Introduction
Accelerating paces of digital business and technology innovations have created massive amounts of data that require continual integration. Ninety percent of the world’s data has been created in the past two years. Examples of new data include cloud, mobile, social media, and IoT data. With diverse formats, sources, and locations, these data have created many challenges. Many organizations have adopted hybrid and multi-cloud strategies to manage data proliferation, gain flexibility, reduce costs, and increase capacities. The cloud storage and data warehouses—such as Azure, Google, Amazon Web Services (AWS), and Snowflake—particularly have gained enterprise acceptance.

Meanwhile, organizations still need to maintain and integrate their mainframes and other on-premises systems. It may surprise many people that 71% of Fortune 500 companies are still using the mainframe to manage their mission-critical workload, and 92 of the world’s top 100 banks continue to use the mainframe. Many mainframe systems contain master data or systems of records, so the mainframe is not going away anytime soon. Consequently, cloud migration often is a multi-year strategy with moving targets.

1 IBM Cloud
2 “Mainframe to Cloud: Design Once, Deploy Anywhere in a Hybrid World.”
However, the adoption of hybrid and multi-cloud without an independent integration layer can lock organizations to cloud vendors and create brittle point-to-point integration. For example, when various lines-of-business (LOBs) within an organization are directly integrating diverse data sources/targets using different integration utilities, each LOB would develop their own logics on data access, mapping, transformation, loading, and management. Data governance is often an after-thought. The end-result is the classic dreaded point-to-point integration, which results in redundant work, increased cost, higher complexity, and deteriorated quality assurance. Moreover, point-to-point integration is a fertile ground for security and governance breaches because of the lack of the coherent supervision. More critically, the cloud vendor lock-in can create a multitude of significant concerns:

- **Losing control over IT infrastructure**: A cloud provider may not meet all the requirements such as availability, performance, and regulatory compliance.

- **Concerns over data security and ownership**: Different cloud providers have different data security and ownership policies. Hefty costs may occur when moving data away from a cloud provider.

- **Lack of competitive advantages**: Some cloud providers directly compete or can become a competitor in a short time.

- **Relying on a single cloud provider**: Serious problems can occur if a cloud provider closes its business, becomes a frequent hacker/terrorist target, or gets in trouble with regulatory bodies.

- **Lessened bargaining power**: Cloud provider lock-in reduces an organization’s bargaining power to negotiate costs, priorities, SLAs, and the like.

To avoid cloud vendor lock-in, integration architects and managers should adopt four best practice guidelines detailed in this white paper:

1. Apply the separation-of-concerns design principle
2. Design three resilient integration architectures
3. Plan your integration strategy
4. Adopt an enterprise integration solution
Best Practice 1: Apply the Separation-of-Concerns Design Principle

Separation-of-concerns, a time-tested design principle, prevents cloud vendor lock-in. It separates computing functions into different layers (or concerns), such as business processes, applications, data management, integration, infrastructure, and so on. Each layer typically represents a distinct class of software.

In the past, separating the data management layer from the application layer resulted in relational databases, which greatly advanced computing capabilities. Similarly, separating the integration layer from applications and data management layers was a milestone that improved system agility, flexibility, scalability, and sustainability. These considerations become more important in hybrid and multi-cloud environments.

Without separation-of-concerns, organizations can face serious vendor lock-in problems. This is often caused by point-to-point integration using custom coding or embedded integration utilities. For example, many organizations used Apache Pig and MapReduce to develop custom code to access and process big data during Hadoop 1.0 stage. They then later switched to YARN to manage workload during Hadoop 2.0 stage. Still, more changes are coming. Some organizations are running Hadoop in containers, such as Docker, in Hadoop 3.0 stage. Technology changes come in waves. With point-to-point integration, each wave requires re-writing, re-deployment and re-optimizing integration logics. In another example, an organization had developed extensive ETL jobs using embedded integration utilities as parts of a reporting software because they wanted to save the costs of acquiring an independent integration tool. However, when its reporting vendor decided to sunset its integration utilities to focus on its core competency—analytics—this organization faced the daunting task of migrating hundreds of ETL jobs to a new software.

Separation-of-concerns and fast-changing technologies require knowledge specialization. IT professionals have increasingly become more specialized in certain domains such as integration, data management, analytics, and application development. This knowledge specialization not only improves skill levels but also enables stronger security and governance.
For example, sensitive and confidential data should only be accessible by appropriate roles and enforced by systems that follow separation-of-concerns. Many recent cyber breaches were caused by losing sight of this important design principle because organizations have mixed various roles, such as business users, database administrators, and software developers.

The diagram below compares two distinct integration architectures: the left side shows a point-to-point integration, which results in cloud vendor lock-in, inflexibility, increased costs/complexity, and security/regulation breaches. In comparison, the right side shows a more resilient integration architecture adopting the separation-of-concerns principle, which provides flexibility, reduced costs, regulatory compliance, and business competitiveness.

**Point-to-Point Integration**
Cloud vendor lock-in, brittle, inflexibility, increased costs/complexity, and security/regulation breaches.

**Separation-of-Concerns**
Resilient, flexibility, reduced costs, regulatory compliance, and business competitiveness.

Precisely data integration software enables you to create an independent integration layer to connect with any data sources and targets. This can empower your organization to manage technology changes from wave to wave.
Best Practice 2: Design Three Resilient Integration Architectures for Hybrid and Multi-Cloud Environments

In a hybrid and multi-cloud environment, there are three integration architectures that follow the separation-of-concerns principle to ensure success and reduce risks. They are as follows:

1. **On-premises to cloud integration:** Organizations typically start migrating LOB systems to the cloud to test the water first. As many organizations are still using the mainframe to manage their mission-critical workload, it is vital to integrate cloud and mainframe data seamlessly to get a complete picture of customers, suppliers, products, and employees. Unlocking valuable mainframe data can be a substantial challenge because of lack of skilled resources, legacy data structures, custom-coding, and proprietary storage formats. A versatile integration tool can help integrate, govern, and optimize data anywhere—from the mainframe to the cloud—in a way that’s easy, fast, cost-effective, and secure.

2. **Cloud to on-premises integration:** Once data and systems are moved to the cloud, they often need to be pulled back to on-premises for analytical and transactional purposes. For instance, one organization first used Precisely Connect’s replication capabilities to move real-time transactional data to their data lake. Both transactional systems and data lake reside on-premises. In the next phase, they used Connect to extract data from AWS S3 back to their on-premises analytics platform. Unlike embedded integration utilities provided by cloud providers that are meant for tactical usages that focus on getting data into their systems, Connect can help build high-performing integration jobs, enforce regulatory compliance, and retain data ownership, no matter where and how you plan to move data.

3. **Cloud to cloud integration:** The multi-cloud strategy combining an independent integration layer helps avoid cloud vendor lock-in, gains flexibility, reduces costs, and increases capabilities. For example, one large insurance company first used Precisely’s Connect to migrate their on-premises data warehouse to AWS in 2016 to 2017. They then added Google Cloud storage later. Precisely provided a future-proof solution that allowed them to integrate AWS with Google Cloud and future data sources easily.

Precisely’s Connect data integration software can build above integration architectures, enabling the design of jobs once and then deploying anywhere—hybrid cloud, multi-cloud, single-server, and distributed platform. It allows the definition and enforcement of governance policies in one central location without worries about data breaches. It also gives a full view of the end-to-end data lineage for change management and regulatory compliance.
**Best Practice 3:**
**Plan Your Integration Strategy**
To build future-proof integration architectures in hybrid and multi-cloud environments, integration architects and managers should use the following key questions for planning:

- **What is your organization’s digital business strategy?**
  Digital business has created new business models or transformed existing ones by fully integrating digital and physical worlds. Innovative technologies have profoundly altered the landscape of many industries. Some organizations thrive in the face of changes, while others perish. While IT professionals do not create digital business strategies, they play a key role enabling and influencing these strategies through powerful digital platforms. One of the key components in digital platforms is a high-performing and future-proof integration tool set.

- **What are your organization’s current and future integration requirements?**
  Inventorying the current integration ‘pains’ is one of the first steps to improve integration and manage changes. For some organizations, the pains include the lack of skills, poor performance, and low availability of systems. For others, their pains include inflexibility, or excessive resources tied up by maintenance. It is crucial to evaluate your organization’s integration tools and architecture. Moreover, staying informed about the digital business strategy helps prepare for future requirements and challenges. For example, if your organization plans to expand from North America to the European Union (EU) in the next three years, you can prepare for it by evaluating integration vendors that meet the EU’s regulations such as the General Data Protection Regulation (GDPR).

- **Where is the data gravity, on the cloud or on-premises?**
  The center of gravity is formed by the primary locations of data mass (i.e. files, documents, databases or data storages). If data gravity is on the cloud, it makes sense to deploy integration jobs on the cloud near the data mass to improve performance and simplify architecture. If the data gravity is in your local data centers, it makes more sense to deploy integration jobs on-premises close to your data mass. Syncsort provides the design of integration jobs once, and the deployment anywhere: on-premises or on the cloud.

- **How has your organization’s funding been changed over the years?**
  We are seeing a trend that IT funding is being shifted towards LOBs from centralized, shared IT groups. With fewer resources and higher demands, it is critical to develop future-proof integration jobs quickly without tying up precious resources in manual coding and rework.

- **What are your governance requirements?**
  New self-service integration personas—for example, data scientists, data engineers, and tech-savvy business users—are emerging as parts of digital business. They often have conflicting integration requirements and priorities. How to harmonize their conflicting requirements, while maintaining enterprise standards and optimizing costs in the long term? The industry best practice is to adopt bi-model IT, meaning two complementary modes of IT delivery. Mode 1 focuses on governance and quality-of-services (QoS) such as performance, availability, data quality, and scalability. Mode 2 focuses on agility and speed. In mode 2, self-service users often create their integration solutions quickly to validate their analytical hypotheses or fulfill ad-hoc transactional needs. Rather than letting self-service users run wild without any governance, the shared IT services can provide governed sandboxes, educate users on technologies, and offer suitable assistance. Once prototypes are proven and have met certain thresholds (such as costs, QoS requirements or the number of jobs/departments involved), the shared IT services can then turn these prototypes into shared assets using an enterprise integration tool and applying the separation-of-concern principle.
Best Practice 4: Adopt an Enterprise Integration Solution

It is easy to see that manual-coding is costly, unsecure, and unsustainable in hybrid and multi-cloud environments. Using multiple integration tools have equal pitfalls too: inconsistent logics, redundant work, and high costs to acquire and maintain tools. If your organization has several integration tools, it is time to re-evaluate and standardize them to a single enterprise solution that is easy to use, cost effective, and future-proof in hybrid and multi-cloud environments.

Precisely provides solutions for over 12,000 customers, including 90 of the Fortune 100. Precisely data integration software helps integrate, govern and optimize data anywhere. Highlights of Precisely’s world-leading software include:

- Precisely’s Connect can be deployed on-premises, on a distributed platform, on a private/public cloud, or on a hybrid cloud. Connect can perform ‘design once, deploy anywhere’ with no need to re-design, re-compile, re-deploy, and re-optimize. Precisely enables applications for emerging frameworks and platforms.

- Precisely has intuitive, drag-and-drop graphic interfaces and high-performance ETL capabilities to integrate all enterprise data sources, including mainframes, RDBMS, IBM i-Series, NoSQL, Hadoop, streaming, and the cloud.

- Precisely has a better total-cost-of-ownership than other integration providers. Gartner Magic Quadrant for Data Integration Tools (2019) highlighted Precisely’s robust tools, positive customer relationships, and broad applicability for data management scenarios.

- Precisely data integration and its associated software offers enterprise data management capabilities: ETL, CDC, data quality, data enrichment, location intelligence, security and more. Precisely’s vision and commitment is to build an industry-leading data integrity platform.
Summary
This white paper has detailed four best practice guidelines to avoid cloud vendor lock-in:
1. Applying the separation-of-concerns design principle
2. Designing three resilient integration architectures
3. Planning your integration strategy
4. Adopting an enterprise integration solution

The table below compares two integration architectures: the left is the point-to-point integration architecture using custom coding or embedded integration utilities provided by cloud vendors; the right is a resilient integration architecture that adopts the separation-of-concerns principle:

<table>
<thead>
<tr>
<th>Comparison Criteria</th>
<th>Point-to-Point Integration Using Custom-Coding or Embedded Utilities</th>
<th>Separation-of-Concern Using Precisely Integration Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoid Vendor Lock-in</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Separation-of-Concern</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Flexibility</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Design Once, Deploy Anywhere</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Governance &amp; Regulatory Compliance</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Skills</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sustainability</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Total Cost of Ownership</td>
<td>X</td>
<td>✓</td>
</tr>
</tbody>
</table>

As has been made evident, the integration architecture following the separation-of-concerns principle helps meet the current and future challenges of a complex data environment.

To learn more about how Precisely Connect can help your business build a design once, deploy anywhere integration plan and avoid vendor lock-in, visit www.precisely.com.
About Precisely
Precisely is the global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 90 percent of the Fortune 100. Precisely’s data integration, data quality, location intelligence, and data enrichment products power better business decisions to create better outcomes.

Learn more at precisely.com.