**Document Framework: Leveraging Microsoft Fabric for Our Future Enterprise Data Platform**

**1. Executive Summary**

* **Brief overview of the current data platform (T-SQL, IDMC, SQL Server Managed Instance on Azure).**
* **Strategic vision for leveraging Microsoft Fabric to support analysis, reporting, and operational needs.**
* **Goals: Simplify the tech landscape, enhance scalability, and phase out Cognos in favor of Power BI.**

**2. Current Challenges and Objectives**

**2.1 Challenges**

* **Complex workflows with redundant and duplicate data processing.**
* **Fragmented tools for analytics, reporting, and operational use.**
* **Dependence on Cognos, limiting scalability and modern visualization capabilities.**

**2.2 Objectives**

* **Unify data processing, storage, and analytics under one platform.**
* **Optimize data workflows with layers like bronze, silver, and gold.**
* **Enable advanced analytics and AI/ML capabilities alongside T-SQL workloads.**
* **Provide consistent, high-quality data for business decisions.**

**3. Microsoft Fabric: Overview and Features**

**Microsoft Fabric is a cutting-edge data platform that provides a unified solution for managing data across engineering, analytics, and business intelligence domains. It also integrates seamlessly with Microsoft Purview to enhance data governance and cataloging capabilities, making it a complete solution for data-driven organizations. Microsoft Fabric is a cutting-edge data platform that provides a unified solution for managing data across engineering, analytics, and business intelligence domains. Below are its key features and additional highlights.**

**3.1 What is Microsoft Fabric?**

* **Integrated platform for data engineering, data science, real-time analytics, and business intelligence.**
* **Built on Azure, providing robust integration with existing Azure services.**

**3.2 Key Features**

* **Data Virtualization: Allows querying data without the need to physically move it, saving time and storage costs.**
* **Integration with Azure Services: Deep integration with Azure Synapse, Azure Data Factory, and other Azure components enhances productivity.**
* **Real-Time Analytics: Offers real-time data processing capabilities for streaming data insights.**
* **Centralized OneLake Storage: Fabric’s OneLake serves as the central storage hub, supporting shortcuts for shared data access across teams and environments.**
* **Built-in Collaboration Tools: Enables seamless collaboration through shared workspaces and integrated tools like Microsoft Teams.**
* **Advanced AI Integration: Built-in support for training and deploying AI/ML models with Spark and Azure Machine Learning.**
* **Enhanced Cost Management: Fabric provides tools for monitoring and optimizing storage and compute costs to align with organizational budgets.**
* **Power BI Integration: Delivers modern reporting and visualization capabilities to meet diverse analytical needs, with features like paginated reports, embedded analytics, and AI-driven insights.**
* **Microsoft Purview Integration: Enhances governance by enabling comprehensive data discovery, lineage tracking, classification, and cataloging to ensure compliance and data integrity.**
* **Data Virtualization: Allows querying data without the need to physically move it, saving time and storage costs.**
* **Integration with Azure Services: Deep integration with Azure Synapse, Azure Data Factory, and other Azure components enhances productivity.**
* **Real-Time Analytics: Offers real-time data processing capabilities for streaming data insights.**
* **Centralized OneLake Storage: Fabric’s OneLake serves as the central storage hub, supporting shortcuts for shared data access across teams and environments.**
* **Built-in Collaboration Tools: Enables seamless collaboration through shared workspaces and integrated tools like Microsoft Teams.**
* **Advanced AI Integration: Built-in support for training and deploying AI/ML models with Spark and Azure Machine Learning.**
* **Enhanced Cost Management: Fabric provides tools for monitoring and optimizing storage and compute costs to align with organizational budgets.**
* **Power BI Integration: Delivers modern reporting and visualization capabilities to meet diverse analytical needs, with features like paginated reports, embedded analytics, and AI-driven insights.**
* **Data Virtualization: Allows querying data without the need to physically move it, saving time and storage costs.**
* **Integration with Azure Services: Deep integration with Azure Synapse, Azure Data Factory, and other Azure components enhances productivity.**
* **Real-Time Analytics: Offers real-time data processing capabilities for streaming data insights.**
* **Centralized OneLake Storage: Fabric’s OneLake serves as the central storage hub, supporting shortcuts for shared data access across teams and environments.**
* **Built-in Collaboration Tools: Enables seamless collaboration through shared workspaces and integrated tools like Microsoft Teams.**
* **Advanced AI Integration: Built-in support for training and deploying AI/ML models with Spark and Azure Machine Learning.**
* **Enhanced Cost Management: Fabric provides tools for monitoring and optimizing storage and compute costs to align with organizational budgets.**
* **Unified Lakehouse Architecture: Combines data lake and data warehouse capabilities.**
* **T-SQL Support: Seamless compatibility with existing T-SQL workloads.**
* **Spark Workloads: Enables advanced data engineering and AI/ML.**
* **Power BI Integration: Modern analytics and visualization tools, ideal for replacing Cognos.**
* **Fabric Layers: Bronze (raw data), Silver (cleaned data), and Gold (curated data).**
* **Shortcuts for Reference and Master Data: Reduce redundancy by centralizing data sources.**
* **Security and Governance: Role-based access, lineage tracking, and compliance tools.**

**4. Proposed Architecture**

**4.1 Data Ingestion**

* **Utilize Fabric’s integration capabilities to process data through IDMC as the ETL tool and ingest it from other sources.**
* **Real-time and batch processing support.**

**4.2 Data Staging Layers**

* **Bronze Layer: Store raw ingested data for traceability.**
* **Silver Layer: Clean and standardize data; remove duplicates and apply validations.**
* **Gold Layer: Curated datasets ready for reporting, AI/ML, and analytics.**

**4.3 Reference and Master Data Lakehouse**

* **Implement a dedicated lakehouse for reference and master data with its own bronze, silver, and gold layers.** 
  + **Bronze Layer: Capture raw reference and master data from source systems.**
  + **Silver Layer: Clean, standardize, and deduplicate reference and master data.**
  + **Gold Layer: Provide curated, validated datasets for consumption by all workloads.**
* **Use shortcuts to share curated data with other lakehouses, ensuring a single source of truth for analytics, reporting, and AI/ML processes.**

**4.4 Environment Segregation**

* **Maintain separate environments for DEV, QA, and PROD to ensure data quality and integrity across the pipeline.**
* **Implement shortcuts between environments for consistent reference and master data access while adhering to governance policies.**

**4.4.1 Environment Separation Options**

**Microsoft Fabric supports multiple strategies for managing DEV, QA, and PROD environments:**

1. **Single OneLake with Separate Workspaces:**
   * **Create distinct workspaces for DEV, QA, and PROD within the same OneLake.**
   * **Use shortcuts to share curated reference and master data across environments.**
   * **Apply role-based access controls (RBAC) to limit permissions for each environment.**
2. **Separate OneLake Instances:**
   * **Use individual OneLake instances for each environment for strict data and resource isolation.**
   * **Data movement between environments must be handled using pipelines or data export/import workflows.**
   * **This approach is ideal for organizations with stringent compliance or security requirements.**
3. **Hybrid Approach:**
   * **Use a shared OneLake for DEV and QA while maintaining a separate OneLake for PROD.**
   * **Simplifies testing and validation while ensuring PROD data remains isolated.**
4. **Governance Best Practices:**
   * **Use Azure AD integration for managing user and group permissions.**
   * **Monitor data lineage and enforce deployment gates to control data promotion between environments.**
   * **Leverage Fabric’s auditing tools to track changes and ensure compliance.**

* **Maintain separate environments for DEV, QA, and PROD to ensure data quality and integrity across the pipeline.**
* **Implement shortcuts between environments for consistent reference and master data access while adhering to governance policies.**

**4.5 Architecture Diagram**

* **Visual Representation: Below is a diagram illustrating the sequential data flow from ingestion through the bronze (raw data), silver (cleansed data), and gold (curated data) layers, along with features like OneLake for centralized storage and shortcuts for cross-environment data sharing.**
* **Include elements such as:** 
  + **Data sources (e.g., external feeds processed through IDMC, other internal/external systems).**
  + **Data staging layers (bronze, silver, gold).**
  + **Analytical and AI/ML processes.**
  + **Power BI dashboards as end-user interfaces.**
  + **Environment segmentation for DEV, QA, and PROD to ensure robust testing and secure production environments.**
* **Visual Representation: Below is a diagram illustrating the sequential data flow from ingestion through the bronze (raw data), silver (cleansed data), and gold (curated data) layers, including the integration of a dedicated lakehouse for reference and master data.**
* **Include elements such as:** 
  + **Data sources (e.g., external feeds processed through IDMC, other internal/external systems).**
  + **Data staging layers (bronze, silver, gold).**
  + **Dedicated reference and master data lakehouse with shortcuts to other workloads.**
  + **Analytical and AI/ML processes.**
  + **Power BI dashboards as end-user interfaces.**

**4.1 Data Ingestion**

* **Utilize Fabric’s integration capabilities to process data through IDMC as the ETL tool and ingest it from other sources.**
* **Real-time and batch processing support.**

**4.2 Data Staging Layers**

* **Bronze Layer: Store raw ingested data for traceability.**
* **Silver Layer: Clean and standardize data; remove duplicates and apply validations.**
* **Gold Layer: Curated datasets ready for reporting, AI/ML, and analytics.**

**4.3 Analytics and Reporting**

* **Leverage Power BI for interactive dashboards, real-time reporting, and deep analytics.**
* **Replace Cognos reports with Power BI to provide a modern and scalable reporting solution.**
* **Use Power BI’s integration with OneLake to access curated datasets from gold layers seamlessly.**

**4.3.1 Best Practices for Power BI Usage**

1. **Model Optimization:**
   * **Design efficient data models using star schema to improve query performance.**
   * **Utilize composite models to combine direct query and imported data for flexibility.**
2. **Data Governance:**
   * **Implement row-level security (RLS) to restrict access to sensitive data.**
   * **Use shared datasets to ensure consistent and governed data across reports.**
3. **Collaboration and Deployment:**
   * **Use shared workspaces for collaborative report development.**
   * **Establish a clear deployment pipeline: DEV → QA → PROD for Power BI reports.**
4. **Performance Monitoring:**
   * **Enable usage metrics and query diagnostics to track performance and optimize reports.**
   * **Use incremental data refresh for large datasets to improve load times.**
5. **Integration with Other Tools:**
   * **Embed Power BI reports into applications like Microsoft Teams or custom web portals.**
   * **Utilize Power BI’s AI features (e.g., natural language queries and AI visuals) for advanced insights.**

* **Leverage Power BI for interactive dashboards and reports.**
* **Replace Cognos reports with modern Power BI alternatives.**

**4.4 Advanced Analytics and AI/ML**

* **Use Spark workloads for predictive analytics and machine learning.**
* **Combine structured and unstructured data for deeper insights.**

**4.5 Architecture Diagram**

* **Visual Representation: Below is a diagram illustrating the sequential data flow from ingestion through the bronze (raw data), silver (cleansed data), and gold (curated data) layers to final consumption points, ensuring that each layer serves specific roles in the pipeline.**
* **Include elements such as:** 
  + **Data sources (e.g., external feeds processed through IDMC, other internal/external systems).**
  + **Data staging layers (bronze, silver, gold).**
  + **Analytical and AI/ML processes.**
  + **Power BI dashboards as end-user interfaces.**
* **Include elements such as:** 
  + **Data sources (e.g., external feeds processed through IDMC, other internal/external systems).**
  + **Data staging layers (bronze, silver, gold).**
  + **Analytical and AI/ML processes.**
  + **Power BI dashboards as end-user interfaces.**

**4.1 Data Ingestion**

* **Utilize Fabric’s integration capabilities to ingest data from IDMC and other sources.**
* **Real-time and batch processing support.**

**4.2 Data Staging Layers**

* **Bronze Layer: Store raw ingested data for traceability.**
* **Silver Layer: Clean and standardize data; remove duplicates and apply validations.**
* **Gold Layer: Curated datasets ready for reporting, AI/ML, and analytics.**

**4.3 Analytics and Reporting**

* **Leverage Power BI for interactive dashboards and reports.**
* **Replace Cognos reports with modern Power BI alternatives.**

**4.4 Advanced Analytics and AI/ML**

* **Use Spark workloads for predictive analytics and machine learning.**
* **Combine structured and unstructured data for deeper insights.**

**5. Benefits of Microsoft Fabric**

**5.1 Operational Efficiency**

* **Streamlined workflows reduce redundant processing and improve efficiency.**
* **Shortcuts for Reference and Master Data ensure single data sources.**
* **Data Virtualization minimizes data movement, speeding up access to insights.**
* **Built-in tools simplify collaboration across teams and workflows.**
* **Power BI provides self-service analytics, reducing dependency on IT for report generation.**

**5.2 Enhanced Analytics**

* **T-SQL and Spark workloads address diverse analytical needs.**
* **Power BI’s rich visualization capabilities provide actionable insights.**
* **AI-driven features in Power BI enable predictive and prescriptive analytics for deeper business understanding.**

**5.1 Operational Efficiency**

* **Streamlined workflows reduce redundant processing and improve efficiency.**
* **Shortcuts for Reference and Master Data ensure single data sources.**
* **Data Virtualization minimizes data movement, speeding up access to insights.**
* **Built-in tools simplify collaboration across teams and workflows.**
* **Streamlined workflows reduce redundant processing and improve efficiency.**
* **Shortcuts for Reference and Master Data ensure single data sources.**

**5.2 Enhanced Analytics**

* **T-SQL and Spark workloads address diverse analytical needs.**
* **Power BI’s rich visualization capabilities provide actionable insights.**

**5.3 Scalability and Flexibility**

* **Elastic compute and storage ensure scalability.**
* **Unified platform supports diverse workloads from data engineering to AI.**

**5.4 Governance and Security**

* **Robust tools for data lineage, versioning, and compliance.**
* **Integration with Azure AD for role-based access control.**
* **Purview for Governance:** 
  + **Centralized data cataloging to organize and discover data assets efficiently.**
  + **Automated lineage tracking for transparency across data pipelines.**
  + **Classification and labeling to ensure compliance with regulatory requirements.**
  + **Policies for secure data sharing and access control across environments.**
* **Robust tools for data lineage, versioning, and compliance.**
* **Integration with Azure AD for role-based access control.**

**6. Implementation Roadmap**

**6.1 Phase 1: Planning and Assessment**

* **Evaluate current workflows and map to Fabric’s capabilities.**
* **Identify datasets for migration and prioritize critical workloads.**

**6.2 Phase 2: Pilot and Testing**

* **Implement pilot projects for T-SQL and Spark workloads.**
* **Develop a proof of concept for Power BI replacing Cognos.**

**6.3 Phase 3: Full Migration**

* **Transition all workloads to Microsoft Fabric.**
* **Establish best practices for using bronze, silver, and gold layers.**

**6.4 Phase 4: Optimization and Expansion**

* **Optimize workflows and implement advanced analytics.**
* **Train users on Power BI and Fabric’s features.**

**7. Conclusion and Next Steps**

* **Recap of how Microsoft Fabric addresses current challenges and meets future needs.**
* **Immediate next steps: finalize the roadmap, secure stakeholder buy-in, and begin pilot projects.**

**8. Appendices**

**8.1 Glossary of Terms**

* **Define key terms like bronze, silver, gold layers, lakehouse, etc.**

**8.2 References**

* **Documentation and links to Microsoft Fabric resources.**

**8.3 Cost Analysis (Optional)**

* **Overview of anticipated cost implications and ROI.**