Base de datos que puede correr en casi cualquier sistema operativo. De Oracle destacamos la abundancia de perfiles con experiencia en esta tecnología y la gran cantidad de herramientas que hay para su administración y monitorización.
- ▶ **IBM DB2**, Suele ser la segunda base de datos más utilizada en entornos Unix/Linux después de Oracle. Es claramente un indiscutible ganador en Mainframe. Hay perfiles profesionales para DB2 pero no tanto como para Oracle. Por otro lado, el perfil de DB2 Mainframe no tiene por qué saber moverse en DB2 linux/unix.

RELACIONAL

- ▶ **Microsoft SQL Server**, Base de datos con compatibilidad únicamente para sistemas Windows. Hay muchos perfiles expertos en SQL Server y no es difícil su adquisición. Su integración con Microsoft Azure ha mejorado mucho su flexibilidad y rendimiento.
- ▶ **Teradata**, Destaca sobre los demás tipos de bases de datos por su capacidad de almacenamiento y de análisis de datos. Suele ser utilizado en grandes instalaciones de Big Data.
- ▶ **SAP Sybase**, Aunque ya no vive los momentos de éxito de hace una década, sigue destacando por su escalabilidad y rendimiento.

RELACIONAL



MINERÍA DE DATOS

Document Database	Graph Databases
 	 
Wide Column Stores	Key-Value Databases
 	    

NO SQL ORIENTADA A DOCUMENTOS

- ▶ Mongo DB, Probablemente es la base de datos más popular. Tiene la posibilidad de trabajar con datos estructurados y no estructurados. Destaca por su gran capacidad de escalado y rendimiento. Suele ser la que más perfiles experimentados tiene.
- ▶ Permiten trabajar por pares de clave-valor y acceder a distintas partes de los datos almacenados.
- ▶ Mongo DB no soporta atomicidad en las operaciones y garantiza integridad eventual. Los cambios se irán replicando a lo largo de todos los nodos pero no se garantiza que todos los nodos tengan el mismo dato a la vez.

NO SQL ORIENTADA A DOCUMENTOS

- ▶ **Couchbase Server**, Base de datos libre creada en el proyecto Apache. Tampoco garantiza integridad de datos al 100%. Destaca su gran consola de administración en la que se puede acceder a muchos datos de forma muy fácil.

- ▶ **Mark Logic Server**, esta base de datos destaca sobre las anteriores por permitir integridad de datos y compatibilidad con XML, JSON y RDF. Sistemas soportados: Windows, Solaris, Red Hat, Suse, CentOS, Amazon Linux y Mac os.

NO SQL ORIENTADA A CLAVE VALOR

- ▶ Ideales cuando se accede a datos por clave. La diferencia de este tipo de base de datos radica en la posibilidad de almacenar datos sin ningún esquema predefinido. Son las más sencillas de utilizar. Suelen ser muy eficientes para las lecturas y escrituras. Los datos suelen almacenarse en estructuras complejas como BLOB.
- ▶ Redis
- ▶ Riak

NO SQL ORIENTADA A COLUMNAS

- ▶ Bases de datos en las que puedes mapear claves a valores y agruparlas en estructuras. Utilizadas en entornos donde hay poca escritura y existe la necesidad de acceder a varias columnas de muchas filas. Muy útiles en procesamiento y análisis de eventos, gestión de contenido y en análisis de datos.
- ▶ Apache Cassandra
- ▶ Apache Hbase
- ▶ Vertica

NO SQL ORIENTADA A GRAFOS

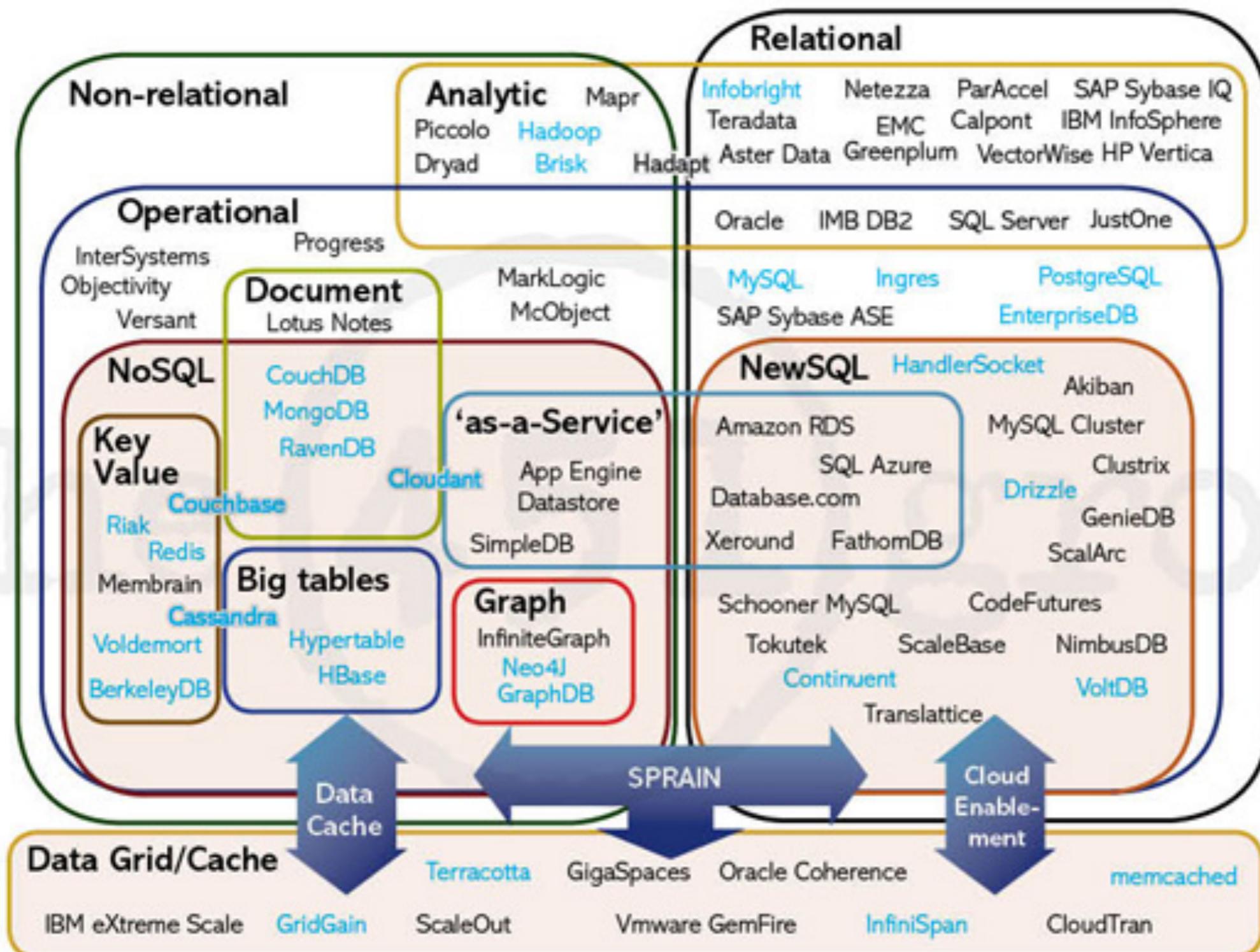
- ▶ Utilización de la teoría de grafos para enlazar los datos de la base de datos. Todo elemento apunta a su elemento adyacente. Estas bases de datos son recomendadas si tus datos están muy relacionadas como en redes sociales, detección de fraude, recomendaciones en tiempo real, etc. Aquí, la base de datos deberá estar normalizada donde cada estructura tendrá una columna y cada relación dos.
- ▶ Neo4j
- ▶ Infinite graph

MODELO HÍBRIDO

- ▶ Cada vez más son las empresas que ofrecen soluciones hibridas en las que utilizan varios motores de bases de datos para dar cabida a varios modelos NoSQL e incluso a motores relacionales.
- ▶ Por ejemplo, CortexDB, Foundation DB y Orient DB ofrecen varios modelos NoSQL.
- ▶ IBM ha extendido su base de datos DB2 para ofrecer la posibilidad de utilizar bases de datos NoSQL con BLU Acceleration. IBM DB2 en su versión extendida permite almacenar datos en XML, JSON y almacenamiento en modo grafo.

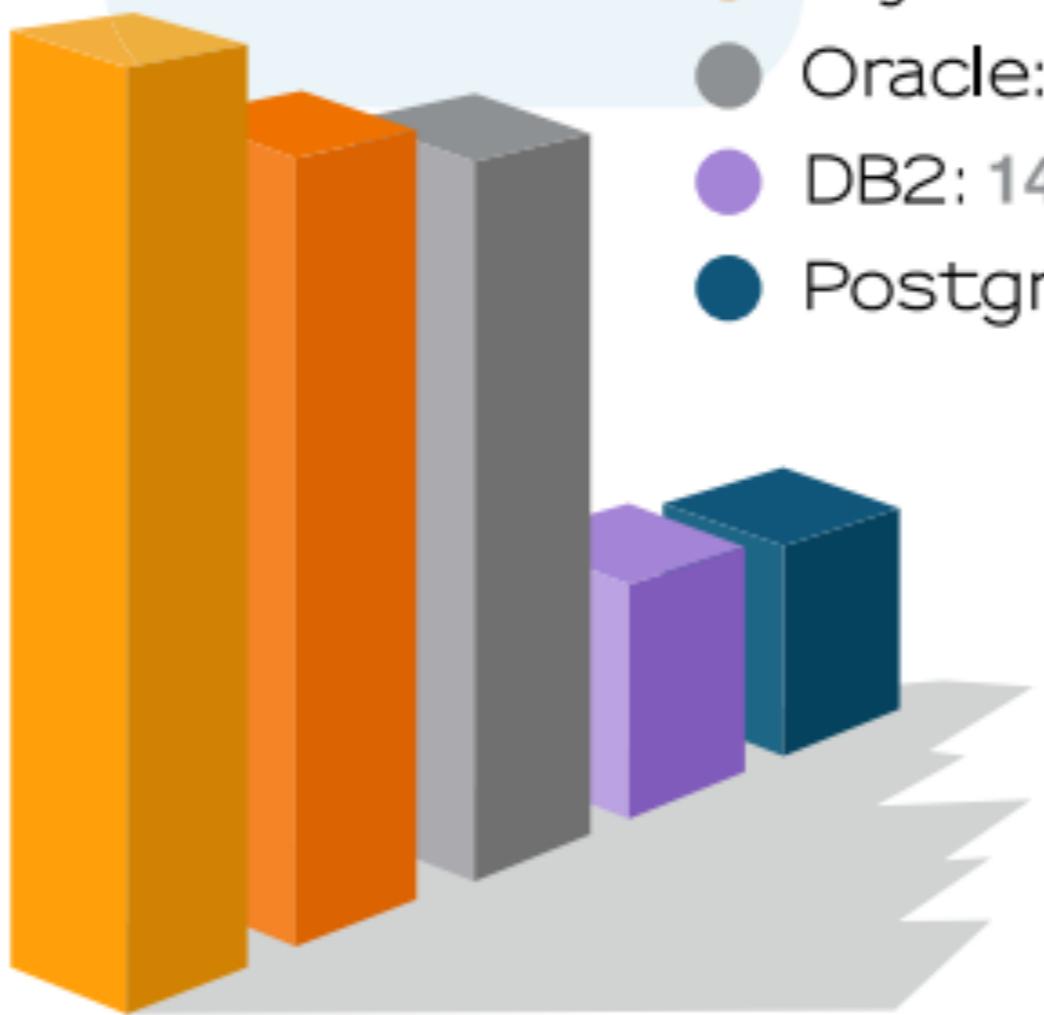
BASES DE DATOS AS A SERVICE

- ▶ Bases de datos que ofrecen sus servicios en la nube. Sólo tienes que leer y escribir tus datos utilizando servicios ofrecidos por el proveedor. Probablemente este tipo de base de datos pase a ser uno de los más demandado por su fácil utilización y aprendizaje junto a la tendencia de ir a la nube por pequeñas y medianas empresas.
- ▶ **Amazon SimpleDB**, Base de datos que como dice su nombre es para bases de datos simples. Si quieres hacer accesos a bases de datos sencillos que no requieran de queries complejas como comparaciones, agrupaciones, Amazon SimpleDB puede ser una buena opción. Los datos son almacenados como textos y se generan estructuras formadas por pares de parámetro valor.



BigData Tools: nSQL Movement

Data Analysis & Platforms	Databases / Data warehousing	Operational	Multivalue database
Hadoop PARACCEL Dremel GridGain HPCC Systems Apache Drill calpont ORACLE Zettaset Hortonworks HD	INFOBRIGHT Hibari BigdataDB orientDB redis	Cassandra APACHE HBASE riak Infinispan Neo4j HYPERTABLE Globals	Rocket Versant JPA MarkLogic mObject IBASE INTERNATIONAL
Business Intelligence	Data Mining	Social	Data aggregation
talend JASPERSOFT Palo	SpagoBI Jedox pentaho KNIME KEEL	RAPIDMINER RAPIDANALYTICS WEKA togaware	Apache Solr elasticsearch
KeyValue	Document Store	Graphs	Multidimensional
EEROSPIKE leveldb GENIE DB Tokyo Cabinet SCALIEN Project Voldemort RAPTORDB STS DB OpenLDAP	mongoDB CLUSTERPOINT JasDB SchematreeDB RaptorDB siadb	COUCHBASE RavenDB EJDB djon deno db	Gephi InfiniteGraph FlockDB GraphBuilder INFO GRID HYPERGRAPHDB GraphSense Gremlin
Object databases	Object databases	Multimodel	XML Databases
db4objects Picolisp HSS Database™	ZOPE STERICOUNTER siaqodb Sterling EyeDB NDatabase	NEOPPOD meronymy	existdb alchemydatabase
Grid Solutions	Grid Solutions	Grid Solutions	Grid Solutions
GIGASPACE	HAZELCAST	Galaxy	Qizx xindice



SQL DATABASES

- SQL Server: 57%
- MySQL: 40%
- Oracle: 38%
- DB2: 14%
- PostgreSQL: 13%

NoSQL DATABASES

- MongoDB: 10%
- Hadoop: 8%
- Cassandra: 4%
- Riak: 2%
- Couchbase: 1%



VARIABLES

VARIABLE

- Magnitud que puede tomar varios valores distintos.
- Una imagen, percepción o concepto que permite ser medido –capaz de tomar distintos valores-.
- Unidad racional de análisis que puede asumir cualquiera de los valores de un conjunto designado de ellos.

Hay que decir que hay una gran parte de la comunidad que defiende que hay variables que no se pueden medir correctamente como los sentimientos preferencias, valores y sensaciones, aunque a veces sí se pueden utilizar mediante indicadores indirectos.

VARIABLE VS CONCEPTO

La principal diferencia entre una variable y un concepto es la **capacidad de ser medido**.

Los conceptos son imágenes mentales o percepciones y por ello su significado puede variar mucho de una persona a otra. Un concepto no puede ser medido utilizando variables corrientes, ya que diferentes personas darán diferentes valores al mismo concepto.

Es por ello que los conceptos deben ser convertidos (reducidos a) en variables, de tal forma que puedan ser “medidos” con un grado de precisión aceptable utilizando escalas de medida.

VARIABLE VS CONCEPTO

Conceptos	Variables
<ul style="list-style-type: none">• Efectividad• Satisfacción• Impacto• Excelencia• Éxito• Amor propio• Riqueza• Violencia Doméstica• etc.	<ul style="list-style-type: none">• Sexo (Masc/fem)• Actitud• Edad• Ingresos• Peso• Altura• Religión• etc.
<ul style="list-style-type: none">- Subjetividad- No existe uniformidad entre los términos y son muy interpretables- Por sí sólos no se pueden medir	La medición depende del grado de precisión de las escalas que se use.

CONVIRTIENDO CONCEPTOS EN VARIABLES

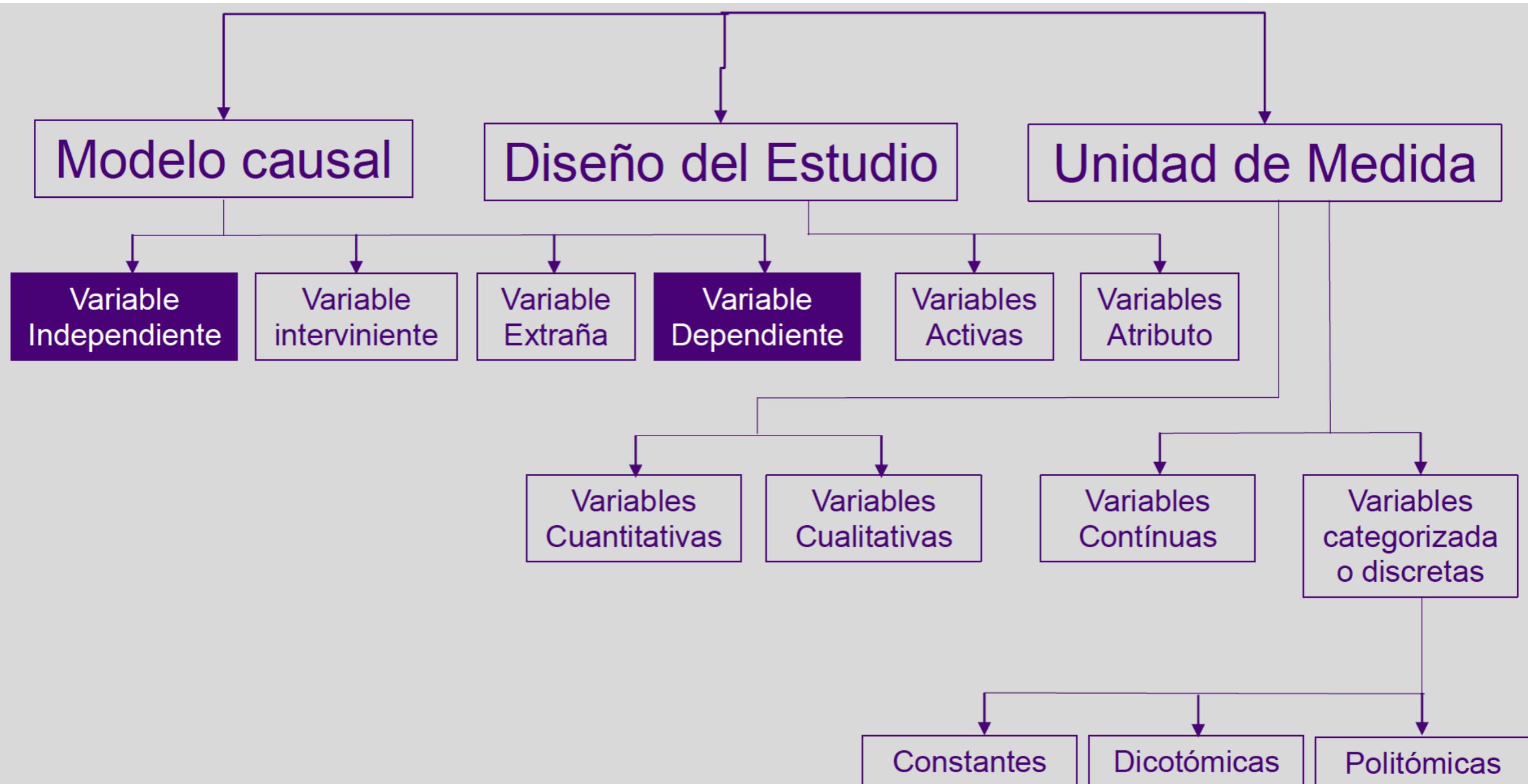
1. Identificar indicadores relativos al concepto.
2. Medir el grado de relación de los conceptos, (son un reflejo del concepto en su totalidad, resume una dimensión del concepto, un concepto puede ser la suma de varios indicadores, etc.)
3. Dividir un concepto en varias dimensiones (Reducciónismo).
Millonario → Riqueza → Bienes Muebles – Bienes Inmuebles
→ Todo se traduce en euros..

Este proceso es relativamente objetivo con variables cuantitativas, con variables cualitativas se necesita posteriormente utilizar pruebas de control y fiabilidad para detectar el grado de objetividad

CONVIRTIENDO CONCEPTOS EN VARIABLES

Conceptos →	Indicadores →	Variables	
		Variables	Definiciones operacionales
Riqueza	1. Ingresos 2. Bienes	1. Ingresos por años 2. Valor total de bienes	Si > 300,000€ No < 300,000€
Nivel Académico	1. Promedio de notas de exámenes 2. Promedio de notas de trabajos 3. Ambos	1. Porcentajes de notas	1. Superior al 75% (Q1)
Efectividad (De un Hospital)	1. Nº Pacientes 2. Cambios en las enfermedades (número o tendencias) 3. Mortalidad (Edades, tendencias). 4. Estado nutricional 1. Peso 2. Episodios	Pacientes/Mes Número de enfermos/1000 habs. Cambios en tipologías de enfermedades Cambios de peso Episodios de enfermos por año	Estudio comparativo con antes, y después del mismo centro y otros centros. Estudio de cada variable con expertos.

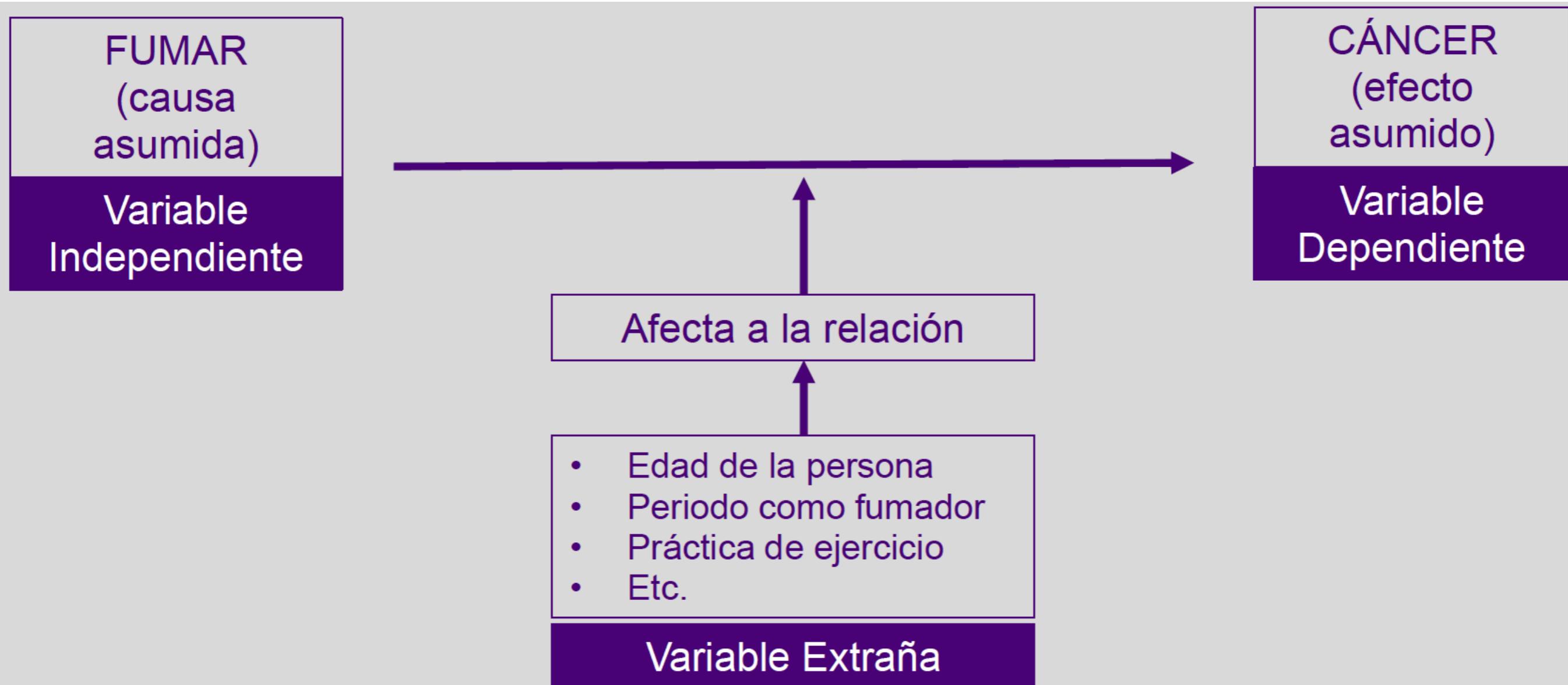
TIPOS DE VARIABLES



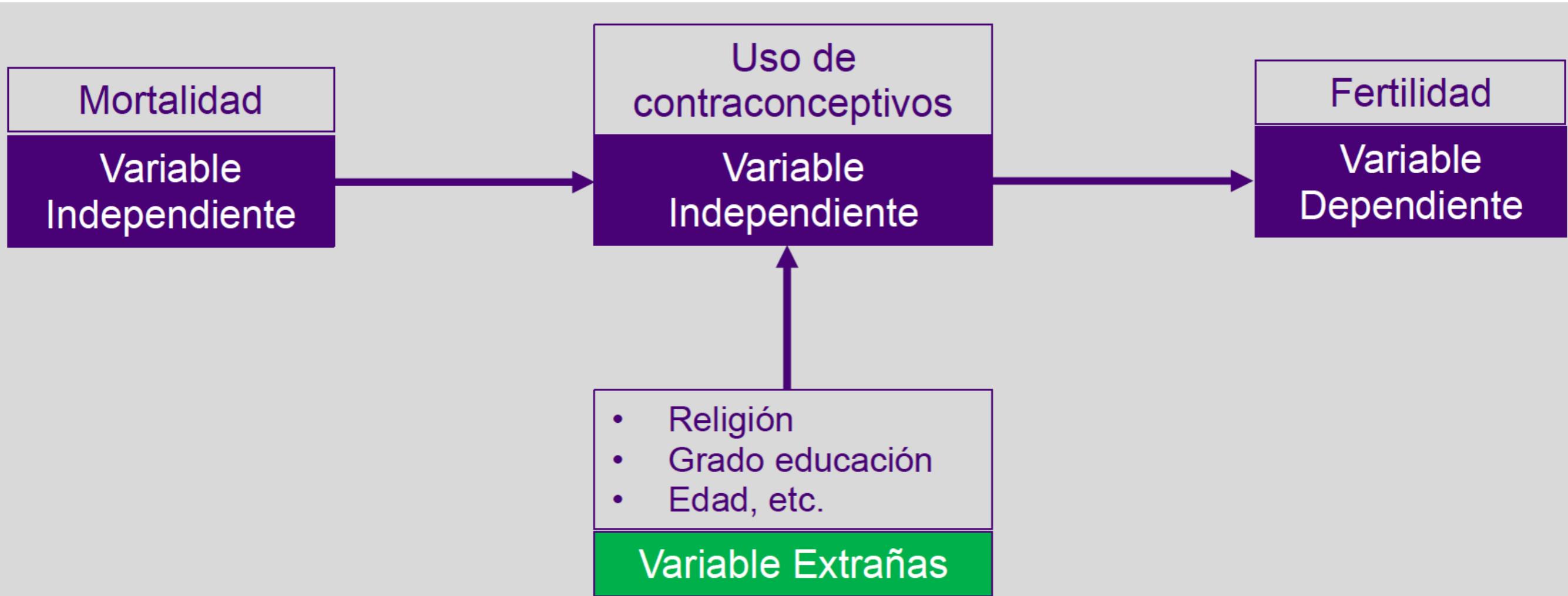
TIPOS DE VARIABLES - CAUSALIDAD

Variable Independiente	La causa por la que se supone que es la responsable original de los cambios en un fenómeno o en una situación
Variable Dependiente	La variable dependiente es aquella que el investigador estudia, la cual varía según los cambios que se produzcan en la variable independiente. Relación de causa efecto .
Variable Extraña	Son factores que afectan a la variable dependiente y que normalmente no son medidos, (aunque una vez identificadas sí son medibles) estos factores influyen en la fortaleza o debilidades de las variables dependientes e independientes.
Variable interviniente	La variable interviniente necesaria cuando existe una relación probada entre dos variables (independiente y dependiente) pero no existe relación teórica. Es el “eslabón perdido” que une a ambas y explica la dependencia indirecta entre ellas.

TIPOS DE VARIABLES - CAUSALIDAD



TIPOS DE VARIABLES - CAUSALIDAD



TIPOS DE VARIABLES - DISEÑO DEL ESTUDIO

Variable Activas

Variables que pueden ser manipuladas, modificadas o controladas. Son las variables que el investigador podrá modificar para ver el efecto que tienen.

Variable de Atributo

Estas variables no se pueden modificar, y son cualidades y características de la población a estudiar.

TIPOS DE VARIABLES

Tipos de variables

Variable: corresponde a la característica de la Unidad de Análisis

TIPOS DE VARIABLES

VARIABLES CUANTITATIVAS

CONTINUA

Intervalo

Toma cualquier valor dentro de un intervalo

Ejemplos: Peso; Estatura; Temperatura, etc

DISCRETA

Toma valores enteros
Ejemplos: Número de Hijos, Número de empleados de una empresa, Número de asignaturas aprobadas en un semestre, etc.

VARIABLES CUALITATIVAS

NOMINAL

Característica o calidad cuyas categorías no tienen un orden preestablecido.

Ejemplos: Sexo, Deporte Favorito, color de ojos, profesión, estado civil, religión

ORDINAL

Hay un orden preestablecido o jerarquía entre las categorías

Ejemplos: grados militares, organigrama de una empresa, escalafón de los profesores universitarios, grados de Disney, estadio de un tumor. Grado de Interés por un tema, etc.

Unidad de Medida: Gramos o Kilos para la variable Peso; Grados C o F para Temperatura

TIPOS DE VARIABLES - UNIDAD DE MEDIDA

Variables
Contínuas

Pueden tomar cualquier valor dentro de un rango (Por ejemplo: número enteros positivos).

Variables
categorizadas
y discretas

Variables que sólo pueden tomar valor de un conjunto cerrado. Pueden expresarse de forma ordinal o numeral.

Variables
constantes

Sólo tienen una categoría o valor, por ejemplo, árbol, taxi, agua, etc.

Variables
Dicotómicas

Sólo tienen dos categorías o valores, Sí-No, Bueno-Malo, Feo-Guapo, Hombre-Mujer, etc.

Variables
Politómicas

Tienen más de dos categorías o valores. Por ejemplo: Religión (Musulman, Cristiano, Hinduista, Budista).

TIPOS DE VARIABLES - UNIDAD DE MEDIDA - ESCALAS

Las variables que se van a estudiar determinarán el tipo de análisis que se va a llevar a cabo.

Según Escalante, medición es la clasificación de casos o situaciones y sus propiedades, de acuerdo a ciertas reglas lógicas.

TIPOS DE VARIABLES - UNIDAD DE MEDIDA - TIPOS ESCALAS

Escala nominal: es el nivel más elemental de medición y consiste en clasificar los objetos de estudio según las categorías (y subcategorías) de una variable: ejemplo: estado marital que se clasifica así: soltero, casado, viudo, divorciado, unión libre. Esta clasificación solo nombra los indicadores de la variable pero no representa ninguna jerarquización, dado que todos los indicadores se encuentran al mismo nivel.

Religiones:

Cristianismo

Islam

Hinduismo

Budismo

Enfermedades Mentales

Esquizofrenia

Paranoia

Depresión

TIPOS DE VARIABLES - UNIDAD DE MEDIDA - TIPOS ESCALAS

Escala ordinal. Este tipo de escala se utiliza para clasificar los objetos, hechos o fenómenos en forma jerárquica, según el grado que posea una característica determinada, sin proporcionar información sobre la magnitud de las diferencias entre los casos así clasificados; ejemplo: tipo de higiene (excelente, buena, regular, mala). En este caso se refleja distancia entre una y otra categoría, o cuánto es mayor una de otra.

<i>Ingresos:</i>	<i>Status Socioeconómico</i>	<i>Actitudes</i>	<i>Actitudes (escala Likert)</i>
<i>Por encima de la media</i>	<i>Clase alta</i>	<i>Muy a favor</i>	0-30
<i>Normales</i>	<i>Clase media</i>	<i>A favor</i>	31-40
<i>Por debajo de lo normal</i>	<i>Clase baja</i>	<i>Indiferente</i> <i>En contra</i> <i>Muy en contra</i>	41-50

TIPOS DE VARIABLES - UNIDAD DE MEDIDA - TIPOS ESCALAS

Escala de intervalo. Es la que clasifica a la variable de acuerdo a un rango numérico. Este tipo de escala mide las variables cuantitativas , siendo igual la distancia entre dos puntos o valores de un continuo. Sin embargo el punto cero es arbitrario y convencional, por lo que no se pueden establecer razones o proporciones, ni comparar dos escalas sin definir el mismo punto de partida. Las variables de inteligencia, rendimiento académico y temperatura, debido a que el punto cero es arbitrario; tomando como ejemplo la temperatura, se puede decir que el cero no representa la ausencia de calor; sin embargo, la distancia entre cualesquiera dos puntos de la escala es igual, o sea, que el cambio de temperatura entre 36° y 37°C es igual al cambio entre 40° y 41°C.

Actitudes (escala Thurstone)

10-20

21-30

31-40

41-50

Temperaturas

10-20

21-30

31-40

41-50

Energía

10-20

21-30

31-40

41-50

TIPOS DE VARIABLES - UNIDAD DE MEDIDA - TIPOS ESCALAS

Escala de proporción razón. Este tipo de escala constituye el nivel más alto de medición para las variables cuantitativas; contiene las características de una escala de intervalo con la ventaja adicional de poseer el cero absoluto, lo cual permite determinar la proporción conocida de valores de la escala. El peso, la talla y número de alumnos son ejemplos de variables de razón o proporción, en las que el cero representa la nulidad o ausencia de lo que se estudia.

Peso

10 kg

20 kg

30 kg

40 kg

Distancias

300 Mtrs

100 km

200 km

40

PROYECTOS