The term “Big Data” refers not only to the volume of data, but the complexity of data processing required to extract knowledge from that data. The old method of gathering data to answer a question or validate a hypothesis is giving way to data-driven exploration and discovery using massive datasets collected in survey mode. Software can explore the high dimensionality of Big Data to discover connections and relationships not known ahead of time; the data itself drives the discovery.

To the user, LSST is not only a telescope, it is a database - a Big Data project tuned to facilitate data-driven explorations of the most fundamental questions of the Universe. LSST will survey the visible night sky relentlessly for ten years, building a 500 PB non-proprietary database of images and a 15 PB catalog of text data describing properties of nearly 40 billion individual stars and galaxies.

Cutting-edge computer applications will be used to hold the data and mine it for scientific discoveries. The data management system being developed must be able to process the nightly alert data, 20,000 expected transient alerts per minute, in near real-time, and construct annual data releases (DRs) at the petabyte scale. Each DR processes image data collected from the start of the survey. LSST focuses on providing the well-calibrated data to the community through educational and high performance computing interfaces. The software development team consists of more than 100 people working in six different sites across the US, with further support and contributions from IN2P3, developing an integrated set of software to realize the LSST science goals.

Analysis of LSST’s multi-dimensional database will reveal answers in four primary science areas:

- Probing Dark Matter and Dark Energy
- Exploring the Transient Optical Sky
- Taking an Inventory of the Solar System
- Mapping the Milky Way