



# Healthcare Solution Provider

LOWERS COSTS & SCALES ANALYTICS WITH APACHE™ HADOOP®

Healthcare Case Study



Louisville, KY based ZirMed, founded in 1999, is a leading provider of healthcare information management solutions. Healthcare providers, including physicians, hospitals and large health systems, use the company's cloud-based revenue cycle management offerings to manage the complex process of billing and collecting revenue from patients and payers.

The company's products allow providers to verify patient eligibility and benefits; submit and track claims; process payments from both patients and payers; and facilitate, manage, and analyze clinical and financial communications among all three constituencies.

By aggregating data from its network of over 1,200 payers, 200,000+ providers and millions of patients, ZirMed has created a national database of healthcare payment information that it uses to help providers reduce administrative burdens, drive patient engagement and increase financial performance.

## Challenge

In 2011, just after being named for the sixth consecutive year to INC magazine's prestigious ranking of the nation's fastest-growing private companies, ZirMed hired CEO Thomas W. Butts, with an eye towards achieving even greater growth.

While the company had made great strides in servicing small and medium providers, Butts, former CEO Optos, PLC and Chief Operating Officer IDX Inc., a 27-year industry veteran, quickly realized that ZirMed's next phase of growth would come from expanding its ability to serve large healthcare enterprises. It was also clear to him that **data and analytics would be the way ZirMed differentiated itself in the market.**

When Butts shared his vision with ZirMed CTO Chris Schremser, Schremser knew he was in for a challenge. The company's current analytics offering would not meet the demands of this expanded vision, even after years of development and several architectural iterations.



## NETWORK SIZE

**1,200**  
*payers*

**200k<sup>+</sup>**  
*providers*

**1m<sup>+</sup>**  
*patients*

[www.hortonworks.com/contact/sales](http://www.hortonworks.com/contact/sales)

Initially launched in 2008 based on traditional proprietary technologies—including a data warehouse, an ETL system, and a reporting tool—“Analytics 1.0” at ZirMed was slow to develop and sluggish to use. “Analytics 2.0,” delivered in 2012, ditched the static reporting tool for an interactive reporting layer, but was still based on the data warehouse. While it offered some relief to frustrated users, the company ran into performance issues with a modest increase in the product’s user base.

And performance wasn’t the only thing Schremser had to worry about.

“Performance issues aside, the cost of our previous analytics solution presented a significant barrier to progress,” said Schremser. “The data warehouse sat on top of enterprise storage and was very expensive. Plus, we needed numerous copies for various test, development, and staging uses, multiplying the cost.”

“As a result of high costs, we could only support loading a subset of the fields we needed into the data warehouse,” continued Schremser. “When we needed to run queries that required data from the missing fields, we had to run them against the unoptimized OLTP systems with a huge performance penalty.”

---

**In spite of the historical challenges, Schremser’s marching orders were clear. He needed to find a way to deliver a scalable, user-friendly analytics system that didn’t break the bank.**

---

## Solution

While Schremser was familiar with the work happening around big data platforms like Apache Hadoop, he initially feared that a big data platform would be out of the company’s reach.

“Much of the local talent base here in Louisville comes from large companies enterprise organizations. As a result, many of these individuals specialize in Windows and .NET, said Schremser. “The typical crop of big data technologies, which rely on Linux, Java and even more ‘exotic’ languages like Scala, would have been a show stopper for us. We couldn’t move fast enough if we had to build on unfamiliar technologies from the OS up.”

After learning about Microsoft and Hortonworks Data Platform (HDP) for Windows at the Microsoft TechEd conference, Schremser realized it offered just what he needed: **The scalability and commodity economics of Hadoop running on a technology platform that his team knew and trusted.**

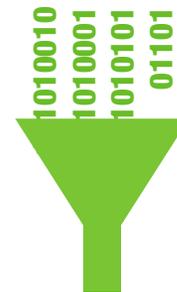
To prove out his hunch, Schremser kicked off a proof-of-concept. Members of the data team at ZirMed quickly deployed a small Hadoop cluster running HDP for Windows 1.3 on 8 modestly configured machines—2 master nodes and 6 workers.

The team’s next win came equally quickly. Two weeks after standing up the cluster, the team completed loading the company’s key transactional data—9 years worth of electronics remittance data—into Hadoop.

Ready to put Hadoop to the ultimate test, the ZirMed team ran a battery of queries against the data in the cluster. The real-world queries they used fell into several buckets: those easily handled by the data warehouse, those that could be handled by the data warehouse but that required new aggregates to be created, and those that couldn’t be run on the data warehouse and had to be run against the source OLTP system. In each case, the Hadoop cluster bested the performance of the incumbent technology platform.



*The typical crop of big data technologies, which rely on Linux...would have been a show stopper for us. We couldn't move fast enough if we had to build on unfamiliar technologies from the OS up.*



*No subsets. We capture all data, for all years.*

The most standout performance gains came from the most challenging queries: those that needed to be run against the OLTP system and that commonly required 24-48 hours to run in that environment. With HDP for Windows, the longest-running of these queries took only 25 minutes!

Now confident in his decision to turn to HDP for Windows, Schremser took the results to his CEO with his recommendation to move forward.

## Results

With the CEO's nod, the team built a production Hadoop cluster running HDP for Windows 2.0. The 29 node cluster contains 1.2 PB of raw storage—of which 420 TB is usable considering replication—at a total hardware cost of \$235,000. The company was used to paying \$300,000 for 100TB of raw enterprise storage alone, (75-80 TB usable) with no processing power, and had invested over \$750,000 when licenses and compute servers were factored in.

“With HDP for Windows we got a really resilient platform with 5 times the amount of usable storage, plus a ton of processing power, for about 30% of the cost of traditional enterprise technologies,” said Schremser.

Queries that took 25 minutes on the proof-of-concept cluster take 5 minutes now in the production environment. (With more data—the company continues to load transactions to the Hadoop cluster.) ZirMed's team thinks they could optimize the query and get it down to 2 minutes, and they're looking forward to adopting emerging Hadoop technologies like YARN and Tez to help them achieve even faster performance.

An additional benefit: because the data on Hadoop is accessible via Hive it is much easier for their team to work with. “It's SQL. And anyone can write SQL,” says Schremser.

“We expect our success with analytics and Hadoop to have a dramatic effect on ZirMed's business, allowing us to increase customer satisfaction and create new revenue streams,” said Butts. “This is the first time in my career that I've seen technology move faster than the business.”

---

**As a result of the company's successful “Analytics 3.0” endeavor, ZirMed has begun to articulate an expanded vision of the next generation of health management, powered by data and analytics.**

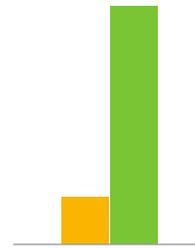
---

“We believe that the future of healthcare is in marrying data from financial and clinical applications with other public and private data sets,” said Butts. “This will allow providers to deliver a higher level of care while enhancing financial results, and we're excited to be at the forefront of this trend.”

## Hortonworks. We do Hadoop.

Hortonworks is a leading commercial vendor of Apache Hadoop, the open source platform for storing, managing and analyzing big data. Our distribution of Apache Hadoop, Hortonworks Data Platform, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions.

Hortonworks is the trusted source for information on Hadoop, and together with the Apache community, Hortonworks is making Hadoop an enterprise data platform. Hortonworks provides unmatched technical support, training and certification programs for enterprises, systems integrators and technology vendors.



STORAGE

**5x**

*usable capacity*

**30%**

*of the cost*

---

QUERIES THEN

**24–48** *hours*

QUERIES NOW

**3–5** *minutes*

To learn more, visit

[www.hortonworks.com/contact/sales](http://www.hortonworks.com/contact/sales)

or call (855) 8-HORTON