Data Platforms Architecture and Pattern based solution architecture

Ashish Gupta
06/01/2020

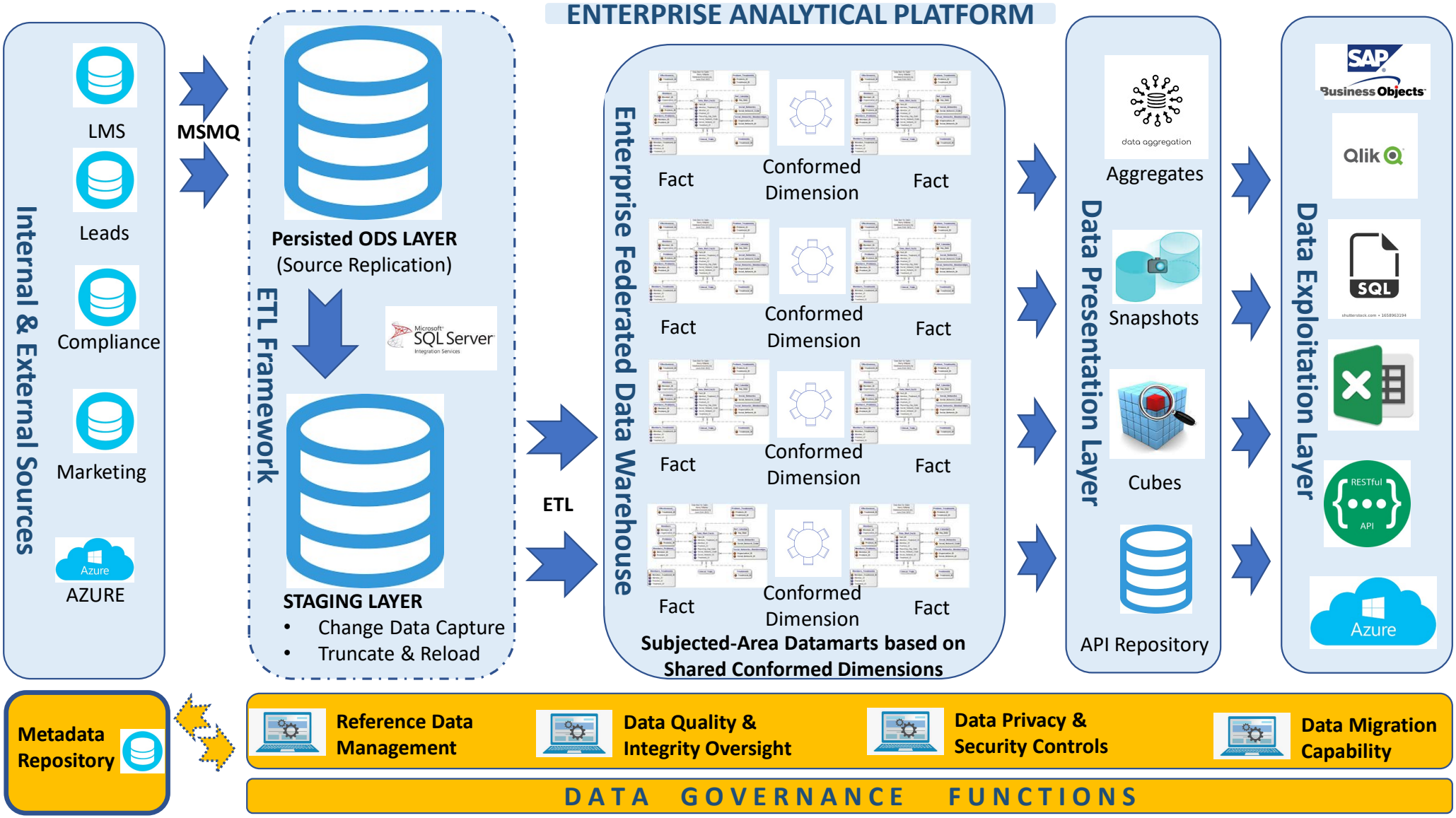
"Describe a large scale data initiative you played a key role in - what were the main architecture layers in that solution"

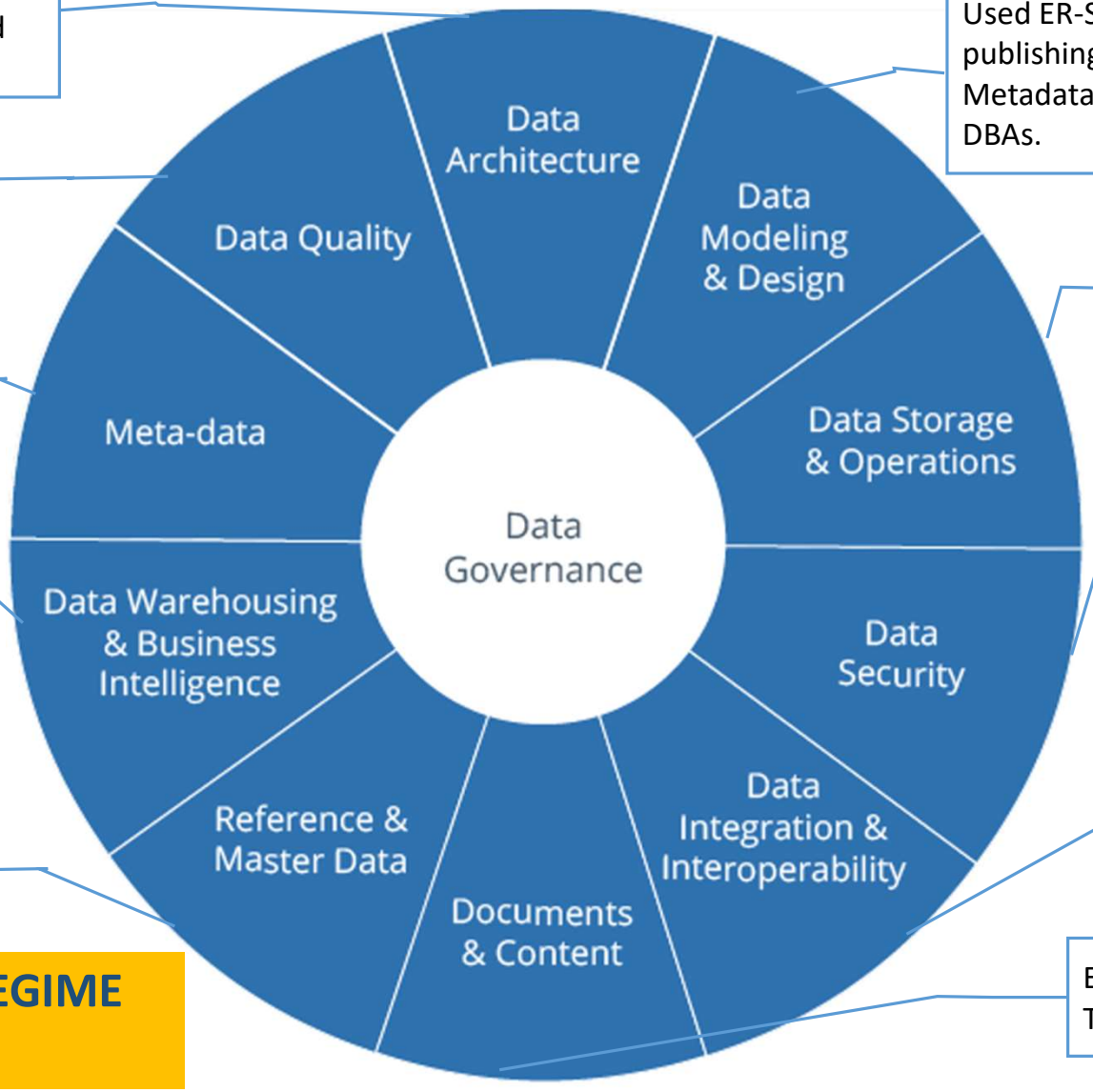
Universal Technical Institute Inc – Enterprise BI Architect

- I joined Universal Technical Institute as Enterprise BI Architect with mandate of establishing an analytical platform to support its analytical and strategical reporting needs.
- I accomplished this goal by wearing several hats. I realized, that for us to realize this vision, we had to establish an enterprise class data warehouse as its foundation. Next slide shows the envisioned and established analytical platform framework.

American Express – Lead Technical Architect / Senior Engineer

- I was brought in to lead the effort for migrating Merchant Risk Portfolio Applications from legacy (Teradata) Platform to Enterprise Big Data Platform.
- Big Data Analytics Platform.Architecture is presented in slide #4.


ENTERPRISE ANALYTICAL PLATFORM
Internal & External Sources
LMS
Leads
Compliance
Marketing
AZURE
MSMQ
ETL Framework
Persisted ODS LAYER
(Source Replication)
Microsoft SQL Server Integration Services
STAGING LAYER
• Change Data Capture
• Truncate & Reload
ETL
Enterprise Federated Data Warehouse
Fact
Conformed Dimension
Fact
Subjected-Area Datamarts based on Shared Conformed Dimensions
Data Presentation Layer
data aggregation
Aggregates
Snapshots
Cubes
API Repository
Data Exploitation Layer
SAP Business Objects
Qlik
SQL
RESTful API
Azure
Metadata Repository
Reference Data Management
Data Quality & Integrity Oversight
Data Privacy & Security Controls
Data Migration Capability
DATA GOVERNANCE FUNCTIONS



Established & Published Artifacts

GENIE - Built automated Metadata-driven Solution

Designed & Established shared repository

Built Analytical Platform

Built automated Metadata-driven Solution

Used ER-Studio for maintaining, publishing Logical & Physical Models, Metadata dictionary, DB Scripts for DBAs.

Dedicated Infrastructure Team

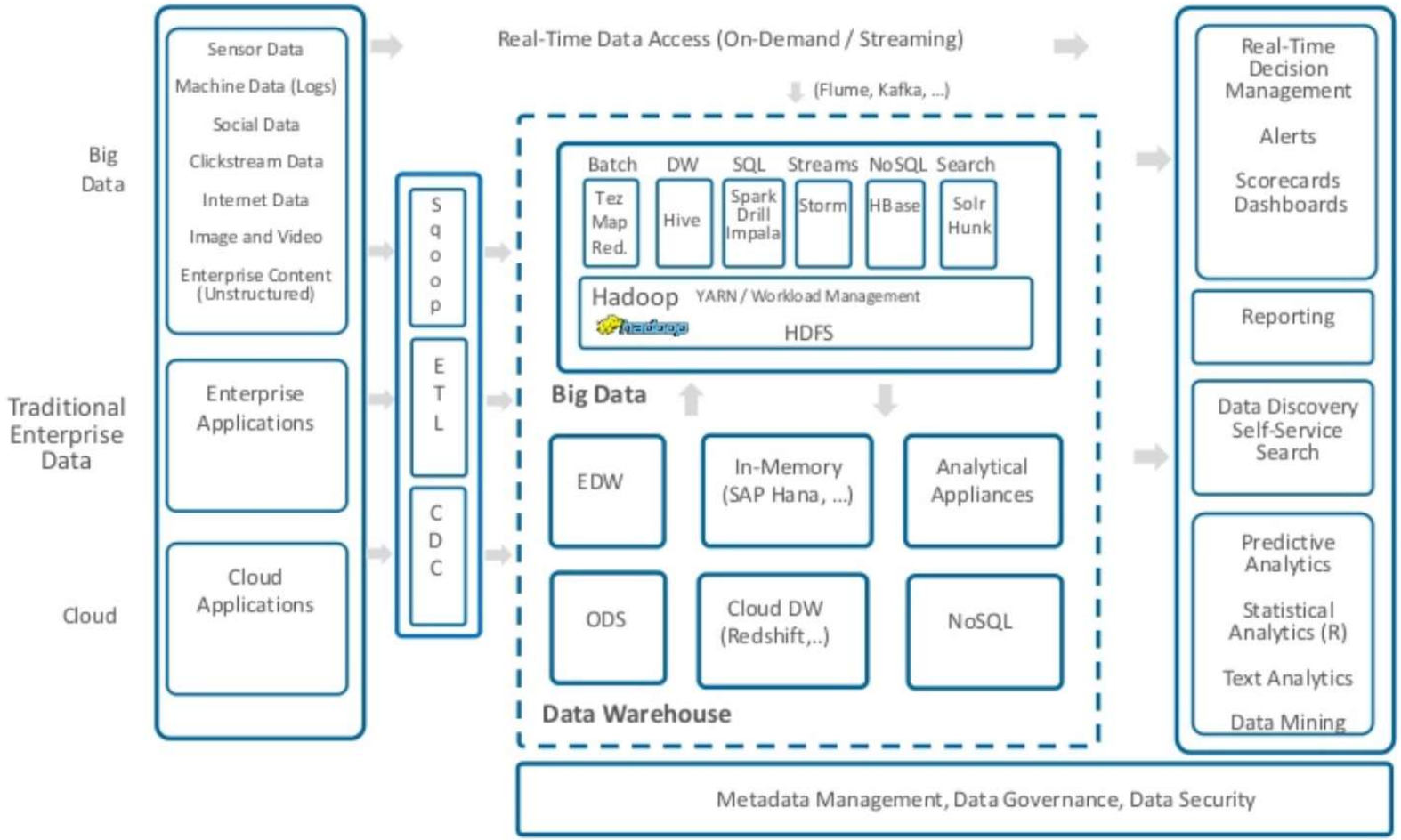
Built automated Metadata-driven Solution

Defined ETL Architecture Methodology

Established Documentation Templates and processes.

DATA GOVERNANCE REGIME ESTABLISHED IN PAST

Big Data Analytics Platform



"Describe an architecture pattern you've used, focus on technical elements of that pattern"

Finding Patterns

- Finding pattern is the key to solving a problem in most efficient, cost-effective, reusable, repeatable manner and with faster response times.

- Pattern leads to abstraction and abstraction leads to meta-data driven application design which can provide an automated solution. The phenomenon can be applied to solve challenges in data architecture and application architecture.

Patterns in Data Architecture/Engineering - Analytical Platform at UTI Inc

My entire analytical platform architecture comprised of several layers:

- **Persisted ODS** – Persisted data objects replicated from various data sources
- **Staging** – Transient Shadow Delta and Error tables for each dimension and fact in EDW Layer.
- **EDW** – Enterprise Federated Data-warehouse comprising of subject-area based data marts tied together through conformed dimensions.
- **Presentation** – For Horizonal and Vertical Aggregates, Point-in-time snapshots, Multi-dimensional Analytical Cubes, data structures to support Rest APIs.
- **Exploitation/Consumption** – Business Objects, Qlikview, SQL Queries, On-cloud applications (AZURE), Rest APIs, Downstream Applications.

Patterns in Solution Architecture

- Pattern leads to abstraction and abstraction leads to metadata driven solution architecture.
- I have used this approach to build several abstract, automated, scalable and metadata driven solutions for various data governance functions. Here are few examples:
  - **Data Quality and Integrity Management**
    - *A metadata driven framework for creating exception rules*
    - *An abstract on-demand rule engine to run all exception rules*
    - *A repository to store all exception results from rule engine runs.*
    - *A metadata driven framework to organize and distribute exception results.*
  - **Reference Data Management**
    - *A unified application to manage and maintain all reference data-sets along with oversight provision from data steward's perspective (Add/Modify/Delete operations with request and approval functionality)*
  - **Data Privacy and Data Security**
  - **Data Ingestion Frame-work**
  - **Referral generation and distribution framework**
  - **Data Migration Framework**

Thank You.