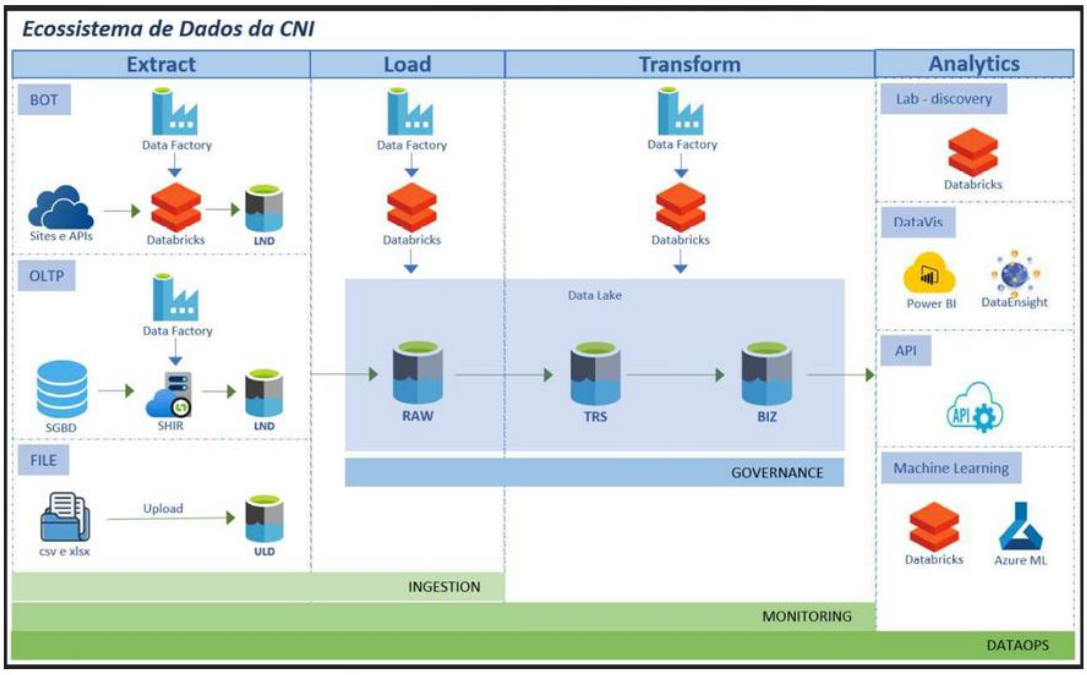
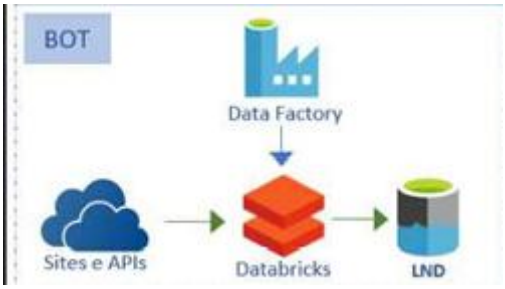
Com suas próprias palavras explique o que significa o conteúdo apresentado na figura a seguir, indicando cenários de uso em uma determinada empresa. Explique também para que servem os componentes da imagem com o intuito de demonstrar seus conhecimentos.




A arquitetura apresenta-se no formato Medalhão (Landing Zone, Bronze, Silver e Gold) usando o processo de ELT ( Extract Load Transform), extrair, carrega num outro momento transforma e disponibiliza os dados, diferente da ETL ( Extract Transform e Load) muito usada em arquiteturas relacionais.

A etapa **EXTRACT** é feito a extração das informações**.**

Com base nisso, tempo:

1 - 


Automação é feita em um notebook (Python / Spark) no Databricks.

As informações extrairdas (sites / api / webserve / wfm) é gravada em uma área de pouso (Land Zone) Azure Data Lake para serem carregadas para as camadas superiores e processadas. O Data Factory é responsável pelo gerenciamento desse pipeline.

2 - 

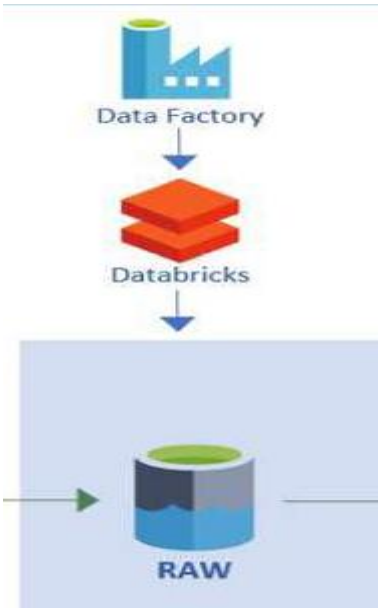

Aqui temos diferente da outra fonte, dados estruturados já processados (deveria— realidade é outra em muitos casos) em formato colunas.

O SHIR ou driver/ conector como em outras Clouds, ele quem faz a conexão com o SGBD (pode ser Oracle, SQL Server, My Sql, Teradata, etc) extrair as informações e grava na Land Zone do Azure Data Lake.O Data Factory é responsável pelo gerenciamento desse pipeline.

3 - 

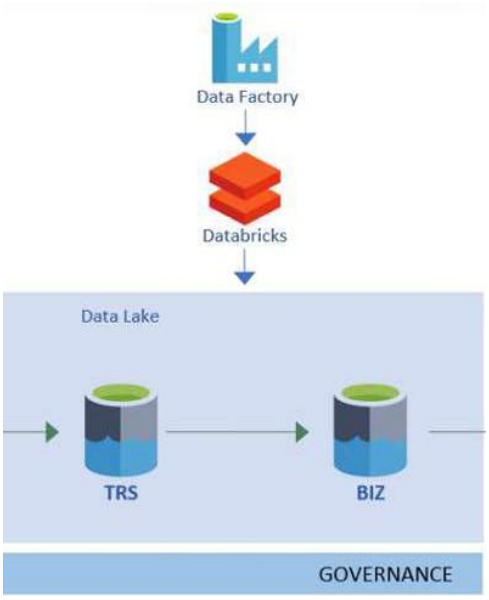

Nesse step, temos a carga de arquivos (csv ,xls, xlsx , txt, pdf , etc) feita de forma manual para uma pasta específica (ULD) do Data Lake Azure. Que também será carregada e processada.

A etapa do **LOAD**

**1-** 

Nessa etapa os dados serão carregados da Landing Zone (camada anterior) para a camada RAW (Camada **Bronze**) do Data Lake, não é feito nenhum tipo de tratamento. Novamente o uso do notebook no Databricks para fazer o papel de executor junto com o Data Factory fazendo o gerenciamento desse pipeline. Vale lembrar que os dados sem tratativa nessa camada são usados também para auditorias, rastreios de informações.

A etapa de **TRANSFORM**

1- 


Na parte de transformação temos o notebook (spark com processamento distribuídos/clusters) no Databricks executando o pipeline e o Data Factory executando o gerenciamento.
Temos mais etapas, onde as informações gravadas na camada RAW (**Camada Bronze**) são extrairdas, vai passar por um processo de higienização, tipagem correta dos dados, remoção de informações duplicadas e salvo em outra camada TRS (**Camada Silver**), sofreram um pré-processamento.

Na camada BIZ(**Camada Gold**) os dados foram transformados, enriquecidos, aplicado as regras de negócio. Estão prontos para serem consumidos pelos usuários (área negócio, área faturamento, área de operações etc.).

Aqui podemos executar as análises, tanto via tabela, quanto por ferramentas, api etc.
Por fim e muito importante a Governança dos Dados (Catalogar os dados, Compliance, Tagueamento para controle de acesso) isso mantém a confiabilidade, segurança a documentação e rastreio das informações processadas.

Etapa final **ANALYTICS**


1 - DATAOPS

Por fim nessa camada final, entrega mais valor.
Pode ser acessada por apis, o Lab-discovery mais voltado para os cientistas de dados e os analistas. O Viz Power bi, Qlink Sense, Looker, DataEnsight para criação de dashboards, relatórios. Também podemos usar como fonte para alimentar modelos de predição, treinamento de modelos de Machine Learning etc.

Todo esse processo encapsulado em práticas de DataOps.