

Big Data Warehousing

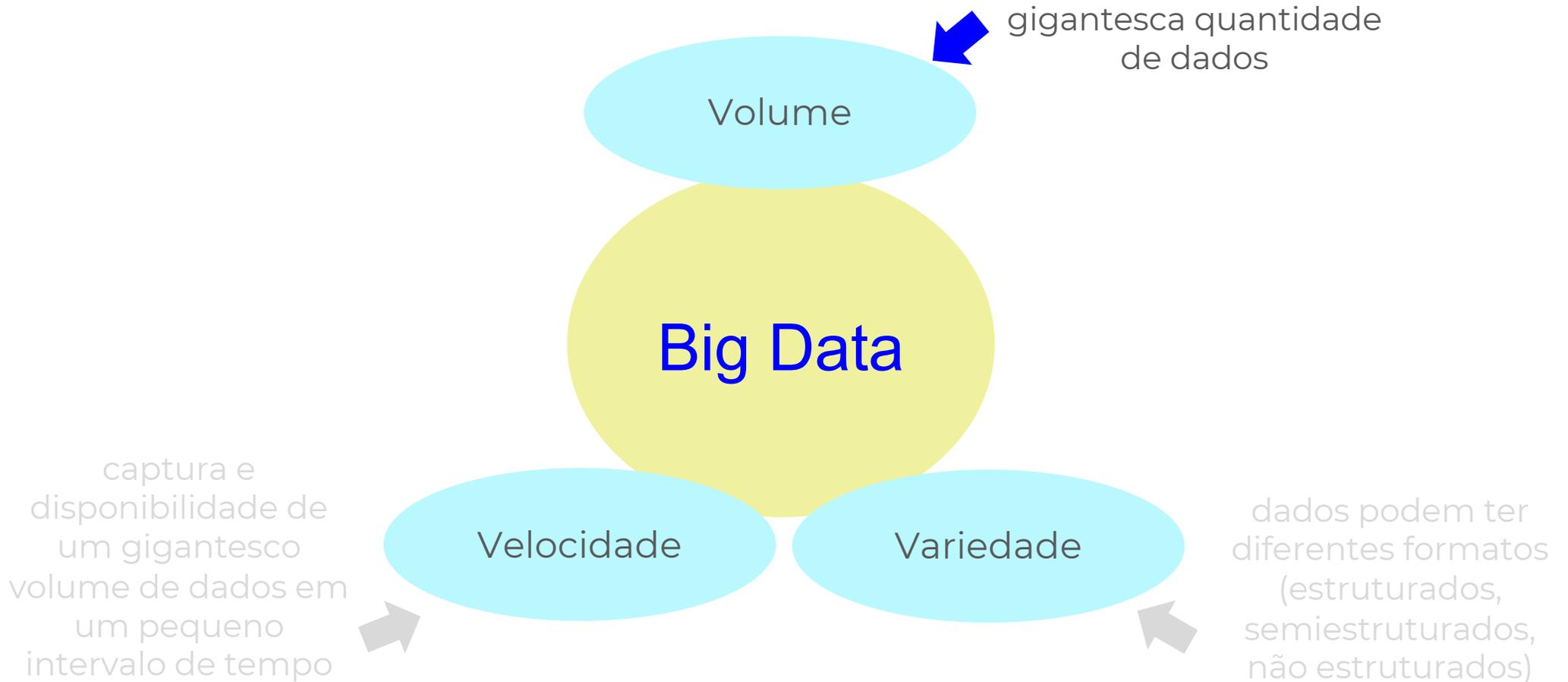
Processamento Analítico de Dados (em Larga Escala)

Profa. Dra. Cristina Dutra de Aguiar

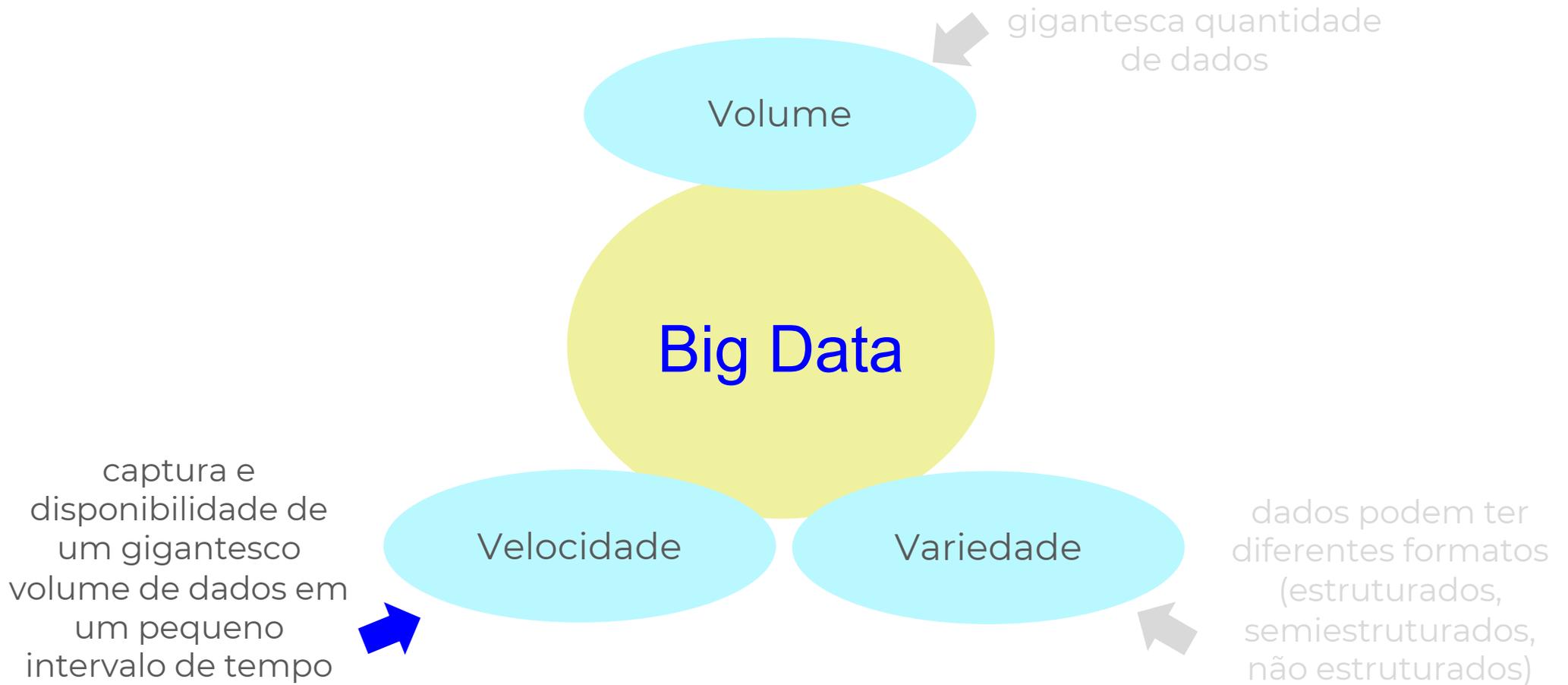
Agenda

- Big Data
 - Definição
 - Desafios
 - Nuvem de Conceitos e Tecnologia
- Data Warehousing
 - Definição
 - Arquitetura
 - Exemplo de Pipeline

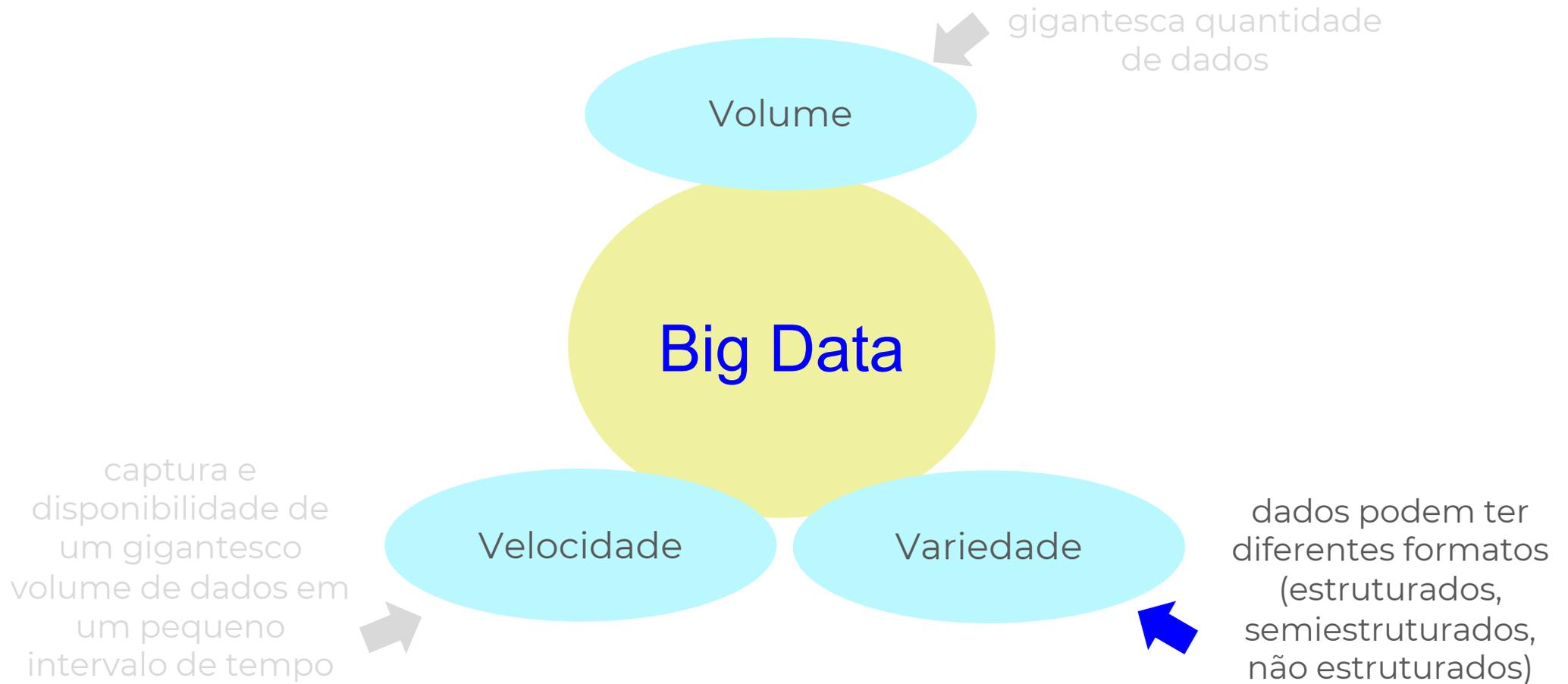
Modelo de 3Vs



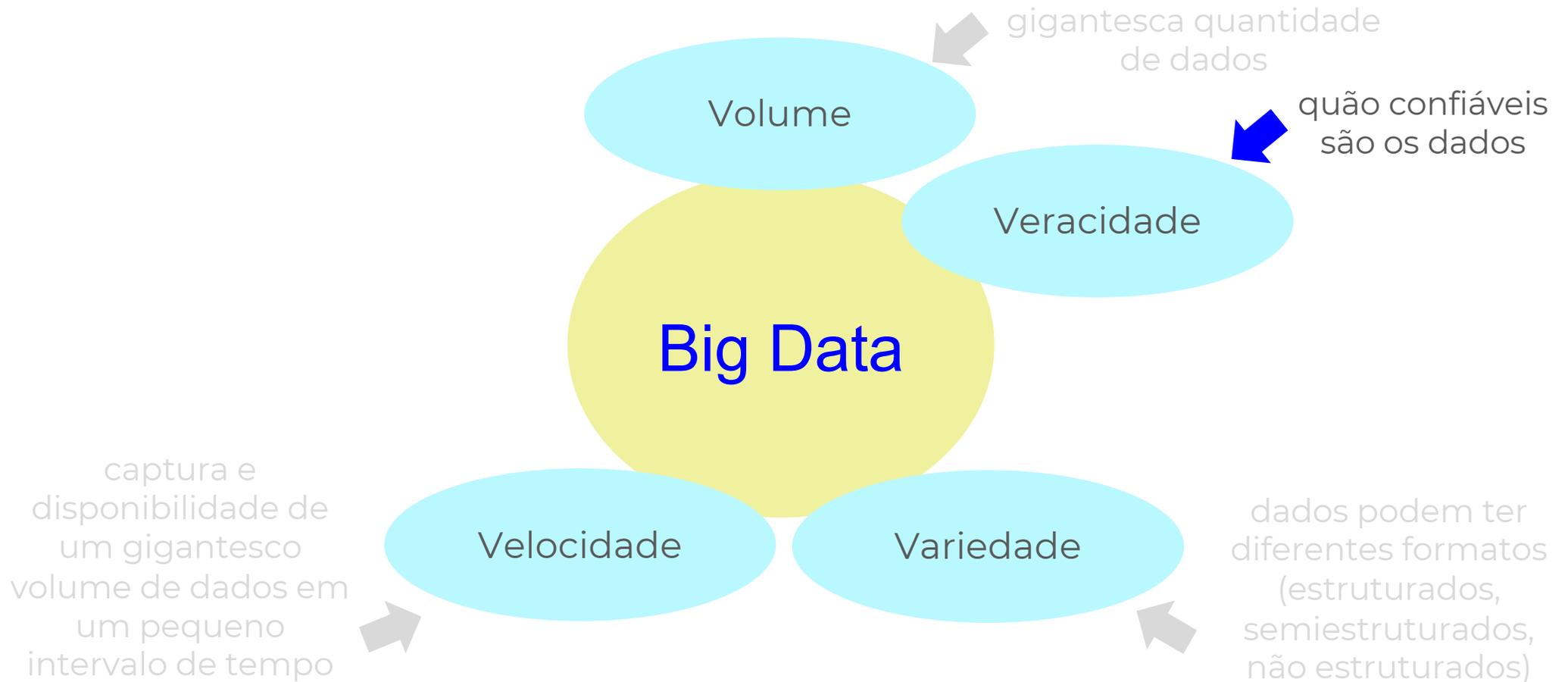
Modelo de 3Vs



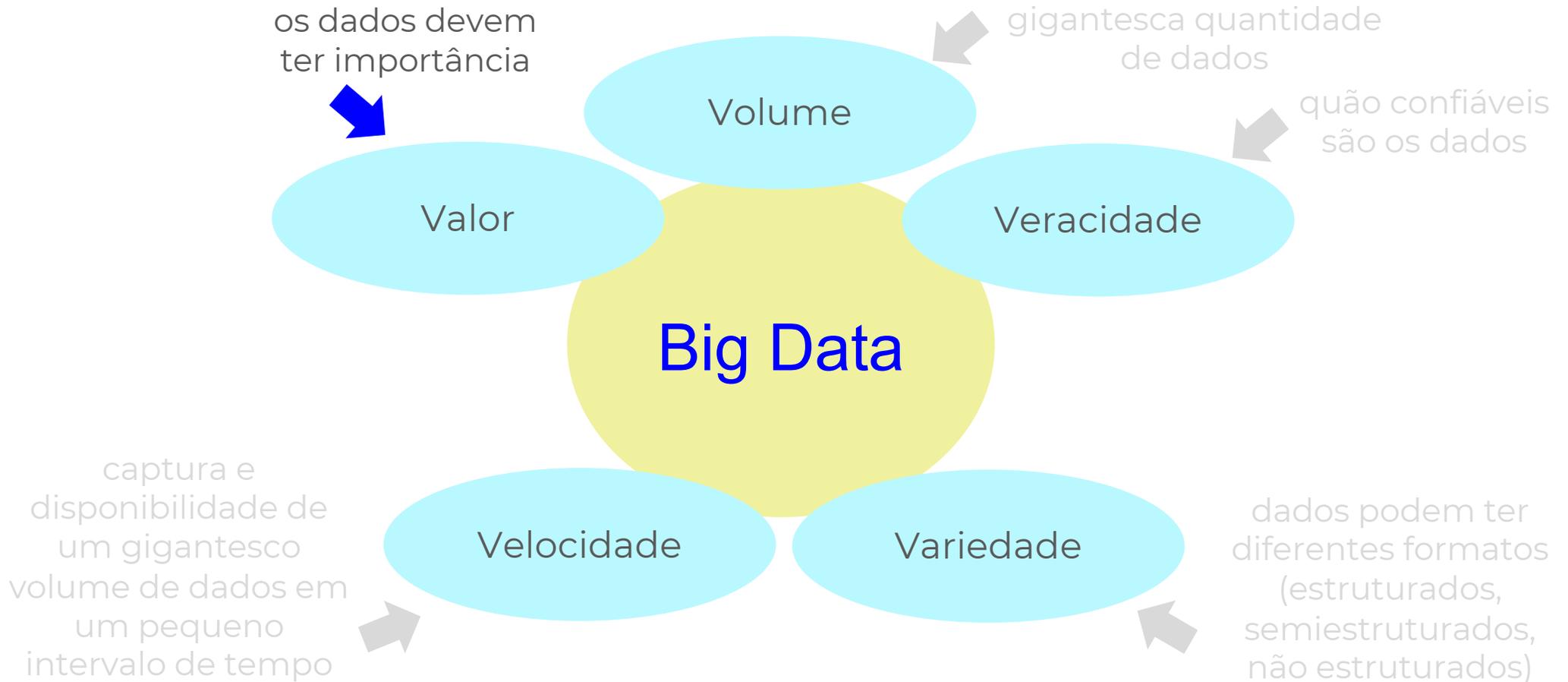
Modelo de 3Vs



Modelo de 4Vs



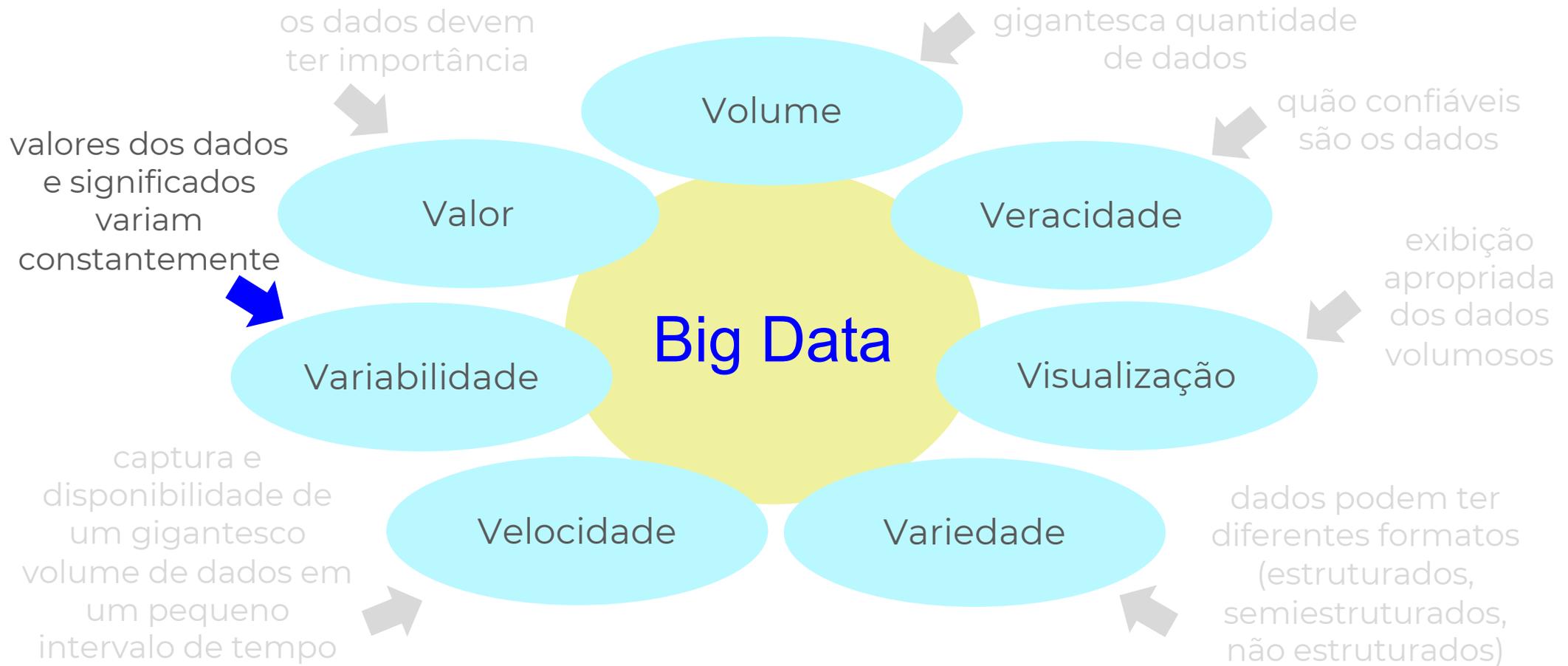
Modelo de 5Vs



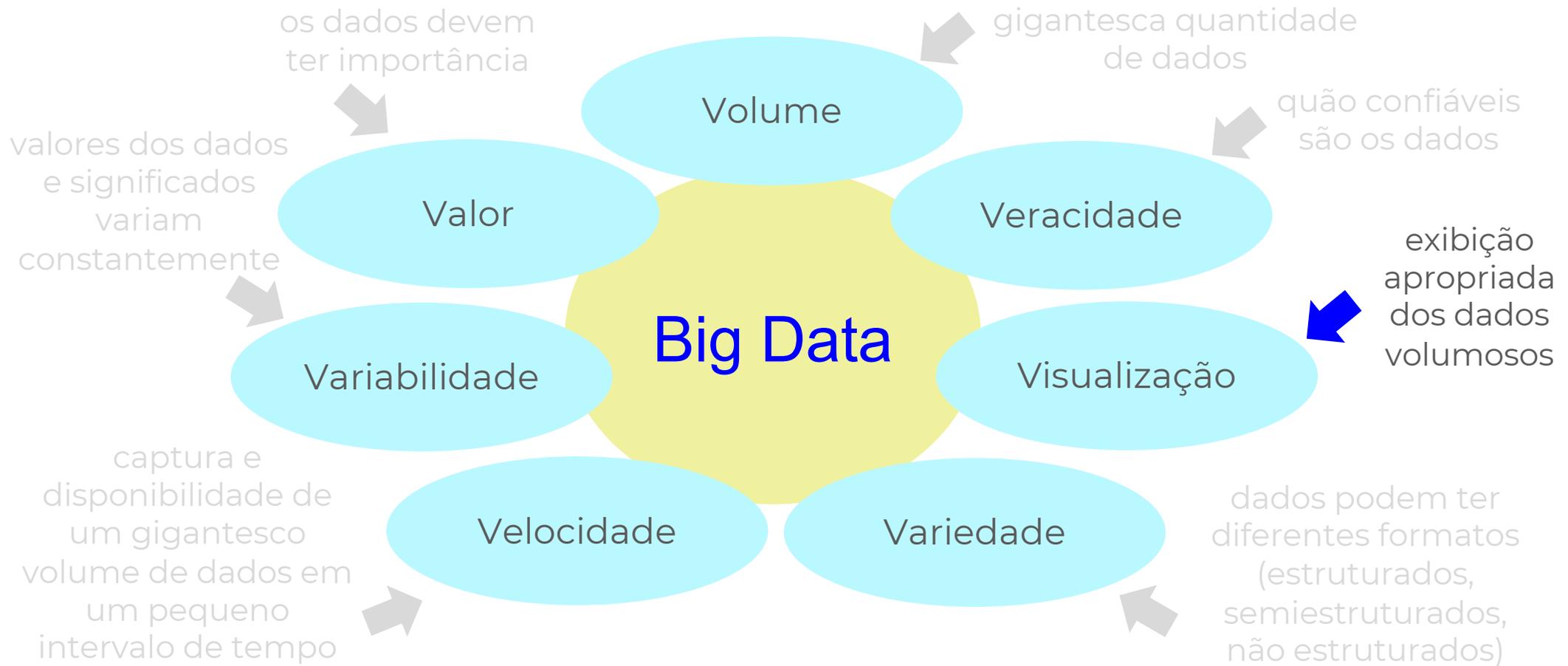
Sharma, S.; Mangat, V. Technology and Trends to Handle Big Data: Survey.

In: *Proc. 5th International Conference on Advanced Computing & Communication Technologies*, p. 266-271, 2015.

Modelo de 7Vs



Modelo de 7Vs



Desafios

- Uso de **ambientes computacionais** com grande capacidade de armazenamento e processamento
 - *Clusters* de computadores
 - Computação em nuvem (*cloud computing*)
- Uso de *frameworks* de processamento paralelo e distribuído para simplificar a interação com os ambientes computacionais
 - Apache Hadoop
 - Apache Spark

Desafios

- Uso de ambientes computacionais com grande capacidade de armazenamento e processamento
 - *Clusters* de computadores
 - Computação em nuvem (*cloud computing*)
- Uso de *frameworks* de processamento paralelo e distribuído para simplificar a interação com os ambientes computacionais
 - Apache Hadoop
 - Apache Spark

Desafios

- Uso de **sistemas de arquivos distribuídos** para prover suporte para o armazenamento de grandes quantidades de dados
 - HDFS (*Hadoop Distributed File System*)
- Uso de bases de dados NoSQL (*Not only SQL*) para introduzir flexibilidade no armazenamento de diferentes tipos de dados
 - Não estruturados
 - Semiestruturados
 - Estruturados

Desafios

- Uso de sistemas de arquivos distribuídos para prover suporte para o armazenamento de grandes quantidades de dados
 - HDFS (*Hadoop Distributed File System*)
- Uso de bases de dados NoSQL (*Not only SQL*) para introduzir flexibilidade no armazenamento de diferentes tipos de dados
 - Não estruturados
 - Semiestruturados
 - Estruturados

Nuvem de Conceitos e Tecnologias

Velocidade JSON
Veracidade Cluster
Pipeline Spark HDFS
Variabilidade Databases
Druid Data Warehouse NoSQL
BIG DATA
Kafka Hive
ELT AWS
Azure ETL
Analytics Hadoop
Visualização Streaming Variedade
Valor Volume
SQL Data Lake
Petabytes
Cloud Computing

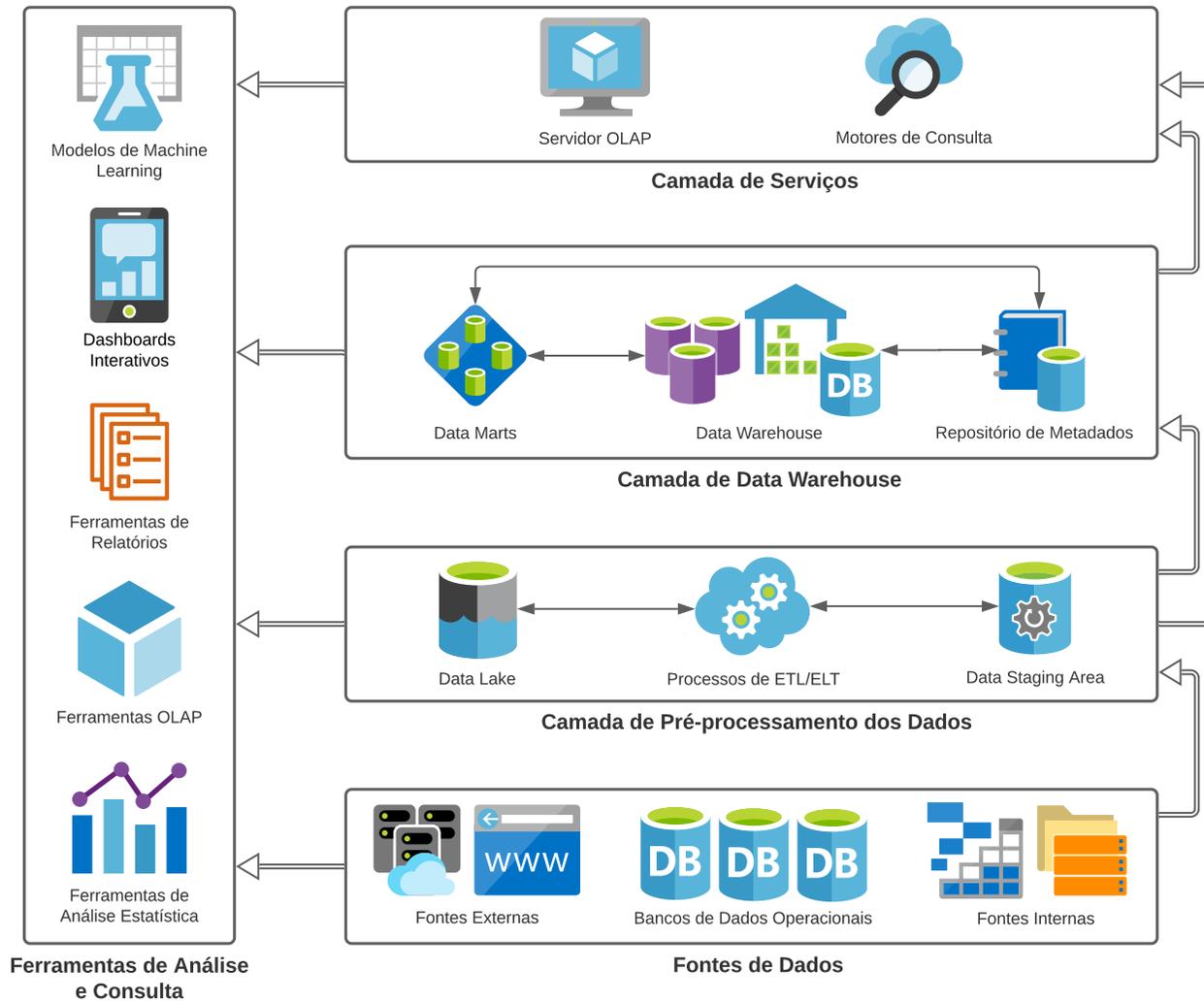
Agenda

- Big Data
 - Definição
 - Desafios
 - Nuvem de Conceitos e Tecnologia
- Data Warehousing
 - Definição
 - Arquitetura
 - Exemplo de Pipeline

Definição

Ambiente que possibilita que dados selecionados de fontes de dados autônomas, heterogêneas e distribuídas sejam armazenados em um único banco de dados, o *data warehouse* (DW), o qual oferece suporte para *consultas analíticas* voltadas à tomada de decisão estratégica

Arquitetura



Exemplo de Pipeline

processamento em lote com big data

