This field guide is brought to you by

Worldwide Community of Information Architects (WWCIA)

In affiliation with IBM Academy of Technology

Data Placement & Storage Solutions

Database Types

Cognitive Enterprise

Database Considerations

Data Characteristics

Technical UseCases

Data Storage and Placement in Hybrid and Multi-Cloud

Field Guide

This exclusive field guide is brought to you by **IBM Worldwide Community of Information Architects (WWCIA)** in affiliation with **IBM Academy of Technology.**

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Foreword

Data no longer resides locked in applications, protected by the logic that surrounds it. It is copied, combined, transformed and delivered to many platforms that are increasing controlled by third parties. How does an organization take control of its data to ensure it is in the right place, at the right time with the right quality whilst protecting its integrity, its business value through controlled access and ensuring the privacy of any data subjects?

These questions are the responsibility of a new type of architect that can design a landscape of application independent data sources fit for many purposes. This field guide captures the key resources and insights used by IBM's top information architects when delivering enterprise grade data services. Its format and style is designed to provide easy access to the information you need. I hope you find it useful and a joy to use

Mandy Chessell

Worldwide Community of Information Architects

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Abstract

This field guide is the outcome of the initiative "Field Guide for Data Storage and Placement in Hybrid and Multi-Cloud" kicked off in April, 2020 by the _Worldwide Community of Information Architects_, an affiliate of _IBM Academy of Technology_ (AoT). The initiative champion is _**Mandy Chessell**_, IBM Distinguished Engineer, CTO Data Platforms, Multi Cloud Leader, and initiative Leader is _**Sushma Singhal**_ and co-leaders are _**Debbie Botha**_ and _**Pat O'Sullivan**_.

There was long need for a field Guide describing which types of architecture patterns for data repositories (including SQL vs Non-SQL) can be used for relevant use cases in the various systems across hybrid and Multi-cloud.

This Field Guide is a result of 10 months long journey of research, interviews, discussions, creations of graphics to visualize the concepts and writing by all participants who volunteered and worked in their own time.

The Field Guide defines various types of data repositories for hybrid and Multi-Cloud environments. It provides a high-level point of view of which workloads are optimized based on the cloud type, data storage and placement and defines the use cases within these workloads and drill down into on which clouds it can run, and the various repositories that can support it with architectural decision points to consider.

Authors and Acknowledgement

Initiative Executive Sponsor: _**Mandy Chessell**_
Initiative Lead: _**Sushma Singhal**_
Initiative Co-Leads: _**Debbie Botha**_ and _**Pat O'Sullivan**_
Authors & Participants: Ahmed El-Shafie, Alok Jain, Aoun Lutfi, Arpan Dan, Debbie Botha, Kashyap Kamdar, Mehdi Charafeddine, Mohammed Abdul Qadeer Moini, Nauman Qureshi, Pat G O'Sullivan, Sonia Mezzetta, Sushma Singhal, Venkateswar Jeedigunta

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Introduction:
Data Repositories and Data Placement in a Hybrid and Multi Cloud environment

It is becoming increasingly challenging and important to understand the role of Data in the Cognitive Enterprise. In this field guide we define universal architectural patterns for how data flows through the organization and provide guidance for architectural decisions around the types of repositories to use and how it should be placed in the Hybrid and Multi Cloud Platforms of today and the future.

It is important to understand the various types of Systems interacting with each other and the architectural patterns depicting how data flows through the Cognitive Enterprise.

What's inside?

This field guide provides an overview of Data Storage and Placement options in Hybrid and Multi-Cloud that an Information Architects or a Data Engineers will come across frequently.

Each tab shown on the right side represents a technical usecases and Illustrate the possible data placement options and acts like a table of content for this guide. When run in slideshow mode these tabs are clickable.

This field guide is intended for the use of Information Architects and other users to help determine which types of architecture patterns for data repositories can be used for relevant use cases in the advanced technologies across hybrid and Multi-cloud. The field guide provides a high-level point of view of which workloads are optimized based on the cloud type, data storage and placement. It further defines the use cases within these workloads and drill down into on which clouds it can run, and the various repositories that can support it with architectural decision points to consider.
The content of this field guide focuses on the following 6 technologies topics (what are called "exponential technologies" in the formal Cognitive Enterprise Specification):

- Artificial Intelligence
- Blockchain
- Decision Support
- Internet of Things
- Intelligent Workflow
- Decision Support

Deciding on the right option is not always easy

The world of hybrid multi-cloud offers several data placement options to Information Architects and Data Engineers as they solution complex data storage needs. It becomes pivotal to understand various characteristics of the data itself before solving a usecase. There are several technical constraints to be understood and considerations to be evaluated for the right database choice.

Both hybrid cloud and Multi-cloud architectures provide operational flexibility to businesses, but in different ways. If cost is the focus of an organization's cloud provider selection, then keep in mind that a hybrid cloud environment will have additional cost requirements due to private cloud infrastructure and bandwidth facilities. Gartner has predicted that within the next two years, over 50 percent of global enterprises are likely to "adopt an all-in-cloud strategy."

Hybrid and Multi-cloud databases are essential, but not every database management platform offers the unique features that fully support a multi-cloud architecture.

This field guide takes into consideration the attributes and expectations from a Multi-Cloud Data Management Platform including the following:
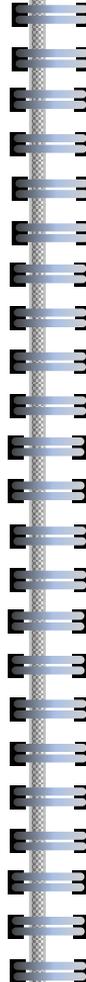
Data autonomy

The ability to own data without vendor lock-in anywhere is imperative in today's world of increasing privacy regulations, evolving public clouds, and risk management. The database management platform should enable full data autonomy for your organization.

Scalability

The database management platform should offer linear scalability and be able to add nodes and compute capacity with scale-out architecture and no downtime. This is one of the key powers of an Active Everywhere database.

Distribution

The platform should span multiple regional and global data centers, clouds, and service provides.

Real-Time Capabilities

A database management platform should provide anytime, anywhere access with instant responsiveness.

Always-on

There should be zero downtime across many years.

Contextuality

The platform should offer contextual relevance by serving intelligent information to the user at the right time, through the right channels.

These are all characteristics of an enterprise data layer that will support a multi-cloud strategy.

Cognitive Enterprise will leverage data platforms to achieve its goals

| Culture | Skills | Ways of Working | Experience |

| Industry Platforms | Transaction Platforms | Cross Industry Platforms |

| Decision Process | Front-Office Process | Back-Office Process |

| Artificial Intelligence | Blockchain | Automation | Internet of Things | 5G |

| Proprietary Data | Licenses Data | Public Data |

| Custom applications | Legacy applications | API-Enabled applications | Cloud Native applications | Digital applications |

| Public Cloud | Private Cloud | On-Premise | Security |

Reference: IBM Institute of Business Value – Building the Cognitive Enterprise

https://www.ibm.com/thought-leadership/institute-business-value/report/build-cognitive-enterprise

As organizations attempt to navigate the market, the next era of business reinvention emerges with platforms at its heart. The digital journeys of the past decade are now becoming cognitive journeys driven by these exponential technologies. As these capabilities are applied at scale, they increasingly result in change to the core of organizations. The business platforms are built on new and dynamic intelligent workflows that connect front- and back-office processes end to end. These workflows are transformed by exponential technologies—AI, blockchain and IoT.

In turn, the business platforms, intelligent workflows, exponential technologies, and data are supported by next-generation applications. These applications span new and legacy solutions made possible by an open, secure, and hybrid multicloud infrastructure. Each of these capability layers is wrapped in an agile, skills-based culture that fosters new ways of working and drives new employee experiences.

One critical aspect of evolving such multi-faceted multi-cloud Cognitive Enterprises is the role of the different data stores underpinning such an ecosystem. This document examines how the different data components flow through the various systems of the Cognitive enterprise and what are the data selection and storage characteristics and considerations.

Finally, this document will look at the specific data considerations in relation to a number of key technology areas of the Cognitive enterprise such as Artificial Intelligence, Blockchain, Intelligent workflows and others .

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)
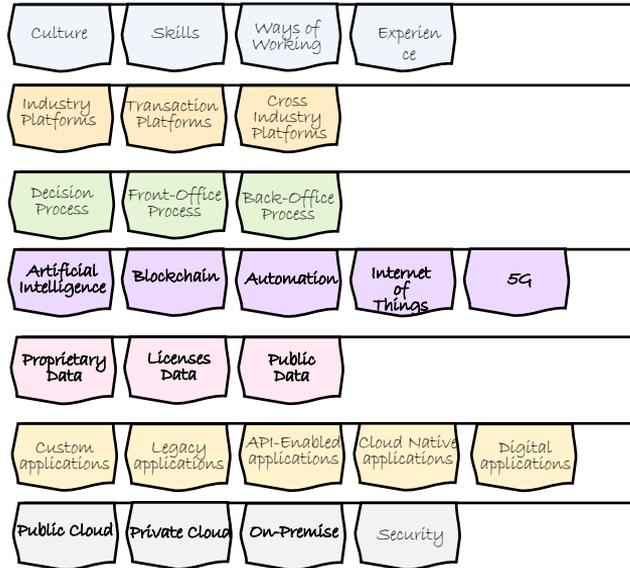
Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Cost of ownership considerations

Given the diverse range of data store requirements needed to fully underpin the data flows in the typical Cognitive enterprise, it is likely that an organization will require to deploy a number of different data store implementations.

A key consideration is to determine the overall cost of ownership for these different data stores and to identify what is the optimal combination.

Clearly there are the obvious differences in ownership characteristics of cloud and on-prem options . Generally, the cloud options allowing the minimization of capital expense and a focus on operating expenses and on-prem options allowing more control on the operating expenses but with a significant capital expenditure.

Irrespective of these cloud/on-prem considerations there are a range of other considerations :

Upfront implementation and ramp up costs – whenever adopting new data stores and associated infrastructure there is likely to be a significant start-up cost in terms of training both for IT and for end users) integration and potential external/consulting support. Such initial implementation costs are likely to be higher where the organization is adopting new technology or Data store solutions from a vendor with which they are not familiar.

Ongoing User and IT training - as the data stores get further adopted within the enterprise it is inevitable that additional training is required.
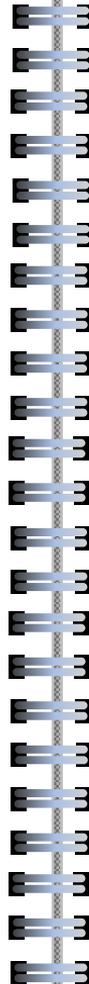
IT support skills – what is the overall cost to the enterprise to maintain and grow the deployment. It is not just the cost of such skills but also the availability of staff with the required skills in the market – this can be particularly an issue when dealing with relatively new technologies.

Vendor lock-in – what is the longer-term costs to the enterprise of dealing with vendors. Will the contractual terms and conditions lead to restrictions when the solution needs to be grown or adapted for new use cases.
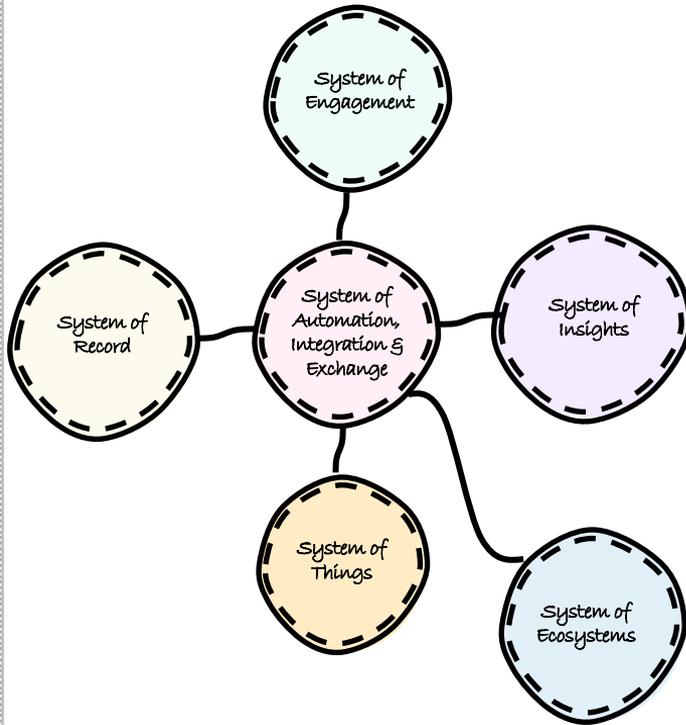
Alignment with the IT strategy – how do data store selections align with the overall IT strategy of the organization, for example in terms of support for specific standards or the stance on Open-Source technology. Another consideration in this area is how well the data stores can be integrated with the rest of the technology ecosystem of the enterprise.

Range of use of Data stores – can an organization make use of one data store to carry out several functions and therefore avoid additional purchases. For example, perhaps the need for Graph technology can be accommodated within the features of a general-purpose RDBMS , rather the incurring the additional cost of acquiring a specialized Graph DB.

Licensing – what are the ongoing software licensing implications for different data stores, especially if such licensing costs are impacted when the number of users of the Data store grows.

In summary, it is important not just to look at the different data stores in terms of the ability to deliver on the different aspects of the cognitive enterprise. It is also necessary to consider how they can be used and combined to minimize the overall cost profile.
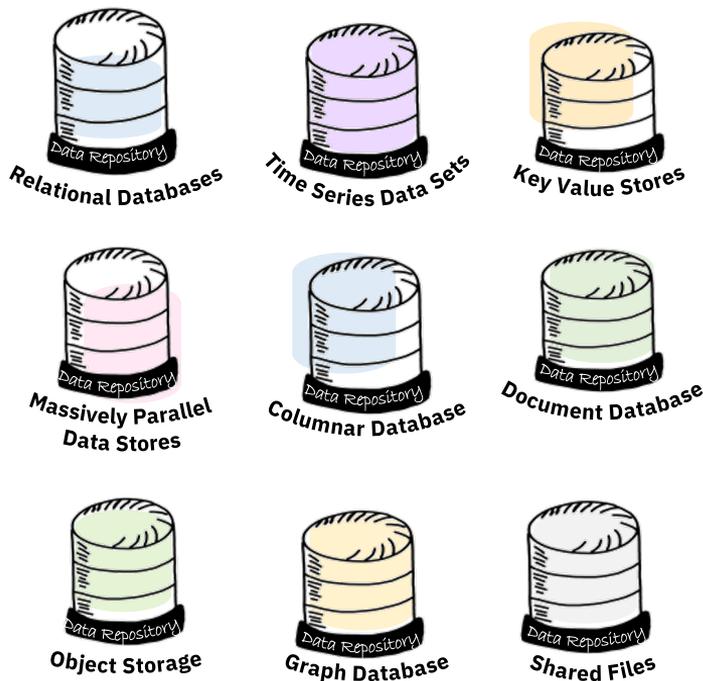
Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

The Cognitive Enterprise has a set of interconnected systems interfacing and sharing data with each other

How data flows in the Cognitive Enterprise

System of Engagement (SOE)	Where an organization Engages with the Customer. Omni Channel and conversational ability to communicate, collaborate, interact and transact.
System of Record (SOR)	Provides the back-end transactional systems. These systems are connected to the SoE and other systems where needed via the SoAIE.
System of Insights (SOI)	Where data from the various Systems are collected, organized and analyzed for insights and decisions. Where Data Science models and AI models are built and then infused in Processes in these systems.
System of Things (SOT)	Where things communicate with each other and the Enterprise. Examples are Operational Technology, Internet of Things, Telematics and Edge Computing, which provides new business models.
System of Automation, Integration & Exchange (SOAIE)	Used to Automate, Integrate and Orchestrate processes and data within and between the various systems in real time. This Abstraction layer with standards and shared services ensures agile implementation of fast changing SoE on top of slow changing SoR.
System of Ecosystem	Ecosystems to which Organizations belong. Examples are Digital Marketplaces, Data Marketplaces, Supply Chains, B2B Data Exchange Platform including Regulatory

Introduction
Artificial Intelligence
Decision Support
Internet of Things (IoT)
Blockchain
Intelligent Workflow
Data Governanc
Conclusion Summary
Systems Diagram

How Data Repository Types are used

Relational Databases

Time Series Data Sets

Key Value Stores

Massively Parallel Data Stores

Columnar Database

Document Database

Object Storage

Graph Database

Shared Files

Examples of popular Data Repository types

Very useful when strong consistency guarantees are important

Relational Databases organize data as a series of two-dimensional tables with rows and columns. Each table has its own columns, and every row in a table has the same set of columns. (SQL) for retrieving and managing data. Implements a transactionally consistent mechanism that conforms to the ACID (Atomic, Consistent, Isolated, Durable) model for updating information.

Good for storing telemetry data. Scenarios include IoT sensors or application/system counters.

Time Series data set is a set of values organized by time, and a time series database is a database that is optimized for this type of data. Time series databases must support a very high number of writes, as they typically collect large amounts of data in real time from many sources. Updates are rare, and deletes are often done as bulk operations. Although the records written to a time-series database are generally small, there are often many records, and total data size can grow rapidly.

Optimized for apps with simple lookups, but are less suitable for querying data across different key/value stores

Key value store is essentially a large hash table. You associate each data value with a unique key, and the key/value store uses this key to store the data by using an appropriate hashing function.

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Extremely quick queries, TBs of data in minutes and seconds. But it's the most expensive option
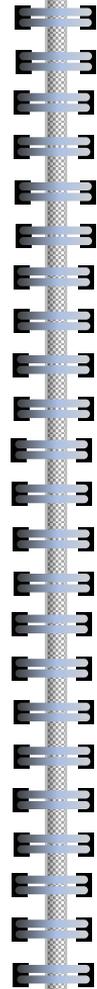
Massive Parallel Data Stores provide massively parallel solutions for ingesting, storing, and analyzing data. This data is distributed across multiple servers using a share-nothing architecture to maximize scalability and minimize dependencies. Extremely fast queries and aggregations. Its one of the most expensive as well.

Store data in key order, rather than by computing a hash. Read / write atomic with a single column-family,

Think of a column-family database as holding tabular data with rows and columns, but the columns are divided into groups known as column families. Each column family holds a set of columns that are logically related together and are typically retrieved or manipulated as a unit. Other data that is accessed separately can be stored in separate column families. Within a column family, new columns can be added dynamically, and rows can be sparse (that is, a row doesn't need to have a value for every column).

Conceptually like a key/value store, except that it stores a collection of named fields and data (known as documents) JSON

Document Database does not require that all documents have the same structure. This free-form approach provides a great deal of flexibility. Applications can store different data in documents as business requirements change.

Very Cheap, not suitable for all data cases, mainly files.

Object Storage is optimized for storing and retrieving large binary objects (images, files, video and audio streams, large application data objects and documents, virtual machine disk images). Objects in these store types are composed of the stored data, some metadata, and a unique ID for accessing the object. Object stores enables the management of extremely large amounts of unstructured data.

Stores two types of information, nodes and edges (relations)

Graph database does not require that all documents have the same structure. This free-form approach provides a great deal of flexibility. Applications can store different data in documents as business requirements change.

Sometimes your needs are simple shared file systems

Sometimes, using simple flat files can be the most effective means of storing and retrieving information. Using file shares enables files to be accessed across a network. Given appropriate security and concurrent access control mechanisms, sharing data in this way can enable distributed services to provide highly scalable data access for performing basic, low-level operations such as simple read and write requests.

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Repository Type / Hybrid Multi-Cloud Option	Object Storage	Document	Relational	Time Series	Graph	MPP
IBM Cloud	Cloud Object Storage	Cloudant DB	Compose for MySQL (Beta), DB2 Hosted, Informix	Informix		Db2 Warehouse on Cloud
AWS Cloud	Amazon Simple Storage Service	Amazon Dynamo DB	Amazon Aurora, Amazon RDS	Amazon Dynamo DB	Amazon Neptune	Amazon Redshift
Azure Cloud	Blob Storage	Document DB	MySQL, PostgreSQL, SQL DB, Edge, Stretch DB	Time Series Insights	Cosmos DB	Data Lake, Synapse Analytics
Google Cloud	Cloud Storage		Cloud Spanner, Cloud SQL	BigTable		BigQuery
Oracle	Oracle Object Storage		Oracle DBaaS			Oracle Exadata
Open source		Cassandra CouchDB MongoDB Redis				
Other Vendors		MarkLogic	Databases for PostgreSQL MySQL Service		Neo4j TigerGraph	

Examples of Data Repository types available in the Hybrid and Multicloud environments

The focus areas
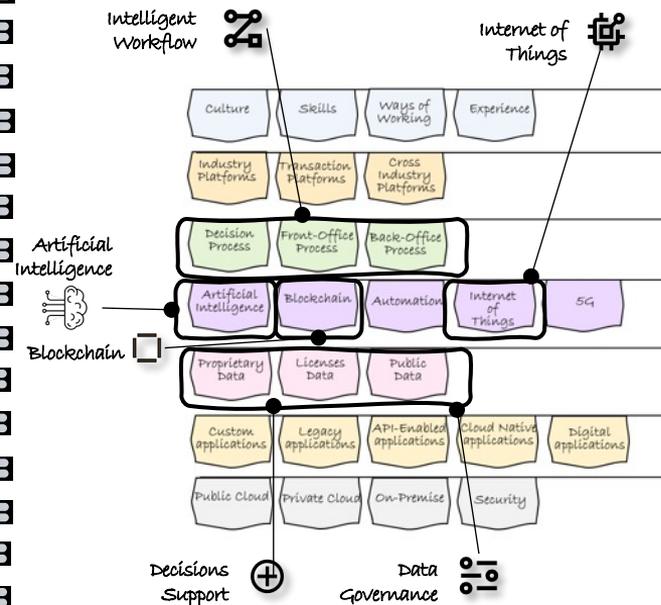
There are potentially dozens of different examples focus areas from across the cognitive enterprise.

Six specific areas were chosen so that it would be possible to provide a wide variety of examples of different areas of consideration when making decisions regarding Data Storage and placement decisions.
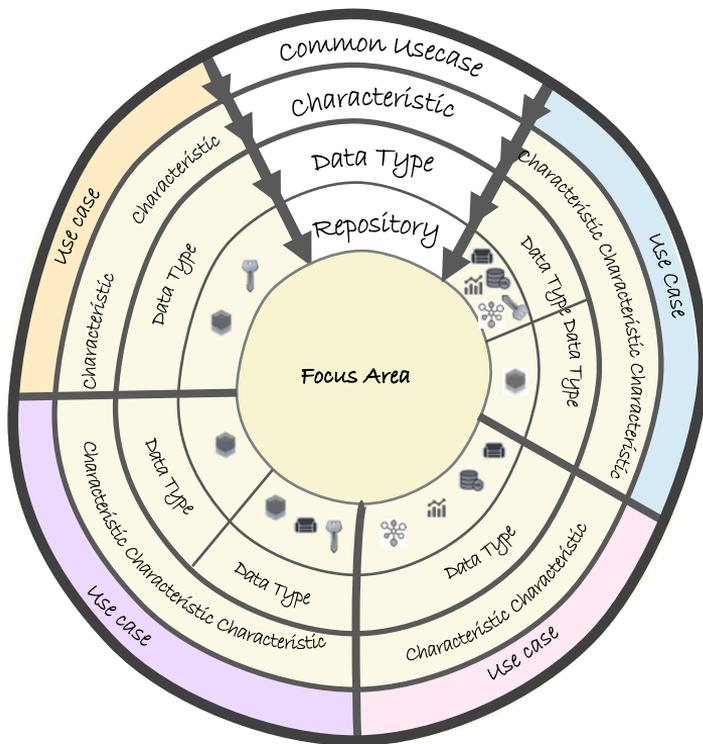
Data Placement Radar Diagram

How we use the Data Placement Radar Diagram

The Radar Diagram* is an innovative way to represent data placement options for a specific technology area or usecase showing multiple dimensions like common usecases, characteristics, data type and repository options all in one summarized view.

The various data repository and placement options are depicted on a Radar Diagram separately for each of the focus areas (Artificial Intelligence, Blockchain, Internet of Things, Decision Support, Intelligent workflow and Data Governance).

We depict the common use cases in the outer layer to group the various types of usage patterns for a focus area.

The Characteristics (second layer) that are needed for the various common use cases are typically found in the repository types (inner layer) and are directly related to the the various data types in the third layer.

The combination of the various layers help with the decision of the relevant data repository.

In the following pages provides the guidelines for deciding on data repositories and data placement for each of the focus areas.

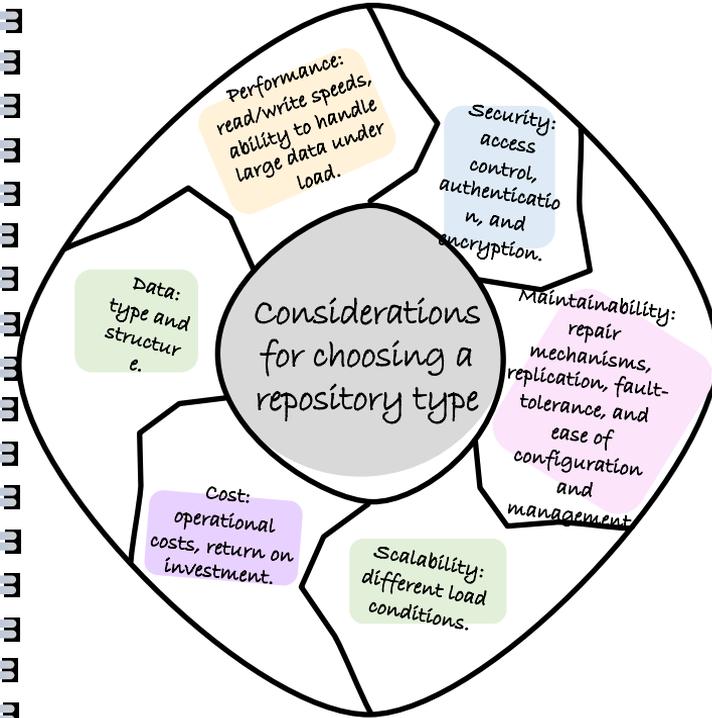**How Data flows in an AI use case**

Data is the "new oil" which enables the use of AI in a business context. Without data and a proper Information Architecture, it becomes very difficult to leverage AI to deliver business value. Throughout the lifecycle of an AI system or use case, data is processed and collected at different levels and serves different purposes; this is best represented in the AI Ladder.

Information Architecture

AI

Machine Learning

Analytics
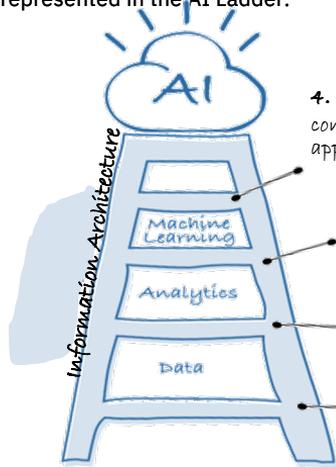
Data

4. Infuse: Where data is consumed in a business application.

3. Analyze: Where data is then procced in a system of insights

2. Organize: Where data processed and index to be then later used.

1. Collect: Entry point, where data comes from a system of record or system of things.

IBM's AI Ladder

Certain considerations are needed for selecting a particular repository type. This is because different use cases and stages in the AI process depend on different characteristics and metrics. Below are the main characteristics:

Considerations for choosing a repository type

Performance: read/write speeds, ability to handle large data under load.

Security: access control, authentication, and encryption.

Data: type and structure.

Maintainability: repair mechanisms, replication, fault-tolerance, and ease of configuration and management.

Cost: operational costs, return on investment.

Scalability: different load conditions.

The first step in Data Science, this stage involves integration and consolidation of different data sources into a storage/ data lake. Since this is long-term and sometimes large-scale, storage *maintainability*, *security*, *performance*, *scalability* are important to ensure the integrity of the data. Here often you get more than one type of database to cater for the different *business requirements* and compatibility with the *source type* of the data, whether that is structured data or unstructured flat files. During this stage, data is also sometimes virtualized to avoid shifting the data to a different location to be used in the next stage.

Data
Science
Use
Case

1. Data Consolidation and Storage (Collect and Organize)

Repository options: All

2. Business Analytics (Analyze)

Repository options: Document Database, Relational Database, Graph Database, Time Series Database

During this step, data is processed, and results are produced. To handle the assets generated in this step, data repositories that are *performant* are used, but in certain situations, *maintainability* is also important. In some cases it may be more suitable to run the analysis close to the original data repository instead of a virtualized source.

24

The models generated and their associated metadata generated in this stage require storage in a *secure*, *scalable*, and *performant* repository. This allows the AI system to be easily monitored and delivered to different applications on demand and with high efficiency. Generally, unstructured data also requires additional structured meta-data to train and deploy the models.
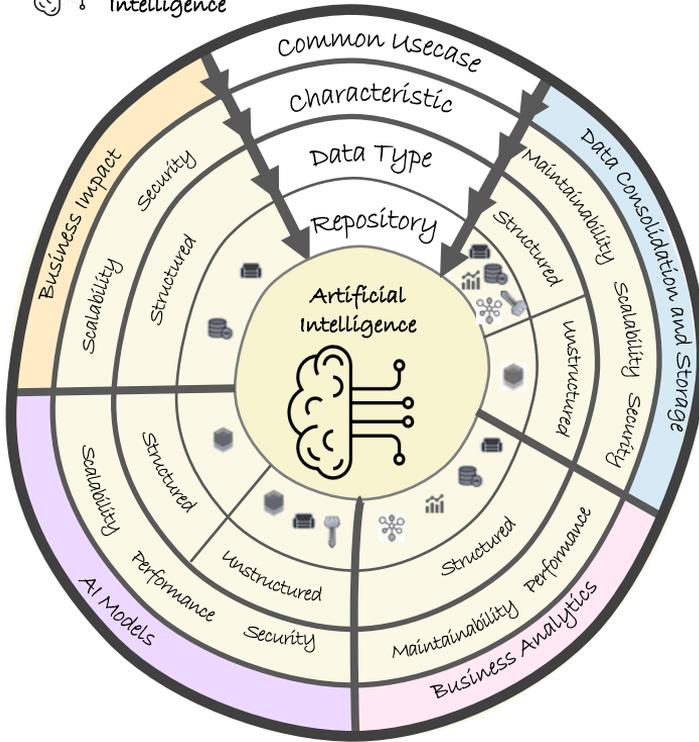
3. AI Models (Analyze)

Repository options: Object/File Storage, Key-value Database, Block Storage, Document

4. Business Impact (Infuse)

Repository options: Document Database, Relational Database

Once AI systems have been deployed and their business impact realized data such as logs, and utilization is generated which needs to be processed. Processing this data requires a *secure*, *scalable* and depending on the *type* of data

The data repository types on each stage indicate only the new data sources used in that phase. Where each phase takes the data input from the data stored in the processed in the previous phase.

25

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Data Placement Radar Diagram for AI

Common Usecase
Characteristic
Data Type
Repository

Artificial Intelligence

Business Impact
Security
Scalability
Structured

Data Consolidation and Storage
Maintainability
Scalability
Security
Structured
Unstructured

AI Models
Scalability
Structured
Performance
Unstructured
Security

Business Analytics
Structured
Performance
Maintainability
Security

Data Repository Legend

Object/Block
Document
Relational
Time
Graph
Key-value

References:

https://www.ibm.com/cloud/architecture/architectures/dataAIArchitecture
https://www.ibm.com/cloud/architecture/architectures/aiAnalyticsArchitecture
https://www.ibm.com/cloud/architecture/architectures/dataArchitecture
https://www.ibm.com/cloud/architecture/architectures/aiAnalyticsArchitecture/reference-architecture
https://www.ibm.com/analytics/journey-to-ai

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

26

27

Decision Support

In an end-to-end solution architecture, often the one size fits all concept is misconstrued as a means of standardization. So is the case of data storage. In the modern age of concepts like lambda and kappa architecture, Cap theorem validations, shared everything vs nothing design, structured vs unstructured etc, storage solutions for data and the different types of data and their persistence has become more and more a topic of discussion and arbitration.

Different kinds of data are best dealt with different data stores. In other words, the applicable use case would need to be aligned with the right solution. The focus of this section is to pick the right backend data structure for the right use case to derive decision support outcomes.

Different domains of data come together in the Decision Support platform and become the first point of business interrogation of the transactional data outside of the operational data stores.

As we look at the broad functional uses, it is apparent that Decision Support provides true value when data domains are stitched together, and aggregations are performed to derive meaningful insights from seemingly disparate and large volumes of data from multiple sources.

For example. do weather changes have impacts on sales, do external seasonal economic factors contribute to rising/declining operational costs, are social media campaigns contributing directly or indirectly to increased clickstream data and so on.

Decision Support use cases fall under the following categories:
- Information Exploration (Data Provisioning)
- Real Time Operational Analytics (Predictive Analytics)
- Traditional Business Intelligence (Prescriptive Analytics / Business Intelligence / Descriptive Analytics / Operational Reporting)

Introduction

Artificial Intelligence

Decision Support

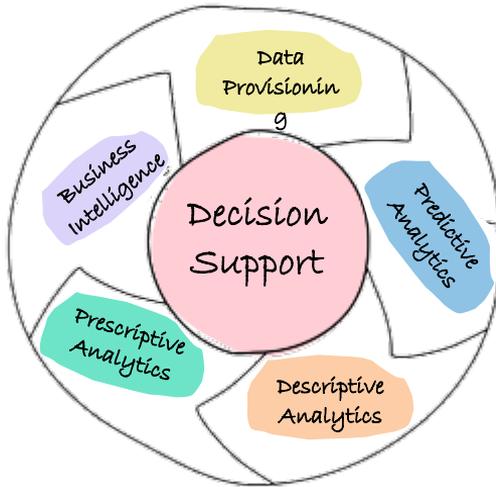Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Data Provisioning: Business users and Data scientists usually try to explore data that is not yet analysed nor governed and usually on a raw system format.

Predictive Analytics: *Understanding the future.* Predictive analytics are about understanding the future. It provides companies with actionable insights based on data. Predictive analytics provides estimates about the likelihood of a future outcome. Companies use these statistics to forecast what might happen in the future.

Predictive Analytics is suited to Big Data repositories and data store like a data lake. Usually, the data is landed as it is from source and Schema on Read is applied.

Descriptive Analytics : *Insight into the past.* Descriptive analytics describes or summarizes raw data and makes it something that is interpretable by humans. They are analytics that describe the past. Descriptive analytics are useful because they allow us to learn from past behaviors and understand how they might influence future outcomes.

Descriptive Analytics is performed primarily on Relational and Columnar type of databases since it needs structured data at the back-end to support this kind of reporting.

Prescriptive Analytics: Prescriptive analytics allow users to prescribe the different possible actions and guide them towards a solution. Usually KPI derivations, calculations, aggregations are performed to define actionable insights from the data in order to advise on possible outcomes before the decisions are made. Prescriptive analytics are often used in what if scenarios and models. Usually, Data Marts are built for the purpose.

When selecting a repository type for Decision Support use cases, the major considerations are as follows:

- Consistency of the data, availability and durability of the data repository
- Latency i.e. the speed at which result is returned to the query
- Replication in multiple locations
- Scalability across different dimensions and under different load conditions.
- Cost of the data repository
- Security considerations such as access control, authentication, and encryption
- Operational analytics and search
- Compatibility with various types of data
- Performance for data retrieval

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram
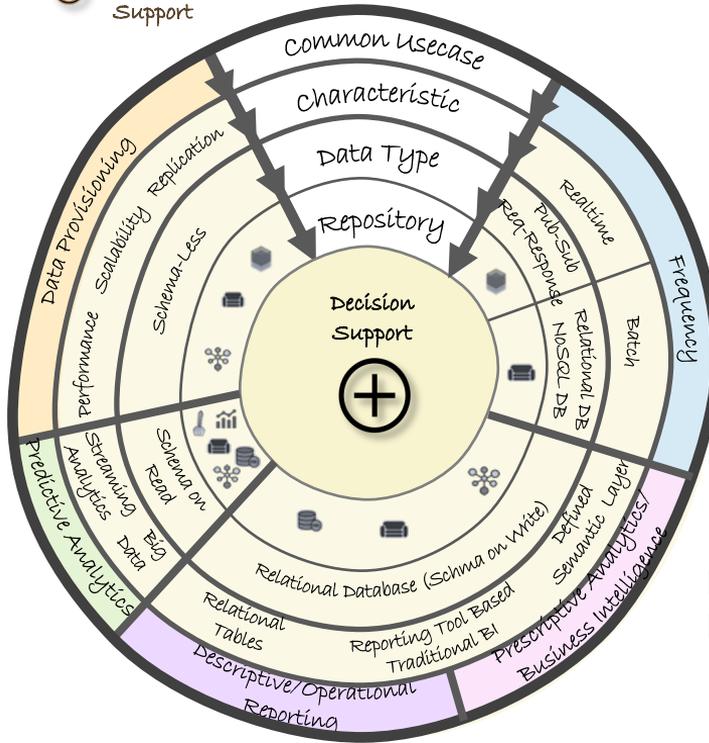
Data Placement Radar Diagram for Decision Support

References:

The CRISP-DM methodology

https://the-modeling-agency.com/crisp-dm.pdf

Internet of Things
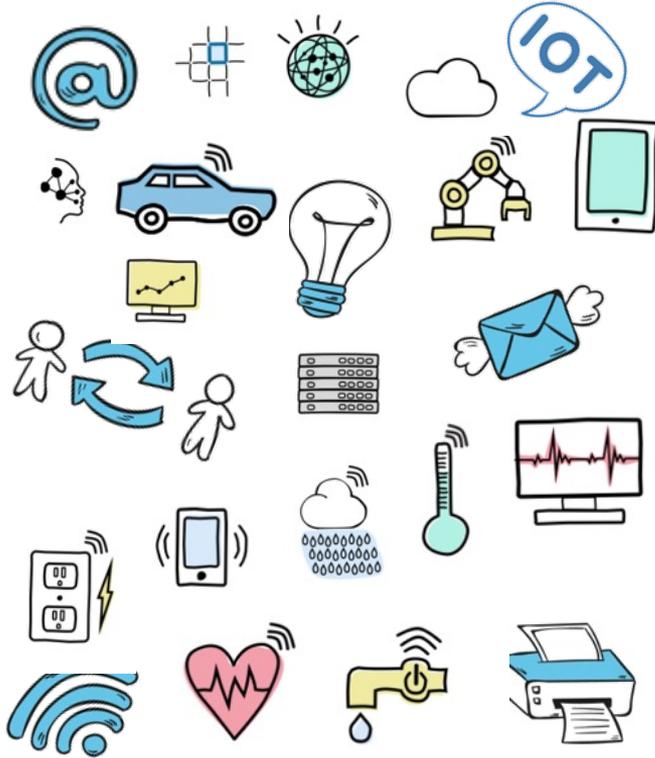
How Data flows in IoT use case

Data driven insights is key to Modern Data Architecture (MDA) which is heavily enabled by IoT data lifecycle from device provisioning, data transmission to device management.

Raw device data
Event and Live device data from IoT devices, which are mostly time series in nature. The recent historical data for your IoT devices. Used for quick view trends in Platform Service boards and cards, threshold-based decision making, alerts and triggering of other services (using serverless most of the times). Used as source of truth for device data. Repository options : **Document DB, No SQL DB, Time series DB**

Event Streams
Ingested Raw data managed by Message brokers for consuming by applications and cloud services. Typical retention is 1- 7 days, with a limit on the message size and # of messages. Repository options : **HDD disks, Page-Cache, In-memory DB**

Data Store (Transactional)
Device emitted data, validated and enriched by the processing applications for transactional processing as per the business function. This data along with other associated application data components are used for business logic and other functional server components. Repository options : **SQL DB, No SQL DB**

Data Warehouse (Business reporting and analytics)
This is mostly fed offline or on scheduler with a ELT or MQT kind of procedures or Cron jobs, to form the data lake or Data store for different business unit reporting, analytics and BI functions. Repository options : **Columnar DB**

Data Lake (Advanced Analytics)

The long term historical data for analytics access. Use analytics tools to extract knowledge from your data. Mostly contains Raw and un-schema data, for analytics by Data scientists. Repository options : **SQL DB, Columnar DB**

Data Historian (Archival)

The complete set of historical data for archival purposes. Mostly used for offline purpose like Long term trend analysis, audit , back up and restore. Repository options : **Object store**

Device Inventory

IoT Device and GW inventory managed individually or along with the SKU (final product model to which it is attached). This is typically managed by ERP systems, from which the Production and QA team would use the data to tag/associate the devices that would be sent to Field. This data holds the life cycle of the Devices, including the repair/return/decommission. Repository options : **SQL DB, Document DB**

Device Management

Holds the Device topology and entity model, Identity register and holds the Device configuration and policy including device type, connectivity, capability and characteristics of the telemetry e.g. frequency of message exchange, size of message. Repository options : **Document DB, Graph DB, SQL DB**

FOTA (Device Firmware)

Firmware of the devices are stored separately and managed by a life cycle process driven by DevOps and QA for various functional releases / versions or fixes. Repository options : **File Server, Blob storage**

Event Rule Management

Rules to be applied on live data as well as on aggregated data, as per the business functionality are managed separately both in Application as well as on the Edge gateway. These can be statically configured on the device/gateway during provisioning or updated later using Device Management. Repository options : **Document DB**

1. Data Type

2. Performance

3. Sovereignty

4. Scalability

5. Security
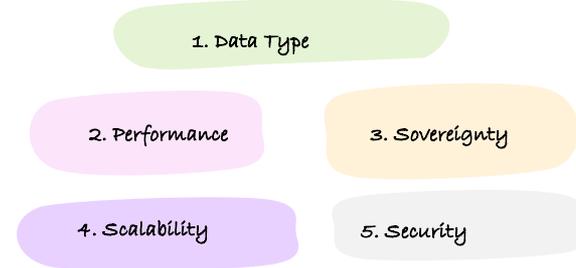
Considerations for landing zone for repository type

Choosing landing zone of IoT solution components in a cloud service type—hybrid, public, private/dedicated, or on-premises (local)—is an important decision. Clearly defined requirements related to Data type, Performance, Sovereignty, Scalability, Availability, and Security are imperative to the decision-making process. The sheer amount of data associated with live data streams from manufacturing sensors or consumer devices means that all aspects of messaging, connectivity, and data management are of the utmost importance.

1. Data Type

Data type plays a key role in deciding the target service and deployment model of IoT data repository. Data types could be SQL oriented, Document oriented, Columnar based , Key Value based, and Big data based for enabling large scale reporting and analytics.

2. Performance

Performance of the data repository, particularly in the IoT event streams is key to throttle the data in real time towards multiple designations like Storage, real time streaming analytics as well as threshold hold decision making. Performance measures in terms of throughput transactions/sec,, CPU cycles need for IO operation, latency in writing and reading the data from a storage are key aspects to choose the final location of the repository in a hybrid multi cloud scenario

3. Sovereignty

Any IoT cloud system must account for data sovereignty rules and store and process data only in those locations permitted by the regulations. This requires the provider cloud to provide the cloud service customer with control over storage and processing locations. This is a key aspect to decide the final location of the data repository type. In a distributed IoT data life cycle across hybrid multi cloud solution, it is important to consider this constraint at arriving at the target landing zone for each of the repo type
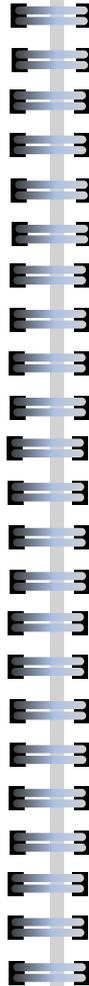
4. Scalability

One of the key architectural decision in an IoT solution, is that the number of concurrent devices and users connecting to the IoT platform must be scalable. That means, the components in the data flow should be elastic through out, to anticipate a newer generation of connected devices that will furnish higher resolution of data streams, as well as sudden outbursts of data. Same time, the system should be able to scale down in non-peak times/days, where in the charges on the repository do not overburden the cost of the solution
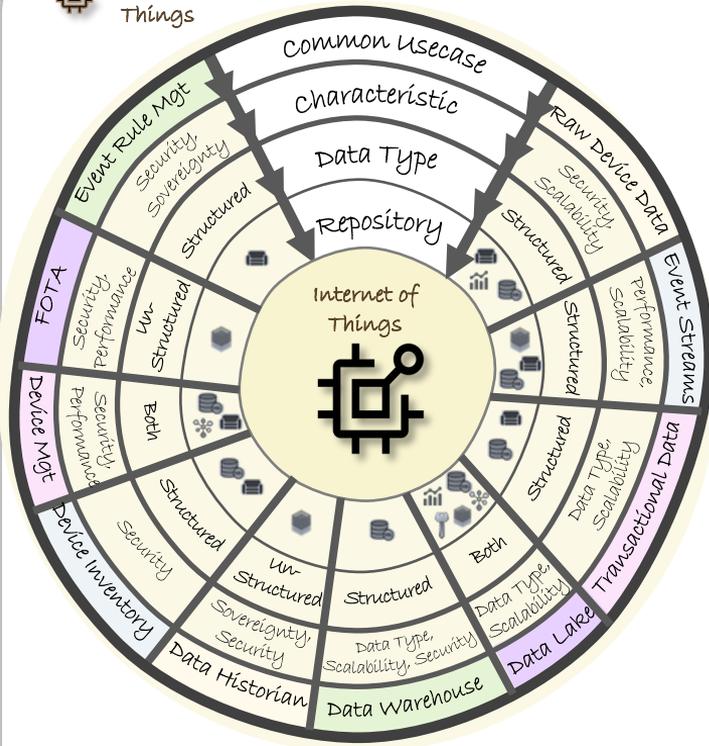
5. Security

IoT data represents not only device data, but it can hint on location, PII, SPI and some of the data related to the usage pattern of the consumer. Applications processing these data e.g., Financial transactions like payments for certain IoT device operations need to ensure appropriate international and regional standards, e.g., GDPR , PCI compliance. The data privacy and identity management of devices and individuals is very vulnerable for misuse, and so these are highest level protected in the data repository e.g., Device ID registry, Device Management repo with strict authentication and authorization of the user/ applications. The target cloud selection enables faster deployment of new compliance and monitoring tools that encourage agile policy and compliance frameworks right from device firmware version , device control/commands authentication token refresh, device data access for analytics and so on.

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Data Placement Radar Diagram for IoT

References:

https://www.ibm.com/cloud/architecture/architectures/
https://www.omg.org/cloud/deliverables/CSCC-Cloud-Customer-Architecture-for-IoT.pdf

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Blockchain

How Data flows in a Blockchain use case

A SmartContract operates as an intermediary between a Client and a Blockchain Ledger where the WorldState maintains the final state of the Blockchain data. Reference: https://hyperledger-fabric.readthedocs.io/en/release-2.2/txflow.html.

You can leverage IBM Cloud Pak for Integration to manage the entire processing at the integration layer through its market-leading API management, streaming, and other capabilities in an efficient, secure, and governed manner.

You can leverage Red Hat OpenShift for building Microservices based Offchain Applications. The selection of a NoSQL or traditional RDBMS for Microservices is proprietary to the business problem and enterprise IT strategy.

You can leverage IBM Cloud Pak for Automation to digitize and automate processes, ensuring the delivery of an integrated ecosystem in compliance with business objectivities and regulatory requirements.

.

Building Blocks of Blockchain

Web Apps	Business Services	Direct vs Indirect	On-chain vs Off-chain
Batch Processes	Role vs Cert Mapping	SC & Endorsement	Document Storage
Dashboard & Reports	Client SDK	Channels	Private Data Collection
Artificial Intelligence	APIs	Network	Data
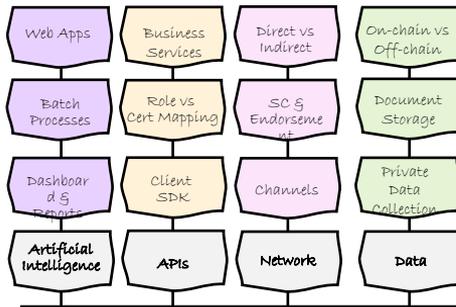
Blockchain Solution

Blockchain Offerings
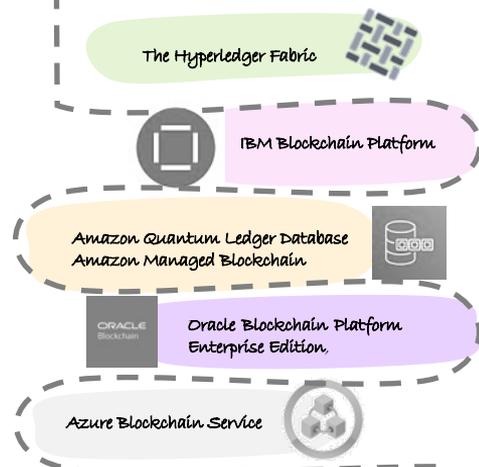
The Hyperledger Fabric

IBM Blockchain Platform

Amazon Quantum Ledger Database
Amazon Managed Blockchain

ORACLE Blockchain — Oracle Blockchain Platform Enterprise Edition,

Azure Blockchain Service

Introduction
Artificial Intelligence
Decision Support
Internet of Things (IoT)
Blockchain
Intelligent Workflow
Data Governanc
Conclusion Summary
Systems Diagram

The Hyperledger Fabric is the modular blockchain framework that has become the de-facto standard for enterprise blockchain platforms. It offers a unique approach to consensus that enables performance at scale while preserving the data privacy enterprises demand. Over 66 organizations building Hyperledger Fabric with 400+ developers to working together.

IBM Blockchain Platform is the enterprise-grade containerized, next generation of blockchain for business. Proven, scalable, flexible and built to run anywhere from hybrid cloud, multicloud to on-prem environments.

Amazon Quantum Ledger Database is a ledger database that offers a ccentralized, immutable, and cryptographically verifiable transaction log. **Amazon Managed Blockchain** is a fully managed services to create and manage scalable blockchain networks and distributed ledger
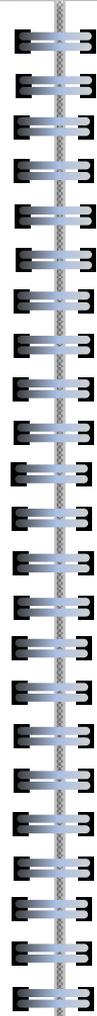
Oracle Blockchain Platform Enterprise Edition, based on The Hyperledger Fabric, enables this flexibility by providing an option to deploy blockchain nodes in customer's data centers or on 3rd party clouds, while still connecting to Oracle Blockchain Platform nodes provided as managed PaaS offering in Oracle Cloud.

Azure Blockchain Service is a fully managed ledger service that enables users the ability to grow and operate blockchain networks at scale in Azure. Hyperledger Fabric consortium available on Azure Kubernetes Service (AKS). Azure also provides the Ethereum POA Consortium solution templates.

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Considerations for choosing a repository type

Blockchain vs. Traditional Application:

Unlike traditional applications, a Blockchain Network would allow the users of a Consortium to have access to the same data available across the Network. Indeed, access to the data at Peer or Node levels can still be prohibited to address specific business requirements.

CouchDB vs LevelDB

The default State Database used for storing the State as Key-Value pairs in Hyperledger is LevelDB. However, I would suggest CouchDB because LevelDB does not support the use of indexes as well as rich queries. CouchDB stores data as JSON objects whereas LevelDB stores data as key-value pairs.

NoSQL vs RDBMS

The decision would typically boil down to Client's preference based on few influencing factors such as schema-less, performance, auto-scaling, replication, and other comparative benefits.

Estimating Storage:

Depending on the transacting hours, for 1 TPS, assuming each transaction of 1 KB, over the course of a year, annual storage required can be between 6.9 GB to 22.6 GB. (Note: Bitcoin, Ethereum, and other leading Crypto operate in a scenario that increases this proposition)

Offchain vs Inchain Data:

The data processed in a Blockchain Network via the Smart Code is referred to as the Inchain Data. Any non-transactional data as Images, PDF, Files, and other documents not required to be stored in a Blockchain by all the Nodes can be referred to as the Offchain Data. Another characteristic of an Offchain Data is that it can be altered, deleted, or changed. For instance, a Patient's medical test report can change over time compared to the Patient ID.

Cloudera vs Hadoop vs EDW vs Any Other Analytical Repository:

Entirely depends on clients' ecosystem and the business problem. The data can be brought-in to Client's existing Enterprise Data Platform for reporting and analytics purposes. The Enterprise Data Platform can have Cloudera, HDFS, Azure, Google, AWS, Enterprise Datawarehouse, Data Lake, or any other type of analytical repository.

46

Introduction

Artificial Intelligenc

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram
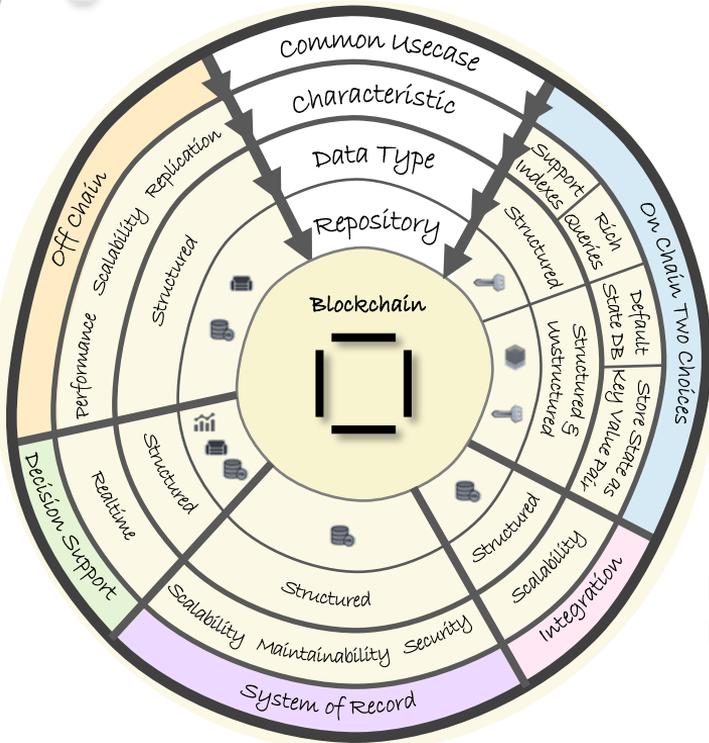
Data Placement Radar Diagram for Blockchain

48

References:

Blockchain architecture for trusted transactions
https://www.ibm.com/cloud/architecture/architectures/blockchainArchitecture

Blockchain reference architecture
https://www.ibm.com/cloud/architecture/architectures/blockchainArchitecture/reference-architecture/

Research leading Blockchain use cases
https://www.ibm.com/blockchain/use-cases/

49

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

To support the speed and flexibility required by the Cognitive Enterprise, it is necessary to reinvent business processes in order to replace, where possible, transactional and decision support activities with automation and cognitive solutions.

Intelligent workflows use digital insights to configure processes and orchestrate emerging technologies to change how work gets done to achieve priority outcomes. Business process platforms provide the backbone for digitally orchestrating these workflows using data, automation and cognitive applications. Data led, technology enabled workflow will be open and span ecosystems and networks.

There is a wide range of potential of business applications that could benefit from such an Intelligent workflow approach. Some primary examples include :
• Cognitive Care in Banking
• Claims Management in Insurance
• Outage Prediction and Prevention in Energy and Utilities
• 5G network optimization in Telco

Intelligent Workflows and Data
Essentially the Intelligent Workflows are the integration point that connects the wealth of data within an organization with the various business activities being undertaken by the enterprise.

50
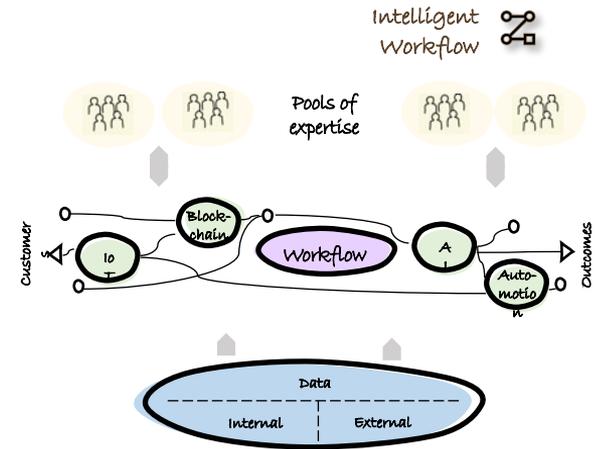
Business Platform for Intelligent Workflow

So the smooth operation of such Intelligent Workflows requires and underpinning of a well organized and governed data storage ecosystem that manages the use of both internal and external data , and often requires to manage and process streaming data in real time. There is also the need to ensure that any associated AI models required by the Intelligent Workflows are stored and managed accordingly.

In terms of understanding how the the operations of these Intelligent Workflows both influence and are influenced by the Data Storage ecosystem, it is best to look at a typical example as the range of different Intelligent Workflow scenarios means that it is difficult to define a singe definitive pattern.

51

Example use case 5G Network Optimization for Telco
A good representative example of an Intelligent Workflow and how it leverages and contributes to different aspects of data stores is 5G Network Optimization.

The business objective of this example is primarily focussed on the ongoing optimization of the 5G network, ensuring that the traffic across the network is managed in an intelligent way , that any likely issues impacting the supply across the network are anticipated and any necessary mitigations taken automatically in a timely manner via an Orchestration component.

An additional feature of this workflow is it reacts to clients behaviour so that appropriate offers regarding replacement of home Wi-Fi with 5G can be sent to clients and accepted offers fulfilled automatically

How the Data Travels

Storage or Raw events – the storage of the relevant initial originating events that are to be processed as part of the 5G Network optimization. This will be a record of a history of all of the interaction associated with the relevant events. This data will also be subject to Service Assurance and Orchestration processes to ensure the appropriate distribution of capacity/traffic across network.

In general the events are composed into different workflows that interact with the appropriate domains (e.g., Product, Billing, CRM) depending how the Domain driven design is set up for the particular use case.

Local/Regional Aggregation of event Data – This is the storage for reporting purposes the various aspects of the events at a Local, Market and Regional level. This information is the gradual aggregation of the event data across different levels from local to national markets.

Need to enable querying of different aspects and aggregations of the events.

Event Metadata– The storage of the metadata and schemas associated with the events.

Interaction with client ecosystem – This is interaction of the event data with other applications such as CRM, MDM, Billing , etc. To enable any client-oriented actions to be undertaken (e.g. offers)

Introduction

Artificial
Intelligence

Decision
Support

Internet of
Things (IoT)

Blockchain

Intelligent
Workflow

Data
Governanc

Conclusion
Summary

Systems
Diagram

Data Storage Considerations

Storage or Raw events
- Storing of Time Series data
- Potentially long term storage of events
- Storage of History of all interactions
- Potentially need to support local "at the edge" analytics
- DB Type :Event Store. Specifically Kafka, however depending on the usage, events could be also stored in MongoDB or Redis.

Aggregation Local/Regional Data
- Storage of different aspect of Events
- Needed for analysis at various levels of aggregation
- Likely fed by Kafka message bus
- DB Type : Use Kafka for for storage of simple aggregations , for more complex aggregations, consider other tools such as MongoDB, or Graph Dbs

Metadata Storage
- Storage of the metadata and schemas associated with the events.
- For example, the Kafka Schema Registry
- DB Type : typically Relational

Customer related data
- Storage of the aggregated events along with relevant Customer Billing data
- Needs to integrate with aggregated client-specific event information
- DB type : mixture of RDMS/NoSQL and event store
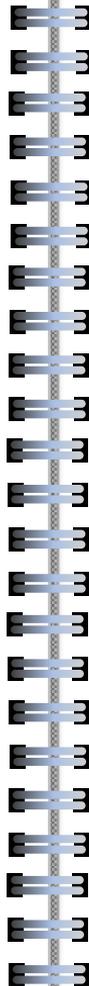
Hybrid/Multi-cloud considerations
- Event stores can be cloud agnostic
- Potential for Client-oriented systems to be stored on Cloud Pak for Data

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram
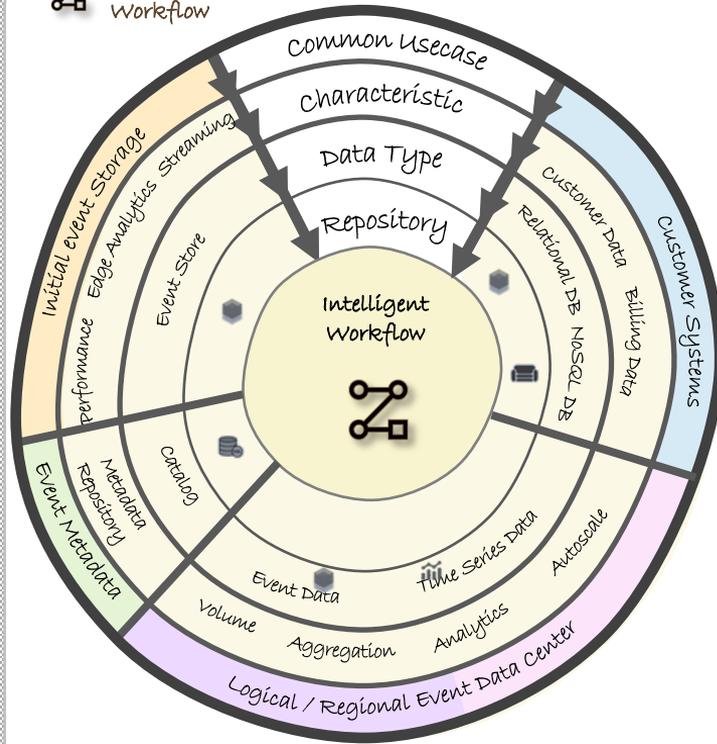
Data Placement Radar Diagram for Intelligent Workflow

Legend items:
- Data Repository Legend
- Object/Block
- Document
- Relational
- Time
- Graph
- Key-value

56

References

- Intelligent Workflows for the Cognitive Enterprise
https://w3.ibm.com/services/lighthouse/spotlight/documents/150650
- Global Smart Buildings Market Report
https://www.globenewswire.com/news-release/2019/05/20/1829041/0/en/Global-Smart-Buildings-5G-AI-Data-Analytics-and-Edge-Computing-Market-Report-2019-2024.html
- Intelligent Workflows "What do we want to be famous for"
https://w3.ibm.com/services/lighthouse/spotlight/documents/153087
- Intelligent workflows 101: Revolutionizing the way your business works
https://www.ibm.com/blogs/services/2019/04/29/intelligent-workflows-101-revolutionizing-the-way-your-business-works/
- Designing intelligent workflows inside IBM Garage
https://uxdesign.cc/intelligent-workflow-design-at-ibm-d1baaeb7cc0b

57
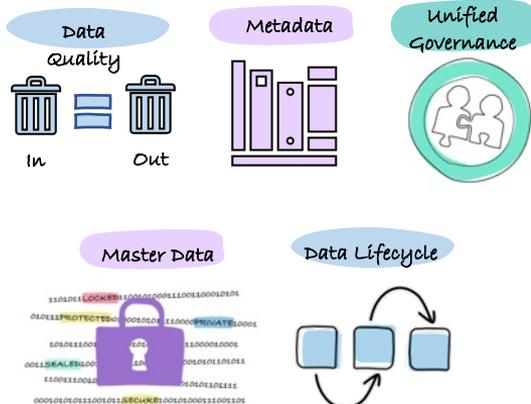
Data Governance

IBM defines Data governance (DG) as "the overall management of data availability, relevancy, usability, integrity and security in an enterprise." Below are some of the core components in Data Governance:

- Data Quality
- Metadata Management
- Master Data
- Data Lifecycle

There are many reasons why Data Governance is critical and a key foundation in data management. Metadata makes data discoverable and therefore usable. It enables data analytics which in turn leads to data monetization. High Data Quality provides trust and reliability in making strategic business decisions. Master Data Management is key in achieving high Customer Lifetime value and Data Lineage is necessary to understand your data lifecycle. Data Governance also ensures that personal information and confidential data is handled correctly across the lifecycle.

There are three (3) focus areas for Data Storage in the context of Data Governance.

Organizations leading Data Governance programs for an Enterprise- **What are the typical database types used?** *This is addressed in the next portion of this section.*

For effective Data management using Data Governance as a foundation-**What database characteristics are most useful?** *This is addressed on the radar diagram of this section.*

For a deeper focus on Data Governance-**What applications or services are available?** *This is discussed later in this section.*

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

There are several **database types**, and each offers its own specialty on where it excels in the support of core data governance functions.
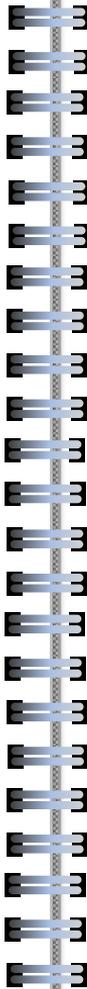
Enterprise level **Relational Databases** pre-dominantly have Data Governance features across the board when compared to other database types. Features that were unique to NoSQL databases, have been added to modern RDBMS such as Graph processing and RDF storage. Scalability, performance and multi-source data replication are other key features.
Databases: IBM DB2, Azure SQL DB, PostgreSQL, Oracle, SQL Server

Document databases do not enforce data integrity in data relationships as RDBMS does (i.e. Primary Key, Foreign Keys). Data that inherently has many-to-many relationships and require many joins are not good use cases for document DBs. Its flexible schema from a Data Governance perspective is a challenge. Data Quality issues may surface if not managed properly with an add on feature(s) or service(s) that focuses on Data Quality. Some document databases support other models such as RDF which could facilitate data lineage.
Databases: Cloudant, Elastic Search (DB & Search engine) MarkLogic, Mongo DB

Graph Databases excel in managing data relationships. Visualization of these relationships make it easier to understand hierarchies and relationships across entities representing business domains. A graphDB is schema less, it provides the flexibility for an evolving data model that would be needed in master data management.

It provides the ability to add new relationships and entities as needed across an enterprise (organizations, business units, data domains, etc.) and to search through them quickly as opposed to a relational database. Some Graph Databases support RDF which is important when doing Metadata management to enable the ability to identify hidden relationships not explicitly defined.
Databases: JanusGraph, Azure Cosmos DB, Neo4j

File storage such as Hadoop and Cloud Object storage must be included in the data scope of a Data Governance program especially as these are commonly used in data lakes. The reality is that executing Data Governance for these storage systems is more complex. Features in this space are limited when compared to structured data and mostly focus on data classification to enable data protection, data loss and backup and recovery. There are applications that offer a SQL engine on file formats such as IBM Db2 Big SQL.
File Storage: HDFS, Cloud Object Storage

Key database characteristics are listed by Data Governance use case (Data Quality, Metadata Management, Master Data Management, Data Lifecycle)

- **Data Quality**: *Data Analysis* and the definition of *Data rules* based on expected data usage such as allowed values are necessary to identify data anomalies. *Data Enrichment/Cleansing* features to then address data anomalies and increase data quality. Ongoing *Monitoring/Alerts* for a proactive approach to minimize business impact. Prevention of data loss and data consistency with features such as *Data Replication/Synchronization* and *ACID* are key.

- **Data Lifecycle**: *Data Lineage* captures the data origin, destination and data changes along the way together with *history* are important aspects in the tracking of data. Especially with financial regulations that stipulate this as a hard requirement. Features such as *Relationship Management* and *Change Data Capture (CDC)* enable Data Provenance.
- **Master Data Management**: *Visual Data Modeling* more easily portrays the several data relationships including hierarchal across the board. A major part of master data management is *data migration, integration and mapping* of data from several sources including *de-duplication* of data to create one version.

- **Metadata Management**: Data discovery is enabled through understanding what data you have and where it's located-*Data Dictionary/Catalog. Data Classification is* based on data content and provides the correct handling of data i.e. sensitive data. Capturing and maintaining Data relationships and definitions can benefit from automation via *Inferencing*. Today's world consists of many data tools and a push button solution for the *Sharing of Metadata* across any vendor tool is key - Egeria addresses this.

Add on services, platforms and applications for Data Governance
The following services or platforms can be used together with a database. It is important to understand that the features a database offers usually isn't sufficient to address overall Data Governance. A software platform, application or service that focus on different aspects of Data Governance is necessary.
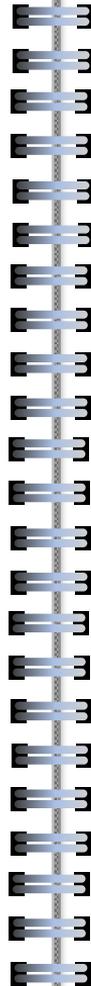
IBM Tools below address Data Analysis/Data Rules, Data Enrichment/Cleansing including De-duplication, Monitoring/Alerts, Data Lineage, Relationship Management, Visual Data Modeling, Data Migration/Integration, Data Mapping, Dictionary/Catalog and Data Classification.

> Watson Knowledge Catalog
> InfoSphere Information Governance Catalog
> InfoSphere Information Analyzer
> InfoSphere Data Architect
> InfoSphere Master Data Management

ODPi Egeria – Metadata management by automatically sharing and exchanging metadata between tools and data platforms, no matter which vendor
SQL Server Data Quality Services – Data Rules/Data Analysis, Monitoring/Alerts
SQL Server Integration Services – Data Migration/Data Integration, Data Rules/Data Analysis
SQL Server Migration Assistance – Data Migration/Data Integration, Data Mappings

- Azure Analysis Services – Visualized Data Modeling
- Azure Data Factory – Data Lineage, Data Migration/Data Integration
- Azure Database Migration Service – Data Migration/Data Integration
- Hasura GraphQL – Visualized Data Modeling, Data Migration/Data Integration, Data Rules/Data Analysis, Monitoring/Alerts, Data Dictionary/Catalog
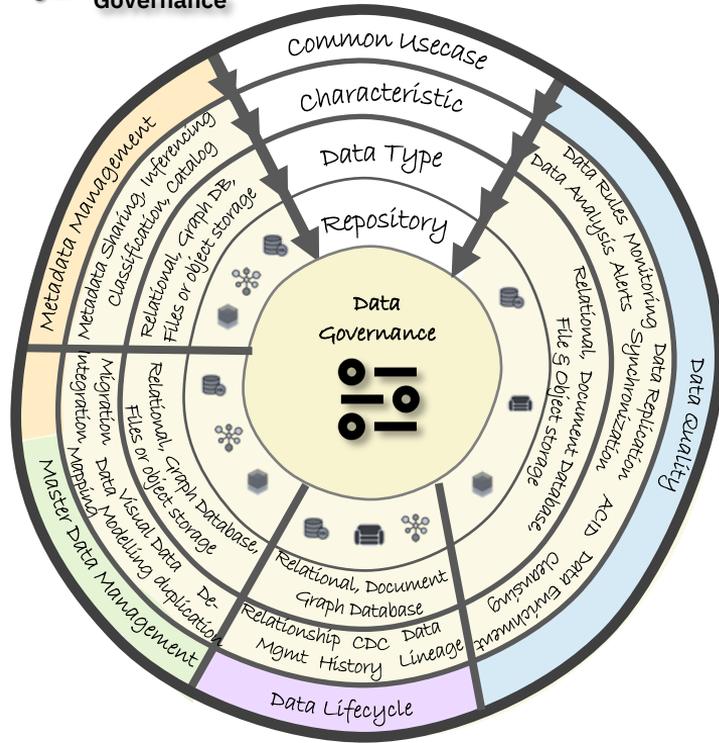
Unstructured Data Governance products
- Cohesity
- Commvault
- Druva

Data Placement Radar Diagram for Data Governance

Common Usecase
Characteristic
Data Type
Repository

Data Governance

Metadata Management
Metadata Sharing, Inferencing
Classification, Catalog
Relational, Graph DB, Files or object storage

Data Rules Monitoring
Data Analysis Alerts
Data Replication Synchronization
Relational, Document Database
ACID
File & Object storage

Data Quality

Migration
Data Visual Data De-
Relational, Graph Database, Files or object storage
Migration Data Modelling duplication
Mapping
Master Data Management

Relationship Mgmt CDC History
Relational, Document Graph Database
Data Lineage
Data Enrichment
Data Cleansing

Data Lifecycle

Data Repository Legend

- Object/Block
- Document
- Relational
- Time
- Graph
- Key-value

Data Placement Radar Diagram for Data Governance
*Backup & Restore and Scalability/Performance apply to all use cases

References:

- https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-database-paas-overview
- https://db-engines.com/
- https://docs.microsoft.com/en-us/azure/postgresql/
- https://www.ibm.com/support/knowledgecenter/SSEPGG_11.1.0/com.ibm.db2.luw.licensing.doc/doc/r0053238.html
- https://docs.oracle.com/en/database/oracle/oracle-database/19/index.html
- https://docs.microsoft.com/en-us/sql/sql-server/
- https://www.marklogic.com/
- https://docs.mongodb.com/manual/introduction/
- https://docs.microsoft.com/en-us/azure/cosmos-db/introduction
- https://neo4j.com/docs/getting-started
- https://documentation.commvault.com/commvault/v11_sp19/article?p=documentation.htm
- https://docs.druva.com/
- https://www.cohesity.com/resource-assets/datasheets/Cohesity-DataPlatform-Datasheet.pdf
- https://www.ibm.com/products/db2-big-sql
- SIGNORETTI, E., (2020). GIGAOM RADAR FOR UNSTRUCTURED DATA MANAGEMENT V1.0, KNOWINGLY, INC.
- https://www.ibm.com/events/in-en/think/skills-summit/assets/cog_enterprise_reinventing_your_company_research_insights_v2.pdf
- https://cloud.google.com/solutions/build-a-data-lake-on-gcp
- https://docs.microsoft.com/en-us/azure/architecture/example-scenario/data/data-warehouse
- https://en.wikipedia.org/wiki/Polyglot_persistence
https://aws.amazon.com/solutionspace/financial-services/data-and-analytics/

Introduction
Artificial Intelligence
Decision Support
Internet of Things (IoT)
Blockchain
Intelligent Workflow
Data Governanc
Conclusion Summary
Systems Diagram

Summary

Enterprises are willing to invest in new Data Platforms and Cloud solutions as they leave traditional data centers due to the increasing cost of aging data infrastructure and the volume of data being generated each day. This creates an urgent need for a new Modern Data Architecture (MDA). A modern data architecture drives the cognitive enterprise and enables advanced technologies in a multi-cloud environment that is fueled by clean and contextual data. A modern data architecture supports the next generation cognitive enterprise to fully exploit data using advanced technologies like pervasive artificial intelligence (AI), automation, Internet of Things (IoT) and blockchain in a hybrid or multi-cloud environment.

Top Seven Characteristics of a Modern Data Architecture are:

- Cloud native and cloud enabled
- Robust, scalable, & portable data pipelines
- Seamless data integration
- Real-time data enablement
- Decoupled & extensible
- Domain-driven, event-based, microservice-enabled
- Balanced

This document has examined the specific needs of a range of these different technology areas across the Cognitive Enterprise to understand what the typical implications are for these technological areas when it comes to considerations for the selection of different data stores and how the different characteristics mentioned on the previous page come into play.

Particular attention is given to the context of how these characteristics and requirements for data storage change as the data travels across the different component areas of the enterprise in support of each of these technological areas.

In addition, there are the considerations and choices about the selection and placement of data stores in the broader context of a multi-cloud landscape that is the focus of an increasing number of organizations.

Overall, the intention is that this document, and the associated references to other material, provides a useful set of guidelines and data points to assist a Data Architect in guiding organizations attempting to build integrated cross-enterprise and multi-faceted cognitive solutions.

Introduction

Artificial Intelligence

Decision Support

Internet of Things (IoT)

Blockchain

Intelligent Workflow

Data Governanc

Conclusion Summary

Systems Diagram

Data Governance is important across the various systems, be it governing the data or providing data repositories to govern the data.

Decision Support collects data from all the systems into the SOI where it is organized and analyzed.

System of Engagement

Mobile
App
Website
Channels

Artificial Intelligence is mostly prepared in the SOI and infused in all the other systems.

System of Insights

Analytic
Data Lake
Data Warehous
ML & AI

System of Record

Core
System
ERP
Loyalt
Files

System of Automation, Integration & Exchange

API
BPM
RPA
Enterprise
Content

Intelligent Workflow ensures that the interaction between the Systems are intelligent, event driven and automated, infused with AI..

Internet of Things handles events between Machines and things in the System of Things and interfaces with al the other systems.

System of Things

Operational
Tech
IoT
Edge
Sensors

System of Ecosystems

Public
Data
Blockchain
Platform

Blockchain processes transactions in the System of Ecosystems while interfacing with all the other systems.

Object/Block   Document
Time   Graph
Relational   Key-value

68   m   Data Repository Legend

Notes:

70

Want to learn more about Information Architecture?

Become a member of

IBM Worldwide Community of Information Architects (WWCIA)

**Affiliated to
IBM Academy of Technology**

Join via Slack Channel
#a-ae-wwcia

Slack Workspace
ibmaot.slack.com

IBM Academy of Technology
https://w3.ibm.com/academy