Authors: Aaron Kesler, Jump Thanawut, Scott Monteith
Version 1.0
Last updated: December 14, 2023

Overview & SnapLogic's Approach to AI / LLM:
Product & Data:
- **Product Features & Scope:**
- **Data Usage & Opt-Out Options:**
  - **Data Usage in SnapGPT:**
  - **Data Processing:**
    - **Architecture:**
    - **Data Flow:**
- **Data Retention & Residency:**
  - **Data Retention:**
  - **Data Residency:**
- **Controls – Admin, Groups, Users:**
- **Guidelines for Secure and Compliant use of SnapGPT**
- **Compliance:**

Security and Data Handling Protocols for SnapGPT

SnapLogic acknowledges and respects the data concerns of our customers. The purpose of this document is to present our data handling and global data protection standards for SnapGPT.

Overview & SnapLogic's Approach to AI / LLM:

SnapLogic utilizes high-quality Enterprise Language Learning Models (LLMs), selecting the most appropriate one for each specific task. Current support includes Azure OpenAI GPT, Anthropic Claude on Amazon Bedrock, and Google Vertex PaLM.

Product & Data:

Product Features & Scope:

SnapGPT offers a range of features, each designed to enhance user experience and productivity in various aspects of pipeline and SQL query generation:

- **Input Prompts:** This feature allows customers to interact directly with the LLM by providing input prompts. These prompts are the primary method through which users can specify their requirements or ask questions to the LLM.
- **Describe Pipeline:** This skill enables users to obtain a comprehensive description of an existing pipeline. It helps in understanding and documenting the pipeline's structure and functionality.
- **Analyze Pipeline:** This feature ingests the entire pipeline configuration and analyzes it to make suggestions for optimization and improvement. It assists users in enhancing the efficiency and effectiveness of their pipelines.
- **Mapper Configuration:** Facilitates the configuration of the mapper snap by generating expressions to simplify the process of mapping input to output.
- **Pipeline Generation:** Users can create prototype pipelines using simple input prompts. This feature is geared towards streamlining the pipeline creation process, making it more accessible and less time-consuming.

- **SQL Generation without Schema:** Tailored for situations where the schema information is not available or cannot be shared, this feature generates SQL queries based solely on the customer's prompt, offering flexibility and convenience.
- **SQL Generation with Schema (coming feb 2024):** This advanced feature generates SQL queries by taking into account the schema of the input database. It is particularly useful for creating contextually accurate and efficient SQL queries.

Data Usage & Opt-Out Options:

At SnapLogic, we recognize the importance of data security and user privacy in the rapidly evolving Generative AI space. SnapGPT has been designed with these principles at its core, ensuring that customers can leverage the power of AI and machine learning while maintaining control over their data. Our approach prioritizes transparency, giving users the ability to opt-out of data sharing, and aligning with industry best practices for data handling. This commitment reflects our dedication to not only providing advanced AI solutions but also ensuring that these solutions align with the highest standards of privacy and data protection.

Data Usage in SnapGPT:

SnapGPT is designed to handle customer data with the utmost care and precision, ensuring that data usage is aligned with the functionality of each feature:

Customer Input and Interaction: Customer inputs, such as prompts or pipeline configurations, are key to the functionality of SnapGPT. This data is used solely for the purpose of processing specific requests and generating responses or suggestions relevant to the user's query. No data is retained for model training purposes.

Feature-Specific Data Handling: Each feature/skill of SnapGPT, like pipeline analysis or SQL generation, uses customer data differently. See the table below for details on each skill.

Skill Name	Description of the Skill	Data Transferred to LLM
Input Prompts	Direct input prompts from customers are transferred to the LLM and tracked by SnapLogic analytics.	Prompt details only; these are not stored or used for training by the LLM.

Describe & Analyze Pipeline	Allows customers to describe a pipeline, with the entire pipeline configuration relayed to the LLM.	Entire pipeline configuration excluding account credential information.
Mapper Configuration	Enables sending input schema information within the prompt to the LLM for the “Mapper configuration” feature.	Input schema information without account credential information.
Pipeline Generation	Uses input prompts to create pipeline prototypes by transmitting them to the LLM.	Input prompts only; not stored or used for training by the LLM.
SQL Generation W/out Schema	Generates SQL queries based only on the customer's prompt in situations where schema information cannot be shared.	Only the customer's prompt; no schema information is used.
SQL Generation W/ Schema (Feb 2024)	Generates accurate SQL queries by considering the schema of the input database.	Schema of the input database excluding any account credentials, enhancing query accuracy.

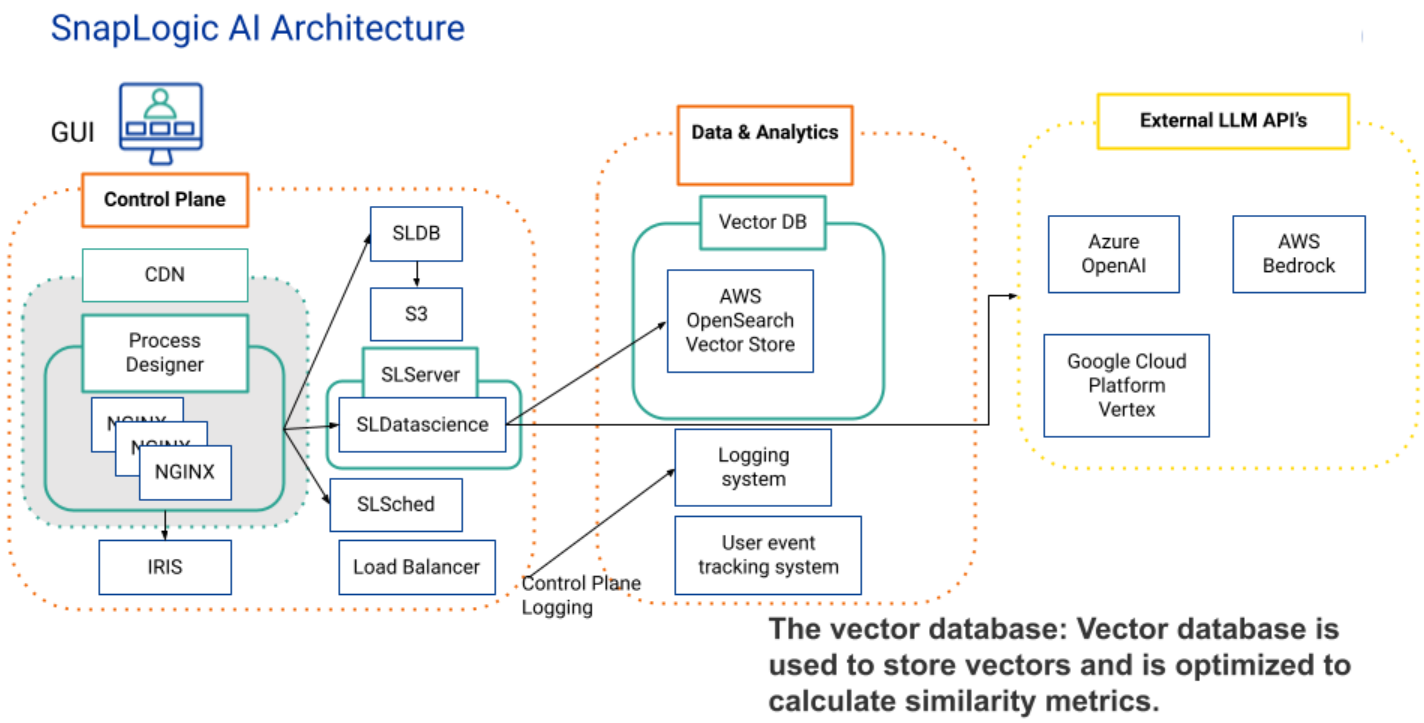
Future Adaptations: In the near future, we intend to offer customers opt-out options. Choosing to opt-out of including any environment-specific data in SnapGPT prompts can impact the quality of response from SnapGPT as it will lack additional context. As of the current version, usage of SnapGPT will include sending the data from the features listed above to the LLMs. We recommend that customers who are not comfortable with the described data transfers to wait for the opt-out option to become available.

Impact of Opting Out: Choosing to opt-out of data sharing may impact the functionality and effectiveness of SnapGPT. For example, opting out of schema retrieval in SQL Generation may lead to

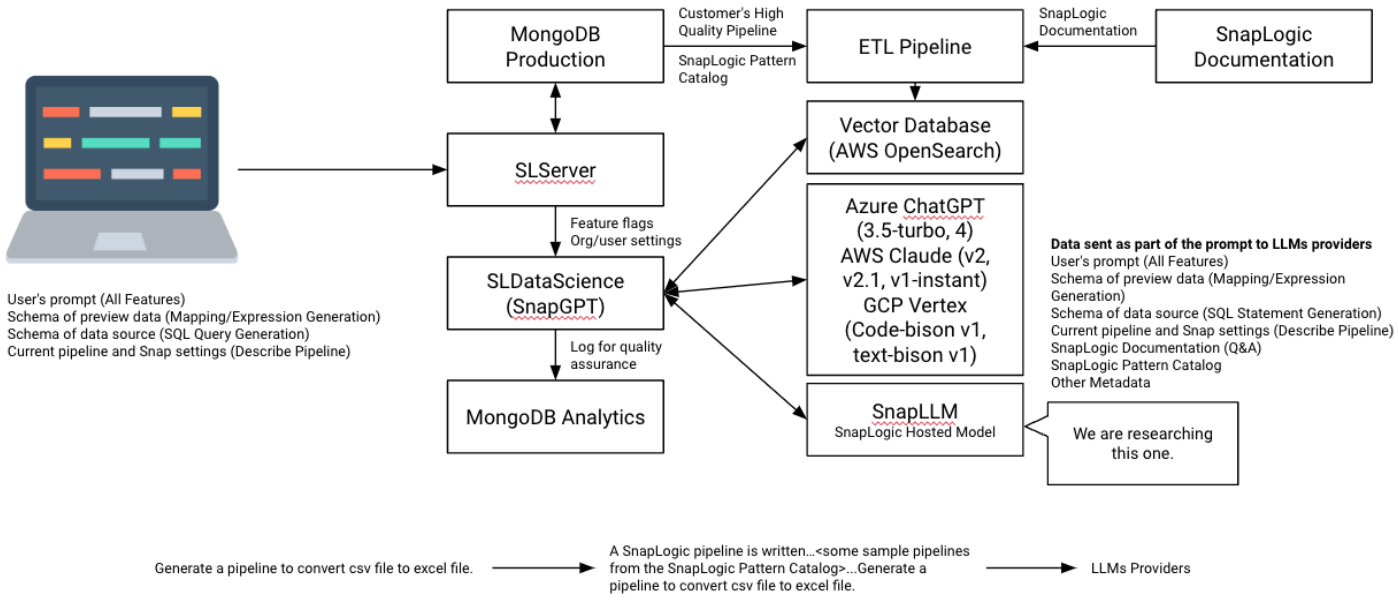
less precise query outputs. Users are advised to consider these impacts when setting their data sharing preferences.

Data Processing:

Architecture:



Data Flow:



Data Retention & Residency:

SnapLogic is committed to ensuring the secure handling and appropriate residency of customer data. Our data retention policies are designed to respect customer privacy while providing the necessary functionality of SnapGPT:

Data Retention:

- **No Retention for Model Training:** SnapGPT is designed to prioritize user privacy. Therefore, no customer data processed by SnapGPT is retained for the purpose of model training. This ensures that user data is not used in any way to train or refine the underlying AI models.
- **Storing Usage Data for Adoption Tracking:** While we do not retain data for model training, SnapLogic stores usage data related to SnapGPT in Heap Analytics. This is strictly for the purpose of tracking product adoption and usage patterns. The collection of usage data helps

us understand how our customers interact with SnapGPT, enabling us to continuously improve the product and tailor it to user needs.

Data Residency:

- **Location-Based Data Storage:** Our control planes in the United States and the EMEA region adhere to the specific data residency policies of these locations. We ensure compliance with regional data protection and privacy laws, offering customers the assurance that their data is managed in accordance with local regulations.

Controls – Admin, Groups, Users:

SnapLogic provides robust control mechanisms for administrators, while ensuring that group and user-level controls align with organizational policies:

- Administrators have granular control over the use of SnapGPT within their organization. They can determine what data is shared with the LLM and have the ability to opt out of data sharing to meet specific data retention and sharing policies. Additionally, admins can control user access to various features and skills, ensuring alignment with organizational needs and security policies.
- **Group Controls:** Currently, groups do not have specific controls over SnapGPT. Group-level policies are managed by administrators to ensure consistency and security across the organization.
- **User Controls:** Users can access and utilize the features and skills of SnapGPT to which they are entitled. User entitlements are managed by administrators, ensuring that each user has access to the necessary tools for their role while maintaining data security and compliance.

Guidelines for Secure and Compliant use of SnapGPT

At SnapLogic, we understand the critical importance of data security and compliance in today's digital landscape. As such, we are dedicated to providing our customers with the tools and knowledge necessary to utilize SnapGPT in a way that aligns with their internal information security (InfoSec) and privacy policies. This section offers guidelines to help ensure that your interaction with SnapGPT is both secure and compliant with your organizational standards.

- **Customer Data Control:** Customers are encouraged to actively manage and control the data they share with SnapGPT. By understanding and utilizing the available admin and user controls, customers can ensure that their use of SnapGPT aligns with their internal InfoSec and privacy policies.

- **Best Practices for Data Sharing:** We recommend that customers review and follow best practices for data sharing, especially when working with sensitive or confidential information. This includes using anonymization or pseudonymization techniques where appropriate, and sharing only the data in prompts and pipelines that is necessary for the task at hand.
- **Integrating with Internal Policies:** Customers should integrate their use of SnapGPT with their existing InfoSec and privacy frameworks. This integration ensures that data handling through SnapGPT remains consistent with the organization's overall data protection strategy.
- **Regular Review and Adjustment:** Customers are advised to regularly review their data sharing settings and practices with SnapGPT, adjusting them as necessary to remain aligned with evolving InfoSec and privacy requirements.
- **Training and Awareness:** We also suggest that customers provide regular training and awareness programs to their users about the responsible and secure use of AI tools like SnapGPT, emphasizing the importance of data privacy and protection.

Compliance:

For detailed information on SnapLogic's commitment to compliance with various regulatory standards and data security measures, please visit our comprehensive overview at SnapLogic Security & Compliance (https://www.snaplogic.com/security-standards). This resource provides an in-depth look at how we adhere to global data protection regulations, manage data security, and ensure the highest standards of compliance across all our products, including SnapGPT.

For specific compliance inquiries or more information on how we handle compliance in relation to SnapGPT, please contact the SnapLogic Compliance Team at Security@snaplogic.com.

For further details or inquiries regarding SnapGPT or any other SnapLogic AI services, please contact our SnapLogic AI Services Team (ai-services@snaplogic.com). For more information on SnapLogic Security and Compliance: https://www.snaplogic.com/security-standards