

# Pentaho+ Platform Advanced Capabilities – Confidential Queries

Pentaho is a powerful business intelligence (BI) and data integration (DI) platform designed to help organizations process, analyze, and visualize data. Its complex capabilities span a variety of data handling and analytics processes. Here's a link providing less than 2-min video overview of the platform components:

- Pentaho Data Integration <https://www.youtube.com/watch?v=D4AYV3MT6zU>
- Pentaho Business Analytics <https://www.youtube.com/watch?v=C3MyInpETE8>
- Pentaho Data Catalog <https://www.youtube.com/watch?v=b39BHeMLjPY>

Pentaho's capabilities can be extended to handle confidential queries and data processing tasks, particularly in sensitive environments such as healthcare, finance, law enforcement or defense. Integration with solutions like Duality Technologies, which specializes in privacy-preserving data collaboration and secure multiparty computation (MPC), enhances Pentaho's ability to handle confidential data securely. These techniques enable multiple parties to analyze and compute on encrypted data without exposing the raw data. Pentaho can function as the data integration and analytics engine, orchestrating the extraction, transformation, and preparation of encrypted data for processing by Duality's platform. The results of encrypted computations are then decrypted to produce insights while preserving privacy.

## Use Case Scenarios

- ⇒ *Cross-Organizational Analytics*: When two or more organizations need to collaborate on sensitive datasets (e.g., patient records in healthcare, customer data in banking), Pentaho can integrate data and prepare it for confidential query execution in Duality.
- ⇒ *Regulatory Compliance*: Pentaho's role as a data integrator ensures traceability, auditability, and adherence to privacy laws like GDPR, CCPA, and HIPAA, while Duality ensures computations remain secure and private.
- ⇒ *Fraud Detection*: Financial institutions can share encrypted data to detect fraudulent patterns collaboratively, with Pentaho managing the pre-processing and Duality executing secure queries.

## Workflow Overview

1. Data Integration:
  - Pentaho extracts data from various sources, such as relational databases, cloud platforms, and legacy systems.
  - Data is encrypted at the source using Duality's encryption mechanisms.
2. Data Preparation:
  - Pentaho performs necessary transformations (e.g., normalization, deduplication) on encrypted data or coordinates with Duality to ensure transformations occur securely.
3. Confidential Query Execution:
  - Pentaho passes the encrypted data to Duality's platform, which processes the data using secure MPC or homomorphic encryption.
  - Duality executes confidential queries, such as aggregations, filtering, and predictive analytics, while maintaining data confidentiality.

#### 4. Decrypted Results:

- Results are returned to Pentaho, where decrypted insights are visualized or used in dashboards and reports, ensuring only authorized users can view the results.

### Advanced Encryption Techniques

- Homomorphic Encryption: Duality allows computations to be performed on encrypted data without decrypting it. Pentaho integrates by feeding encrypted datasets and retrieving results post-processing.
- Secure MPC: When multiple parties are involved, each contributes encrypted data. Duality computes jointly on the encrypted inputs without exposing individual contributions.
- Differential Privacy: Duality applies noise to results to prevent re-identification of individuals, while Pentaho ensures that the output remains actionable.

### Advantages of Integration

- End-to-End Security: Pentaho ensures secure data movement and integration, while Duality focuses on secure computation.
- Data Privacy Compliance: The integration allows organizations to comply with stringent data privacy regulations while conducting meaningful analytics.
- Scalable Architecture: Pentaho's scalability allows processing of large datasets, and Duality ensures secure computations scale seamlessly.
- Interoperability: Pentaho's ability to work with various data formats and systems complements Duality's focus on cryptographic computation, enabling a seamless workflow.

### Potential Applications

- Healthcare: Enabling collaborative research on encrypted patient data across hospitals without exposing sensitive health information.
- Financial Services: Securely analyzing encrypted transaction data across institutions to detect fraud or assess credit risk.
- Public Sector: Sharing and analyzing encrypted intelligence data across agencies while maintaining national security.
- Supply Chain: Securely sharing and analyzing proprietary data between partners for better demand forecasting and inventory management.

### Implementation Considerations

- Performance: While Duality's encryption and computation add overhead, Pentaho's optimization capabilities can streamline pre-processing tasks to minimize delays.
- Security Configuration: Proper role-based access control (RBAC) and audit trails in Pentaho ensure that only authorized users can trigger secure queries or view sensitive data.
- Cost and Scalability: Organizations must balance computational costs associated with secure queries against the need for confidentiality.

## Conclusion

Pentaho's combination of data integration, BI tools, and advanced analytics makes it a versatile solution for enterprises looking to unify and leverage their data effectively. Pentaho's integration with privacy-preserving technologies like Duality represents a significant step forward in secure data analytics. This partnership allows organizations to unlock valuable insights from confidential datasets while adhering to stringent privacy regulations. It enables opportunities for secure collaboration in sensitive industries where data privacy and security are paramount.

Let us know if you'd like to explore any specific capability in more detail!