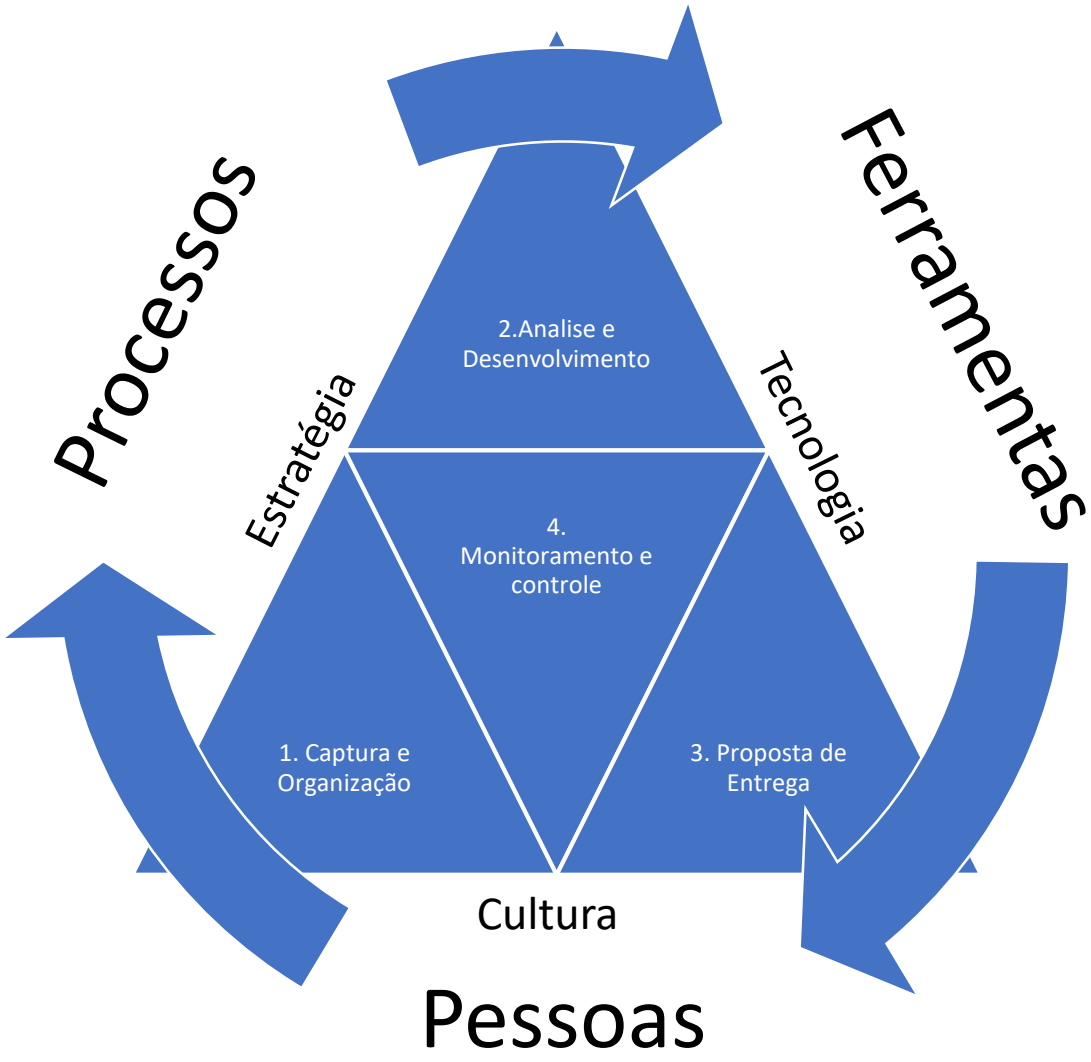
Brintell

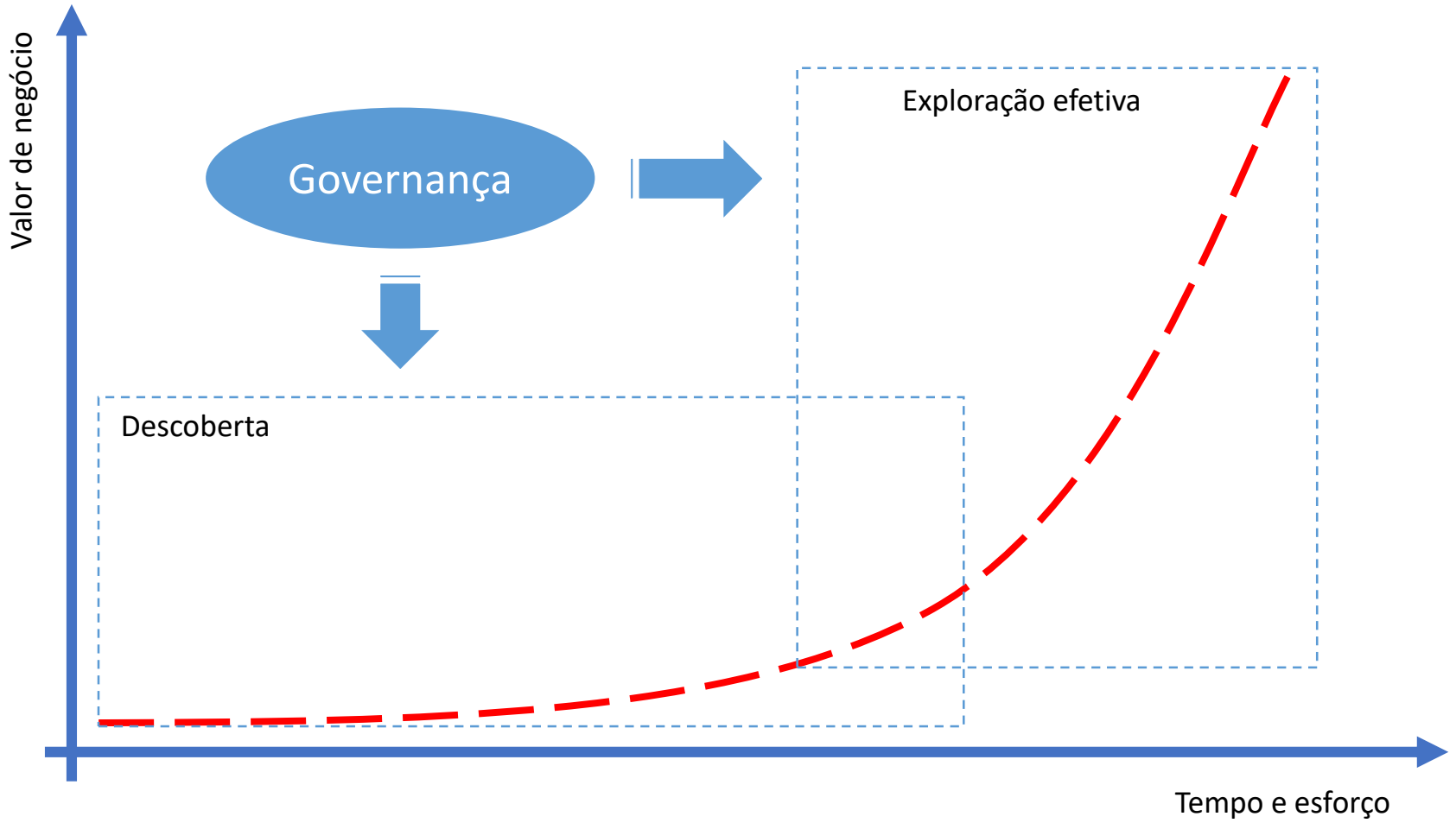
Intelligent Analytics

BIG DATA

Brintell
Intelligent Analytics

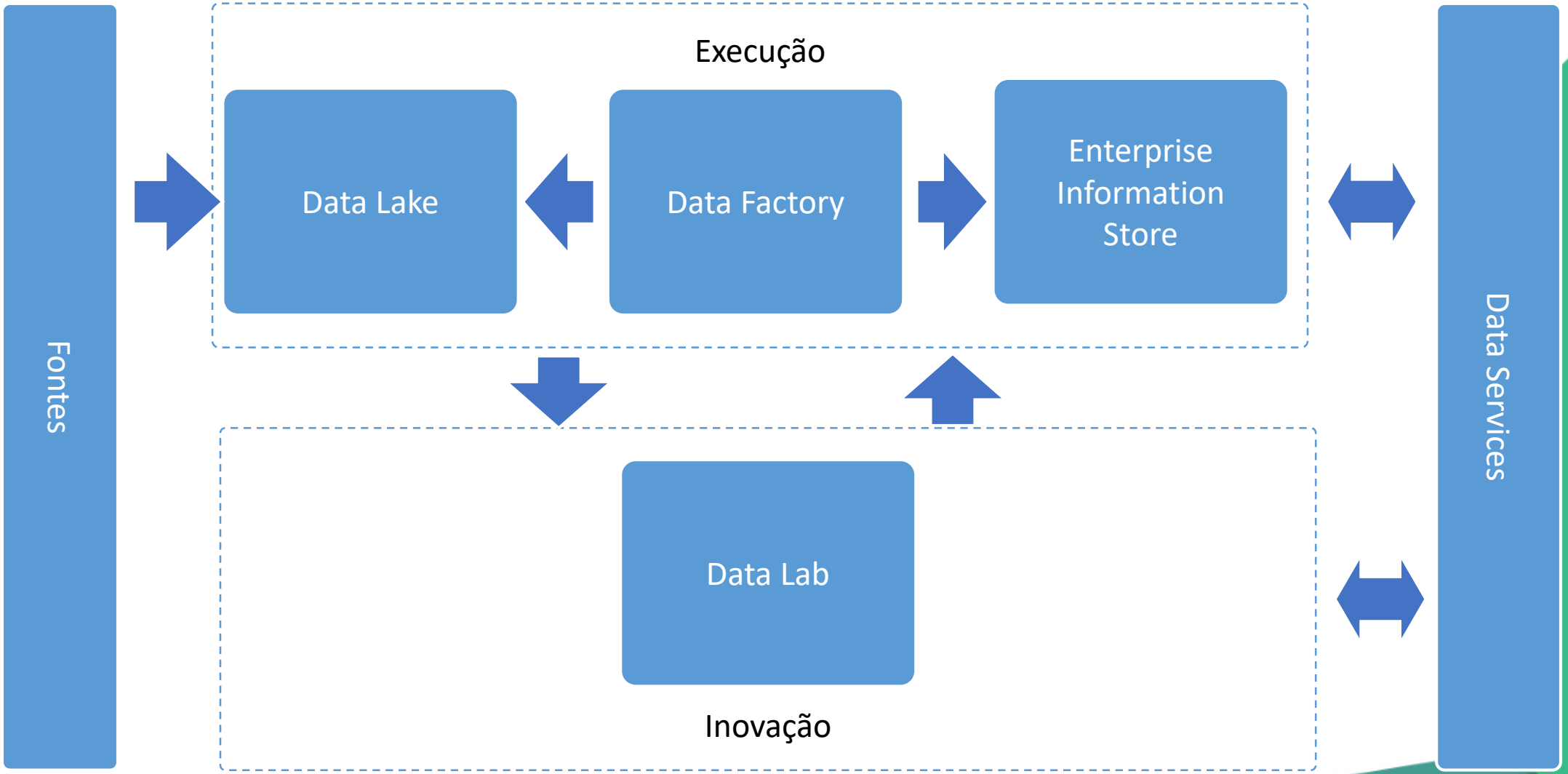


Arquitetura Lambda

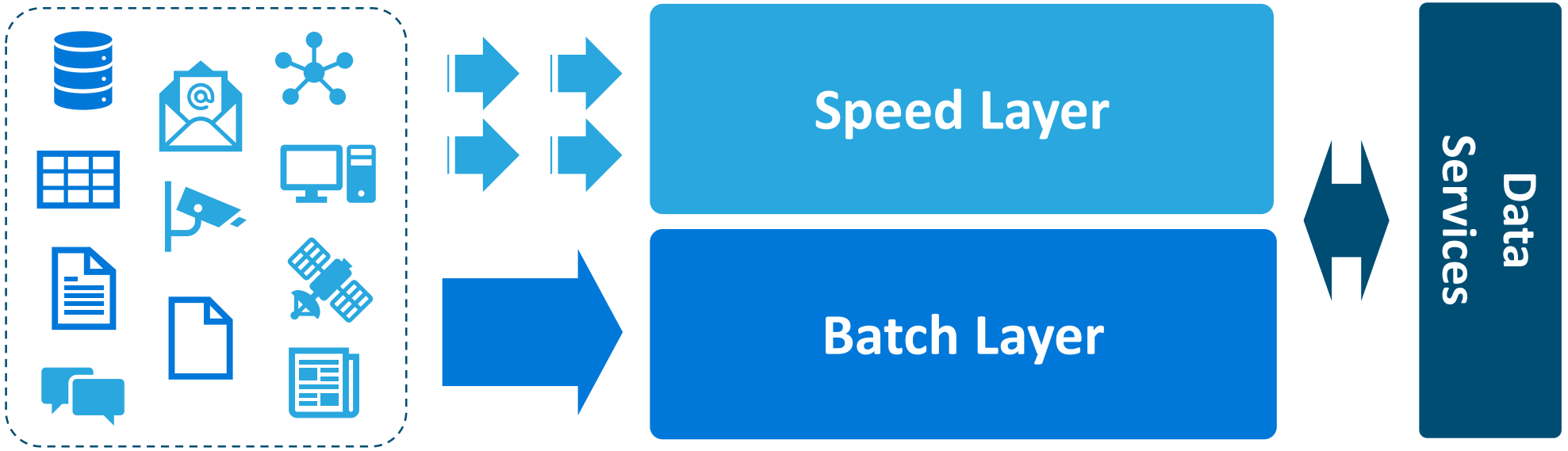


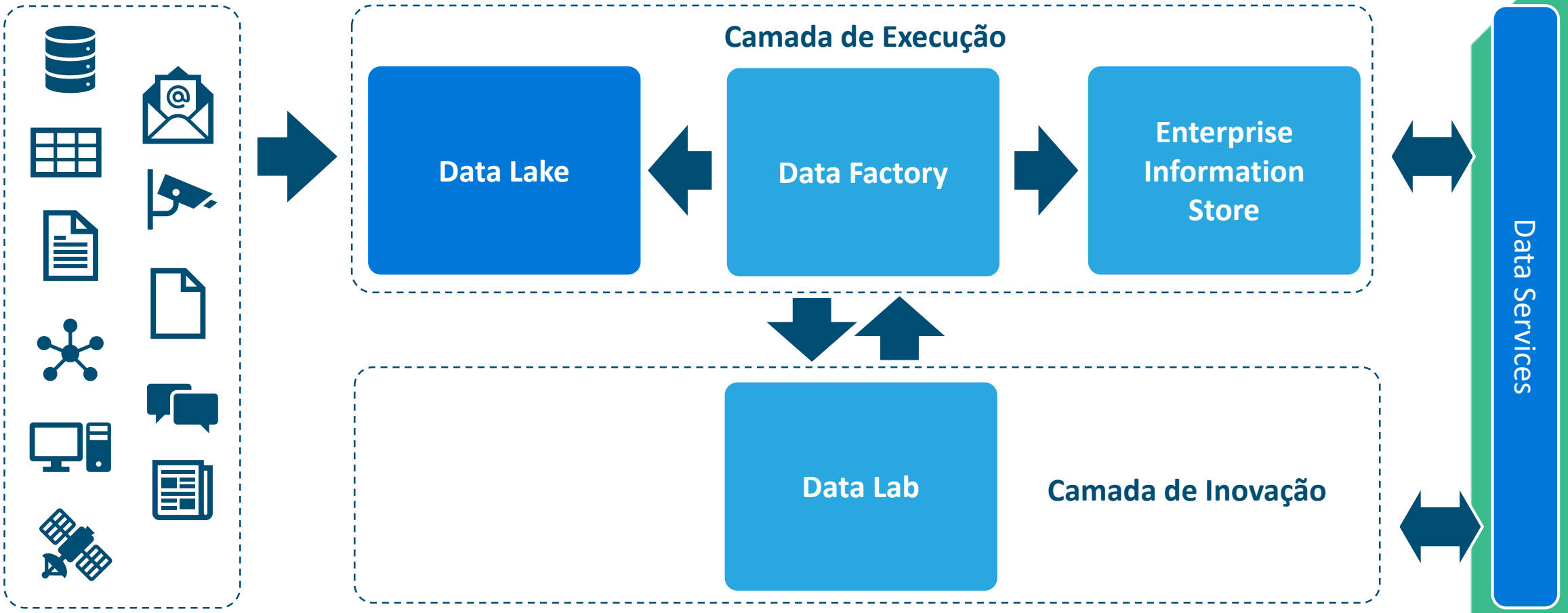
Brintell

Arquitetura conceitual para Gestão da Informação

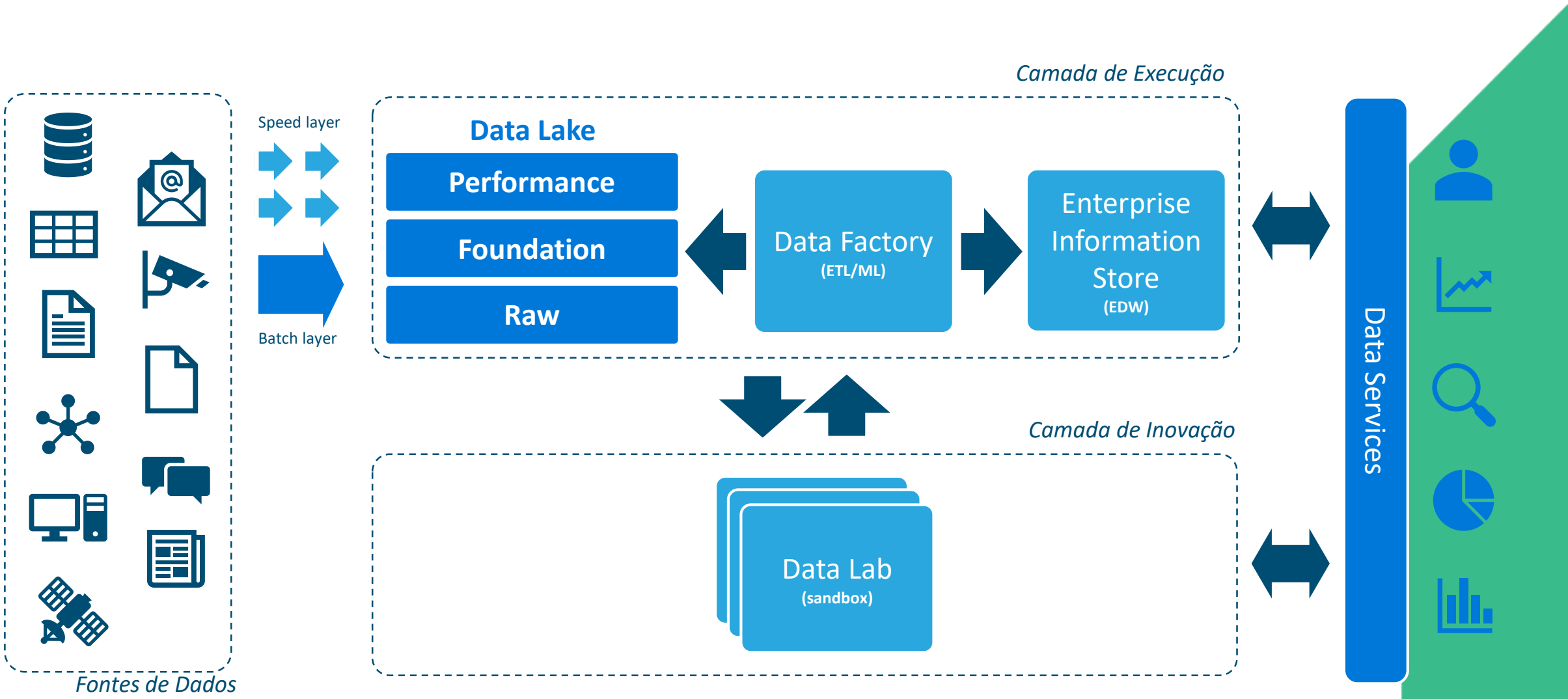


Arquitetura Lambda

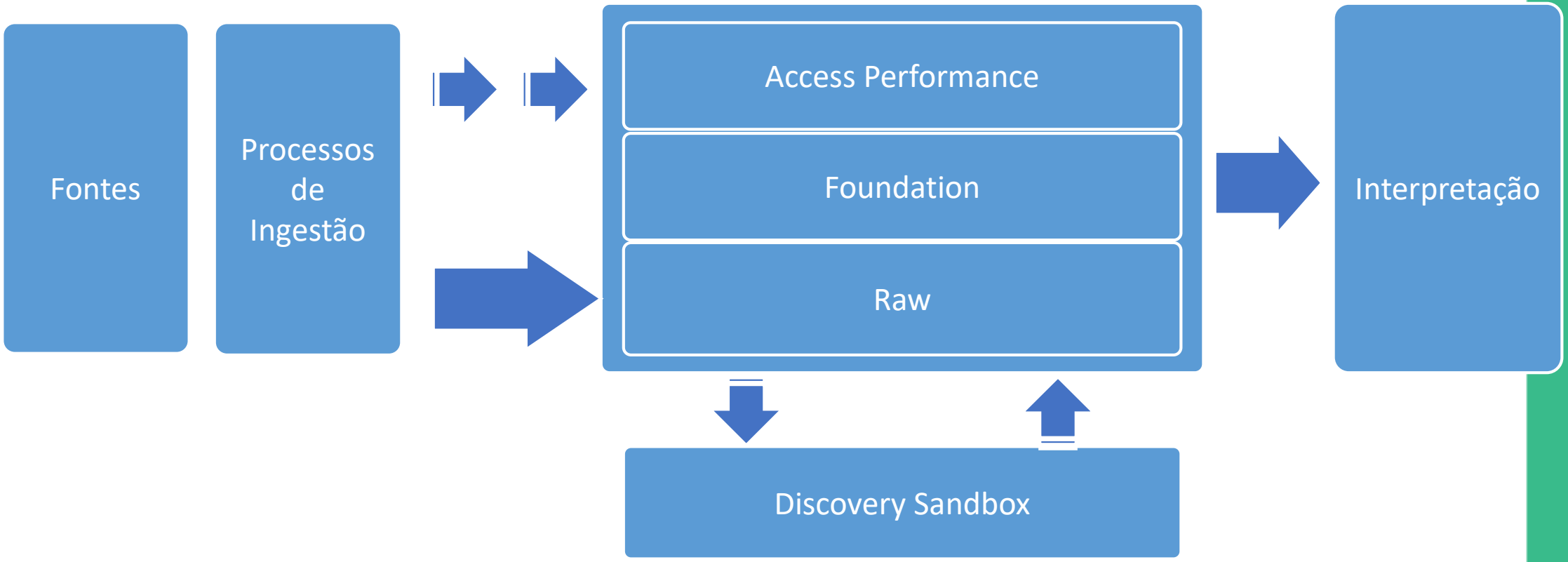



Camada de Execução
Data Lake
Data Factory
Enterprise Information Store
Data Lab
Camada de Inovação
Data Services

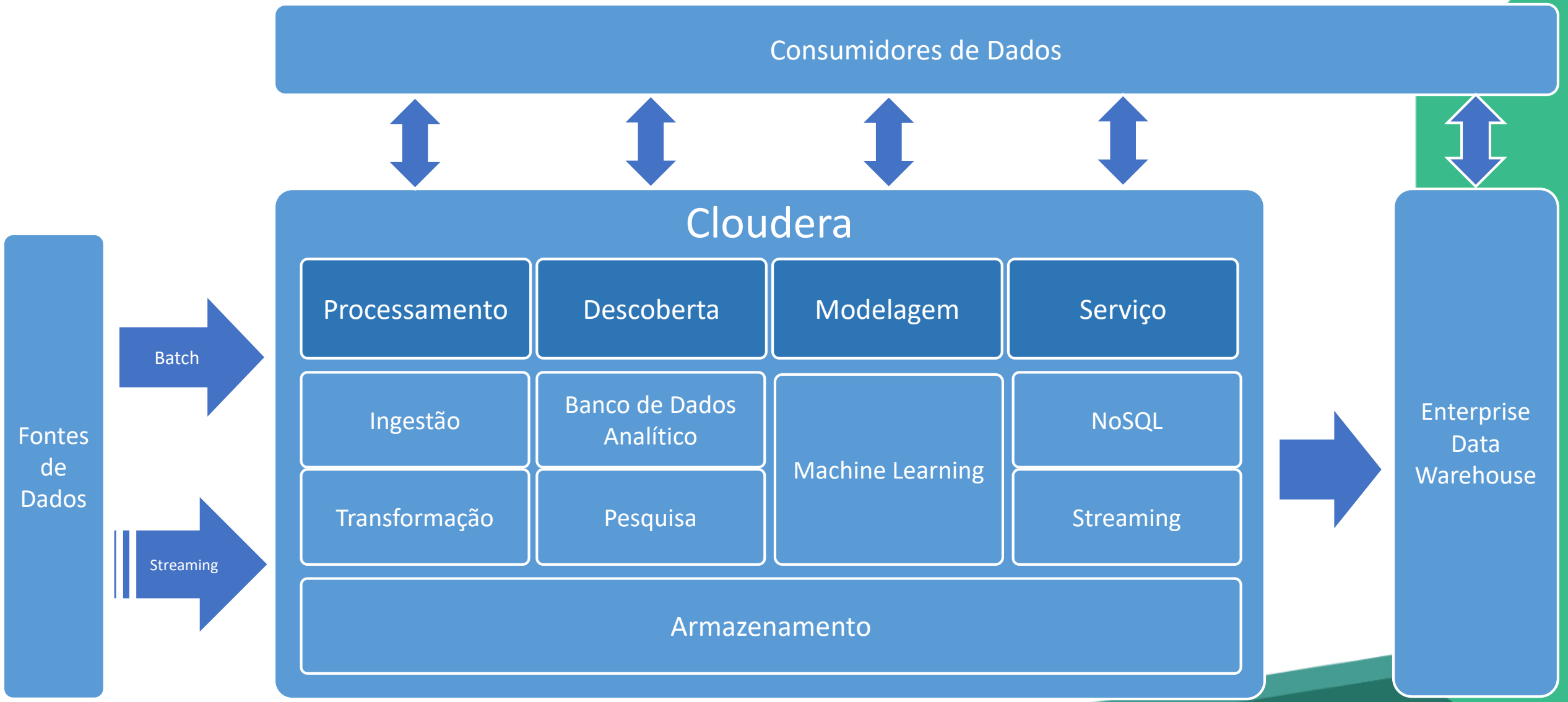
Brintell
Intelligent Analytics



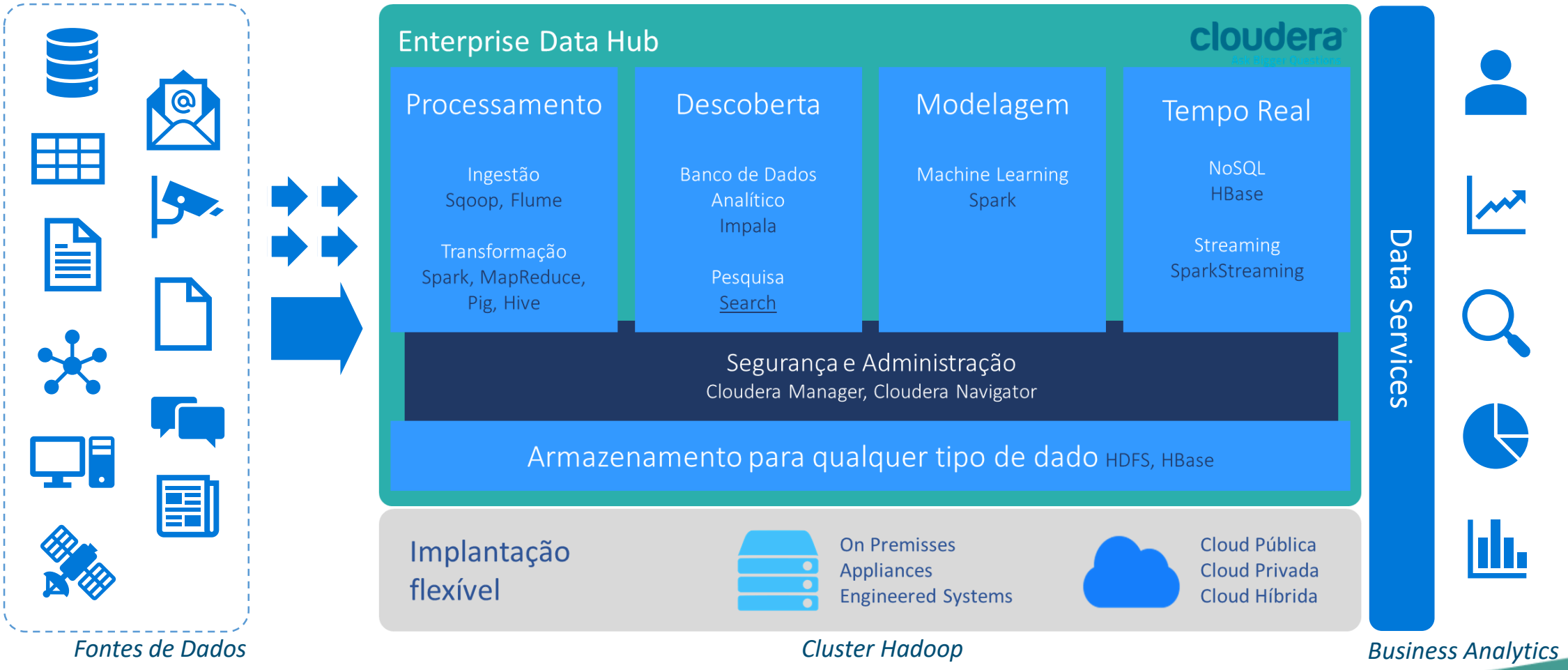
Arquitetura Lambda



Arquitetura conceitual para Enterprise Data Hub com Data Warehouse



Brintell
Intelligent Analytics



Fontes de Dados

Cluster Hadoop

Business Analytics

Macro processos

Dados	Aquisição	Organização	Análise	Decisão

Brintell

Arquitetura Física

Worker Template

Roles: (DN[N]+NM[N])+(ID[N]|HRS[N]+SS[3])
Resources: [Cores >= drivers, 256+ GB]
OS: [see supported versions]
Hostname:
IP/Subnet: 192.168.0.[0-0]/24 (bounded 2 TOR)
Rack:/rack[01-NN]

/(root) [500GB+,RAID1]
/data01 [xfs|ext4,JBOD]
/data02 [xfs|ext4,JBOD]
/dataNN [xfs|ext4,JBOD]

Worker Instance

Roles: DN[1-10]+NM[1-10]
Resources: [16 Cores, 68 GB]
OS: [CentOS 7.3]
Hostname: wn[1-10]-edh1.Cloudera.com
IP/Subnet: 192.168.0.[0-0]/24 (bounded 2 TOR)
Rack:/rack[01-NN]

/(root) [500GB+,RAID1-A]
/(root) [500GB+,RAID1-B]
/data01 [1TB ext4,JBOD]
/data02 [1TB ext4,JBOD]
/data03 [1TB ext4,JBOD]
/data04 [1TB ext4,JBOD]
/data05 [1TB ext4,JBOD]
/data06 [1TB ext4,JBOD]
/data07 [1TB ext4,JBOD]
/data08 [1TB ext4,JBOD]
/data09 [1TB ext4,JBOD]
/data10 [1TB ext4,JBOD]
/data11 [1TB ext4,JBOD]
/data12 [1TB ext4,JBOD]

RM

Total recursos workers

Quantidade de nós: 10
Total Processadores: 160 cores
Total Memória: 680 GB
Total Discos: 120 TB

Reserva para o sistema

Sistema Operacional:	1 core / 8192 MB
Task overhead:	0 core / 8192 MB
CM Agent:	1 core / 1024 MB
HDFS DN:	1 core / 2048 MB
YARN RM:	1 core / 1024 MB
NON DFS:	10%

Recurso disponível por nó

1 contêiner por processador e disco

Total Processadores: 12 cores
Total Memória: 48 GB
Total Discos: 12 TB – NON DFS

Recurso disponível para os contêineres

(Até 12 por nó)

Total Processador: 1 core
Total Memória: 4 GB

Map

Container capacity

Map task

Map Heap

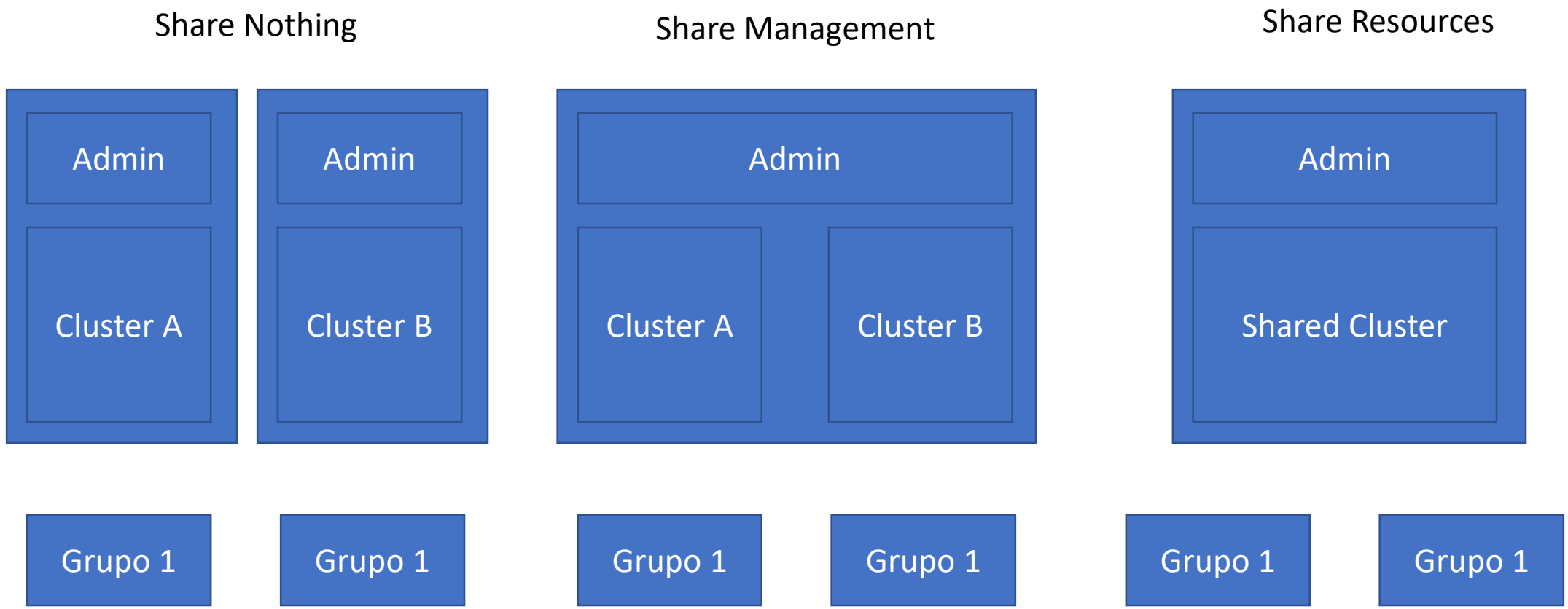
NM Process

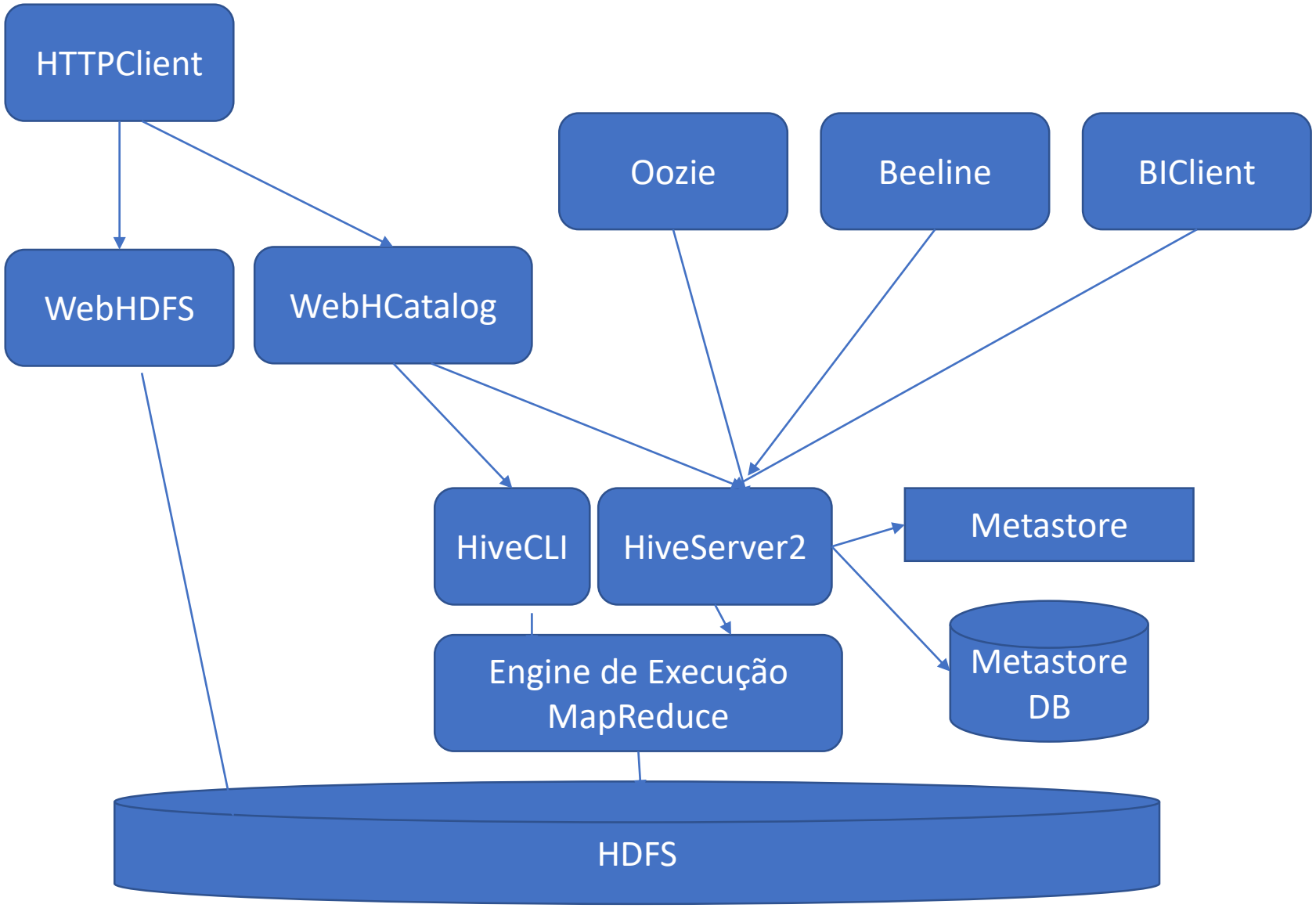
Reduce

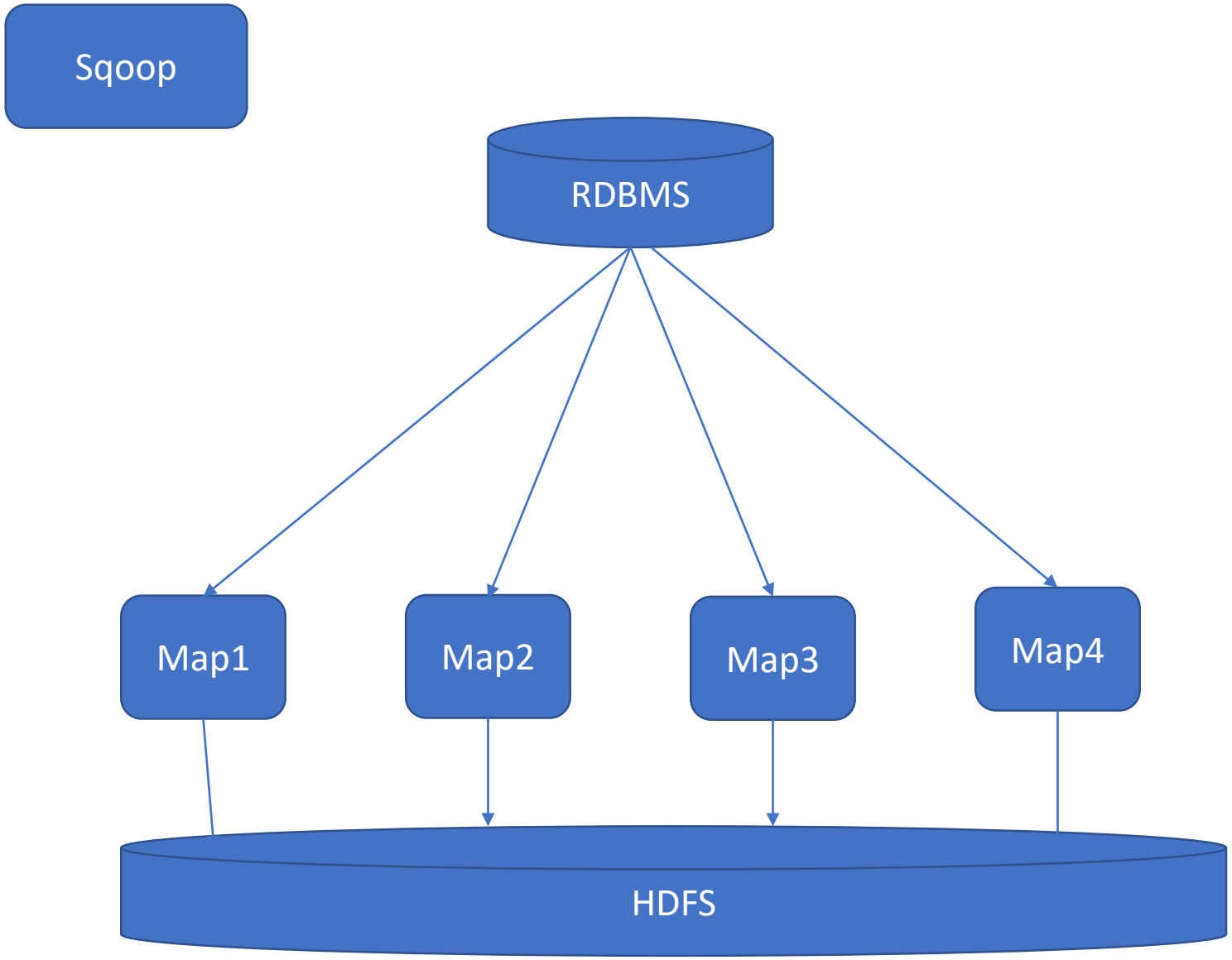
AM

Brintell
Intelligent Analytics

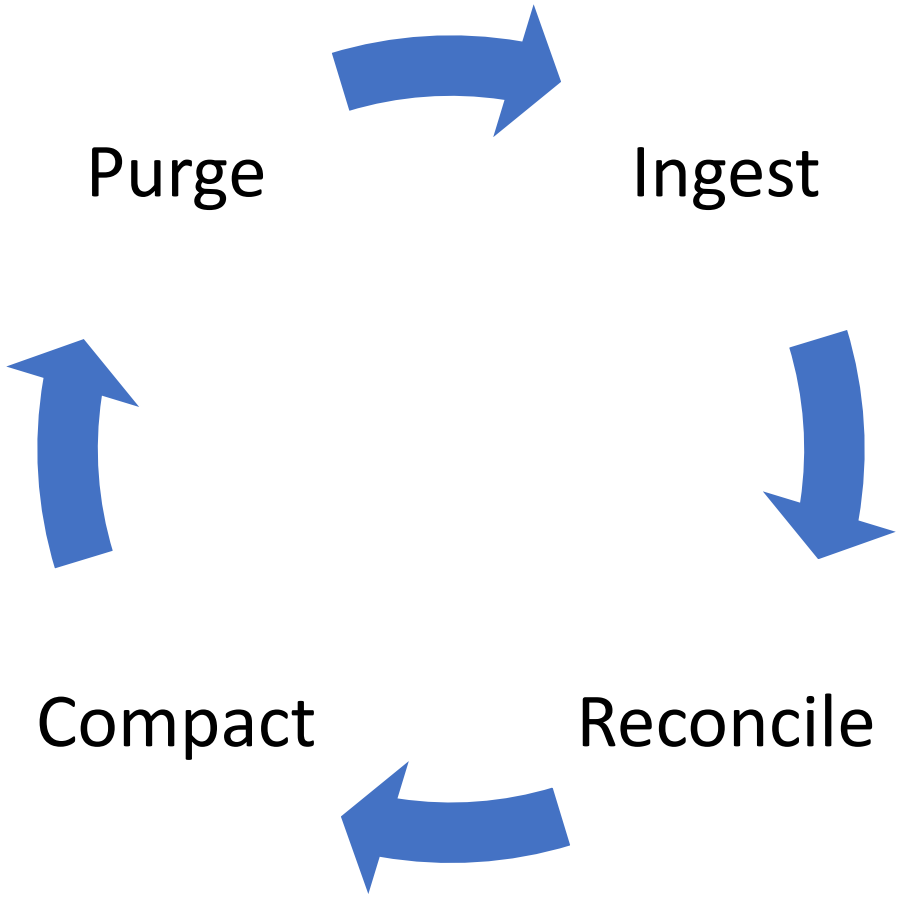
Modelos para alocação de recursos para Ambientes de desenvolvimento homologação e produção




HTTPClient
Oozie
Beeline
BIClient
WebHDFS
WebHCatalog
HiveCLI
HiveServer2
Metastore
Metastore DB
Engine de Execução MapReduce
HDFS


Sqoop
RDBMS
Map1
Map2
Map3
Map4
HDFS

Incremental Update



Quando não usar Hadoop

Se for trabalhar apenas com informações estruturadas

Se for trabalhar apenas com arquivos pequenos e baixo throughput

Se todos os dados cabem em apenas um nó

Se não for usar paralelismo

Não pode ser usado para substituir a tecnologia atual, mas para complementar

Se não tiver equipe capacitada para manter

Se não tiver como sustentar o investimento com infra (on premisses/cloud)

Se não tiver condições de lidar com o mundo open-source

Quando não usar Hadoop

Se não tiver necessidade de escalar

Se for necessário apenas a execução de tarefas simples em SQL

- SELECT G(...) FROM table GROUP BY F(...)
- collection.flatMap((k,v) => F(k,v)).groupBy(_._1).map(_.reduce((k,v) => G(k,v)))

Se o problema pode ser resolvido em apenas um banco relacional

Se precisar garantir ACID

Se for necessário apenas fazer um upgrade no hardware atual

Brintell
Intelligent Analytics

Obrigado!