**PERSISTÊNCIA**

**DE DADOS**

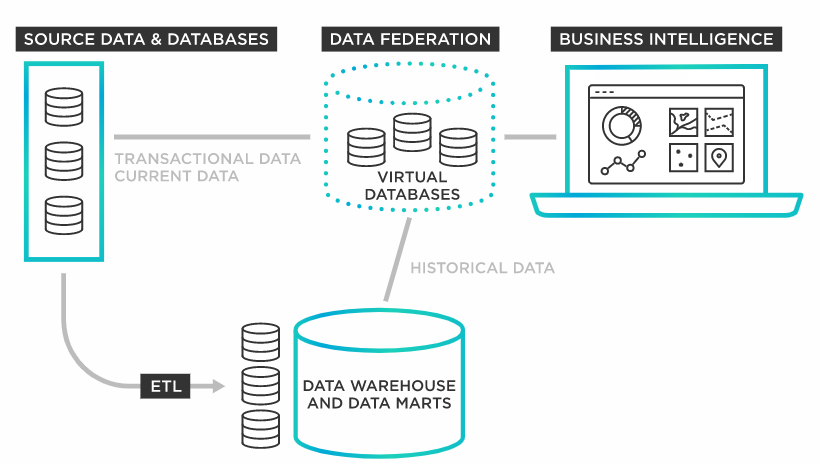
Tratando-se de um solução complexa e integrada voltada para o monitoramento periódico de frotas, fluxo de passageiros, tempo de viagens, segurança dentro dos transportes, dentre diversos outros aspectos vinculados ao IoT e centrais de monitoramento especializados, é fundamental que exista uma persistência desses dados, para que seja garantida a confiabilidade, desempenho, durabilidade e a garantia de que essa massiva quantidade de dados sejam recuperadas com todo seu histórico para análises preditivas e de tomada de decisão das grandes empresas de transporte público (ônibus) e órgãos federais associados.

A persistência de dados deve envolver o salvamento de dados em um sistema de armazenamento não volátil para que o valor dos dados seja recuperado de forma confiável posteriormente. Os dados assumem formatos estruturados, envolvendo o uso de tecnologias de armazenamento projetadas para preservar os diferentes tipos de dados em sua estrutura adequada, incluindo quaisquer metadados que descrevam a origem, o formato ou o histórico desses dados. Para a solução, a persistência será feita por meio do banco de dados seguindo o modelo dimensional (sql server), data warehouse em nuvem e data lake, garantindo assim:

* Tomada de decisão adequada
* Dados consolidados de várias fontes
* Análise de dados históricos
* Qualidade, consistência e precisão de dados
* Separação do processamento analítico dos bancos de dados transacionais, o que melhora o desempenho dos dois sistemas

A persistência é fundamental para a ciência de dados e aprendizado de máquina porque o combustível para análise origina-se da coleta de conjuntos de dados abrangentes que representam o comportamento histórico, bem como a entrada operacional atual. Embora os dados possam ser armazenados localmente em uma plataforma, eles geralmente residem em armazenamentos de dados internos ou externos, são consolidados em um data lake ou podem ser acessados ​​de fontes de dados virtuais federadas.





Fonte: TIBCO – Federação de dados

As dimensões e fatos serão feitas de maneira incremental, ou seja, tabelas físicas que estão em constante atualização e crescimento, sendo exigido assim dados não voláteis. Dessa maneira, utilizaremos a persistência desde o momento da obtenção dos dados brutos (stages) até as fatos que servirão de análise para a solução.

A camada persistente contém um conjunto de tabelas persistentes que registram o histórico completo de alterações nos dados da tabela/consulta que é a origem da tabela persistente. A origem pode ser uma tabela/arquivo de origem, uma consulta de origem, outra tabela de preparo ou uma visualização na camada de transformação.

Em uma tabela persistente, pode haver várias versões de linha para cada linha encontrada na fonte. Cada versão de linha tem uma data de vigência e uma data de término que marcam o intervalo de datas em que essa versão de linha estava em vigor (ou existia).

Tecnicamente falando, uma tabela persistente é uma tabela bitemporal. Ela permite realizar consultas em duas linhas de tempo: tempo válido e tempo de transação. A hora válida é a hora em que uma linha é efetiva (ou seja, a linha efetiva e a data e hora de término). O tempo de transação denota a hora em que a versão da linha foi registrada no banco de dados. A tabela persistente oferece suporte ao tempo de transação marcando cada versão de linha com um ID de execução em lote inserido e atualizado. A execução em lote está associada a uma data e hora no banco de dados em lote.