

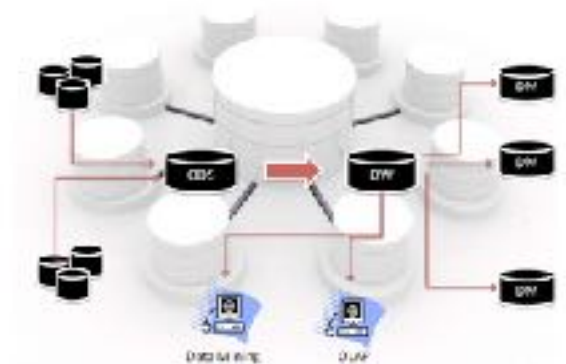
A decorative graphic on the left side of the slide consisting of a grid of squares in various shades of blue and purple, arranged in a stepped pattern.

Fundamentos de *Business Intelligence* (6º Encontro)

Prof. MSc. Fernando Siqueira

Projeto de *Data Warehouse*

Dicas





Projeto de Data Warehouse

Dicas de Revisão

- Granularidade
- Dimensão Tempo
- *Slow Changing Dimension*
- Conformidade dos Dados
- Chaves Substitutas (SKs)
- Modelo Dimensional
- Qualidade de Dados



Granularidade

- O grão é uma das mais importantes definições na modelagem de dados do DW.
- Trata-se do menor nível da informação e é definido de acordo com as necessidades levantadas no início do projeto.
- Quanto maior for a granularidade, menor será o detalhe (ou maior será a sumarização).
- Quanto menor for a granularidade, maior será o detalhamento (ou menor será a sumarização)



Dimensão Tempo

- A Dimensão temporal (ou dimensão data para alguns) é a mais importante perspectiva para a análise dos dados.
- Sem ela é difícil fazer a averiguação descritiva dos fatos.
- É imprescindível para a análise das ocorrências durante o tempo.
- É importante que seja avaliado a conformidade, analisando os níveis de detalhes exigidos para essa dimensão, mantendo a coerência de definição entre os diversos Data Marts.



Slow Changing Dimension

- O SCD tem papel fundamental para a visualização de informações históricas e armazenamento de versões dos dados de acordo com variações no tempo.
- Retrata as Dimensões que sofrem atualizações em seus campos e os classifica pelo tipo de mudança existente em cada uma delas.
- Exemplo:
 - Em modelo dimensional você tem a possibilidade de avaliar as vendas de cada filial e de cada vendedor.
 - Se um vendedor é transferido para uma outra filial, deve-se ter o tratamento dessa mudança para não ter uma avaliação incorreta do montante de uma venda para cada filial.



Conformidade dos Dados

- A conformidade propicia o compartilhamento semântico de uma mesma Dimensão entre vários *Data Marts*, mantendo a consistência das informações geradas pelas diferentes análises.
- Centraliza as alterações, e permite uma otimização no uso dos recursos computacionais.
- A conformidade de Dimensões é imprescindível para um projeto evolutivo de DW exitoso.



Chaves Substitutas (*Surrogate Keys*)

- Chaves substitutas ou artificiais (surrogate keys), são chaves desprovidas de inteligência de negócio.
- Utilizadas como artifício para o versionamento dos dados e junção entre tabelas Fatos e de Dimensões.



Modelo Dimensional

- Revise também o modelo dimensional resultante na busca de erros que só são evidentes após a implementação do DW.
- Verifique a existência de relacionamentos desnecessários e duplicidade de informações sem propósito e que oneram a performance das análises e consultas.



Qualidade de Dados

- A qualidade dos dados é um dos aspectos mais importante no processo de construção do DW.
- Devemos analisar os dados armazenados a procura de inconsistências e deformidades.
- Caso seja detectado alguma anormalidade nos dados, o processo de ETL deverá ser verificado e revisado para garantir que as informações armazenadas possuam valor para a utilização na tomada de decisão empresarial.



Qualidade de Dados

- A qualidade dos dados é um dos aspectos mais importante no processo de construção do DW.
- Devemos analisar os dados armazenados a procura de inconsistências e deformidades.
- Caso seja detectado alguma anormalidade nos dados, o processo de ETL deverá ser verificado e revisado para garantir que as informações armazenadas possuam valor para a utilização na tomada de decisão empresarial.

Self-service BI

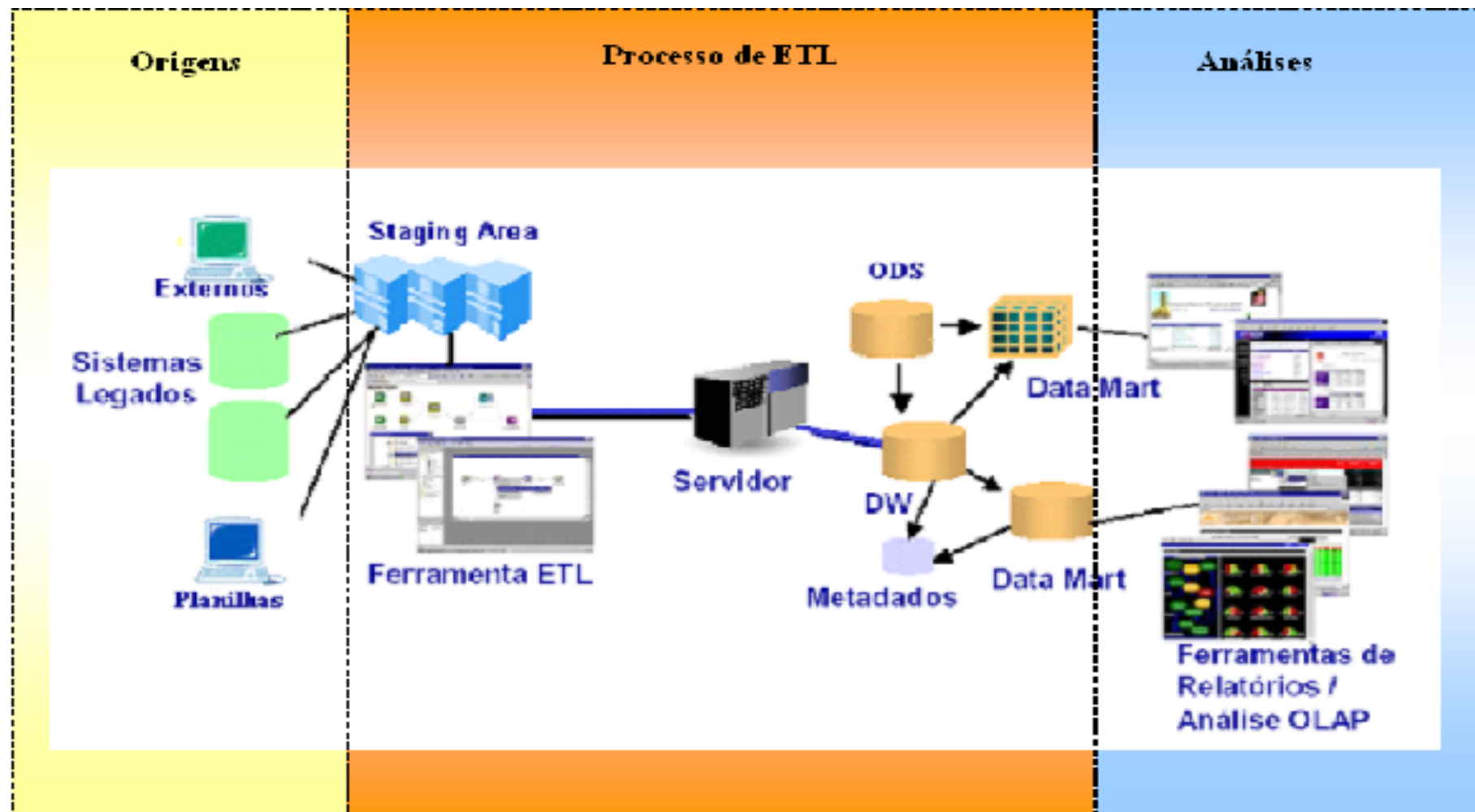




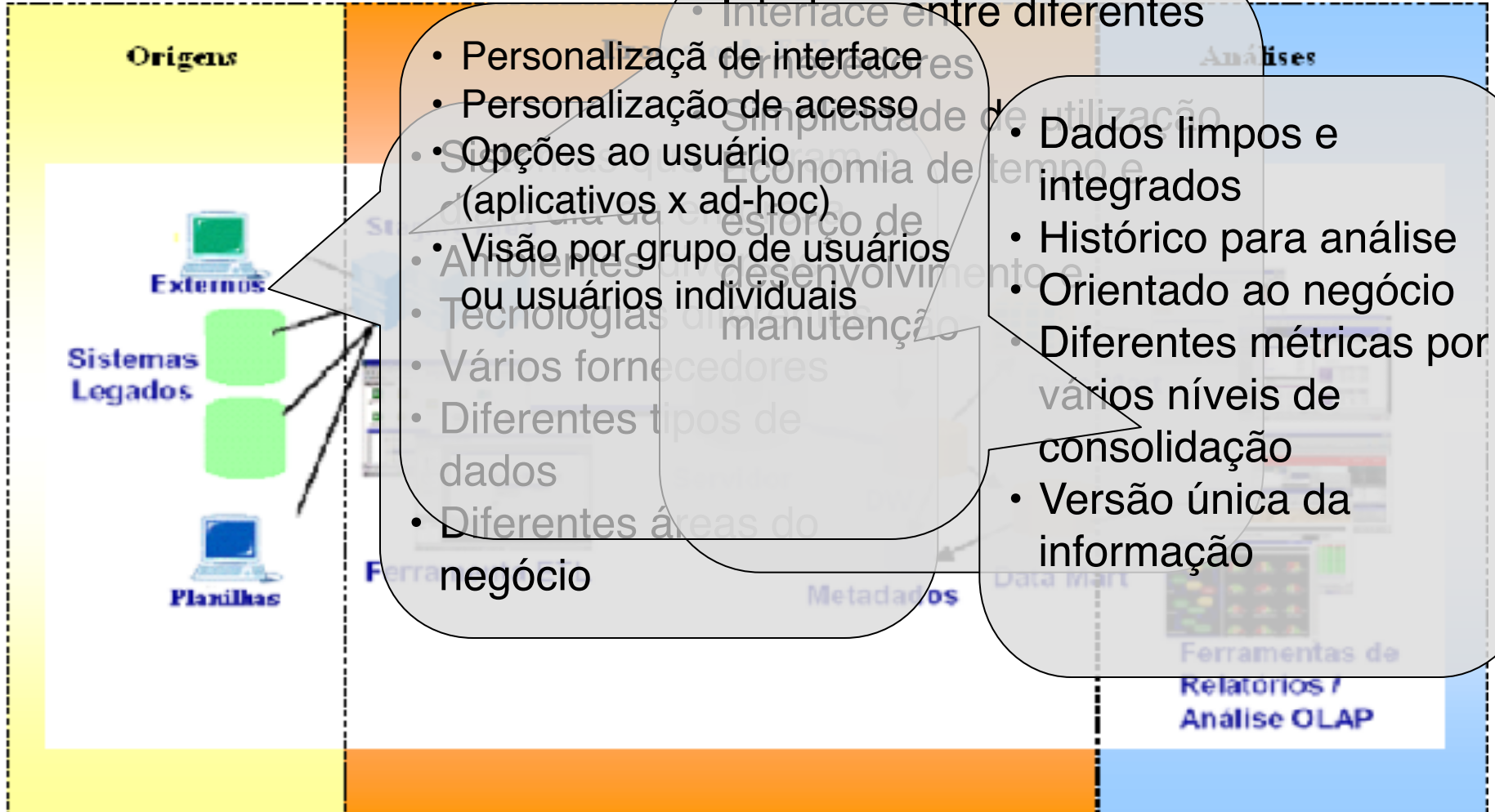
Business Intelligence

Visão Geral

O Processo BI ...



O Processo BI ...



Self-service BI



Self-Service BI

Conceito

- Criar e implantar suas próprias análises apoiado na estruturação de uma arquitetura corporativa e de ferramentas que entreguem autonomia na concepção e modelagem para a implementação de relatórios e *dashboards*.



Self-Service BI

Contexto



- A **complexidade do negócio** cresceu além da capacidade de organizações e indivíduos para compreendê-lo e gerenciá-lo.
- A necessidade de velocidade **analítica**, flexibilidade e inovação exigidas pelo negócio digital mudou o investimento analítico para unidades de negócios dos departamentos de TI tradicionais, mas a mudança em si acrescenta ainda mais complexidade.
- O negócio digital exige análises efetivas e inteligência de negócios para permitir decisões de negócios rápidas e confiáveis. Muitas empresas lutam para conseguir esse objetivo aparentemente simples.

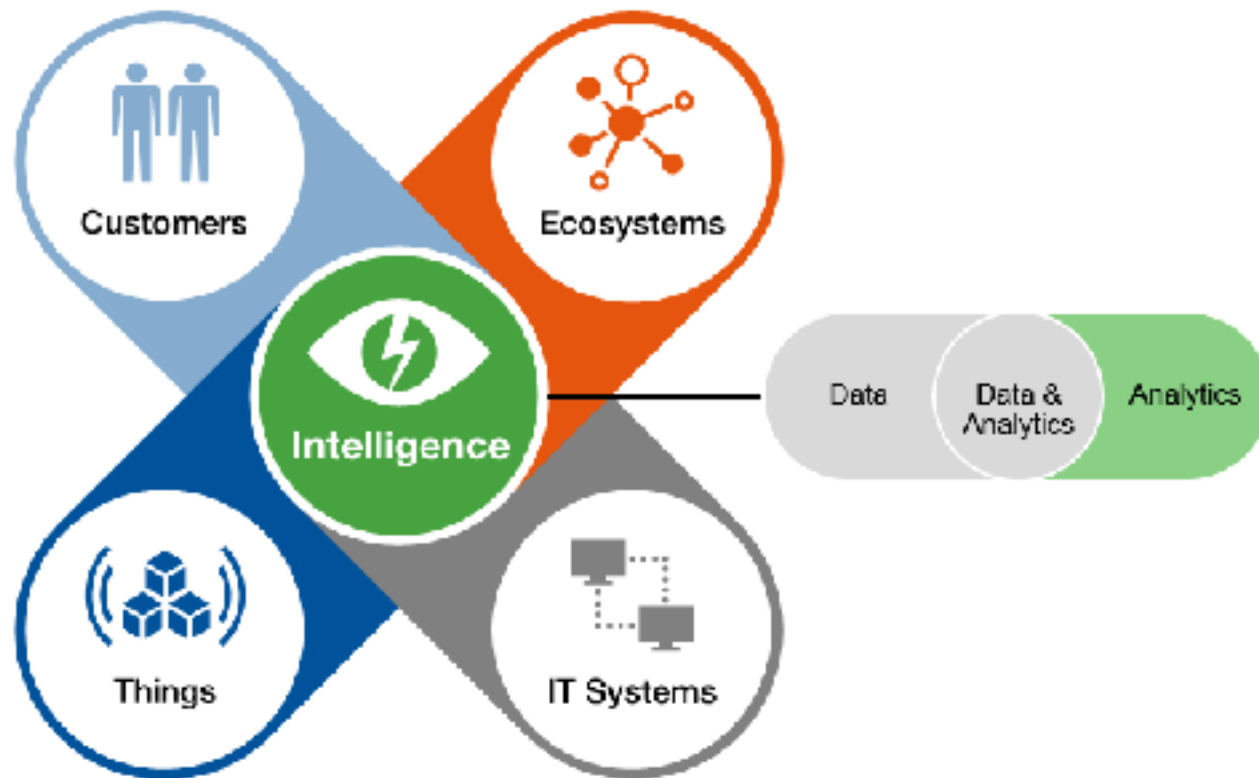
Self-Service BI

Desafios



- Líderes de dados e análise identificaram três desafios principais:
 - **Primeiro**, eles precisam verificar quais recursos de análise são essenciais para oferecer valor comercial.
 - Em **segundo** lugar, eles devem gerenciar os papéis, habilidades e conhecimentos emergentes que permitirão que as equipes de análise façam seu próprio trabalho e treinem usuários empresariais para realizar suas próprias análises quando apropriado.
 - Em **terceiro** lugar, eles são dificultados por processos e tecnologias existentes na criação de soluções de análise ágeis que suportam diretamente iniciativas empresariais digitais.

Self-Service BI



Self-Service BI

Pontos de Atenção



- As necessidades de cada empresa é que direcionarão a solução e as questões como segurança, integridade e disponibilização da informação, que devem ser cuidadosamente analisadas.
- É preciso planejar a adoção de soluções self-service, assim como implantar processos de utilização e governança consistentes. Porque a competitividade de sua organização no futuro depende da definição dessas políticas hoje.

Self-Service BI

Pontos de Atenção



- Self-service BI não significa necessariamente que o usuário buscará as informações nas fontes de dados, consolidará as fontes distintas, criará modelos que integre tais informações.
- A proposta é que o usuário facilmente, e de forma planejada e projetada, “sirva-se” de gráficos, dimensões e medidas para elaborar seus painéis com liberdade e autonomia, em um ambiente seguro e com governança.
- A TI é responsável pela construção do ambiente analítico e na condução da governança.

Self-Service BI

Pontos de Atenção



- Custo das ferramentas
- Redundância de Dashboards/Relatórios
- Perda de Desempenho
- Ausência de Governança



Ferramentas de Visualização de Dados

Ferramentas de BI

MicroStrategy

 Microsoft

 **birst**

 **SAP**

 **sas**

 **alteryx**

 **tableau**
SOFTWARE

PROGNOZ

 **Information
Builders**

 **Logi**
ANALYTICS

 **TIBCO**

 **DATAWATCH**

 **pentaho**
A HCL Group Company

 **Qlik**

ORACLE

Ferramentas de BI

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Visualização de Dados

"O uso de representações visuais para explorar, dar sentido e comunicar dados".

- Visualização de dados vs. Visualização de informações
 - Informação = agregação, resumo e contextualização de dados
- Relacionado a gráficos de informação, visualização científica e gráficos estatísticos
- Muitas vezes inclui gráficos, ilustrações, ...

Visualização Analítica

- Um termo recentemente cunhado
 - Visualização de informações + análise preditiva
- Visualização de informações
 - Descritivo, orientado para trás
 - "O que aconteceu" o que está acontecendo “
- Análise preditiva
 - Preditivo, focado no futuro
 - "O que vai acontecer" "por que isso vai acontecer“
- Há um forte movimento em direção à análise visual

visual analytics

O Crescimento da *Visualização de Dados e Analítica*

- Quadrante Mágico para Plataformas de *Business Intelligence e Analytics*
- Muitas empresas de visualização de dados estão no 4º quadrante
- Existe um movimento para a visualização

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2018)



O Crescimento da *Visualização de Dados e Analítica*

- Principais *players*
 - Tableau, QlikView, Power BI, ...
- Aumento do foco dos *big players*
 - MicroStrategy melhorando Visual Insight
 - Microsoft reforçando PowerBI
 - IBM evoluindo o Cognos Insight
 - Oracle adquirindo Endeca
 - SAP reforçando o Lumira
 - SAS evoluindo Visual Analytics



Ferramenta de Visualização Prática

Prática

- Tableau
 - <https://www.tableau.com/pt>
- Power BI
 - <https://powerbi.microsoft.com/pt-br/>
- Qlik
 - <https://www.qlik.com/pt-br/>
- Portal da Transparência
 - <http://www.portaltransparencia.gov.br/downloads/>





LDW – Logical Data Warehouse



LDW

Introdução

- Nas últimas décadas, o DW se tornou uma ferramenta essencial para que as organizações compreendam os clientes, melhorem os processos e até executem as operações diárias.
- No entanto, forças convergentes aceleram as mudanças de negócios e tecnologia que desafiam a eficácia do DW tradicional.
- Como resultado, muitos pressupostos do DW e práticas comuns não podem mais ser considerados como adequados para o atual processo de tomada de decisão.

LDW

Motivação

- Os seguintes pontos propuseram oportunidades e desafios para a evolução do DW:
 - As mídias sociais: embora a mídia social seja uma mina de ouro para análise de sentimentos, o DW não consegue processar e integrar redes sociais devido a barreiras causadas por dados não estruturados e volumes de dados maciços.
 - Computação em nuvem: infraestrutura como serviço (IaaS) e plataforma como serviço (PaaS) menor custo de entrada, diminui os requisitos de habilidades e entrega soluções mais rápidas. Em contraste, as análises "out-of-the-box" oferecidas pelo software como aplicações de serviço (SaaS) reduzem a necessidade do DW.



LDW

Motivação

- Os seguintes pontos propuseram oportunidades e desafios para a evolução do DW:
 - Mobilidade: novos dispositivos móveis podem capturar instantaneamente dados de alta fidelidade - por exemplo, tempo, localização e identidade. Eles também permitem que *business intelligence* (BI) atinja uma base de usuários mais ampla. O desafio é processar dados capturados rapidamente e fornecer insights relevantes para os contextos dos usuários.

LDW

Motivação

- Os seguintes pontos propuseram oportunidades e desafios para a evolução do DW:
 - Informações: grande análise de dados apresenta vantagens competitivas e novas oportunidades de negócios. No entanto, 3Vs - de características importantes dos dados impõem grandes desafios ao DW.
 - Primeiro, o volume de dados está crescendo exponencialmente a cada ano, e o TDW muitas vezes não consegue atender a SLAs.
 - Em segundo lugar, as organizações devem trabalhar com uma variedade cada vez maior de formatos de dados - por exemplo, feeds, mídia social e informações contextuais; O TDW não pode processar uma variedade de formatos de dados.
 - Em terceiro lugar, o negócio de hoje requer informações em tempo real (velocidade de dados), mas a maioria dos TDWs ainda operam em um mundo em lote.



LDW

Conceito

- “O *Logical Data Warehouse* (LDW) é uma arquitetura de gerenciamento de dados para análise que combina os pontos fortes dos armazéns de repositório tradicionais com gerenciamento de dados alternativo e estratégia de acesso.”



LDW

Benefícios

- Os principais benefícios do LDW são melhorar a tomada de decisões e proporcionar uma vantagem competitiva.
- O LDW aproveita o potencial da inteligência de negócios em dados em repouso e em movimento, a fim de fornecer informações poderosas para ambientes operacionais e estratégicos.
- Novos paradigmas de computação, como Hadoop e CEP, utilizados pelo LDW, têm potencial para ajudar as organizações a atender às crescentes demandas de 3Vs.



LDW

Benefícios

- O LDW também melhora a agilidade e reutilização do *data warehouse* seguindo uma arquitetura em camadas.
- Os sistemas podem ser alterados de forma mais fácil e rápida usando novos padrões de integração de dados, como a federação de dados e a integração orientada para mensagens.
- Esses padrões são mais flexivelmente acoplados e flexíveis do que o tradicional ETL.



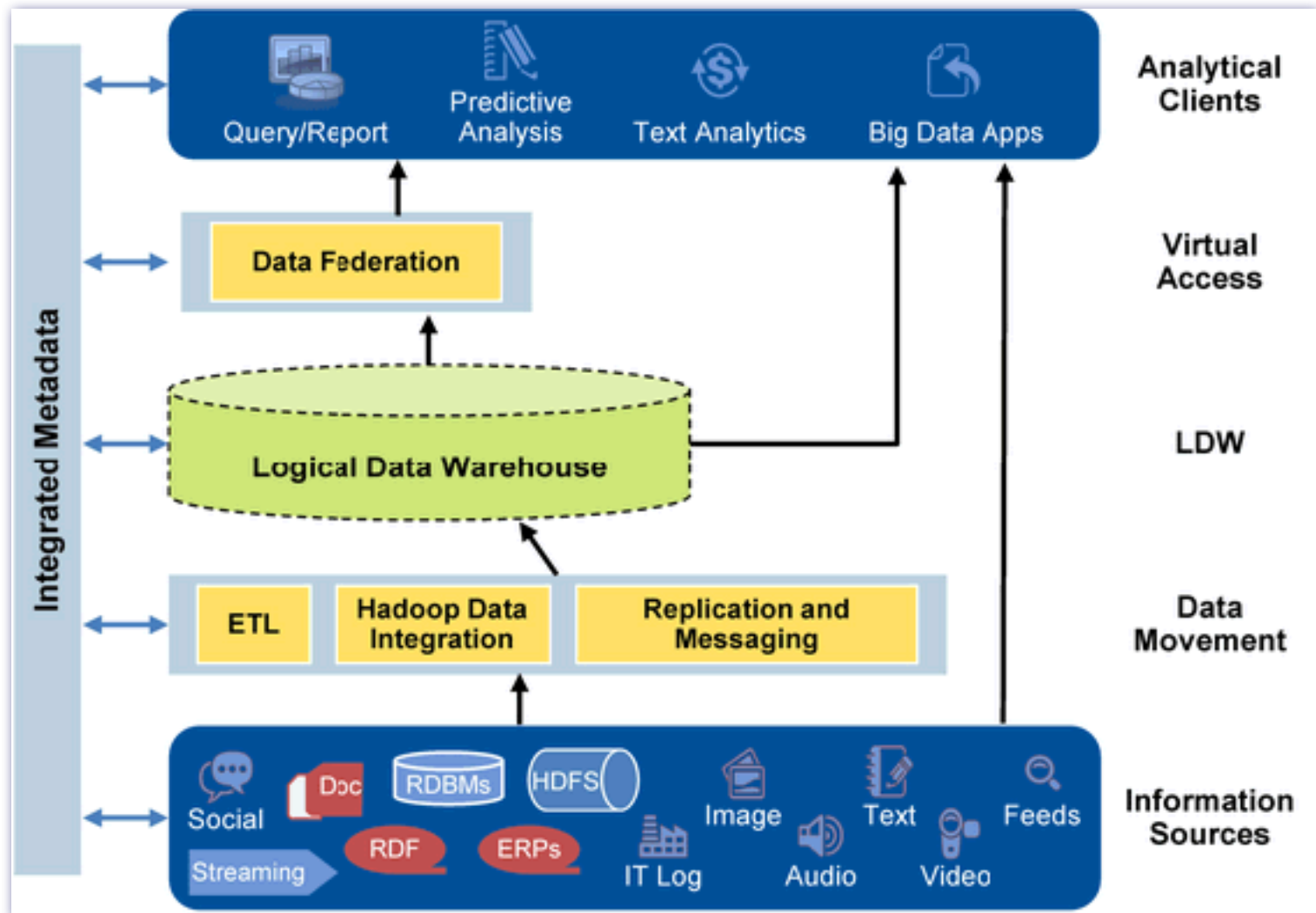
LDW

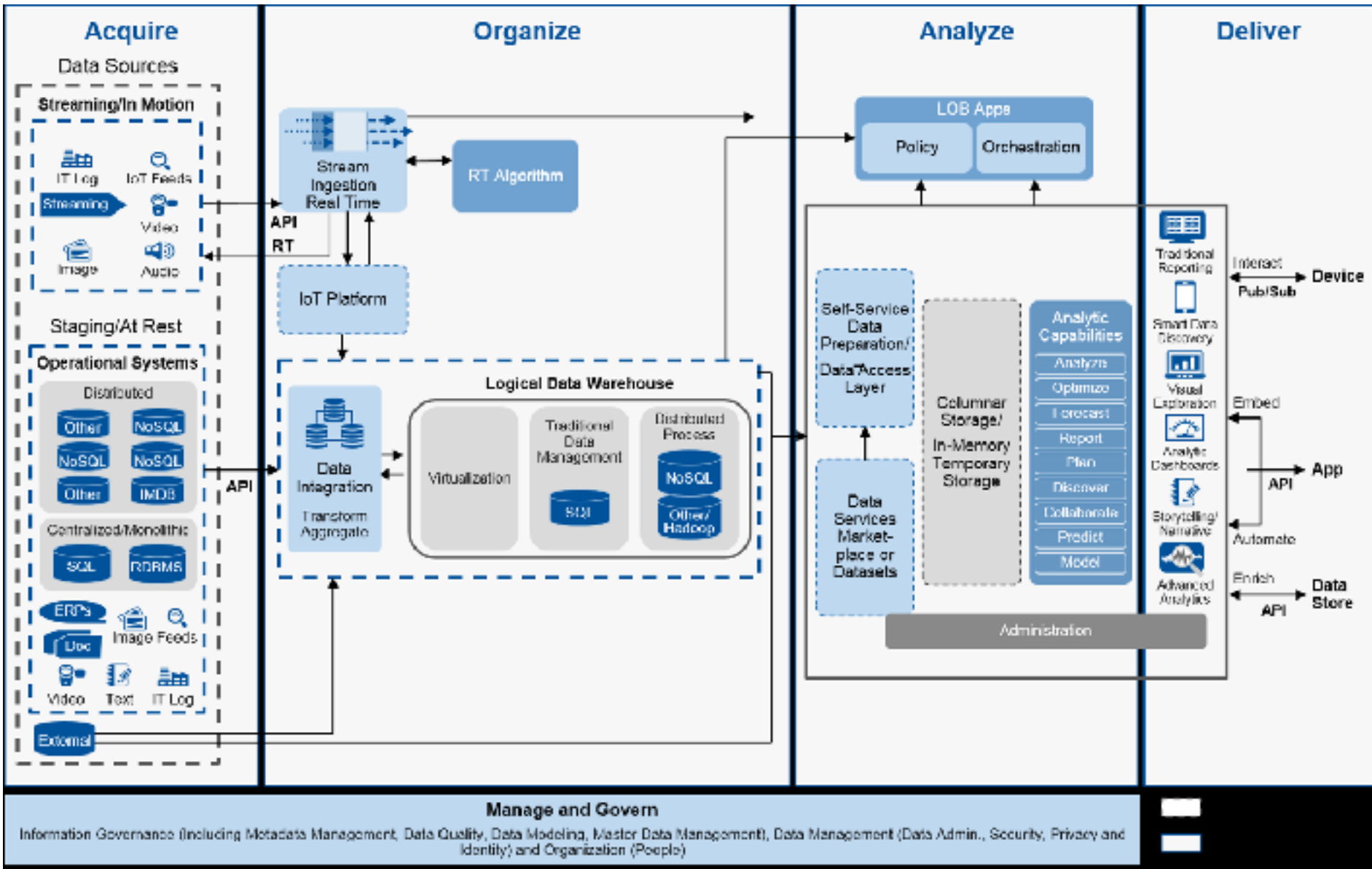
Componentes

- Repository Management
- Data Virtualization
- Distributed Processes
- Auditing Statistics and Performance Evaluation Services
- Metadata Management
- SLA Management

LDW

Arquitetura





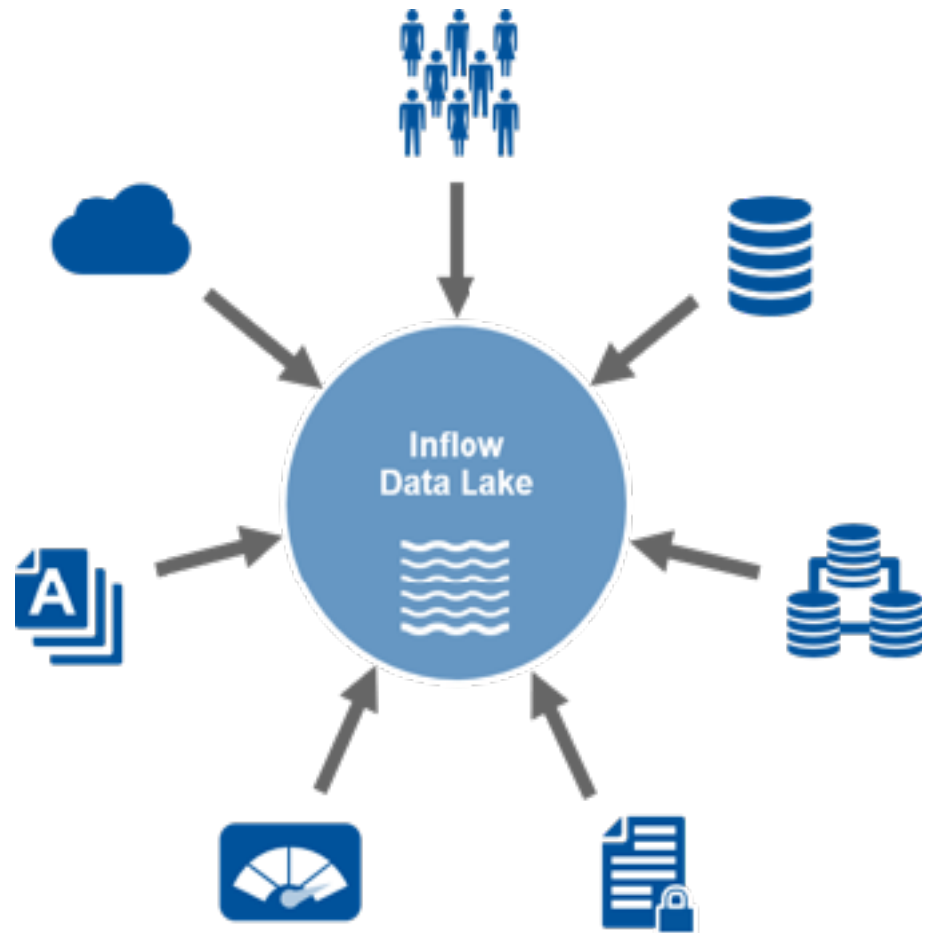
Data Lake



Data Lake

Conceito

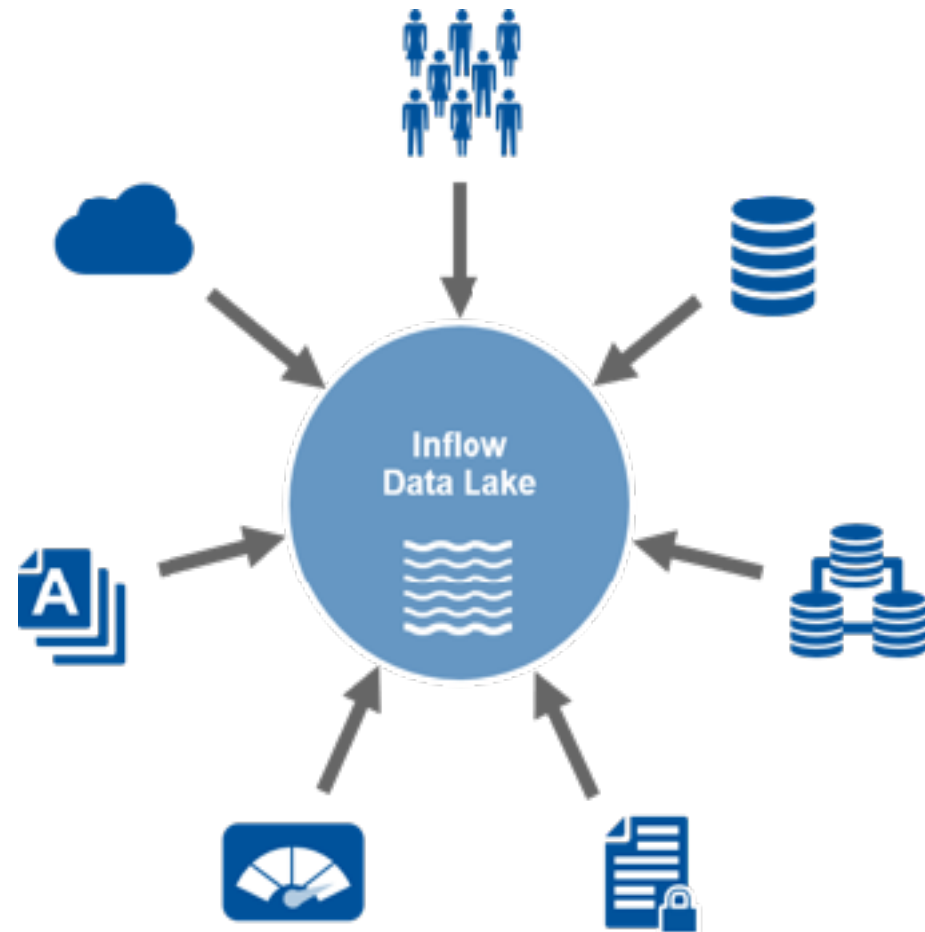
- **Data Lake** é um termo criado pelo CTO (Chief Technical Officer) do [Pentaho](#), James Dixon, para descrever um componente importante no universo da análise de dados e do [Big Data](#).
- A ideia é ter um único repositório dentro da empresa, para que todos os **dados brutos** estejam disponíveis a qualquer pessoa que precise fazer análise sobre eles.



Data Lake

Conceito

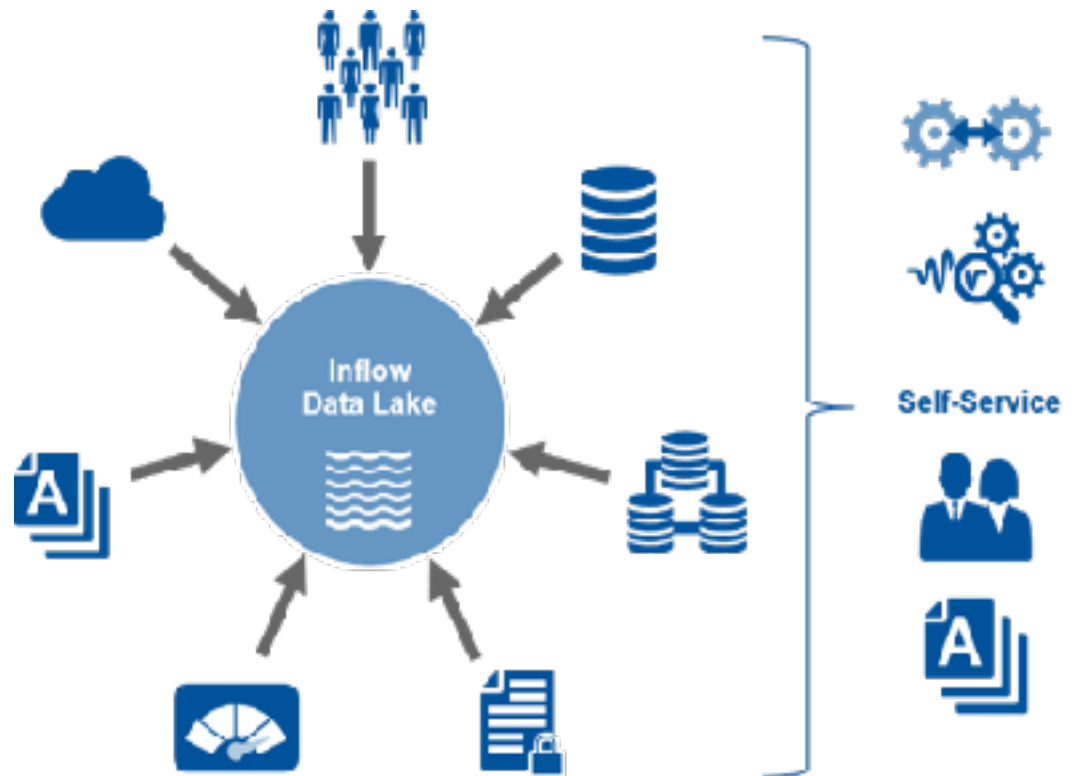
- A **data lake** contém dados não refinados quando a estrutura de dados é desconhecida antecipadamente ou quando as organizações querem aumentar a análise e a agilidade operacional, complementando seus sistemas de registro com sistemas de percepção.
- Hoje, o Gartner indica três estilos de arquiteturas de **data lake**, que são complementares.
- Um **data lake** bem sucedido começa com um único estilo de lago de dados e, em seguida, adiciona características de outros estilos à medida que amadurece.



Data Lake

Arquitetura – Inflow Data Lake

- **The inflow style** é uma arquitetura de **data lake** melhor para silos de informações.
- Acomoda uma coleção de dados de várias fontes diferentes que estão desconectados fora do "lago", mas podem ser usados em conjunto colocando-se em um único lugar.



Data Lake

Arquitetura – Inflow Data Lake

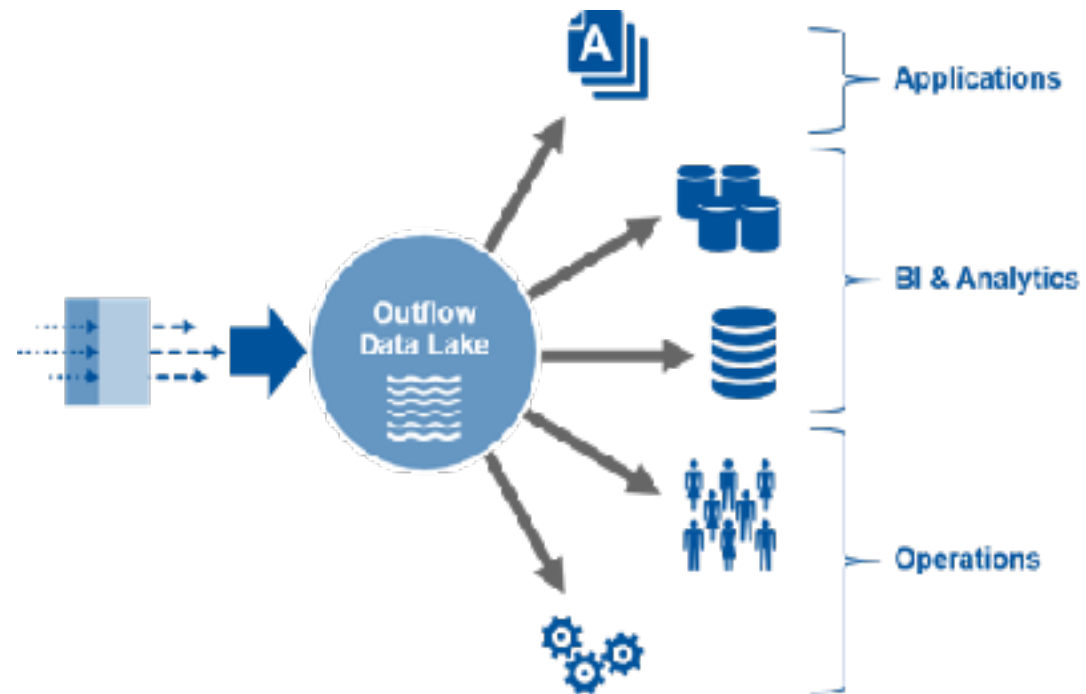


- O **Inflow Data Lake** é muitas vezes referido como "um hub de dados". Bom para o **Self-Service BI**. Usuários avançados de análise podem encontrar níveis de disponibilidade de informações sem precedentes para novas oportunidades de negócios.
 - Por exemplo: um fabricante pode se beneficiar armazenando todos os dados sobre seus produtos em um lago de dados, incluindo conteúdo tão diverso como especificações de projeto, dados de produção de produtos, informações de armazenagem, pedidos de clientes, histórico de pagamentos e dados sociais sobre o consumo de produtos. Ser capaz de analisar todas essas visões pode descobrir alguns padrões anteriormente ocultos, que podem levar a aprimoramentos de produtos bem-sucedidos e a novas formas de vender produtos.
- O **Inflow Data Lake** está mais próximo do conceito de EDW, a diferença é na capacidade de armazenar dados não refinados e desconhecidos, o que não seria prático para manter no EDW.

Data Lake

Arquitetura – Outflow Data Lake

- **The outflow style** é melhor para obter os dados mais rapidamente. É uma área de pouso (staging/ ODS) para dados recém-chegados com o intuito de ter acesso imediato.
- Emprega o conceito de schema-on-read para a interpretação e refinamento de dados *downstream*.
- Um **Data Lake** nesse estilo geralmente não é o destino final dos dados, mas pode manter os dados brutos a longo prazo para preservar o contexto para armazenamentos de dados *downstream* e aplicativos e das aplicações.



Data Lake

Arquitetura – *Outflow Data Lake*

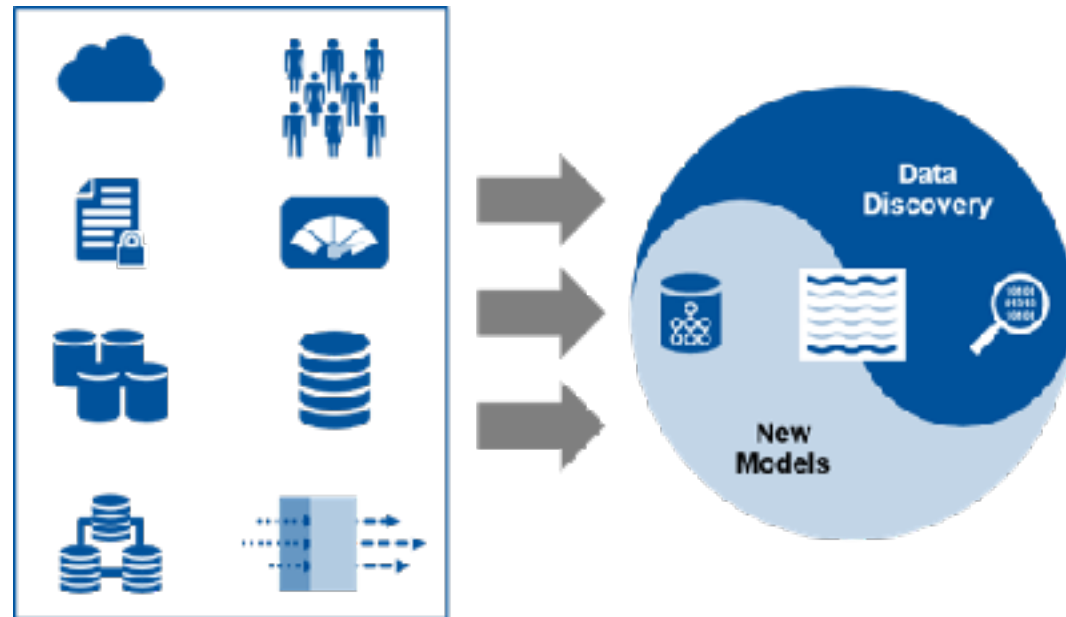


- Este estilo é o melhor para obter dados mais rápido. Assemelha-se a um ventilador porque ele bombeia os dados de seus consumidores. Para fazer isso com sucesso, requer uma governança de dados ágil que pode acomodar as necessidades de consumidores específicos, bem como deixar espaço suficiente para a interpretação localizada do esquema em leitura.
- O **Outflow Data Lake** é uma "loja de dados" operacional tradicional (ODS) e / ou uma área de teste, com dados mais granulares e mais históricos.
- As organizações podem encontrar muitos usos para o alto volume de dados, tanto de forma analítica como operacional, mas também não querem duplicar a preparação de dados em todos os processos de consumo *downstream*. A comunidade de dados ajuda os diferentes consumidores a concordar sobre o que estão procurando. Isso é muito parecido com a necessidade de um ODS para fornecer dados estruturados, com baixa latência e semelhança de significado para múltiplos processos operacionais e analíticos.
- O **Outflow Data Lake** geralmente hospeda os dados que acabarão no EDW, depois que ele for transformado no processo ETL, mas enquanto isso oferece flexibilidade para a análise imediata ou uso operacional. Este é o estilo predominante para capturar os dados IoT.

Data Lake

Arquitetura – Data Science Lab

- O **data science lab** é o estilo apropriado para habilitar a inovação de novas maneiras. É semelhante aos estilos de arquitetura **inflow** e **outflow**, mas é usado para uma finalidade mais restrita, como a segurança cibernética, a visão de cliente de 360 graus ou a análise de um motor a jato.
- Este tipo de lago é mais adequado para a descoberta de dados, o desenvolvimento de novos modelos de análise avançada, aumentar a vantagem competitiva da organização através de novos *insights*.



Data Lake

Arquitetura – Data Science Lab



- O **data science lab** é o melhor para habilitar a inovação. O desenho de sua arquitetura se assemelha ao yin e ao yang: contém elementos dos fluxos de entrada e de saída, mas comparados com eles, requer pouca ou nenhuma governança, apenas guardas para garantir que os cientistas de dados não violem a segurança, a ética e os regulamentos básicos, como a privacidade, leis para localidade de dados ou conformidade com a indústria.
- O **data science lab** também possui um escopo mais estreito de casos de uso em comparação com outros estilos de arquitetura, por exemplo, um **data lake** para pesquisas médicas.
- O **data science lab** é o mais próximo do conceito de data mart: disponibiliza um escopo específico de dados, muitas vezes com curadoria, para usuários específicos.
- Como um data mart, que se destina a um propósito específico ou a um conjunto de propósitos, este **Data Lake** não tenta resolver todos os problemas da empresa.
- Comparado com um data mart tradicional, o laboratório de ciência dos dados pode armazenar dados novos, não disponíveis anteriormente, para a descoberta de dados e para o desenvolvimento de novos modelos avançados de análise, que geralmente são conduzidos com novas ferramentas.
- Devido à natureza avançada da análise, o **data science lab** está disponível apenas para um grupo de usuários pequeno e altamente qualificado.
- A principal característica dos seus dados é o volume, porque mais dados geralmente revelam novos insights ou uma agulha no palheiro. Por exemplo, a aprendizagem de máquinas em mais dados pode produzir resultados de alta qualidade sem recorrer a algoritmos sofisticados. O volume de dados é alto, mas, eventualmente, apenas uma pequena porção dele será em uso real.
- Neste tipo de **data lake**, a perspicácia empresarial é muito importante, e os cientistas de dados geralmente trabalham lado a lado com especialistas que podem fazer as perguntas certas e interpretar os resultados.

Data Lake

Arquitetura – Data Science Lab



Abaixo estão os exemplos de dados em lagos de dados específicos da indústria:

- Banking:** Accounts, households, savings, credit, payments, ATM, risk, CRM, clickstream, mobile app, social media, call center notes.
- Public sector:** Usually data about citizens, social security, demographics, labor statistics and public policy combined with specialized information for a particular purpose, such as public health or regulatory supervision.
- Telecommunications:** Call detail records (CDR), geolocation, CRM, data from mobile devices, data about network operations.
- Manufacturing:** Materials, factory equipment (both real-time and historical), documentation, warranties, test results, ERP, bills of material, shipping, process monitoring logs and alarms, regulations.
- Healthcare:** Data from medical devices, insurance, electronic medical records, treatment protocols, data on experience of using drugs or devices, medical libraries like [PubMed](#).
- Higher education:** Data for scientific research.

Data Lake

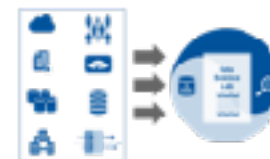
Arquitectura – Requisitos



1. Inward Lake



2. Outward Lake



3. Data Science Lab Lake

Requirements

Requirements	1. Inward Lake	2. Outward Lake	3. Data Science Lab Lake
Latency			
High latency			
Low latency			
Big Data Characteristics			
Volume			
Velocity			
Variety			
Information Management			
Data modeling			
Data governance			
Fit into the existing EA			
Scalability			
Optimized for data input			
Optimized for data output			
User concurrency			
Consumption Patterns			
BI			
Advanced analytics			
Operations			

Less Applicable



More Applicable



Data Lake

Objetivos

- O objetivo de um **Data Lake** é apresentar uma visão não refinada dos dados.
- O **Data Lake** é um conceito, não uma tecnologia, e pode contar com qualquer ferramenta escalável que possa armazenar dados em seu formato exato ou quase exato.
- Cabe aos usuários do **Data Lake** interpretar os dados e determinar a melhor aplicabilidade para os casos identificados.
 - Por exemplo, os mesmos arquivos de log podem ser interpretados para a segurança cibernética de uma forma e para automação de processos de outra forma.
- Os dados nos “lagos” são altamente contextuais, e sua interpretação requer gerenciamento de metadados sólidos e governança da informação.



Data Lake

Mitos

- Um **Data Lake** é Hadoop.
- Os **Data Lakes** são baratos de implementar.
- Obter todos os dados que você pode no **Data Lake**.
- Ao construir um **Data Lake** nenhuma modelagem de dados é necessária antecipadamente.
- Em virtude de manter todos os dados em um só lugar, você obtém uma única fonte da verdade.
- Todos podem usar o **Data Lakes**.
- Os **Data Lakes** contêm petabytes de dados brutos.
- Um **Data Lake** é o novo EDW.
- Um **Data Lake** é um método de integração de dados.
- Um **Data Lake** pode escalar para milhares de usuários.
- Se construirmos um **Data Lake** as pessoas irão usá-lo.

Data Lake

Recomendações

- Arquitecte seu **Data Lake** com uma compreensão clara de como sua organização derivará o valor dos dados que você armazena no lago. Estabeleça um roteiro para implementar um **Data Lake** no ritmo de sua empresa é capaz de consumir os resultados.
- Aplique a modelagem de dados e direcione os recursos de gerenciamento de metadados imediatamente.
- Assegurar uma governança de informação adequada (melhor descrita como advocacia de dados) para prevenir ou minimizar erros devido à má interpretação de dados. Implemente políticas e padrões de dados para evitar a "poluição nos lagos".
- Desenvolva habilidades diferentes através de treinamento e práticas que ampliam a experiência atual das equipes e contratam mentores de curto prazo para aprimorar essas habilidades.
- Engage especialistas em infraestrutura - especialistas em nuvem, armazenamento, rede e segurança - no projeto de **Data Lakes** desde a sua criação.



Data Lake

1^a. Abordagem

Data or Zone Areas:

- **Landing Zone:** which can be categorized as staging/ raw
- **Governed Enterprise Zone:** Advanced Analytics
- **Ungoverned Zone:** Advanced Analytics for data scientist



Data Lake

2^a. Abordagem

Data or Zone Areas:

- **Stage:** Persistent zone where source data is untransformed or has light transformation. Objects are Normalized.
- **Hub:** Data is persisted, transformed, and organized into business subject areas meant for heavy re-use. i.e Customer, Contract, People, etc.
- **Rpt:** Data is organized specifically for application / output needs.
- **Sandboxes:** Open to Analysts to create any data objects they wish by using data in any of the above zones



Data Lake

3ª. Abordagem

Data or Zone Areas:

- **Ingested data Zone:** this is the area where all data being ingested are landed (as system of record replica) and they are time variant. - this zone also acting as staging area for our data DWH - in relate with the DWH ETL processes
- **Analitical construct zone:** we call it data lab or sandbox are for the data scientists o do their things and stuff with data, where some data based on their analytical activities being copied to seperate partition on the Data lake seperated from the ingested area - and data realm in this area has its lifetime depend on the analytical construct activities.
- **Archive zone:** this is obviously as DWH offload area.
- **Operation Application Zone** (we called EDS - Enterprise Data Services), this is where the data lake in serving the operational system.



Data Lake

4^a. Abordagem

Data or Zone Areas:

- A **staging area** where ingested data is placed and further prepared. Only technical transformations and privacy controls are initiated on this. Data in this area should not be there for long, but in case transformations or privacy controls fail this is where the "fixing" occurs.
- A **data area** (or raw data) where the data is stored for longer periods. This is the source where the rest of the transformations work from.
- A **foundation area** where data is kept mapped to a conceptual data model. Focus here is on high-quality, high-value data.
- An access area where optimized datasets are exposed. Focus here is on particular exposure requirements, such as performance, or only having a particular view active (like latest data, rather than historized data)
- A **vault area** where sensitive data, or data that needs to be kept for litigation purposes, is kept.
- A **temp data** area which is used for intermediate data storage in processing pipelines. This data is (or should be) cleared at the end of processing steps. Data stored here is not accessible by anyone except for cleanup jobs and the generating job itself.



Data Lake


Considerações Finais

- **Data Lake** tem diferentes, muitas vezes conflitantes, definições e interpretações. Em algumas interpretações, um **Data Lake** é apenas sinônimo de "dados importantes" e é usado em vez desse termo cansado e em excesso. Em outras interpretações está com uma função enganada como um substituto para um data warehouse corporativo (EDW).
- No entanto, ambos são princípios de organização de dados, não tecnologias para si mesmos.

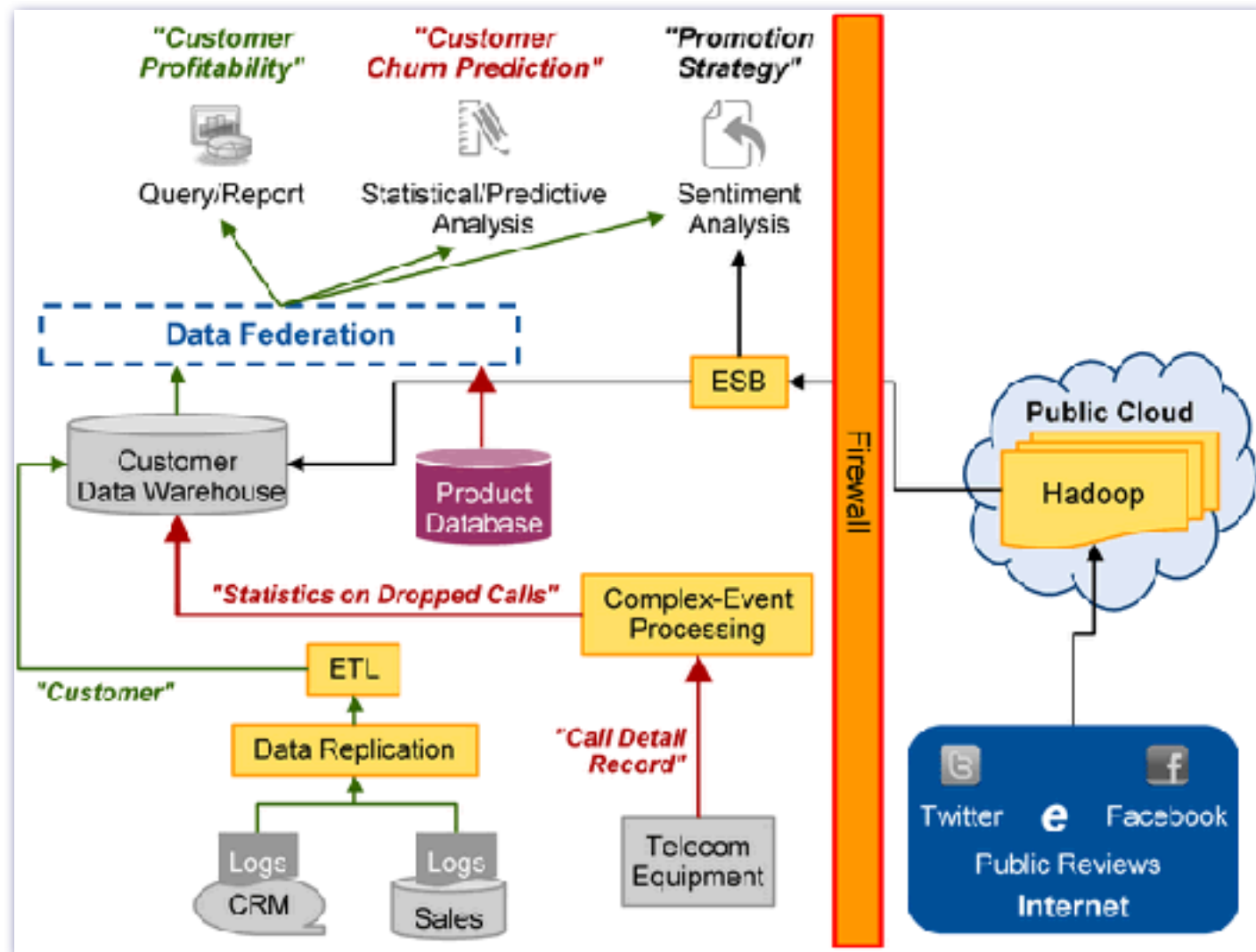


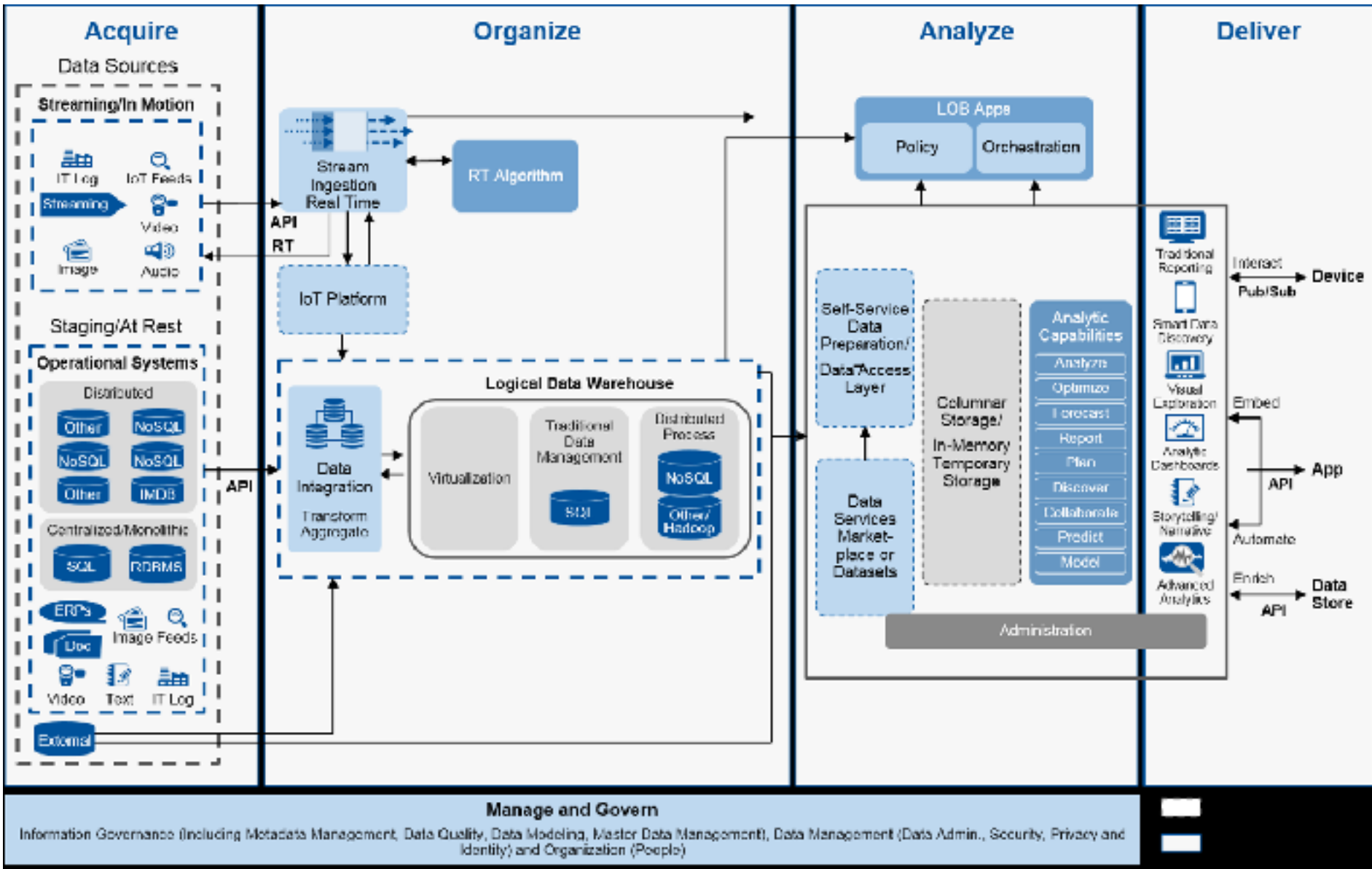
Fundamentos de *Business Intelligence*

A Nova Plataforma do BI

 Analytic Workflow Component	Traditional BI Platform	Modern BI Platform
Data source	Upfront dimensional modeling required (IT-built star schemas)	Upfront modeling not required (flat files/flat tables)
Data ingestion and preparation	IT-produced	IT-enabled
Content authoring	Primarily IT staff, but also some power users	Business users
Analysis	Predetined, ad hoc reporting, based on predetined model	Free-form exploration
Insight delivery	Distribution and notifications via scheduled reports or portal	Sharing and collaboration, storytelling, open APIs

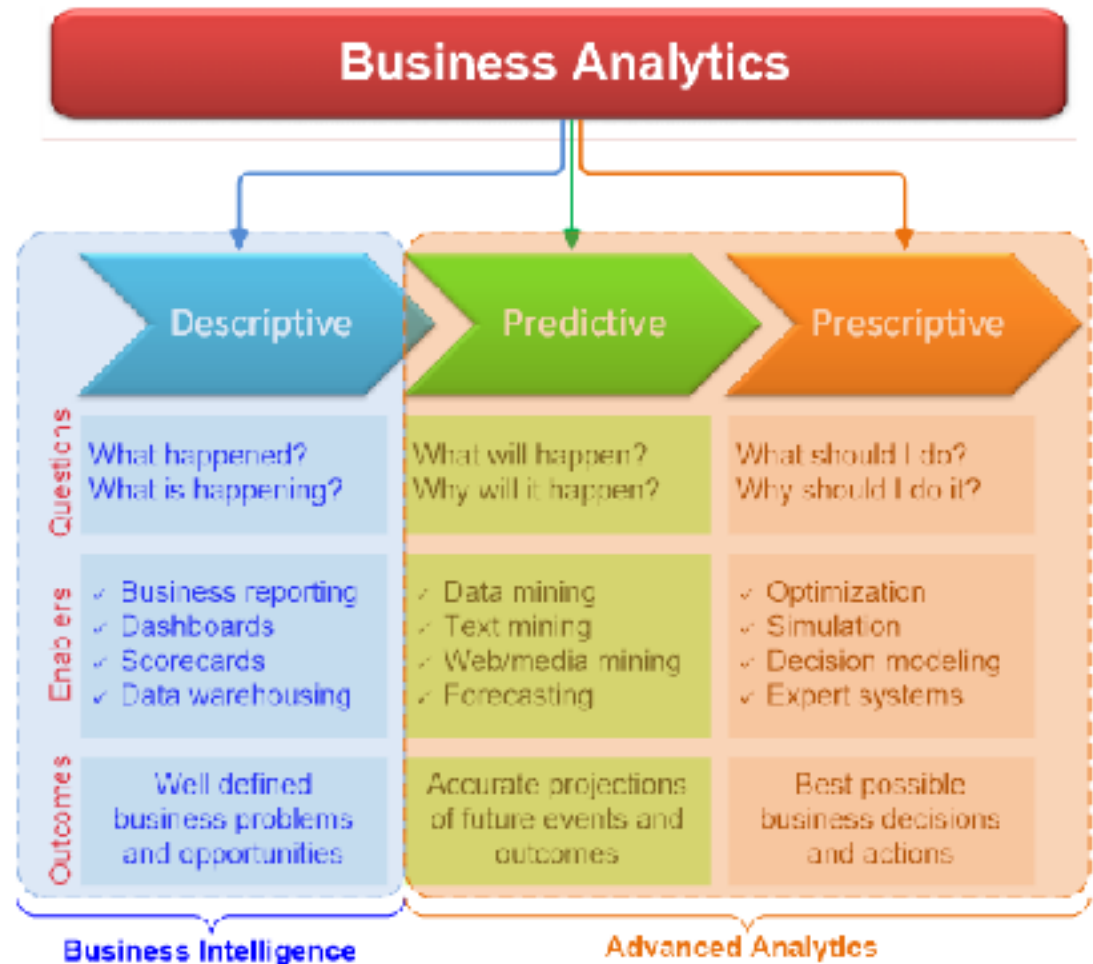
A Nova Plataforma do BI





Business Intelligence e Data Warehouse

- O BI costumava ser tudo relacionado ao uso de dados para suporte de decisão gerencial
- Agora, é uma parte do *Business Analytics*
 - BI = Análise descritiva





Gestão de Dados

- “Controlar e alavancar eficazmente o uso dos ativos de dados e sua missão e objetivos são atender e exceder às necessidades de informação de todos os envolvidos (stakeholders) da empresa em termos de disponibilidade, segurança e qualidade.”

(DMBOK, 2009)

Gestão de Dados



Especialização em Ciência de Dados com Big Data, BI e *Data Analytics*



NUNCA DESISTA DOS SEUS SONHOS



Prof. Fernando Siqueira
fernando.siqueira@uni7.edu.br