1)What is Flume NG?

Flume NG (Next Generation) is a huge departure from Flume OG (Original Generation) in its implementation although many of the original concepts are the same. If you're already familiar with Flume, here's what you need to know.

* You still have source*s* and sink*s* and they still do the same thing. They are now connected by *channels*.
* Channels are pluggable and dictate durability. Flume NG ships with an in-memory channel for fast, but non-durable event delivery and a file-based channel for durable event delivery.
* There's no more logical or physical nodes. We call all physical nodes agent*s* and agents can run zero or more sources and sinks.
* There's no master and no ZooKeeper dependency anymore. At this time, Flume runs with a simple file-based configuration system.
* Just about everything is a plugin, some end user facing, some for tool and system developers. Pluggable components include channels, sources, sinks, interceptors, sink processors, and event serializers.

2)Can Flume provides 100% reliability to the data flow?

Yes, Flume has 100% plugin-based architecture, it can load and ships data from external sources to external destination which seperately from Flume. So that most of the bidata analysis use this tool for sreaming data.

3)can Flume can distributes data to multiple destinations?

Yes. It support multiplexing flow. The event flows from one source to multiple channel and multiple destionations, It is acheived by defining a flow multiplexer.

4) Explain about the different channel types in Flume. And which channel type is faster?

The 3 different built in channel types available in Flume are-

MEMORY Channel – Events are read from the source into memory and passed to the sink.

JDBC Channel – JDBC Channel stores the events in an embedded Derby database.

FILE Channel –File Channel writes the contents to a file on the file system after reading the event from a source. The file is deleted only  after the contents are successfully delivered to the sink.

MEMORY Channel is the fastest channel among the three however has the risk of data loss. The channel that you choose completely depends on the nature of the big data application and the value of each event.