SEI Training

Big Data Architectures and Technologies

About the Course

Scalable "big data" systems are significant long-term investments that must scale to handle ever-increasing data volumes and workloads. These are high-risk applications in which the software and data architectures provide the foundation for success. This course is designed for architects and technical stakeholders such as product managers, development managers, and systems engineers who are involved in the development of big-data applications. It focuses on the relationships among application software, data models, and deployment architectures, and how specific technology selection relates to all of these. While the course touches briefly on data analytics, it focuses on distributed data storage and access infrastructure, and the architecture tradeoffs needed to achieve scalability, consistency, availability, and performance. We illustrate these architecture principles with examples from selected NoSQL product implementations.

"Good overview of 'gotchas' of different technologies."

"Very knowledgeable instructors."

"Liked the incorporation of real-world products and examples."

— Course Attendees

Objectives

At the completion of the course, learners will understand:

- What "big data" is, how and why it has evolved, and the technologies that have emerged to address its complexities in the realm of computer science and software engineering
- The basics of distributed systems, including durability, transactional consistency, and replica consistency
- The quality attributes important in distributed systems and how they are achieved in practice
- Specific technologies, such as NoSQL and NewSQL databases
- Data modeling and the common types of data that need modeling
- Performance considerations in data modeling
- Distributed data processing frameworks employed in big data systems, such as Hadoop and its associated HDFS file system, which support downstream activities
- The newly emerging distributed data processing frameworks
  - Distributed computations with Spark
  - Stream processing with Storm
- Architectural issues present when building big data systems
- Big data system design tactics
- Software engineering heuristics to achieve effective, reliable, and scalable software systems

Attendees' evaluation of course strengths:

- Relation to software architecture concepts with technology
- Need for NoSQL technology assessments
- Exercises for modeling using different NoSQL DBs

Who Should Attend?

- Architects
- Technical stakeholders involved in the development of big data applications
- Product managers, development managers, and systems engineers

Topics

- The major elements of big data software architectures
- The different types and major features of NoSQL databases
- Patterns for designing data models that support high performance and scalability
- Distributed data processing frameworks

Prerequisites

There are no prerequisites for this course. It is recommended that participants have some experience in the fields of development of software-intensive systems.

Materials

Students will receive the complete set of slides and recommendations for related papers and reference materials.

Three Ways to Attend

1. Public, instructor-led offering at an SEI office
2. Private, instructor-led training at customer sites
3. eLearning

Big Data and Software Architecture Educational Resources

Search the SEI Website for "Big Data" and find podcasts, blog posts, webinars, and more.

For More Information

To learn more and to register for the course, visit sei.cmu.edu/go/big-data

Training courses provided by the SEI are not academic courses for academic credit toward a degree. Any certificates provided are evidence of the completion of the courses and are not official academic credentials.

About Us

For four decades, the Software Engineering Institute (SEI) has been helping government and industry organizations to acquire, develop, operate, and sustain software systems that are innovative, affordable, enduring, and trustworthy.

Contact Us

Software Engineering Institute
4500 Fifth Avenue, Pittsburgh, PA 15213-2612

Phone: 412.268.5800 | 888.201.4479
Web: www.sei.cmu.edu | www.cert.org
Email: info@sei.cmu.edu