When designed well, a data lake is an effective data-management design pattern for supporting many use cases. In this study, both new and old, and at large scale. Organizations are adopting the data lake design pattern (whether on Hadoop or a relational database) because lakes provision the kind of raw data that users need for data exploration and discovery-oriented forms of advanced analytics. In the recent TDWI Best Practices Report, “Data Lakes – Purposes, Practices, Patterns, and Platforms,” TDWI analysts discuss emerging best practices, and take a look at user trends and challenges for older, mature, and at least at small-scale data lakes.

Here are several of the key survey results.

**BENEFITS**

When do you expect to have a data lake in production?

- **Within 24 months:** 24%
- **Within 12 months:** 21%
- **Within 36 months:** 23%
- **Within 24 months:** 15%
- **Never:** 7%

**Reasons for Choosing Hadoop-based Data Lakes**

- Enable new analytics and expand old ones
- Can be extended
- Enable fraud prevention
- Enable real-time processing

**DATA LAKE OWNERS**

- Analysts and developers: 33%
- Data scientists: 29%
- IT Managers: 17%
- Application and business units: 16%
- Third party (e.g., cloud provider): 6%
- Central IT: 4%
- Research groups: 25%
- Other: 6%

**DATA LAKE ADOPTION**

- Nearly a quarter of respondents already have a data lake in production, up from 14% in the 2015 TDWI Best Practices Report.
- Of the 72 respondents who have data lake experience, Hadoop is preferred over relational databases for managing data’s exploding diversity and scale.
- When designed well, a data lake is an effective data-driven design pattern for supporting many use cases.

**DATA LAKE OWNERS**

- **Application and business units:** 25%
- **IT Managers:** 17%
- **Research groups:** 16%
- **Other:** 6%
- **Central IT:** 4%
- **Data scientists:** 33%
- **Analysts and developers:** 33%

**DATA LAKE OWNERSHIP**

- **Engineers and architects:** 43%
- **Data scientists:** 39%
- **Analysts and developers:** 33%
- **DBAs, DI specialists:** 12%
- **Managers:** 6%

**DATA LAKE USE CASES**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%
- **Enable real-time analytics:** 29%
- **Reduce time to market:** 24%
- **Enable business unit innovation:** 21%
- **Enable business unit innovation:** 15%

**DATA LAKE ACTIONS**

- Mostly raw data with a few areas for restructured data: 56%
- Exclusively raw data: 32%
- Mostly restructured data: 12%
- Don’t know: 2%

**DATA LAKE PRODUCTION**

- Already in production: 24%
- In production: 23%
- Within 24 months: 21%
- Within 36 months: 19%
- Never: 7%

**DATA LAKE TRANSFER**

- **Hadoop:** 43%
- **Relational database:** 29%
- **Both:** 14%
- **Don’t know:** 4%

**DATA LAKE_TASKS**

- **For respondents with data lake experience, Hadoop is preferred over relational databases for managing data’s exploding diversity and scale.**

**DATA LAKE DRIVERS**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE PLATFORM ISSUES: HADOOP VS. RELATIONAL DATABASES**

- **Lack of data governance:** 31%
- **Lack of data management tools and skills for Hadoop:** 29%
- **Inadequate skills for Hadoop:** 25%
- **Inadequate skills for relational databases:** 16%
- **Lack of compelling business case:** 15%
- **Inadequate tools and skills for relational databases:** 4%

**DATA LAKE PURPOSES**

- **Application and business units:** 33%
- **IT Managers:** 33%
- **Research groups:** 25%
- **Other:** 6%
- **Central IT:** 6%

**DATA LAKE SKILLS**

- **Very technical people and, on average, at least one person with more than two years of experience:** 61%
- **Very technical people and, on average, at least one person with more than two years of experience:** 61%

**DATA LAKE BENEFITS**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE USE CASES**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE SAVINGS**

- **30%:** 40%
- **25%:** 39%
- **20%:** 18%
- **15%:** 3%
- **10%:** 0%

**DATA LAKE ADAPTATIONS**

- Mostly raw data with a few areas for restructured data: 56%
- Exclusively raw data: 32%
- Mostly restructured data: 12%
- Don’t know: 2%

**DATA LAKE USE CASES**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE BENEFITS**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE USE CASES**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%

**DATA LAKE BENEFITS**

- **Enable new analytics and expand old ones:** 41%
- **Can be extended:** 34%
- **Enable fraud prevention:** 32%
- **Enable real-time processing:** 31%