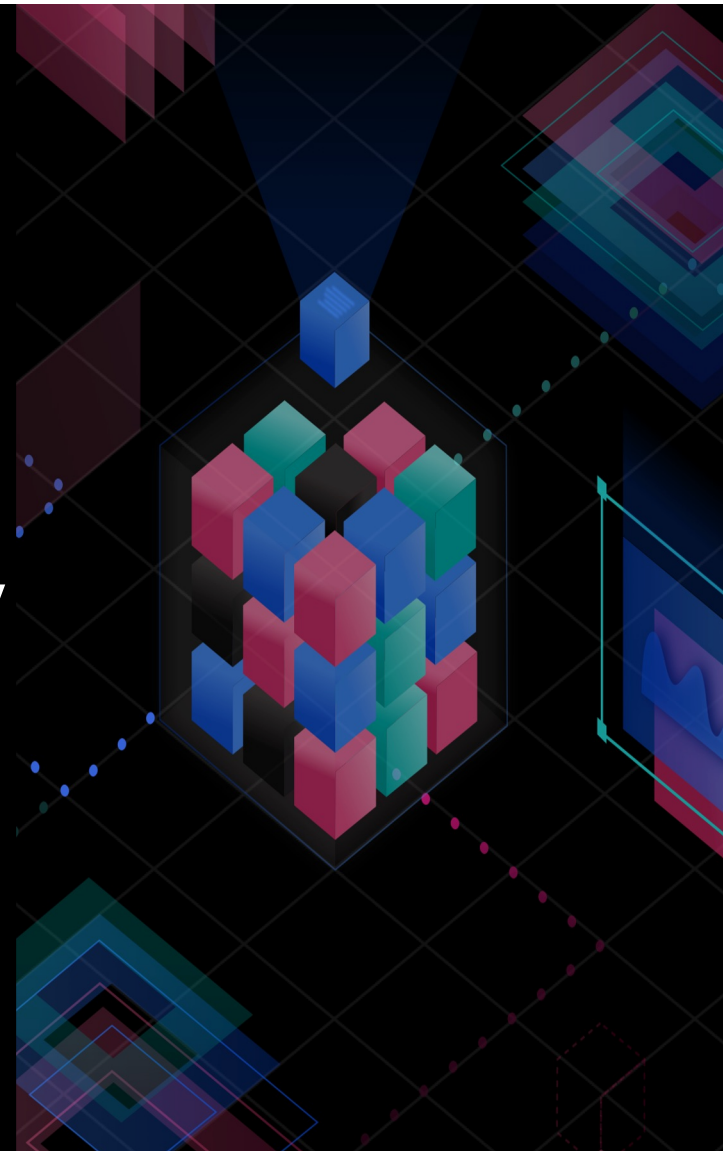# Lab1 -
Data Virtualization, Governance & Privacy

Workshop – CPD 4.5.0

CPD 4.5.0 – Analytics Modernization Workshop Purpose

Goal:
>   Provide a hands-on Cloud Pak for Data 4.5.0 environment that demonstrates the end-to-end connected CPD capabilities, via labs, with a consistent data narrative relatable to everyone (Human Resource Department data scenarios).

Key Message:
>   CPD is the one stop shop for all analytical and data needs, centralized administration of security, data governance, analytics, data science

Capabilities currently showcased:
>   Data Governance – WKC, Watson Query
>   Cognos Analytics – Data Models, Reports, Dashboards, Explorations
>   Planning Analytics – Workforce Planning Module
>   Trustworthy AI – Attrition Prediction with Watson ML, Watson Studio, OpenScale, OpenPages

To create a robust foundation for AI, businesses need IBM's data fabric for data governance and privacy

	Access	**Provides** quick access to data across complex hybrid cloud environments

	Governance	**Provides** high-quality information to the right people at the right time

	Compliance	**Protects** personal and confidential information to ensure privacy and compliance

What You Will Learn

You'll learn how to build an end to end, virtualized, catalogued, protected and governed data platform for your customers.

Capabilities/Technologies Showcased

Key Message: Truly protecting sensitive data is hard enough when you're dealing with a single source; organizations today have hundreds or thousands of disparate sources, and data privacy regulations are getting more stringent. That's a serious challenge to any organization.

Story/Lab: Bring data from completely different database types and locations together, and then automate the process of understanding where the sensitive data is hiding within that new, virtual data set.

- Create a virtualized data set joining data from DB2 and Postgres without moving it

- Leverage Automated Data Enrichment to highlight SPI, PII, etc.

- Use data protection rules to specify what types of users can see what types of data.

- Visualize the data with 2 personas: 1 that can see PII, and on that cannot, without doing any explicit coding.

Watson Query and Watson Knowledge Catalog (2 hours)

Why You Need This Capability

Truly protecting sensitive data is hard enough when you're dealing with a single source; organizations today have hundreds or thousands of disparate sources, and data privacy regulations are getting more stringent.

Simple manual or even tool based data profiling can no longer handle the complexity of today's data, including the importance of finding all of your sensitive data, wherever it lives. For any offering to deliver on this, it needs to have *automated AI and ML infused data discovery and enrichment*.

IBM's **Watson Query and Knowledge Catalog** deliver critical automation and virtualization capabilities to solving this problem, ensuring you can deliver the data your customers need in a secured fashion, ***from any tool.***

Watson Query and Watson Knowledge Catalog (2 hours)