# The SnapLogic Integration Cloud: A Technical Overview

A SNAPLOGIC WHITEPAPER

snapLogic

# Table of Contents

# Introduction

The SnapLogic Integration Cloud delivers a fast, multi-point, and modern integration platform as a service (iPaaS) built to elastically address both real-time application and batch-oriented data integration requirements.

### Faster Integration

The SnapLogic Integration Cloud balances design simplicity with platform power so you can get up and running with a faster time to value. The SnapLogic Designer, Manager, and Monitoring Dashboard are delivered as a multi-tenant cloud service built for the "citizen integrator."

### Multi-point Integration

The SnapLogic Integration Cloud provides the industry's widest variety of pre-built intelligent connectors, called Snaps. Whether you need to connect SaaS applications, analytics, Big Data, ERP, identity management, social media, online storage, or technologies like SFTP, Oauth, and SOAP, there's usually a Snap for that. If not, you have the ability to build your own custom Snaps using our Java-based Snap SDK.

### Modern Integration

The SnapLogic Integration Cloud is purpose-built for the cloud, meaning there are no legacy components that prevent the platform from running at cloud speed. Data is streamed between applications, databases, files, social and big data sources via the Snaplex, a self-upgrading execution network. The SnapLogic Integration Cloud is 100% REST-based so developers and ISV partners can easily embed it into applications and platforms.

This whitepaper is designed to provide a technical overview of the SnapLogic Integration Cloud. For more information visit www.SnapLogic.com or contact us at 1-888-494-1570.

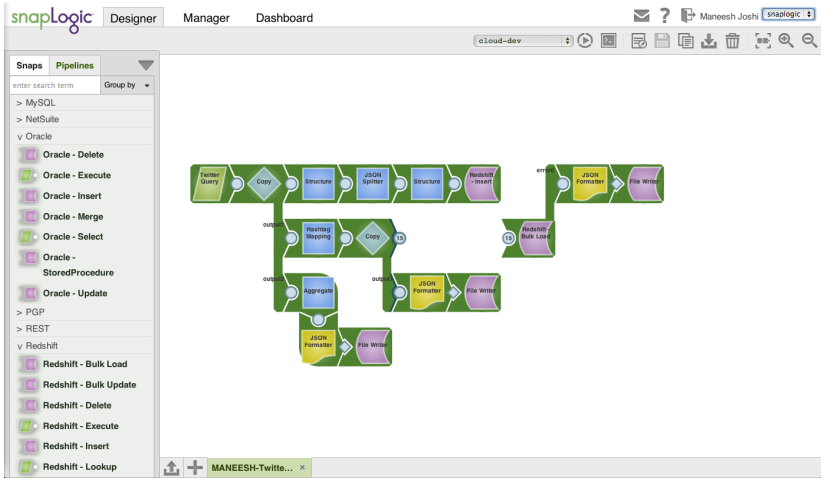
# Integration Cloud Platform Components

The SnapLogic Integration Cloud has five primary components: the Designer, Manager, Monitoring Dashboard, the Snaplex and Snaps. The Designer, Manager, and Monitoring Dashboard are 100% HTML5-based web applications that provide visual interactivity with the SnapLogic Integration Cloud infrastructure, and the Snaplex is an elastic execution network

that processes data. The Designer is targeted towards the integration developer persona who has the semantic understanding of the integration pipeline that is being built. The Manager and the Monitoring Dashboard are built for the integration administrator who manages the integration infrastructure. Snaps are modular collections of integration components built for a specific application or data source. Snaps shield both business users and developers from much of the complexity of the underlying application, data model, and service.

## The Designer

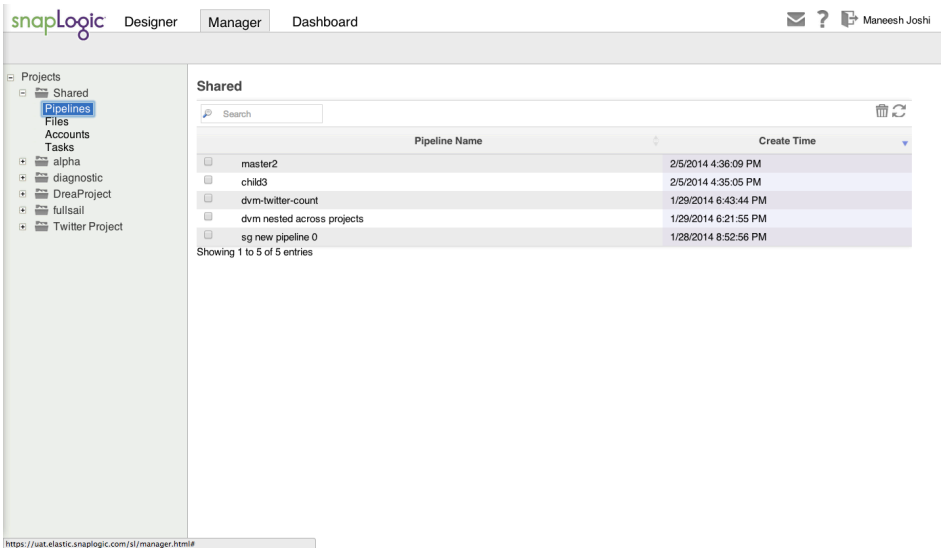
The SnapLogic Integration Cloud Designer is purpose-built for cloud-speed. Building integration workflows, called pipelines, is a simple drag, drop and configure exercise that can be done either in a browser or on a tablet. The cloud-based Designer makes multi-point integrations possible for both advanced and "citizen integrators" with no coding necessary.




*The Designer canvas allows you to build multi-point integrations in a snap.*

## The Manager

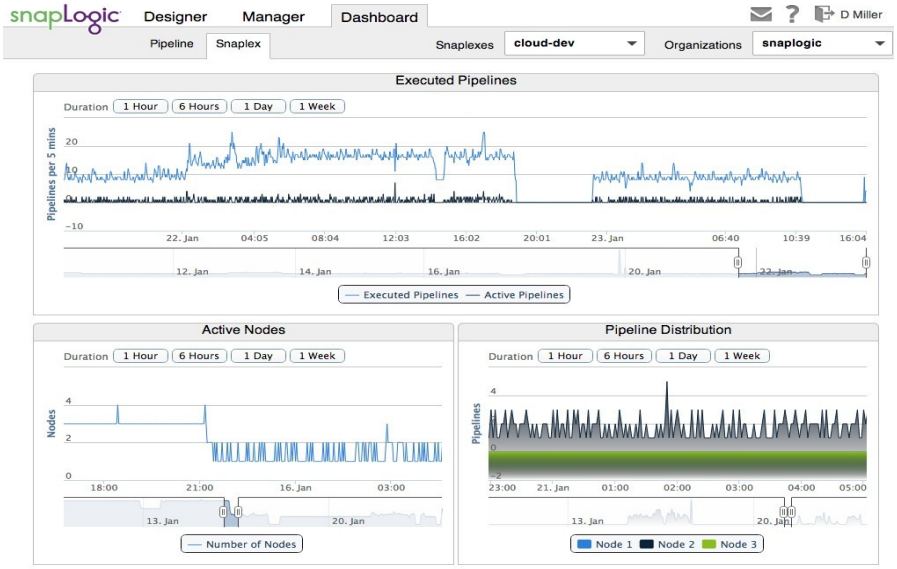
The SnapLogic Integration Cloud Manager allows you to administer your environment and set access controls for users and groups. Manage the lifecycle of data and process flows from development to test, to staging and to production and take advantage of out-of-the-box cluster management, job scheduling, failover, notification and alerting. It is also where projects, files, accounts, tasks, and connections are managed.




*Manage projects, files, accounts and tasks in the administration console.*

## Monitoring Dashboard

The SnapLogic Integration Cloud Monitoring Dashboard allows you to track performance of your integration workloads. The Monitoring Dashboard provides secure visibility into the health of your integrations with system performance dashboards, drill-down capabilities and triggered event notifications. Using the Monitoring Dashboard, administrators can manage their infrastructure from any device. This means you have complete remote visibility into your real-time and scheduled integrations.




*Get immediate insight into your real-time and scheduled integrations and Snaplexes.*

## Snaps

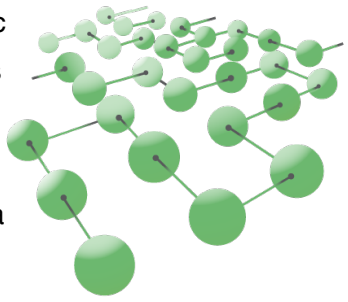
The SnapLogic Integration Cloud offers industry's broadest and richest variety of pre-built integration components, called Snaps. Each Snap acts as a building block of an integration pipeline and performs a single complete function such as read, write, or act on data. The SnapLogic Integration Cloud augments this wide range of Snap offerings with a custom Snap Development Kit (SDK) that allows customers to build their own Snaps that are tailored to their custom data actions or endpoint connectivity.

Read or write Snaps can consume data from a variety of systems (file, databases, etc.), applications (Salesforce, SAP, Workday, ServiceNow, Oracle ERP), and protocols (HTTP, REST, SOAP, FTP). These read/write Snaps provide users the ability to visually introspect and browse endpoint services and objects, and easily connect or consume them. Snaps that act on data are typically transforming, enriching, or cleansing data.

Unlike point-to-point integration tools, SnapLogic allows easy orchestration across multiple endpoints in a single flow, called a pipeline. This pipeline can be triggered based on events or scheduled jobs that are called via REST APIs or invoked programmatically via the SnapLogic application programming interface (snAPI).

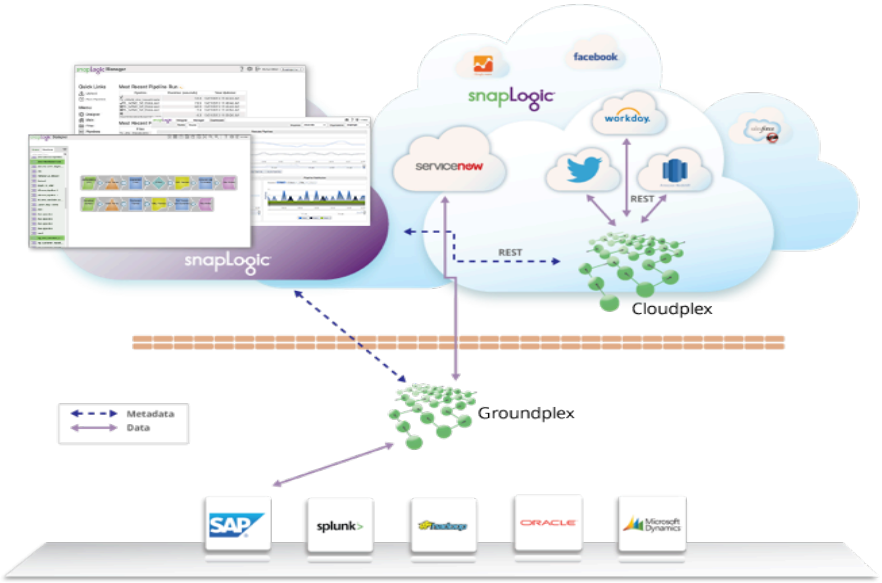
## Snaplex

A Snaplex is the data processing component of the SnapLogic Integration Cloud. Customers can deploy one or many Snaplexes as required to run pipelines and process data.



Snaplexes come in two flavors – On-premise Snaplex (aka "Groundplex") and Cloud Snaplex (aka "Cloudplex").

All Cloudplexes run inside the SnapLogic Integration Cloud. Groundplexes run behind the firewall, either on private data centers or virtual private data centers. Customers use the Manager and the Monitoring Dashboard to administer their Snaplexes. The SnapLogic DevOps team administers the infrastructure key performance indicators (KPIs) such as uptime, etc. of the Cloudplexes. Groundplexes are managed remotely by the SnapLogic Integration Cloud control plane where the software is self-upgrading or restarts are automatic in the events of failure.




*The SnapLogic Integration Cloud Deployment Architecture: Data is streamed not stored.*

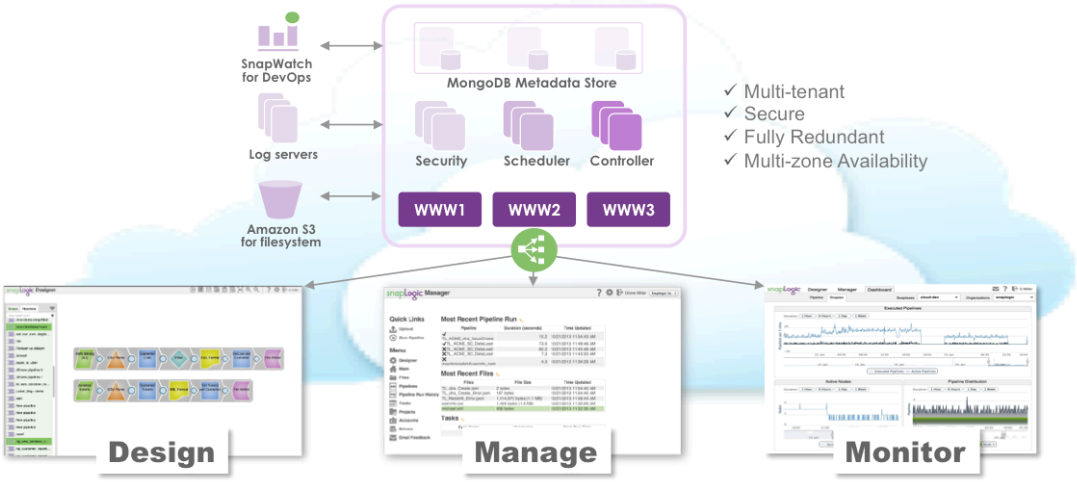
## Elastic Integration Platform Architecture

The SnapLogic Integration Cloud is architected on the concepts of software-defined networking (SDN)[1] . The system is decoupled into two main areas: a control plane and a data plane. The control plane controls where and how data is processed based on user configuration and preferences and some optimization algorithms. The data plane does the actual processing of data as per instructions received from the control plane.

### Control Plane

The "control plane" is 100% multi-tenant service that is hosted on the Amazon Web Service (AWS) infrastructure. The control plane consists of components that constitute the three web applications (Designer, Manager, Dashboard) and several key subcomponents that collectively manage the data plane.

[1] http://en.wikipedia.org/wiki/Software-defined_networking




*The SnapLogic Integration Cloud "control plane" is a multi-tenant service running on AWS.*

The control plane consists of the following components:

1. **Hardware and Software Load Balancers:** Route incoming requests to the appropriate component based on the type of incoming request (e.g. a designer request gets routed to the designer web app, or an incoming pipeline request gets routed to the appropriate data plane).
2. **Security:** Manages the authorization and authentication of users accessing the web applications.
3. **Scheduler:** Manages scheduled tasks and jobs of integration pipelines.
4. **Controller:** The control point where elasticity, lifecycle, software updates, etc. of all runtime components are managed.
5. **Metadata Repository:** Stores pipeline metadata in a fully redundant, secure database (MongoDB). This is where the integration pipeline metadata such as mapping data, configuration data, etc. is stored. This Metadata Repository is configured for backup and recovery to ensure multi-zone availability by leveraging Amazon Web Services' Disaster Recovery services.
6. **Amazon S3 for File System and Log Servers:** Stores system files and pipeline-generated log files.
7. **SnapWatch:** The management and monitoring tool for the SnapLogic DevOps team to administer the entire SnapLogic Integration Cloud infrastructure.

The control plane is multi-tenant and supports a multi-organization configuration. This means each organization (or "org") provisioned in the SnapLogic Integration Cloud gets a view into its

integration pipelines and configurations, which are managed and run independently of other tenants. Customers can also create and manage sub-organizations within a parent organization and manage them as separate departments with fine-grained access control. Sophisticated access control capabilities also allow administrators to group users and grant them group-level permissions to collaborate on integration projects.

## Data Plane

The "data plane" is where the actual business data is processed. Depending on the integration use cases, it consists of one or many Snaplexes. Customers needing to run integrations that orchestrate across cloud/SaaS applications with no on-premises connections will not require any software to run behind their firewall. A Cloudplex will run these integrations. Customers needing on-premise connectivity (e.g. SAP, Oracle, Microsoft Dynamics AX, etc.) will need a Groundplex.

The Cloudplex comes pre-installed and pre-configured inside the SnapLogic Integration Cloud. The Groundplex requires initial installation on your data center servers. After this initial set-up, the rest of Groundplex lifecycle (such as heartbeat monitoring, software upgrades, etc.) is remotely managed by the control plane.

The Snaplex can elastically expand and contract based on data traffic flowing through it. The unit of scalability inside Snaplex is a Java virtual machine (JVM), referred to as a "Node". The control plane has built-in smarts to automatically scale the Snaplex out and in order to handle variable traffic loads. For instance, each Snaplex is initialized with a configurable number of Nodes (say one, for example). Once the utilization of this one node reaches a certain configurable threshold (say 80%) due to spike in traffic, a new Node is automatically spun up to handle any additional workload. Once this excess data traffic has been processed and the second Node is sitting idle, it gets torn down to scale back the Snaplex to its original size.

In terms of AWS concepts, every Cloudplex node is an EC2 instance of the m1.large type. This, however, can be tailored to customer workloads where the instance can be either memory-optimized, compute optimized, or storage optimized.

### Web Standards: JSON, REST, APIs, and Mobile

JavaScript Object Notation (JSON) and Representational State Transfer (REST) are regarded as the key building blocks of the infinitely scalable web architecture. Purpose built for the cloud, the SnapLogic Integration Cloud has embraced these modern technologies and made them a native part of the platform.

The internal representation of data inside a SnapLogic Integration Cloud pipeline is JSON (JavaScript Object Notation) format. This lightweight data-interchange format is also the lingua franca of the web, making SnapLogic an ideal integration platform for the world of cloud and web standards. This format provides the platform the flexibility to handle structured as well as unstructured data.

Each deployed Snap by default is eligible for invocation with the REST protocol. Which also means that every SnapLogic pipeline or its subcomponent is eligible to be exposed as a REST API. An administrator exposing any pipeline as an API is a matter of flipping a switch. The administrator will need to grant requisite permissions (authentication and authorization) to clients. Typical clients of these APIs will be trading partners and mobile consumers looking to consume business data or business processes. For example, a trading partner may need to have real-time insight into inventory to ensure that they can make commitments to their customers that expects a certain product inventory levels. Or, a customer may want to check the status of their order through their mobile application; this lookup involves querying your shipping module with the shipment ID.

## Security

### Inter-Component Communication Security

The communication between the control plane and the Groundplex is via an SSL encrypted link (using port 443, the standard secure port used for secured HTTP traffic, just like any other secure link used for Salesforce or other enterprise or secure web application). This port is normally open outbound (from Groundplex to the control plane), as it is the standard port used by services using SSL. Once the link is established the protocol used over that link is also the standard web socket protocol (standardized by the IETF as RF6455 in 2011). This enables the full range of control communication, which happens over the link between the Groundplex and the control plane in the cloud. No additional ports or channels are required to be opened.

The control plane communicates with the Cloudplex over a secure HTTPS protocol.

### Runtime Data Security

The SnapLogic Integration Cloud is a stateless engine, which means it does not store runtime or business data. Any data read from say Teradata and loaded into Amazon Redshift is not stored anywhere along the way in the SnapLogic infrastructure. Snaps leverage the endpoint security provided by the application, database, file, etc.. For instance, it can read data from a database using a secure JDBC connection, or it can invoke a secure, HTTPS-based REST API as part of the integration pipeline. If the endpoint supports data encryption, the Snap can be configured to send and receive encrypted data as well.

### SnapLogic Design Time Infrastructure Security

Customer integration metadata and log files are stored on Amazon Web Services infrastructure (MongoDB, system files on S3, log files). Being 100% Amazon Web Services (AWS) based, SnapLogic inherently leverages the security and compliance capabilities offered by AWS.

## Getting Started: Building the First Pipeline

1. Within SnapLogic Designer, click the plus sign (+) in the lower left hand corner. This creates a new pipeline tab labeled <initials> new pipeline 0.
2. To change any pipeline information, such as name and location where it is run, click **Edit Pipeline Properties** on the Workspace Controllers. Make the necessary changes and click **Apply**.
3. Drag a Snap onto the workbench from the palette on the left hand side, then click on it to open the info box and configure the Snap. Repeat for other Snaps you want in your pipeline.
   Note: the shape of the Snap connector (circle vs. diamond) indicates which other Snaps it can connect with.


Several more examples and samples are available here.

## Summary

Purpose-built for the social, mobile, analytics and cloud (SMAC) era, the SnapLogic Integration Cloud transforms the on-boarding process onto cloud/SaaS applications into a simple drag-and-drop experience with no coding skills necessary. This state-of-the-art integration platform as a service speeds up integration projects by as much as 3 to 5 times, at a much lower cost than traditional integration offerings. With 160+ pre-built intelligent connectors and an SDK, no data source is out of reach. Lastly, with its modern elastic integration, customers now can handle data at any volume, variety, and velocity, without having to do rigorous capacity planning and provisioning.

For more information, visit www.SnapLogic.com or contact us at info@snaplogic.com.